

MATEMATIKA И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2013
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2013
*Proceedings of the Forty Second Spring Conference
of the Union of Bulgarian Mathematicians
Borovetz, April 2–6, 2013*

**ЗА УЧЕБНОТО СЪДЪРЖАНИЕ ПО ВЕРОЯТНОСТИ
И СТАТИСТИКА В ПРОГИМНАЗИАЛНИЯ
И ГИМНАЗИАЛЕН ЕТАП***

Пламен Матеев, Марусия Славчова-Божкова

Предлагаме за запознаване и обсъждане нашето виждане за учебното съдържание по Вероятности и Статистически анализ на данни в прогимназиалния и първи и втори гимназиален етап, задължителна и профилирана подготовка.

1. Увод. Да отбележим от самото начало, че 2013 година е обявена за Международна година на статистиката и това не е случайно. Навършват се 300 години от издаването на *Ars Conjectandi* на Яков Бернули [1]. В него се определя какво е вероятност и се доказва връзката ѝ с наблюденията, твърдение популярно сега като „Закон за големите числа“. Възможен превод на заглавието от латински е „Изкуството да си предвидлив“. Бернули определя предвидливостта като знанието на вероятността на нещата, което според него на гръцки се нарича „стохастика“.

Все по-осъзнатата и обективно проявяваща се роля на статистиката в математиката и редица сфери на ежедневието намираме в изказвания на влиятелни личности като Hal Varian, икономически директор на Google [4], и носителя на Филдсов медал от 1974 година Мъмфорд [3]. Примерите са многобройни: като започнем от държавното управление, управлението на една фирма, през анализите на състоянието в банковата сфера, на пазара на имоти, и изобщо на пазара, за да стигнем до различните научни области, като физика, откъдето всъщност са започнали някои сериозни статистически анализи, биология, медицина, икономика, и т.н. В световен мащаб мястото на статистиката в образователните цели на държавата все повече нараства, което мотивира и намира израз и в нашите образователни програми. Нашето мнение [2] е, че статистиката като всепроникваща във всички области на познанието и като част от приложната математика трябва да бъде застъпена по-отрано, като акцентът е върху работа с данни и моделирането им.

В следващите параграфи ще изложим представата си за *началните знания*, които трябва да се предадат на възможно най-ранен етап, *основите знания*, с които трябва да разполага „всеки информиран гражданин“, и накрая – *разширено съдържание* за ученици с изразени интереси към математиката.

*Работата е частично финансирана от средствата, отпуснати целево от държавния бюджет на СУ „Св. Климент Охридски“ за научни изследвания през 2012 г. (договор № 125/2012).

2. Начални знания. Изхождайки от разбирането за вероятността като модел на данните, предлагаме в образованието по статистика да се започне с начините за описание, представяне и тълкуване на данните. Това от своя страна предполага изграждане на умения, свързани със събиране, организиране и представяне на данни за един признак в таблица; разчитане и интерпретиране на данни, зададени с таблици и диаграми; намиране на средноаритметично и мода и използването им за анализ на данни; преминаване от графична информация в числова и обратно; таблично или графично представяне на получени или зададени емпирични данни. Като следваща стъпка с цел по-плавен преход към математическите корени на вероятностите считаме, че е удачно учениците да развият умения за изброяване на възможности по правилото за събиране и умножаване на възможности, както и за определяне на множество и намиране на сечение, обединение на множества, допълнение и подмножество на множество. Те са обусловени от въвеждането на понятието *вероятност на случайно събитие*. В края на този етап учениците трябва да са придобили умения да решават задачи, като използват, че сборът от вероятностите на събитие и неговото отрицание е 1 като първа стъпка към пресмятане на вероятност на събитие.

3. Основно съдържание. Този етап започва с изучаване на комбинаторните съединения – пермутации, вариации и комбинации без повторение, които е препоръчително да бъдат въведени с примери от вида на избор на топки от урна или с други добре известни подходящи схеми, чрез които се моделират реални ситуации. Това е с цел въвеждане на класическа вероятност на събитие и възможността да се пресмята не само като отношение на благоприятните възможности към всички възможности, но и комбинаторно. По-нататък като естествено продължение се въвежда *вероятност на сложни събития* и формулите за сума на събития, както и понятието *условна вероятност* с цел намиране вероятност на сечение на две събития. След това чрез достатъчно общи примери, учениците изграждат умения да разпознават и прилагат модели на многократни опити с два възможни изхода в конкретни практически ситуации, както и да разбират и пресмятат разпределение на вероятностите със сума 1. Опит да се покаже, че имаме и други примери за вероятност, е запознаването с геометричната вероятност върху правата и равнината, което е и първи преход към идеята за непрекъснати случайни величини. Това е най-общо съдържанието по вероятности и комбинаторика.

В частта статистика се цели да се развият умения за работа с данни – въвежда се понятието *генерална съвкупност*, илюстрирано с примери от реални ситуации, както и *медианата* в допълнение на *средноаритметичното* и *модата*. Също така се въвежда петчисленото представяне на данните, както и хистограма, полигон и таблица на кумулативните честоти на данните. Това съдържание има за цел да изгради основни умения за тълкуване и анализиране на данни и е въведение в онази част от статистиката, известна като *описателна статистика*.

4. Разширено съдържание. Разширеното съдържание допълва основните знания по вероятности и анализ на данни с формулите за пълната вероятност и формулата на Бейс, както и с понятието *случайна величина*, математическият модел за анализ на данни от наблюдения. Допълват се някои понятия от описателната статистика. Илюстрация на идеята за статистически изводи се поднася върху два примера. Първият е биомно разпределение като модел за големината на пропорция

(в наблюденията на променлива, определяща две допълващи се категории). Вторият е нормалното разпределение като модел на средната стойност на количествена променлива. Понятие за корелация и идеята за регресионна зависимост присъстват във всички училищни програми от този тип.

Последната част включва класическото приложение на вероятностите при определяне на застраховки и осигуряване, свързани с продължителността на живот.

4.1. Случайни величини. *Случайна величина* се въвежда с примери за разпределение на вероятностите върху няколко числа: това може да са броят на точките върху зар, броят на правилните отговори на класа в проведен тест, дължината на класната стая, измерена от учениците в сантиметри, педи, стъпки. С примери се показват и елементарните свойства на случайните величини (умножение с число и сума на две случайни величини). *Средната стойност* (математическото очакване) и *стандартното отклонение* се въвеждат като характеристики на разпределението на вероятността. Показателна илюстрация е вероятността на симетричен интервал спрямо математическото очакване, измерен в стандартни отклонения.

Непосредствено след въвеждането на понятието *случайна величина* се представя приложението му като модел при анализ на данните, основано на началните и основни знания. Описателната статистика се допълва с емпирична функция на разпределение и нейния вариант *огива*, с който се облекчава определянето на медианата и най-вече квантилите. Разширява се и опита с определяне типа на променливите и графично представяне на данни с две и повече променливи.

Тук следва да се доразвият умения за работа с данни – събиране, таблично и графично представяне, описателни статистики и анализ на възможните причини за поява на систематична грешка при екстраполацията на изводите от извадката върху генералната съвкупност. Полезно е да се въведат новите понятия като асиметрия на разпределението и наблюдения „изключения“, наричани още „груби грешки“ или „аутлаери“ (outliers).

4.2. Биномното разпределение, модел на данни и изводи за пропорции. Схемата на Бернули и биномното разпределение са традиционен избор за илюстрация на основните положения в статистиката. Освен това ситуацията, при която това е подходящият модел, са много и разнообразни – от социологически и маркетингови анкети до контрол на дефектни изделия в партида и образователни тестове. В тази схема Яков Бернули доказва сближаването на емпиричната до неизвестната дедуктивна вероятност с нарастване броя на експериментите, твърдение известно сега като „Закон за големите числа“. В тази схема може да се покаже и възможността за безкрайно множество от резултати от експеримента с разумно и резонно определени вероятности върху примера на геометричното разпределение.

Ще илюстрираме възможно представяне на идеите за статистически изводи за вероятността на единичен опит при схемата на биомно разпределение (или дела на благоприятните случаи). Нека си представим следната ситуация: Лице се представя за екстрасенс, който може да познава дали кутия е празна или в нея има скрита бележка.

Нека приемем, че той е прав дори да греша в един от пет случая и се е съгласил да се проведе експеримент с последователно показване на 20 еднотипни кутии. Експериментът трябва да се проведе така, че всеки път отговорът да не зависи от

предишните опити.

Нека експериментът е проведен и резултатът е k правилни отговора. Приемлив модел на експеримента е, че броят на правилните отговори X е случаен и има биномно разпределение с параметри $n = 20$ и $p = 0.8$. Непосредствено се проверява (аналитично или числено с функцията BINOMDIST в електронна таблица), че най-вероятната стойност за X е 16 и вероятността ѝ е 0.218. Почти толкова са вероятни и съседните стойности, т.е. вероятността за стойност 17 е .205 и за 15 е 0.175. Сумата на трите вероятности е 0.598 и следователно вероятността сумарно за останалите стойности е 0.402 (което не е малко).

Въпросът е за кои стойности на резултата k можем да твърдим, че полученото от експеримента значително се различава от първоначалното ни предположение (познава в 1 от 5 опита). Отговорът се определя от вероятността да бъде получен резултат X , който е същият и дори по-противоречащ на предположението: $X \geq k$ (в конкретния пример). Колкото по-малка е тази вероятност, толкова сме по-сигурни, че се наблюдава малко вероятно събитие, от което следва, че разликата с предположението ни, е може би значителна. Обикновено се фиксира предварително някакво малко число, например 5%, което е допустимата стойност α за значимостта на различието.

Доверителна (горна) граница за вероятността p (с която екстрасенсът познава) може да се определи със стойността, над която различието би било значимо (за фиксираната стойност на значимостта $\alpha = 5\%$). Нивото на доверие е вероятността, допълваща до 1 фиксираното число α .

4.3. Нормално разпределение, модел на данни и изводи за средната.

Естественото въвеждане на нормалното разпределение е чрез близостта му до биномното при нарастване броя на експериментите. Равномерното разпределение също показва, че непрекъснатите разпределения са полезно и удобно приближение на дискретните, когато те са с „много и сравнително гъсти“ стойности. Основание за въвеждането на последното е използването на функцията RAND за генериране на (псевдо-)случайни числа, при създаване на виртуални (изкуствени) данни за разнообразни илюстрации на идеи и методи – например Централната гранична теорема.

Аналогично на изложеното за биномното разпределение и приложението му като модел на наблюдения може да се илюстрира приложението на нормалното разпределение за данни за количествена променлива, средната им аритметична стойност и математическото очакване.

4.4. Корелация и регресия.

Регресионният модел отразява зависимостта между количествени променливи. Приложението му може да се илюстрира с редица опити от физика, химия, биология, в които известна или предполагаема зависимост трябва да бъде уточнена с експеримент, при който неминуемо присъстват грешки при измерването. Многобройни са и случаите при изследвания на количествени променливи в естествените и хуманитарните дисциплини, при които не е възможно да се определи точна зависимост, но съществува обща тенденция, около която гравитират отделните наблюдения.

Идеите за корелационна зависимост и определяне уравнението на регресионна права могат да се илюстрират върху данни за две количествени променливи. Това включва определяне на зависима (отклик) и независима променлива (предиктор),

съставяне и решаване на „нормалната система“ от две уравнения с две неизвестни. Изчисляването на остатъците и анализ на разпределението им с подходящи графики и диаграми, откриването на силно различаващи се наблюдения (аутлаери) и отстраняването им е необходимо преди да се използва полученият модел за прогнозиране на отклика.

4.5. Елементи от актюерната математика. Полезни знания от ежедневието и традиционно свързани с вероятностите са актюерните изчисления. Знания и задачи от проста и сложна лихва обикновено са включени в основното съдържание на гимназиалния курс. Акцентът тук е върху финансови отношения, свързани с продължителността на човешкия живот. Това са задачите за определяне на цената на една доживотна рента, каквато е пенсията в известен смисъл и различните варианти на застраховки „Живот“.

Естествено гимназиален курс не може да съдържа съвременните модели от „висшата финансова математика“, които се прилагат в актюерството. Целта е начално запознаване и представяне на стохастичния характер на финансовите процеси, което ще е полезно особено за насочилите към актюерската професия, за чието търсене се очаква значителен ръст.

4.6. Възможности за междупредметни връзки. Силно препоръчително е знанията от областта на вероятностите и статистиката да се представят и илюстрират със средствата на съвременните информационни технологии. Подходящ за целта е широко разпространен софтуер от типа електронни таблици (MS Excel, Open Office Calc, Google Spreadsheet Document), какъвто традиционно се изучава в дисциплината Информационни технологии.

Множеството стандартни вградени команди (функции) и графични възможности могат да се използват при въвеждане на данни от наблюдения и експерименти в табличен вид; генериране на изкуствени данни – категорни данни с желано разпределение, като се използват команди RAND, RANDBETWEEN, както и споменатите по-горе BINOMDIST, NORMDIST, и други от групата „Статистически“ (Statistical), както и команди (функции) от групата „Финансови“ (Financial).

Придобитите знания в областта намират приложение в решаването на конкретни практически задачи от всички области на естествените, социалните и хуманитарните науки. Подходяща форма за демонстрирането им е подготовката на проекти, стимулиращи и работата в екип. Резултатите биха могли да намерят реални приложения и признание в регулярни междуучилищни и международни съревнования, като организирани от ISLP [5].

5. Заключение. Вероятностите и особено анализът на данни или статистиката са от областта на сложната математика или „физико-математиката“ по [6]. В съвременната терминология това са трудните „задачи с думи“. Убедени сме, че именно тези задачи са от жизнена важност в ежедневната практика за вземане на обосновани решения.

ЛИТЕРАТУРА

- [1] ЯСОВ BERNOULLI. Ars Conjectandi. Basel, 1713. В: Бернули, Лаплас, Колмогоров, Вероятности, (ред. Б.Пенков) Наука и изкуство, София, 1982.

- [2] ПЛ. МАТЕЕВ, М. СЛАВЧОВА–БОЖКОВА, СН. МАТАКИЕВА. Дискусия „Обучението по комбинаторика и статистика – реалност и перспективи“. *Математика и математическо образование*, **41** (2012), 118–119.
- [3] D. MUMFORD. The Dawning of the Age of Stochasticity. In: *Mathematics: Frontiers and Perspectives*, ed. V. I. Arnol'd et al., Amer. Math. Soc., 2000, 197–218.
- [4] HAL VARIAN. Hal Varian on how the Web challenges managers, McKinsey Quarterly, January 2009, <https://www.mckinseyquarterly.com> (11.11.2012).
- [5] International Statistical Literacy Project (ISLP), <https://www.stat.auckland.ac.nz/~iase/islp/> (11.11.2012).
- [6] Ж. Ф. ОЛИВИЕ. Геометрия, праволинейна тригонометрия и статика. Русе, Н. Петров, 1871.

Пламен Матеев
 Факултет по математика и информатика
 Софийски университет „Св. Кл. Охридски“
 бул. Дж. Баучер № 5
 1164 София
 и
 Институт по математика и информатика
 Българска академия на науките
 ул. Акад. Г. Бончев, бл. 8
 1113 София
 e-mail: pmat@fmi.uni-sofia.bg

Марусия Славчова–Божкова
 Факултет по математика и информатика
 Софийски университет „Св. Кл. Охридски“
 бул. Дж. Баучер № 5
 1164 София
 и
 Институт по математика и информатика
 Българска академия на науките
 ул. Акад. Г. Бончев, бл. 8
 1113 София
 e-mail: bojкова@fmi.uni-sofia.bg

ON THE CURRICULUM OF PROBABILITY AND STATISTICS IN SECONDARY EDUCATION

Plamen Mateev, Maroussia Slavchova–Bojkova

Our vision about the educational content in the field of Probability and Statistical data analysis in the junior high school and in the first and secondary stage of secondary–school education is presented for information and discussion.