



ДРУШТВО МАТЕМАТИЧА СРБИЈЕ

АКРЕДИТОВАНИ СЕМИНАР:

345

ДРЖАВНИ СЕМИНАР О НАСТАВИ  
МАТЕМАТИКЕ И РАЧУНАРСТВА  
ДРУШТВА МАТЕМАТИЧАРА СРБИЈЕ

Компетенција: К1

Приоритети: 3

ТЕМА:

**РАД СА ПОДАЦИМА  
ВЕЛИКЕ МОГУЋНОСТИ И ВЕЛИКИ ИЗАЗОВИ  
У МАТЕМАТИЧКОМ ОБРАЗОВАЊУ**

РЕАЛИЗАТОРИ СЕМИНАРА:

**проф. др ЗОРАНА ЛУЖАНИН,  
МАРИЈА РАДОЈИЧИЋ**

БЕОГРАД,  
09. – 10. 02. 2019.

## Садржај

УВОД	1
I део: ПОДАЦИ (СТАТИСТИКА)	2
Шта су то подаци и значај података?	3
Како доћи до података?	6
Сређивање података и графичко представљање	7
Груписање података и табеларни приказ	7
Графичко представљања података	8
Основни показатељи скупа података	11
II део: ШАНСЕ (ВЕРОВАТНОЋА)	13
Појам шансе (вероватноће)	13
III део: ПРИМЕРИ, ЗАДАЦИ...	15
Пратећи задаци	15
Примери из медија	15
Могућности - истраживачки и пројектни задаци	16

## УВОД

Подаци су постали основа за доношење валидних одлука у готово свим областима и околностима. Разумевање и коришћење података постаје једна од кључних компетенција у данашње и будуће време. Зато је *Рад са подацима и информација* нашао место у међупредметним компетенцијама и за основно и за средње образовање. Ученик би требало да ову компетенцију развија кроз целокупно школовања и то, у највећој мери, кроз наставу математике.

*Обрада података* чини једну од пет области у образовим стандардима за крај основног образовања за наставни предмет математике, док су *вероватноћа* и *статистика* једна од четири области у стандардима за крај средњег образовања.

Без обзира што на овај начин делује да је рад са подацима заступљен у знатној мери, реалност је, врло често, сасвим другачија. Као пример наводимо да се садржаји из области вероватноће и статистике у општем средњем образовању реализују на самом крају четвртог разреда и то у знатној мери само информативно. Нажалост, појам вероватноће, односно шанси, се не помиње ни у једном програму наставе за основну школу. Слична ситуација влада и у средњим школама све до завршних часова из предмета математике.

Са друге стране ученици се сусрећу са појмовима из области статистике и вероватноће свакодневно („шанса за оздрављење се смањује“, „ризик се удвостручује кад“, „вероватноћа сваким даном расте“, „подаци показују штетност“, „просечни животни век“,...). Иако интуитивно имају слику шта би поједине синтаксе требало до значе, врло често их оне наводе на погрешне закључаке и неразумевања последица.

У XXI веку израчунавање статистичких показатеља и исцртавање графичког приказа постали су веома једноставни, што повлачи њихову веома честу употребу. Међутим ова употреба, у значајном броју случајева, садржи намерне или ненамерне грешке које могу имати изузетно велик утицај на одлуке појединаца (сетимо се примера са ММР вакцинама). Баш због овог феномена, неопходно је започети или појачати рад кроз формално образовања на разумевању и правилном коришћењу

результата добијених обрадом података. Настава математике је право место за иновације које ће омогућити развој ове изузетно важне компетенције код сваког ученика. Имајући у виду значај разумевања и интерпретације података у различитим сферама друштва као и значај за развој критичког резонувања и доношење одлука, рад са подацима треба да представља саставни део математичке писмености.

Приступ који је представљен у наставку омогућује примену обраде података и разумевање кључних појмова из статистике и вероватноће у целокупном доуниверзитетском школовању. Проблемски и истраживачки задаци се природно намећу као једно од решења. Такође је важно напоменути да, без обзира на разумевање других области математике, ученици врло често лако могу да напредују у раду са подацима. Рад са подацима ствара **нове могућности** за наставника математике јер отвара велики простор за приближавање наставе математике стварним ситуацијама у којима се ученици налазе или ће се налазити. Добар одабир примера и метода рада са великом вероватноћом ће покренути или повећати мотивисаност ученика за наставу математике. Додатна могућност, кроз рад са подацима, јесте опипљива корелација са великим бројем других предмета и области.

Без обзира што програми наставе, из нама непознатих разлога, избегавају употребу речи статистика, овде ћемо обраду података посматрати као део статистике што она формално и јесте.

Материјал је конципиран на следећи начин. Теоријски оквир са кратким напоменама за реализацију наставе (сива поља) се налазе подељени у две целине, подаци и вероватноћа. Трећи део чине примери задатака или пројеката који су доступни у електронској форми. Материјал садржи и велики број линкова на корисне додатне примере или занимљивости.

## I део: ПОДАЦИ (СТАТИСТИКА)

### *Шта је статистика?*

Давне, 1933. године, Н.Г.Веллс (1866 - 1946) је изјавио „*Статистички начин мишљења једног ће дана за свакодневни живот грађана постати једнако неопходан као знање читања и писања.*” (Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.) Тај дан је по свему судећи дошао и зато је потребно код сваког ученика развијати, у највећој мери кроз наставу математике, овај начин мишљења који је саставни део аналитичког и критичког мишљења.

Сложеност саме статистике огледа се и кроз велики број одговора на питање *Шта је статистика?* Наводимо неке од њих.

**Статистика је област математике** која се бави сакупљањем, анализом, интерпретацијом, објашњавањем и презентацијом података.

**Статистика је научна метода** којом се прикупљају, уређују, анализирају и тумаче разноврсни подаци о појавама и процесима у природи и друштву.

**Статистика је врста анализе података**, која укључује планирање и интерпретацију обсервација система, често праћено предвиђањем будућих дешавања засновано на математичком моделу система који се прати.

**Статистика је метод** квантитативног истраживања података.

Данас се статистика промењују у свим областима природних и друштвених наукама као једна од незамењивих и поузданих метода у извођењу прецизних закључака из сакупљених података и ради доношења закључака имајући у виду неизвесност на бази статистичке методологије. Такође је доношење одлуке на свим нивоима, од одлуке појединца, до одлука битних за развој друштва, ваљано само ако се заснива на подацима. Примена модерних рачунара је омогућила извођење статистичких прорачуна великих размера, као и развој нових метода које не би било практично спроводити ручним путем. Ипак, до најбоље одлуке долази се комбинацијом доступних доказа заснованих на подацима и критичког мишљења.

Наводимо неколико чињеница којих морамо бити свесни сваког тренутка, а посебно у избору примера и задатака за ученике:

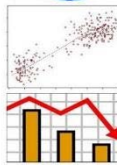
„Једино статистички обрађени подаци могу имати стручну или научну вредност.”

„Статистика може бити само помоћ, али никада замена за здрав разум.”

„Ако неко уз помоћ статистике објави неку лаж, није за то крива статистика: статистика не лаже, него лажу људи.”

# Statistics

“Statistics is the science of collecting, organizing, presenting, analyzing and interpreting numerical data to assist in making more effective decisions.”



<http://marketbusinessnews.com/financial-glossary/what-is-statistics-what-are-statistics/>

## Шта су то подаци и значај података?

### Основни појмови статистике

Предмет испитивања статистике је популација. Популацију чине субјекти, испитаници, случајеви, елементи или појаве... На популацији испитујемо неку карактеристику (особину) коју називамо обележје (променљива, варијабла). Можемо на једној популацији посматрати и више обележја.

**Обележје, варијабла, променљива:** „нешто чије вредности варирају”

**Подаци** су вредности мерења обележја.

пример:

**Старост** испитаника је **33** године.

Старост је обележје, а 33 је податак, односно вредност посматраног обележја.

Испитаник **зна** да **користи рачунар**.

Коришћење рачунара је обележје, а зна је податак.

### Табела (матрица) података

	пол	старост	брачно стање	радни статус	број деце	станује у згради или кући	користи рачунар	висина [cm]	телесна маса [kg]
1.	М	39	у браку	запослен	2	зграда	да	180	75
2.	Ж	33	у браку	незапослен	2	зграда	да	169	60
3.	М	22	самац	запослен	0	зграда	да	190	78
4.	М	69	разведен	пензионер	1	кућа	не	175	78

Колона – променљива, варијабла, обележје,...

Ред (врста) – испитаник, субјект испитивања, експеримент...

Поље (ћелија) – податак (реализована вредност обележја)

Први појам који је неопходан да ученици усвоје јесте **податак** у смислу напред наведене дефиниције. Значи податак је добијен мерењем, експериментом, анкетом и може узимати различите вредности. Такође је код уношења вредности веома важно назначити да за сваки податак мора бити јасно дефинисане вредности које може да узме. Нпр. ако је податак радни статус онда је важно одредити скуп свих могућих реализација. У наведеном примеру скуп могућих реализација је {запослен, незапослен, пензионер}. На овај начин ученици интуитивно усвајају појам обележја и скупа могућих реализација, а податак доживљавају као вредност која се мења. Такође је код бележења података важно напоменути да се у пракси често користе скраћенице (М и Ж за мушки, односно женски пол), или се бројеви користе као замена за речи (1 за мушки пол, а 0 за женски пол). Важно је да ученици разумеју да су у том случају бројеви само ознаке и да служе за лакше баратање подацима, али се никако не могу третирали као бројеви. Као пример података могу се преузети подаци који прате тренутно актуелан догађај, као што је на пример нека кошаркашка утакмица или тениски меч.

Без обзира што ученице не усвајају појам обележја, у примерима и задаци може да се употреби ова реч на интуитивном нивоу.

## Врсте обележја

У претходном примеру учавамо да посматрана обележја могу бити нумеричка (старост, висина, тежина, број деце) или атрибутивна (пол, радни статус, брачно стање...) Нумеричка обележја изражавају се бројем, а атрибутивна описно. Ово је једна од могућих подела обележја. Нумеричка обележја се даље деле на непрекидна и дискретна. Непрекидна обележја су она обележја која могу добити било коју вредност на бројевној оси или неком њеном делу (пример висина). Дискретно обележје може имати само одређене вредности на бројевној оси, односно само вредности из коначног или пребројивог скупа (пример број деце). Подела обележја на нумеричка непрекидна, нумеричка дискретна и атрибутивна, назива се и подела према садржини.

Друга подела обележја врши се на основу примењене мерне скале. Разликовање скале која се користи за мерење реализације обележја је један од најважнијих задатака у правилном одабиру статистичке анализе за обраду.

Приликом мерења најчешће се користе четири мерне скале (нивои мерења), а то су: номинална скала, ординална скала, интервална скала и скала односа.

**Номинална скала** је скуп категорија (вредности, стања) које нису уређене, односно рангиране. Код номиналне скале користе се имена или бројеви (као симболи) за класификацију. На пример, пол испитаника, боја очију, омиљени спорт, крвна група, ...

**Ординална скала** је скуп категорија које су уређене, односно рангиране, али где није познато растојање између категорија и где растојање између две категорије може бити различито од растојања другог пара категорија. Код ове скале мерења вредности се сведе на њихово рангирање према значају и утврђеним критеријумима. Другим речима, бројеви означавају ранг, али не показују величину њиховог односа (нпр. оцена може бити незадовољава, задовољава и истиче се). Често се категоријама у ординалној скали придружују бројеви који показују ранг сваке категорије

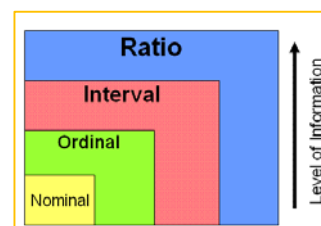
**Интервална скала** поседује јединицу мере, па је растојање између сваке две суседне категорије једнако. Лењир и термометар користе интервалну скалу, лењир милиметре или центиметре, а термометар степене. Примена интервалне скале нам омогућује да упоредимо две реализације.

**Скала односа** (рацио) је интервална скала која има праву (физичку) нулу. Овом скалом постиже се највиши степен мерења. Лењир је пример скале односа, јер 0 cm значи непостојање дужине, односно у питању је реална нула. Из тог разлога можемо рећи да је цвет чија је висина 24 cm, три пута виши од цвета чија је висина 8 cm. Са друге стране, Целзијусова скала јесте интервална скала, али није скала односа. Температура од 0<sup>0</sup> C не значи потпуно одсуство топлоте. Из тих разлога, ако је у једном граду температура 5<sup>0</sup> C, а у другом 10<sup>0</sup> C, не можемо рећи да је у другом граду дупло топлије.

Напоменимо да се прве две скале, номинална и ординална, користе за мерење атрибутивних обележја. Нумеричка обележја се мере на интервалној или скали односа. У пракси се често користи ординална скала која ранговима додељује бројеве (нпр. оцене у школи). Међутим изузетно је важно, у том случају, водити рачуна шта од статистичке анализе можемо, односно смемо да користимо. Неодговарајућа примена статистичког алата може да наведе на погрешне закључке.

Scale	Description	Example
Nominal	Numbers Assigned to Runners	7, 8, 3 (Finish)
Ordinal	Rank Order of Winners	Third place, Second place, First place (Finish)
Interval	Performance Rating on a 0 to 10 Scale	8.2, 9.1, 9.6
Ratio	Time to Finish, in	15.2, 14.1, 13.4

Слика 1.<sup>1</sup>



Слика 2.<sup>2</sup>

<sup>1</sup> <https://www.slideshare.net/AhsanKhanEco/business-research-methods-measurement-questionnaire-and-sampling>

<sup>2</sup> <http://www.six-sigma-material.com/Data-Classification.html>

Сваки задатак који садржи податке омогућује прожирење, односно додатни осврт који се односи на мерење (добијање) података. Ученици би требали да разумеју да се сваки подаци добијају мерењем, односно упоређивањем са скупом могућих реализација.

### Етапе статистичког проучавања

Свако статистичко проучавање почиње тек када се дефинисани циљ и сврха истраживања или анализе. На слици 3 дат је шематски приказ статистичког проучавања

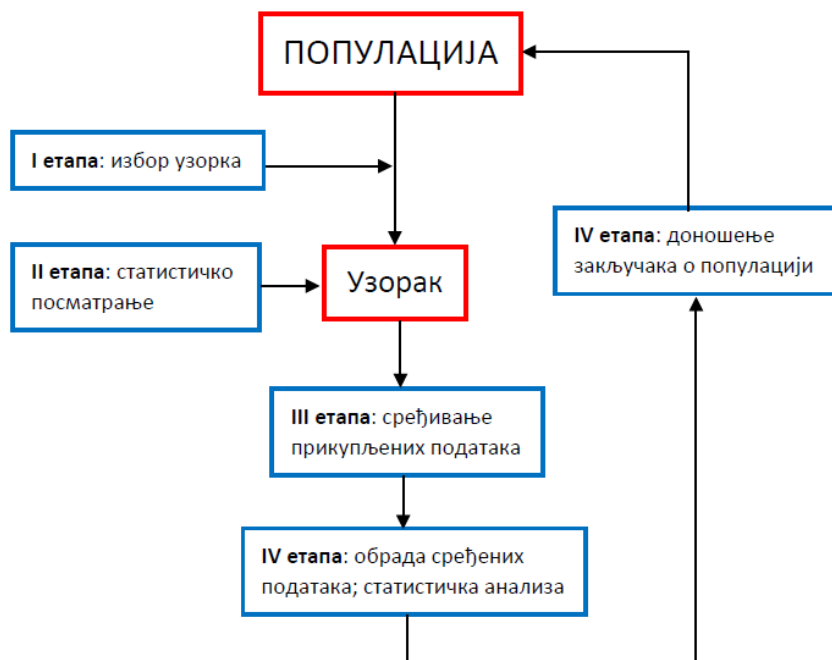
Очигледно је готово немогуће да посматрано обележје региструјемо код сваког елемента популације. Физички је неизводљиво измерити приходе сваког домаћинства или број прочитаних књига сваког осамнаестогодишњака у Србији... Такође је немогуће применити нови лек против високог крвног притиска на сваком пацијенту са повишеним притиском.

Прва етапа у статистичком проучавању представља избор узорка над којим ће бити извршено посматрање, односно прикупљање података. То значи да из популације издвојимо  $n$  елемената (јединки, статистичких јединица) које чине узорак. Број  $n$  називамо обим узорка. Ово је један од најтежих и најкомплекснијих проблема, јер на основу добијених резултата на узорку доносимо закључак о целој популацији.

Друга етапа састоји се у статистичком посматрању изабраног узорка. У овој етапи врше се анкетирања или прикупљање података на терену или неки други вид мерења над елементима узорка у циљу прикупљања потребних података.

У трећој етапи групишемо и сређујемо податке добијене при статистичком посматрању. У случају када узорак није репрезентативан, ово је последња етапа која се реализује.

Четврта етапа се састоји из обраде сређених података и стручне анализе података. У овој етапи применом сложеног статистичког апарата доносимо закључке. Резултати добијени на основу обраде узорка придружују се целој популацији. Уопштавање резултата испитивања узорка на целу популацију из које је узет узорак, могуће је извести, јер постоји тесна повезаност између узорка и популације. Та повезаност произилази из чињенице да је расподела посматраног обележја елемената репрезентативног узорка приближна расподели истог обележја на елементима популације.



Слика 3

Ученици у току школовања парцијално сређу поједине сегменте статистичког проучавања, а да до краја средње школе не добију целовити увид у овај поступак. Међутим, врло једноставно се, већ ученицима седмог и осмог разреда, могу приближити описане етапе.

## Како доћи до података?

Више се не доводи у сумњу чињеница да се закључци и одлуке донете на основу релевантних података и њиховој анализи могу сматрати објективним и ефикасним. Међутим, ова чињеница има и свој значајан недостатак. Ако подаци „нису добри” закључци или одлуке ће имати исту особину. Из тих разлога се прикупљању података мора посветити велика пажња.

Коришћење експеримента и узорка јесу најпознатије методе за добијање података о некој појави или популацији. Са појмом експеримента ученици се срећу у другим предметима, пре свега у физици, хемији и биологији.

Настава математике се лако може обогатити вршењем експеримената као што су бацање коцкице, новчића, окретање спинера или извлачење куглице. Тренутно готово да не постоје примери извођења оваквих експеримената. Сам експеримент се прескаче или се очекује да их учених реализује само „мисаоно“ тј. да их само замисли. У многим земљама експериментална вероватноћа представља важну компоненту у процесу учења, посебно када је реч о увођењу концепта из вероватноће. Спровођење експеримената захтева активност ученика и тако их укључује у процес учења и доприноси разумевању шанси и исхода. Такође овакав вид учења мотивише ученике да изводе закључке и критички резонују. Укључивање експеримента у наставу математике (редовној, додатној или математичкој секцији) омогућило би боље усвајање појмова вероватноће, независности догађаја, условности, закона великих бројева и слично. Све ове појмове ученици би упознали пре свега искуствено и на нивоу интуиције, али би им такав приступ омогућио много боље разумевање података и донетих закључака.

Чак и кад смо у прилици да испитујемо целу популацију, често се одлучујемо да селекујемо само њен мањи део (узорак) и да на основу њега доносимо одлуке или закључке. Постоје многобројни разлози за то. Узорковање може обезбедити поуздане информације за много мање новца од пописа (напомена, под пописом се подразумева прикупљање података над целом популацијом). Подаци добијени на основу узорка се брже и једноставније прикупљају.

У настави математике су често присутни задаци и проблеми у којима се појављује узорак, тј. подаци добијени из узорка. Међутим, лако се у наставу могу укључити и задаци у којима је потребно да ученици изаберу узорак и посматрају неку појаву. Најједноставнији задаци могу бити проширени тако што би ученици прво прикупили податке, па их онда обрадили. Погледајмо следећи пример.

*Од 100 анкетираних грађана 50 је изјавило да је годишњи одмор провело на мору, 17 на планини, а 33 код куће. Представи на кругу резултате овог истраживања, као што је започето...* (ЗБИРКА ЗАДАТАКА ИЗ МАТЕМАТИКЕ ЗА ЗАВРШНИ ИСПИТ У ОСНОВНОМ ОБРАЗОВАЊУ И ВАСПИТАЊУ)

Код прикупљања података корисно је ученике поделити у групе и поставити захтев да свака група анкетира 50 грађана. На тај начин би свака група добила различите податке и графиконе, што пружа додатну могућност да ученици увиде неке особине узорка. Пре самог истраживања потребно је ученике припремити за истраживачке задатке. Такође пожељно је организовати кратка излагања ученичких радова како би се сви упознали са проблемом, а и како би се ученици навикли да износе своје мишљење и презентују идеје пред другима. Потребно је заједно са ученицима евалуирати ученичке радове и дати смернице за унапређење истих.

Званична статистика обезбеђује, на непристрасним основама, бројчане и репрезентативне податке и информације о масовним економским, демографским и друштвеним појавама и о појавама из области радне и животне средине, и то за све кориснике: привредне субјекте и њихова удружења, државне органе, органе аутономних покрајина и органе јединица локалне самоуправе, културне, образовне и научне институције, као и за најширу јавност.

У припрема проблемских задатака важно је укључивати и податке добијене кроз званичну статистику. Највећи произвођач званичне статистике у Републици Србији је Републички завод за статистику <http://www.stat.gov.rs/>

Коришћењем на овај начин доступних података настава математике може додатно допринети развоју свести код ученика о важним показатељима за цело друштво, односно државу. Пример публикације са подацима о демографским показатељима <http://publikacije.stat.gov.rs/G2018/Pdf/G201814015.pdf>

Сваким даном број извора података расте и зато је веома важно усмеравати ученике на различите изворе података са обавезним освртом на поузданост извора. Као примери, поред Републичког завода за статистику, могу се користити сајтови званичних спортских организација (пример: светске тениске асоцијације, <https://www.atptour.com/en/stats>), сајтови који обухватају временске податке (пример: Републички хидрометеоролошки завод, <http://hidmet.gov.rs/>), итд.

Посебан осврт потребно је направити на поузданост података и поузданост различитих истраживања. Неретко се у медијима може чути: „На основу једне америчке студије...“, „Доказано клиничком студијом...(ситним словима пише на 13 испитаника)“, „Научно је доказано...(па нису дати извори)“, „Три месеца коришћења креме, умањује 2 године изгледа бора, а 6 месеци 7 година изгледа бора (измерио независни дерматолог)“. Све су то информације са којима се свакодневно сусрећемо. Корисно је развијати код ученика критички став према истраживањима. Усмерити ученике на битне сегменте истраживања како би могли да оцене поузданост и валидност истраживања. За вежбу је пожељно користити актуелна истраживања и резултате доступне у медијима и свакодневном животу.

## Сређивање података и графичко представљање

Када су подаци прикупљени потребно је приступити сређивању прикупљеног материјала. Сређивање „сирових“ података прикупљених при статистичком посматрању има за циљ да олакша упоређивање, да истакне најзначајније карактеристике тако да буду лако уочљиве и да омогући статистичку анализу.

Резултат ове етапе статистичког анализе је добијање статистичких табела, основних графичких приказа и основних статистичких показатеља. Бројеви који произилазе као резултат сређивања имају саме по себи велики практични значај. У неким истраживањима погодна класификација и табелирање података могу да дају тако јасну слику о посматраном обележју да даљи рад није ни потребан. У другим случајевима сређивање података није довољно и потребно је применити статистичку анализу. Ипак, сваки озбиљан истраживач, односно аналитичар ће сваки закључак који се евентуално намеће на основу увида у податке, верификовати или одбацити применом статистичке анализе.

У поступку сређивања података изузетно је важно проверити квалитет податка што подразумева анализу евентуалних „чудних“ или „сумњивих“ реализација. Постоје статистички поступци који могу да укажу на овакве реализације.

## Груписање података и табеларни приказ

Груписање (класификација) података подразумева разврставање прикупљених података у групе или класе према одговарајућим вредностима обележја реализованих на елементима узорка. У овој фази статистичког истраживања прикупљени статистички материјал претвара се у бројчане информације о посматраном скупу формирањем статистичких серија и табела. Сређивање статистичких података представља уређивање података по сваком посматраном обележју. У ту сврху користи се поступак груписања што представља рашчлањивања скупа од  $n$  података у  $k$  подскупова који се међусобно не преклапају.

Статистичке серије представљају низове података сређених по вредностима обележја (које називамо модалитетима обележја) или хронологији. У зависности од начина формирања и садржаја, оне се деле на две групе: *серије структуре* (ако су уређене по обележју) и *временске серије* (груписање података по моментима времена и по интервалима времена).

У зависности од тога колико серија садржи и како су оне дате и међусобно комбиноване, имамо:

- статистичке табеле – „класичне“ статистичке табела у којима су статистичке јединице (елементи узорка) разврстане према једном обележју;
- таблице контингенције – статистичке табеле за приказивање статистичких података када су статистичке јединице групуране истовремено према више од једног обележја

У току школовања ученици се срећу са великим бројем података који су већ сређени у табели. Међутим, врло је корисно да ученици овладају вештином и правилним табеларним приказом. Уз то ученике треба упутити и на одговарајући софтвер за чување, приказ и обраду података.

статистика Коби Брајана (временска серија): <https://www.basketball-reference.com/players/b/bryanko01.html>



## Графичко представљања података

### Основни принципи графичког представљања

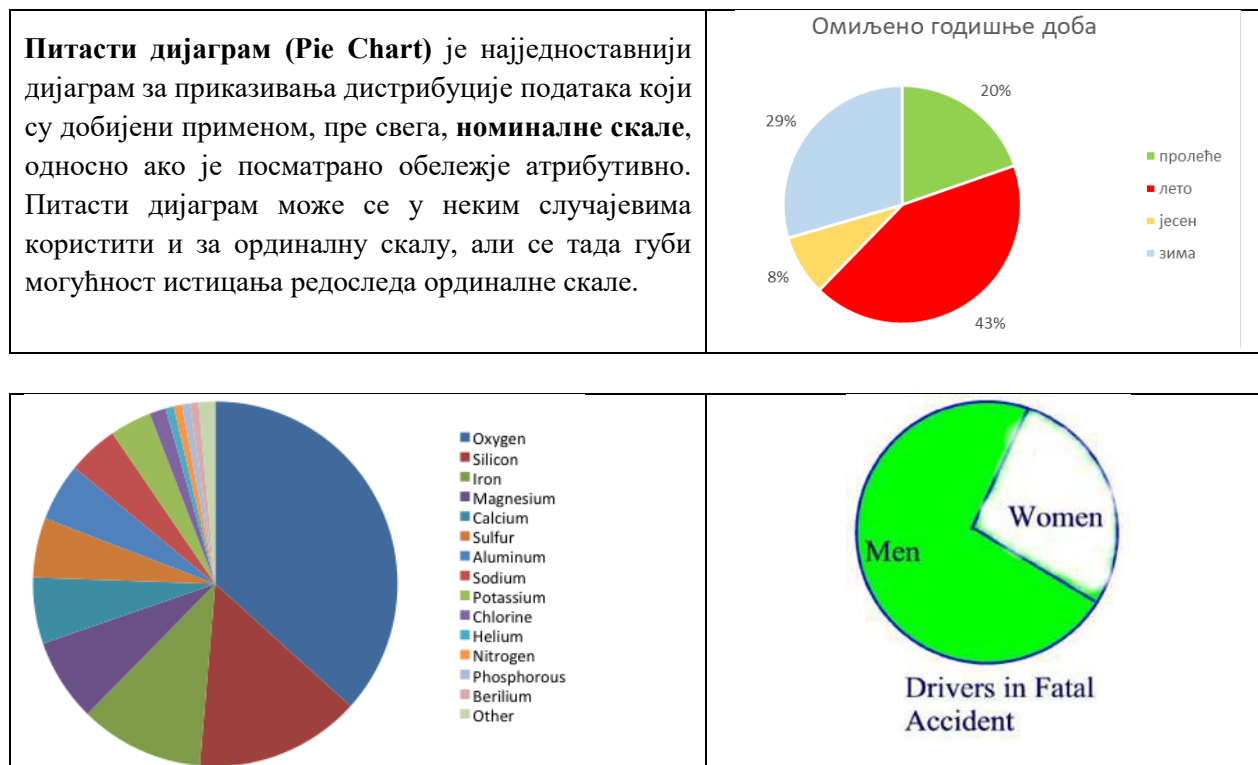
Пре свега напоменимо неколико битних чињеница о графичком представљању података:

- Графички приказ може да буде добар само колико су добри подаци које се приказују.
- Немогуће је направити добар графички приказ из проблематичних података.
- Графикони никако не би смели да приказују искривљену слику о подацима која је на њима представљена. Појава приказивања искривљене слике позната је као дисторзија. Дисторзија може бити намерна или случајна.
- Графички прикази не би требало да буду сложенији од података који се на њима приказују. Ако је „прича” сложена, потрудите се да на графику буде једноставнија.
- Непотребна сложеност може настати због непотребне декорације, као што је употреба боја или 3Д ефеката. У жаргону су такви графикони познати као "chartjunk.". Ако је „прича” једноставна, нека и графикон остане такав.

### Најчешће графички прикази (врсте графичког представљања)

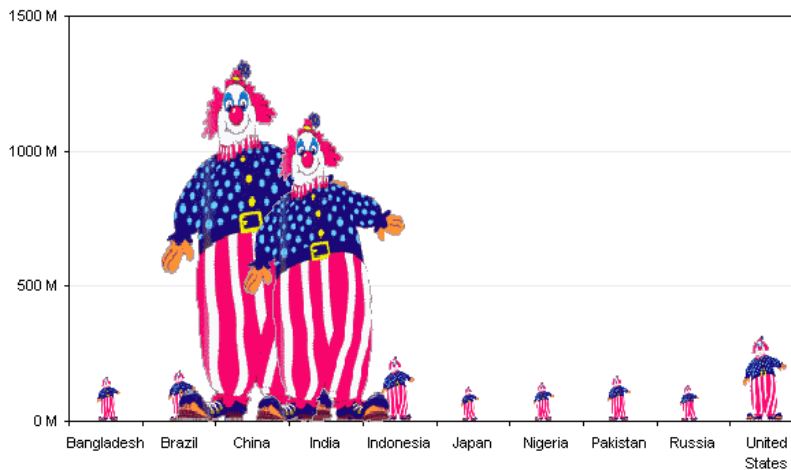
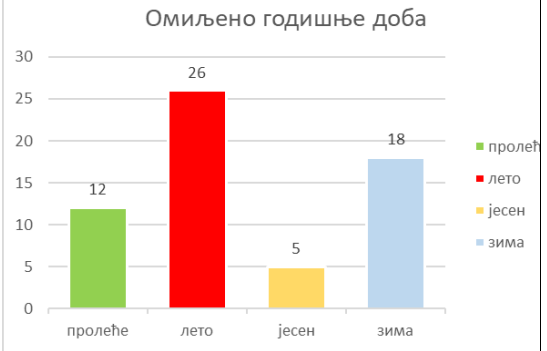
Статистички подаци приказују се графички у виду одговарајућих геометријских облика, ознака на географским картама или помоћу сликовито популарних приказа (слика и фигура). Примена рачунара и готових статистичких програма за графичко приказивање пружа мноштво погодних начина да се статистички подаци прикажу применом различитих геометријских облика и у различитим димензијама. На тај начин омогућава се стицање правих информација о нивоу одређених појава, њиховој структури или променама у времену и простору.

У наставку наводимо најчешће коришћене графичке приказе. Важно је напоменути да и приликом графичког представљања података не смемо да заборавимо којој врсти обележја припадају, атрибутивном или нумеричком, и на којој скали мерења су прикупљени подаци.



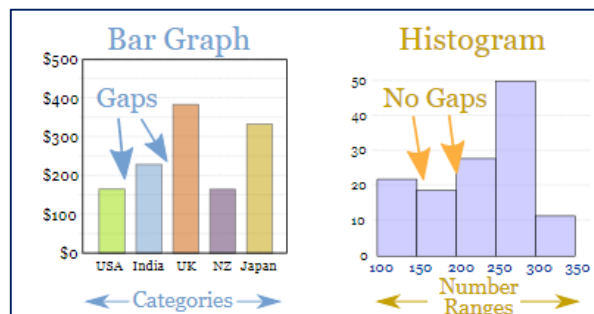
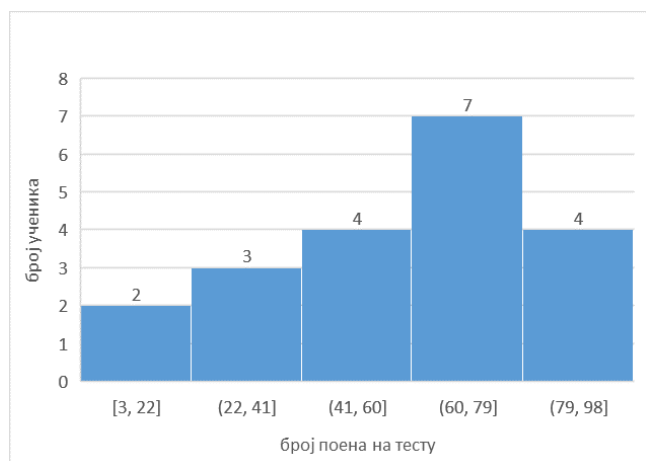
Слика 4: Погрешна употреба питастог дијаграма

**Стубичасти дијаграм** служи за представљање фреквенција обележја мерених, пре свега на номиналној или ординалној скали. Односно, погодни су за коришћење када је посматрано обележје атрибутивно или дискретно нумеричко. Код стубичастиг дијаграма висина „стубића“ одговара апсолутној или релативној фреквенцији појаве модалитета. **Стубиће се по правилу цртају раздвојено.**



Слика 5: Погрешна употреба стубичастиг дијаграма

**Хистограм** служи за приказивање података добијени на интервалној скали или скали односа, односно када је посматрано обележје непрекидно нумеричко. Хистограм фреквенција се састоји из низа повезаних правоугаоника. Овај дијаграм конструише се тако што се на апсиси наносе величине интервала (које би требало да буду једнаке) модалитета обележја, а на ординати вредности апсолутних или релативних фреквенција. Основу сваког правоугаоника представља величина интервала модалитета, а висину одговарајућа фреквенција. Појам хистограма је уско повезан са појмовима **полигон фреквенција** и **крива фреквенција**.

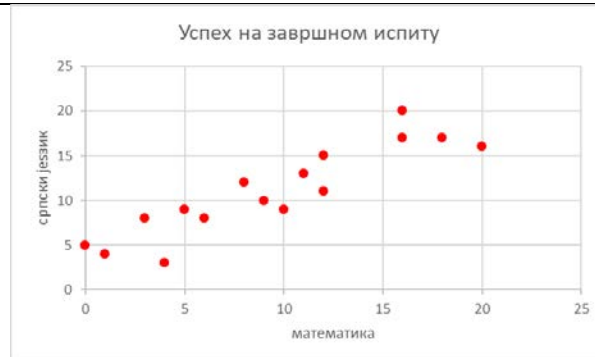


Слика 5: Разлика између стубичастиг дијаграма и хистограма

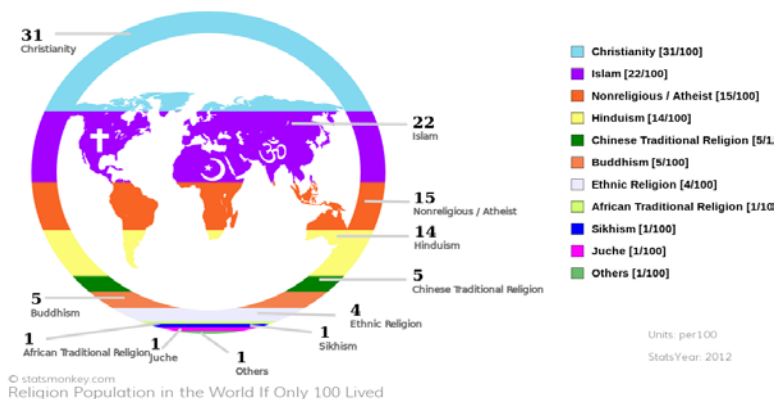
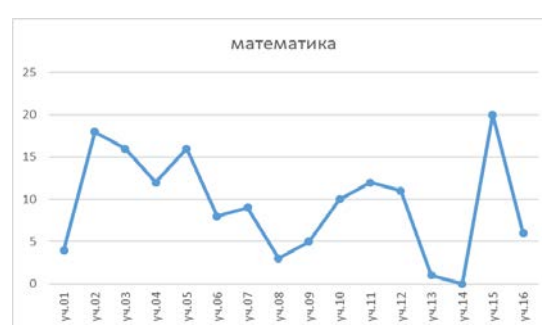
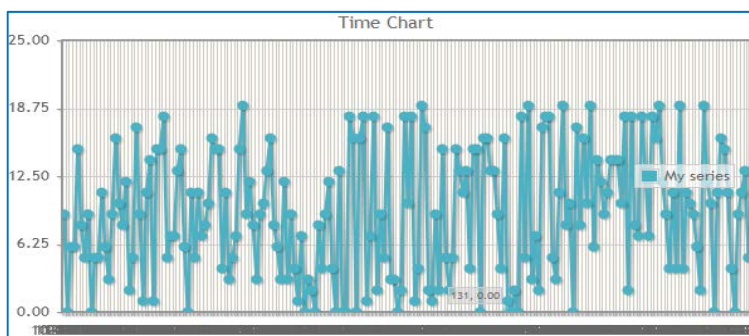
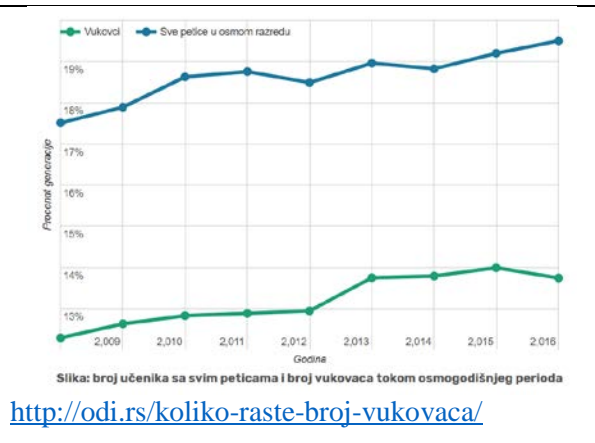
**Дијаграм тачака** једног обележја је у основи сличан стубичастом дијаграму. У овом случају сваком елементу узорка придружујемо једну тачку која се уцртава изнад одговарајуће вредности. Овај дијаграм је погодан за ицртавање мањег скупа података.



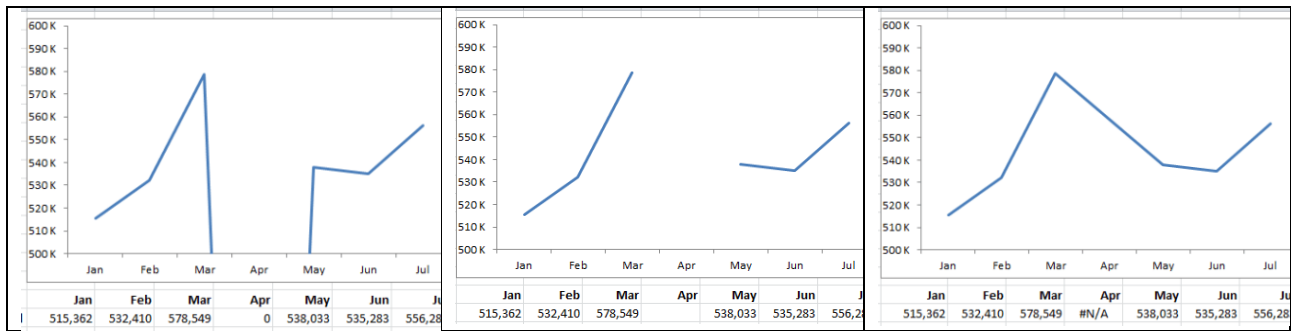
**Тачкасти дијаграм** или **скатерплот** служи за представљање димензионалних података и може бити од велике користи када је потребно установити да ли између два обележја постоји зависност или не, као и облик зависности ако постоји.



**Времески дијаграм** служи за представљање временских серија. На хоризантолној оси налази се време, док се на вертикалној оси налази вредност посматраног обележја. Уцртавањем тачака  $(x_i, y_i)$  и спајањем правом линије две хронолошки узастопне вредности добијамо дијаграм. Времески дијаграм омогућава сагледавање апсолутних промена нивоа појаве током времена. Овај дијаграм може да се користи и за упоредно приказивање две и више појава које су изражене у истим јединицама мере под условом да те појаве имају приближно једнаке почетне вредности.



Слика 6: Лоши примери графичког представљања (непрегледни и погрешни)



Слика 7: Графици са недостајућим подацима

Примери различитих графичких приказа

<https://ourworldindata.org/happiness-and-life-satisfaction> (добри и лоши)

<http://old.briarcliff.edu/departments/cis/excel/2013/006-charts-bad%20examples.htm> (лоши)

## Основни показатељи скупа података

### (Узорачке) мере центрираности

Свакодневно се сусрећемо са појмом просек или просечна вредност. Под тим се обично подразумева нека средња вредност која најбоље представља појаву (обележје) коју посматрамо. Често се у пракси тражи да одредимо само неке карактеристике појаве (обележја које посматрамо). Уколико посматрамо висину стогодишњих храстова, некада нам је довољно да одредимо само један број који ће, у неком смислу, добро представљати посматрано обележје. Бројеве који на неки начин презентују посматрано обележје називамо **параметрима обележја**. Параметри који се најчешће одређују су параметри који описују просечну или најчешће очекивану вредност посматраног обележја. Овакви параметри називају се **мере центрираности обележја**.

За меру средње вредности најчешће се користи аритметичка средина, што није увек најбољи избор. Тако рецимо када се представља висина плате у Србији најчешће се говори о просечној плати, мада би било корисније да сазнамо и медијану и моду за ширу слику. Или пак када се раде анализе о величини стамбеног простора грађана, колико нам говори просечна вредност површине стамбених јединица, а колико мода површине стамбених јединица. Имајући у виду значај средње вредности аритметичка средина, мода, медијана се изучавају у осмом разреду основне школе, а аритметичка средина се уводи у петом разреду.

Ученици углавном немају проблем да спроведу рачун и одреде аритметичку средину док се јављају проблеми када треба да тумаче исту и примењују у реалним животним ситуацијама. Да проблем не подразумевања аритметичке средине није само везан за ученике говоре и поједини текстови у новинама. Па тако можемо прочитати да **просечна жена рађа са...** Шта је заправо просечна жена? Како смо добили просечну жену? Још један од новинских чланака везан је за бракове говори да се најчешће жене венчавају са 28 година, а заправо реч је о просечној старости жена када први пут ступају у брак, док о броју жена које ступају у брак са 28 година не можемо много рећи. Такође ових дана актуелна тема је и коликом брзином треба возити другу половину пута ако смо прву возили 150km/h тако да на крају пута просечна брзина буде 130km/h. Неретко се може чути како је потребно да се друга половина вози 90km/h како би просечна брзина била 130km/h. А колика је хармонијска средина? И ако можда мало чудно али истинито је да више од 99% људи на планети има надпросечан број ногу. Много је оваквих примера у којима се погрешно интерпретира одабрана средња вредност или напосто та вредност не говори много о подацима.

У овом делу навешћемо три функције узорка помоћу којих можемо на основу реализованог узорка (података) оценити мере центрираности обележја. Коју ћемо функцију применити, највише зависи од типа обележја одређеног на основу мерне скале.

**Аритметичка средина.** Нека је дат низ података, тј. нека је дат узорак узорак  $x_1, x_2, \dots, x_n$  нумеричког обележја  $X$ . Аритметичка средина или средња вредност података (узорка) је

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Аритметичка средина сматра се најважнијом мером неког обележја и има велики значај за сложенију статистичку анализу и доношење статистичких закључака. За аритметичку средину важе следеће особине:

- 1) Аритметичка средина се увек налази између најмање и највеће вредности.
- 2) Ако посматрамо два узорка обима  $n_1$  и  $n_2$  и на оба узорка посматрамо исто обележје, тада ако је  $\bar{x}_1$  аритметичка средина првог узорка, а  $\bar{x}_2$  аритметичка средина другог узорка тада се аритметичка средина узорка добијеног спајањем оба узорка може добити коришћењем формуле

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

- 3) Збир одступања реализација у узорку од средње вредности узорка једнак је нули.
- 4) На вредност аритметичке средине велики утицај имају удаљене вредности.

Аритметичка средина има смисла за податке добијене применом интервалне скале или скале односа. У случају примене ординалне скале, ако су категорије одређене бројевима, може се израчунати аритметичка средина, али требало би бити веома пажљив при њеном тумачењу.

Како се за сваку реализацију узорка добија конкретан број за  $\bar{x}$  и како различити реализовани узорци најчешће дају различите аритметичке средине, ученици би требало да интуитивно прихвате стохастичку природу различитих појава, односно да разумеју зашто се не добијају исте вредности. Овде је прилика за наставника да кроз различите примере „наметне“ ученицима идеју да би већина узорка требало да има „сличну“ вредност јер су подаци узети из исте популације.

**Медијана.** Нека су реализације у узорку уређене по растућим вредностима

$$x_1 \leq x_2 \leq \dots \leq x_n$$

Медијана,  $x_{med}$ , је вредност обележја (податак) које се налази на средишњем месту у узорку уређеном по растућим вредностима, односно медијана дели низ датих бројева на два једнака дела у односу на број узорачких вредности (број података). Медијану можемо срачунати за сва обележја која користе ординалну скалу, интервалну скалу или скалу односа. У случају када је број елемената у узорку паран, може да се деси да медијана не буде једнака ни једној од вредности из узорка јер се тада за медија узима аритметичка средина средишње две вредности. Геометријски, медијана је тачка  $x_{med}$  на  $x$ -оси са особином да права  $x = x_{med}$  дели хистограм на два дела једнаких површина.

**Модус (мода),**  $x_{mod}$ , је она вредност обележја која се најчешће појављује у узорку. Модус не мора бити једнозначно одређен, тј. у узорку може постојати више вредности обележја чије су фреквенције максималне. Модус се може одредити за сваки узорак без обзира која је мерна скала коришћена. Међутим, модус код непрекидних нумеричких обележја мерених помоћу интервалне или скале односа нема сврхе израчунавати због постојања других, јачих показатеља центрираности. Геометријски, модус је вредност на  $x$ -оси за коју полигон фреквенција достиже максимум.

**Min-max средина** је једноставна мера центрираности која представља аритметичку средину највеће и најмање вредности у узорку (у подацима)

$$x_{minmax} = \frac{x_{min} + x_{max}}{2}$$

Нека је дат низ података, тј. нека је дат узорак узорак  $x_1, x_2, \dots, x_n$  нумеричког обележја  $X$  чије су све вредности позитивне. **Геометријска средина** таквих података (узорка) је број

$$g_m = \sqrt[n]{x_1 x_2 \cdots x_n}$$

Геометријска средина се често користи када се упоређују различите ставке - проналажење једне „вредности“ за ове ставке - када свака ставка има више својстава која имају различите нумеричке опсеге. Примери геометријске средине: <https://www.mathsisfun.com/numbers/geometric-mean.html>, <https://towardsdatascience.com/on-average-youre-using-the-wrong-average-geometric-harmonic-means-in-data-analysis-2a703e21ea0>

**Хармонијска средина** података (узорка) који су позитивни јесте број

$$h_m = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

За аритметичку, геометријску и хармонијску средину важи релација:

$$h_m \leq g_m \leq \bar{x}$$

Пример:

	подаци (узорци, обима 2)		
	(1,1000)	(10,100)	(20,20)
аритметичка средина	$\frac{1+1000}{2} = 500.5$	55	20
геометријска средина	$\sqrt{1 \cdot 1000} = 31.62$	31.62	20
хармонијска средина	$\frac{2}{\frac{1}{1} + \frac{1}{1000}} = 2.00$	18.18	20

Аритметичка средина и друге мере центрираности често нису довољне за осликавање посматраног обележја. У сваком узорку уочава се разлика појединачне вредности од аритметичке средине или друге мере центрираности. Те уочене разлике у узорку потребно је на неки начин измерити. За то служе **мере растурања** које се још зову и **мере одступања** или **мере варијације**. Код ових мера пожељно је да њихова вредност зависи од свих вредности реализованог узорка, да има неко конкретно значење са јасним својствима и да су погодне за даља израчунавања у статистичкој анализи.

**Ранг** је најједноставнија мера одступања или варијације који се дефинише као разлика највеће и најмање вредности у узорку,

$$rang = x_{max} - x_{min}.$$

Основни недостатак ранга је што не зависи од свих вредности из узорка, већ само од екстремних вредности. Често те екстремне вредности могу бити изузетне и веома удаљене од највећег дела реализованих вредности.

У циљу бољег разумевања, а а самим тиме, и трајнијег знања, било би веома корисно да ученици на истим подацима увежбавају графичко представљање и израчунавање мера центрираности. То подразумева да се ученицима обезбеде посебни материјали (радни листове) које би попуњавали у више наврата.

## II део: ШАНСЕ (ВЕРОВАТНОЋА)

### Појам шансе (вероватноће)

Појам вероватноће јесте веома комплексан појам и спада у садржаје више математике. Међутим, изучавање вероватноће је могуће на различитим нивоима образовања и са децом различитих могућности, а резултати различитих студија показују да изучавање вероватноће има позитиван ефекат

и на математичко резоновање (Farnworth, 1991; Fischbein & Schnarch, 1997; Freudenthal, 1973; Gardner, 1989; Jones, 1995; Koshy, Ernest & Casey, 1999; Shaughnessy, 1992; Sobel & Maletsky, 1988, Tsakiridou, H., & Vavyla, E. 2015)<sup>3</sup>. Интуитивно усвајање концепта вероватноће могуће је започети врло рано. Већ ученици нижих разреда основне школе могу разликовати шта значи да нешто има веће или мање шансе. Многа истраживања попут (Tsakiridou, H., & Vavyla, E. 2015, Fischbein et al. 1984, Davies 1965)<sup>3</sup> сведоче о могућностима деце да још у млађим разредима основне школе одреде исходе појединих догађаја и поређају их према вероватноћи, наглашавајући да вокабулар и ситуације морају бити прилагођени узрасту ученика. Правилним одабирима примера у настави математике може се поставити добар темељ за увођење формалне дефиниције вероватноће на крају средњег образовања или у току високог образовања.

Експерименти у настави математике у основној школи, а рачунарске симулације у средњој школи добра су прилика за увођење статистичке дефиниције. Важно је користити примере који су ученицима блиски, као што је шанса да се роди девојчица или дечак, шанса да у игри ЛОТО буде извучена тачно одређена комбинација, шанса да се појави грб приликом бацања новчића, шанса да падне шестика приликом бацања коцкице... Међутим најчешће коришћени примери (бацање новчића, бацање коцкице, извлачење карте...) подлежу дефиницији класичне дефиниције вероватноће, тј. подразумевају да су сви елементарни исходи једнаковероватни. Форсирањем оваквих примера намећемо ученицима хипотезу да када постоји више различитих исхода да су они једнаковероватни. То наравно не мора бити тачно и зато је важно у настави математике кроз целокупно школовање омогућити ученицима да искуствено покушају да дођу до исправних закључака. Такође би статистичка дефиниција вероватноће морала да заузме значајно место и у настави математике с обзиром да њену примену ученици срећу у другим предметима и областима.

Приликом решавања следећег задатка значајан број ученика бира одговор в) уз образложење да постоје само два могућа исхода!?

*У једној компанији 70% радника ради са дипломом инжењера информационих технологија, а 30% радника ради са дипломом правног факултета.*

*На случајан начин одабран је један радник из те компаније. У питању је Милан, млад успешан радник у тој компанији. Врло је амбициозан и перспективан. Воли пливање и редовно вежба.*

*Колике су шансе да је Милан инжењер информационих технологија?*

*а) 70%*

*б) 30%*

*в) 50%*

Да би се унапредило резоновање ученика на овом пољу потребно је дати ученицима прилику да се суоче са проблемима подложним мисконцепцијама. Што подразумева да се ученици чешће суочавају са оваквим типовима проблема, да се подстиче критичко резоновање ученика и указивање на могуће мисконцепције. Мењање постојеће хеуристике резоновања новом представља један не тако једноставан процес. Па стога потребно је да ученици изнесу своје мишљење и став на почетку о проблему, а потом да се суочавајући се са аргументима потврди или оповргне њихов и став и мишљење. Било би добро да се организује рад у малим група тако да свака група добије различит задатак подложен мисконцепцији

<sup>3</sup> Референце су доступне на линку

<https://drive.google.com/file/d/1EuPu0PUYue0BHFF4JaLpL0DfWXMAOigO/view?usp=sharing>

### III део: ПРИМЕРИ, ЗАДАЦИ...

У циљу конкретизације предлога који смо изнели у материјалу у последњем делу наводимо задатке и примери који могу да буду коришћени на настави математике и то, како кроз редовну наставу, тако и кроз додатну наставу, а посебно кроз рад математичке секције.

#### Пратећи задаци

Три оригинална задатка аутора намењих за рад у основним и средњим школама и који представљају саставни део овог материјала, доступни у лектронској форми, су:

**Задатак 1.** обухвата мере центрираности са графичким представљањем и представља стандарди задатак за увежбавање и закључивање на основу израчунатих вредности и добијених графикана. Садржајно одговара и основној и средњој школи.

<https://drive.google.com/file/d/1ZeKXdO7ZLFRXNFzoggzqi0eQqCv9LQwN/view?usp=sharing>

**Задатак 2.** је истраживачки задатак који обухвата прве три етапе статистичког закључивања. Садржајно одговара ученицима основне школе.

<https://drive.google.com/file/d/1Kt7-zta-5qKFq3VWHJY4MqIjB2N6wArY/view?usp=sharing>

**Задатак 3.** је истраживачки задатак који захтева коришћење података прикупљених од надлежних институција, а обухвата сређивање, представљање и анализу података. Садржајно одговара ученицима средњих школа.

<https://drive.google.com/file/d/1zFCOTFyARO6D2gVC4K39a49qKoiGLRHv/view?usp=sharing>

#### Примери из медија

У данашње време највећи број информација ученик добија путем медија и друштвених мрежа. То је додатни разлог зашто је потребно унапредити ученичке компетенције који се тичу аналитичког мишљења и разумевања података. То је једини начин да заштити ученика, а самим тим и друштво, од нежељених ефеката које могу настати. Подсећамо на пример вакцине ММР. У наставку наводимо три примера у којима се појављује погрешно коришћење података или појмова везаних за рад са подацима.

Просечна жена на Врачару прво дете рађа са 32 године, а жена на југу Србије први пут одлази у породициште са 23 године. Интересантно је колико је Србија мала, а колико су њене културолошке разлике велике.

Извор: <http://www.politika.rs/sr/clanak/404445/Srbiji-neophodno-trece-dete>

U Sarajevu se dnevno konzumiraju dva kilograma heroina, a i 66% policajaca uzima neku drogu. Kasnije u tekstu: U policiji je bilo i nekoliko slucajeva samoubistava, ali o tome svi šute. Moram istaci da smo ipak uradili jedno testiranje prije nekoliko godina. Testirali smo tri policajca na drogu. Jedan je bio pozitivan na heroin i marihuanu, drugi samo na marihuanu, a treći je bio čist. Dakle, 66 posto testiranih pripadnika MUP-a pozitivno je na drogu - tvrdi Murga.

Извор: <http://depo.ba/clanak/70128/u-sarajevu-se-dnevno-konzumiraju-dva-kilograma-heroina-a-i-66-posto-policajaca-uzima-neku-drogu>

In 1995, the UK Committee on Safety of Medicines issued a warning that third generation oral contraceptive pills increased the risk of potentially life threatening thrombosis twofold. The news provoked great anxiety, and many women stopped taking the pill, which led to unwanted pregnancies and abortions—some 13 000 additional abortions in the next year in England and Wales—and an extra £46m (€55m; \$71m) in costs for the NHS. 5 Yet how big was the twofold risk? The studies revealed that for every 7000 women who took the earlier, second generation pills, one had a thrombosis, and this number increased to two in women who took third generation pills. The problem of misleading reporting has not gone away. In 2009, the BMJ published



two articles on oral contraceptives and thrombosis; one made the absolute numbers transparent in the abstract, whereas the other reported that “oral contraceptives increased the risk of venous thrombosis fivefold.

Извор: <https://www.healthnewsreview.org/2010/12/leading-risk-comm-guru-gigerenzer-argues-that-absolute-risk-communication-is-a-moral-issue/>

## Могућности - истраживачки и пројектни задаци

### Српски и страни језик

Посматрајмо латинично писмо. Тада српски и енглески писани језик имају групу заједничка слова. Потребно је проценити у ком језику је слово А заступљеније. Које слово је најзаступљеније у српском језику, а које у енглеском?

Проценити да ли у српском језику постоји више речи које почињу словом р или које садрже слово р?

### Биологија

Немају све животињске врсте исти однос женки и мужјака. Потребно је одредити шансу да се из кокошијег јајета излеже петао.

### Социологија

Процени шансу да случајно изабрани ученик твоје школе има мање од 2 налога на различитим друштвеним мрежама.

### Екологија, географија

Истражити загађеност ваздуха у Србији у једном месецу. Где прикупити податке? Како одабрати узорак? Која су ограничења студије? Шта се може закључити?