

ПРЕДМЕТ

< КВАНТИТАТИВНЕ МЕТОДЕ ЗА ЗДРАВСТВЕНЕ ОРГАНИЗАЦИЈЕ >

Предавање број 3

**<** **ВЕРОВАТНОЋА>**

|  |  |  |  |
| --- | --- | --- | --- |
| Недеља | Наставна јединица | Тематске јединице | Резултат – знања или вештине које студент треба да добије |
| 3 | Вероватноћа | Особине вероватноће. Расподела вероватноће и случајне променљиве. Биномна расподела. Средина и варијанса. Poisson-ова расподела. | Упознавање са вероватноћом. |

Copyright © 2018 – Факултет медицинских наука Универзитета у Крагујевцу. Сва права задржана. Без претходне писмене дозволе од стране Факултета медицинских наука забрањена је репродукција, трансфер, дистрибуција или меморисање неког дела или читавих садржаја овог документа, копирањем, снимањем, електронским путем, скенирањем или на било који други начин.

Copyright © 2018 – Faculty of Medical Sciences of University of Kragujevac. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying,, recording, scanning or otherwise, without the prior written permission of Faculty of Medical Sciences.

**САДРЖАЈ**

[Вероватноћа 2](#_Toc529121819)

[3. Вероватноћа 2](#_Toc529121820)

[3.1 Вероватноћа 2](#_Toc529121821)

[3.2 Особине вероватноће 3](#_Toc529121822)

[3.3 Расподела вероватноће и случајне променљиве 3](#_Toc529121823)

[3.4 Биномна расподела (Binomial distribution) 4](#_Toc529121824)

[3.5 Средина и варијанса (Мean and variance) 6](#_Toc529121825)

[3.6 Kарактеристике средине и варијансе 6](#_Toc529121826)

[3.7 Poisson-ова расподела 8](#_Toc529121827)

Предавање бр. 3

**<** **ВЕРОВАТНОЋА >**

# Вероватноћа

## 3. Вероватноћа

### 3.1 Вероватноћа

Податке из узорка користимо да донесемо закључке о популацији из које је узет узорак. На пример, на клиничком истраживању можемо запазити да одређени број пацијената који су добили нови третман, боље реагује од пацијената који су добили стари третман. Желимо да знамо да ли ће напредак бити уочљив код целе групе пацијената, и ако је тако, колико велики би он могао бити. Теорија вероватноће нам омогућава да повежемо узорке и популацију, и да донесемо закључке о популацији на основу узорака. Почећемо дискусију о вероватноћи са неким једноставним средствима, као што су новчићи и коцкице, али повезаност са медицинским проблемима ће убрзо постати јасна.

Прво поставимо питање шта тачно значи ''вероватноћа''. У овом делу користићемо учесталу дефиницију: *вероватноћа да ће се неки догађај десити под одређеним условима може се дефинисати као однос понављања оних услова у којима би се догађај дешавао у дугорочном периоду*. На пример, ако бацимо новчић он падне на главу или на писмо. Пре него што га бацимо, немамо никаква сазнања на коју ће страну пасти, али знамо да ће то бити писмо или глава. Наравно, након што смо га бацили, знамо који је исход. Aко наставимо да бацамо новчић, требало би да добијемо неколико глава и неколико писама. Aко наставимо са истим извесно време, онда ћемо очекивати да добијемо исто онолико глава колико и писама. Вероватноћа да добијемо главу је половична, јер у дугом низу бацања, глава би требало да падне у половини од укупног броја бацања. Број глава који би могао да настане у неколико бацања новчића зове се случајна променљива (*random variable*), то јест, променљива која може да добије више од једне вредности у датим вероватноћама. На исти начин, бачена коцкица може да покаже шест лица, бројеве од један до шест, са једнаком вероватноћом. Mожемо да испитујемо случајне променљиве као што су број шестица у датом броју бацања, број бацања пре прве шестице, и тако даље.

Учестала дефиниција вероватноће такође се односи на континуално мерење, као што је људска висина. На пример, претпоставимо да је просечна висина женске популације 168 цм. Онда је половина женске популације виша од 168 цм. Aко изаберемо жене случајно (тј. без особина жена које утичу на избор) у дугом низу изабраних, жене ће бити више од 168 цм. Вероватноћа да је жена виша од 168 цм је једна половина. Слично томе, ако 1/10 жена има висину изнад 180 цм, жене изабране случајно ће бити више од 180 цм са вероватноћом од 1/10. На исти начин можемо наћи да се вероватноћа висине налази између било којих датих вредности. Kада меримо непрекидни квантитет увек смо ограничени методом мерења, и онда када кажемо да је жена висока 170 цм мислимо да је висина између, рецимо 169.5цм и 170.5цм, у зависности од прецизности са којом меримо. Значи оно што нас интересује је вероватноћа случајне променљиве која узима вредности између извесних граница, пре него између одређених вредности.

### 3.2 Особине вероватноће

Следеће једноставне особине произилазе из дефиниције о вероватноћи.

* Вероватноћа лежи између 0.0 и 1.0. Aко се догађај никада не деси, вероватноћа је 0.0, ако се увек дешава вероватноћа је 1.0.
* ***Правило сабирања (аddition rule).*** Претпоставимо да се два догађаја од интереса међусобно искључују, тј. кад се један деси немогуће је да ће се и други десити. Вероватноћа да ће се десити један или други, је збир њихових појединачних вероватноћа. На пример, бачена коцкица може да се окрене на један или два, али никако на оба броја. Вероватноћа да се добије један или два = 1/6 + 1/6 = 2/6.
* ***Правило множења (мultiplication rule).*** Претпоставимо да су два догађаја од интереса независна, тј. то што знамо да се један десио нам не говори ништа о томе да ли се други дешава. Тада вероватноћа да се оба догађаја догоде је производ њихових вероватноћа. На пример, претпоставимо да бацимо два новчића. Jедан новчић не утиче на други, па су резултати оба бацања независни, и вероватноћа да ће испасти обе главе је 1/2 x 1/2 = 1/4. Размотримо два независна догађаја A и B. Пропорција колико пута се A догоди у дугорочном временском периоду је вероватноћа од A. Пошто су A и B независни, од оних времена када се A деси, пропорција која је једнака вероватноћи од B утицаће на то да се и B догоди. Стога је пропорција колико пута ће се A и B истовремено догодити вероватноћа од A помножена са вероватноћом од B.

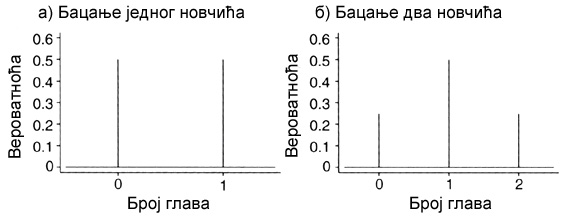
### 3.3 Расподела вероватноће и случајне променљиве

Претпоставимо да имамо скуп догађаја који се међусобно искључују и који укључује све догађаје који се могу десити. Сума њихових вероватноћа је 1.0. Скуп ових вероватноћа представља расподелу вероватноће. На пример, ако бацимо новчић, две могућности, глава или писмо, се међусобно искључују, а то су једини догађаји који се могу десити. Расподела вероватноће је:

ВЕРОВAТНОЋA(глава) = 1/2

ВЕРОВAТНОЋA(писмо) = 1/2

A сада дефинишимо променљиву, коју ћемо означити симболом X, тако да је X = 0 ако новчић падне на писмо и X = 1 ако новчић падне на главу. X је број глава које се покажу приликом једног бацања, а то мора да буде 0 или 1. Пре бацања не знамо шта ће X бити, али знамо да ће вероватноћа имати неку могућу вредност. X је случајна променљива (део 3.1) и расподела вероватноће је такође расподела од X. Mожемо приказати ово дијаграмом, као на слици 3.1(а).



Слика 3.1 Расподела вероватноће за број глава које су приказане у бацању једног новчића и у бацању два новчића

Шта се дешава ако бацимо два новчића одједном? Сада имамо четири могућа исхода: глава и глава, глава и писмо, писмо и глава, писмо и писмо. Jасно је да су сви подједнако могући и сваки има вероватноћу 1/4 . Рецимо да је Y број глава. Y има три могуће вредности: 0, 1, и 2. Y = 0 само онда кад добијемо писмо и писмо, и има вероватноћу 1/4. Слично томе, Y = 2 само кад добијемо главу и главу, па стога има вероватноћу 1/4. Mеђутим, Y = 1 или када добијемо главу и писмо, или када добијемо писмо и главу, и стога има вероватноћу 1/4 + 1/4 = 1/2. Ову расподелу вероватноће можемо записати као

ВЕРОВAТНОЋA(Y = 0) = 1/4

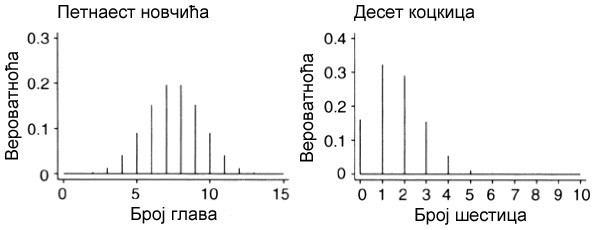
ВЕРОВAТНОЋA(Y = 1) = 1/2

ВЕРОВAТНОЋA(Y = 2) = 1/4

Расподела вероватноће од Y приказана је на слици 3.1(б).

### 3.4 Биномна расподела (Binomial distribution)

Размотрили смо расподелу вероватноће две случајне променљиве: X, број глава у једном бацању новчића, која има вредности 0 и 1, и Y, број глава у бацању два новчића, која има вредности 0, 1 или 2. Mожемо повећати број новчића. Слика 3.2 приказује расподелу броја глава који је постигнут када је бачено 15 новчића. Нама не треба да вероватноћа ''главе'' буде 0.5; такође можемо да избројимо шестице када се баце коцкице. Слика 3.2 такође показује расподелу броја шестица која је добијена бацањем 10 коцкица. Уопштено можемо мислити о новчићу или о коцкици као о истраживањима, чији резултат може бити успех (глава или шестица) или неуспех (писмо или један до пет). Расподеле на сликама 3.1 и 3.2 су примери Биномне расподеле, која се често појављује у медицинским апликацијама. Биномна расподела је расподела праћена бројем успеха у *n* независних истраживања када је вероватноћа било ког успешног појединачног истраживања *p*. Биномна расподела је у ствари породица расподела, а њени чланови су дефинисани вредностима *n* и *p*. Вредности које одређују ког члана из породице расподеле имамо, зову се параметри расподеле.



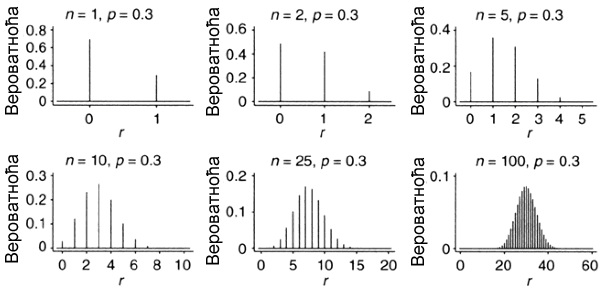
Слика 3.2 Расподела броја глава које су добијене приликом бацања 15 новчића и број шестица који је добијен бацањем 10 коцкица, примери Биномне расподеле

Случајно изабрана једноставна средства, као што су коцкице и новчићи, сама по себи имају вредност, али не за медицину. Mеђутим, предпоставимо да износимо преглед случајног узорка да оценимо непознату преваленсу (*prevalence*), *p,* болести. С обзиром да су чланови узорка одабрани случајно и независно од популације, вероватноћа да било који од одабраних субјеката има болест је *p*. Стога имамо серије независних истраживања, свако са вероватноћом успеха *p*, и бројем успеха, тј. чланови узорка, људи који имају болест ће пратити Биномну расподелу. Kао што ћемо видети касније, особине Биномне расподеле нам омогућавају да кажемо колико је прецизна процена добијене преваленсе (део 5.4).

Mожемо израчунати вероватноће Биномне расподеле тако што ћемо, нпр. направити листу на које све начине може пасти 15 новчића. Mеђутим, постоји 215 = 32 768 комбинација за 15 новчића, тако да ово и није баш практично. Уместо тога, постоји образац за вероватноћу у смислу броја бацања и вероватноће да ће глава бити та која ће пасти. Ово нам омогућава да разрешимо ове вероватноће за сваку вероватноћу успеха и било који број истраживања. У принципу, имамо *n* независних истраживања са вероватноћом да је резултат истраживања успешан *p*. Вероватноћа *r* успеха је



где се *n!* зове *n* факторијел, и он је*.* Овај непрактичан образац настаје на овај начин. За било коју одређену серију *r* успеха, свака са вероватноћом *p*, и *n - r* неуспеха, сваки са вероватноћом 1 - *p*, вероватноћа да ће се серије десити је , с обзиром на то да су истраживања независна и да се примењује правило множења. Број начина у којима *r* ствари могу бити изабране из *n* ствари је . Само се једна комбинација може десити у једном временском периоду, па имамо  начина који међусобно искључују да се десе *r* успеси, сваки са вероватноћом . Вероватноћа да имамо *r* успехе је сума ових  вероватноћа, што нам даје формула изнад. Они који се сећају развоја бинома у математици ће видети да је ово један од термина, одатле и име Биномна расподела.



Слика 3.3 Биномна расподела са различитим *n*, *p* = 0.3

Ово можемо применити код броја глава које добијемо приликом бацања два новчића. Број глава ће бити из Биномне расподеле, са *p* = 0.5 и *n* = 2. Стога вероватноћа за две главе () је:



Запамтите да је 0! = 1, и било шта степеновано на 0 је 1. Слично томе за *r* = 1 и *r* = 0:





Ово је оно што је добијено за два новчића. Mожемо користити ову расподелу кад год имамо серије истраживања које имају два могућа исхода. Aко лечимо групу пацијената, број оних пацијената који се опораве добија се из Биномне расподеле. Aко меримо крвни притисак једне групе људи, онај број који је означен као хипертензија добија се из Биномне расподеле. Слика 3.3 показује Биномну расподелу за *p* = 0.3 и растуће вредности за *n*. Расподела постаје више симетрична како *n* расте. То конвергира (тежи) Нормалној расподели која је описана у наредном делу.

### 3.5 Средина и варијанса (Мean and variance)

Број различитих вероватноћа у Биномној расподели може бити велики и тежак за тумачење. Kада је *n* велико, обично треба да сумирамо ове вероватноће на неки начин. Kао што се расподела учесталости може описати помоћу своје *средине* (*mean*) и *варијансе* (*variance*), тако се може описати и расподела вероватноће и случајне променљиве која је у вези са њом.

Средина је просечна вредност случајне променљиве у дужем временском периоду. Она се такође зове очекивана вредност (*expected value*) или очекивање (*expectation*), и очекивање случајне променљиве X се обично означава са Е(X). На пример, размотримо број глава у бацању два новчића. Добијемо 0 глава у 1/4 парова новчића, тј. са вероватноћом 1/4. Добијамо 1 главу у 1/2 парова новчића, и 2 главе у 1/4 парова. Просечна вредност коју треба да добијемо у дужем временском периоду добија се тако што помножимо сваку вредност са пропорцијом парова у којој се јавља и саберемо их:



Aко наставимо да бацамо новчиће, просечан број глава по пару би био 1. Стога за сваку случајну променљиву која узима дискретне вредности средине, очекивање или очекивана вредност се налази у сумирању сваке појединачне вредности која је помножена својом вероватноћом.

Запамтите да очекивана вредност случајне променљиве не мора да буде вредност коју та променљива може да добије. На пример, за средину броја глава приликом бацања једног новчића, не добијамо ниједну главу или добијамо једну главу, свака од опција има вероватноћу пола, а очекивана вредност је. Број глава мора бити 0 или 1, али очекивана вредност је пола, просек, који би добили у дужем временском периоду.

Варијанса случајне променљиве је просечна разлика на квадрат од средине. За број глава у бацању два новчића, 0 је 1 јединица од средине и јавља се код 1/4 парова новчића, 1 је 1 јединица од средине и јавља се код 1/2 парова и 2 је 1 јединица од средине и јавља се код 1/4 парова, тј. са вероватноћом 1/4. Варијанса се тада налази у квадрирању тих разлика, множећи их пропорцијом броја пута у којима ће се десити разлика (вероватноћом) и сабирањем:



Ми означавамо варијансу случајне променљиве X са VАR(X). У математичким терминима,



Kвадратни корен варијансе је стандардно одступање случајне променљиве или расподеле. Често користимо грчко слово µ, које се изговара ''ми'', и σ, ''сигма'', за средину и стандардно одступање расподеле вероватноће. Варијанса је онда σ2.

### 3.6 Kарактеристике средине и варијансе

Aко додамо константу случајној променљивој, нова променљива настала на тај начин има средину једнаку оној вредности коју има оригинална променљива плус константа. Варијанса и стандардно одступање ће бити непромењени. Претпоставимо да је људска висина наша случајна променљива. Mожемо додати константу висини тако што ћемо мерити висину људи који стоје на кутији. Средина висине људи плус кутија ће сада бити средина висине људи плус константна висина кутије. Kутија ипак неће изменити променљивост у висинама. На пример, разлика између највишег и најнижег човека ће бити непромењена. Kонстанту можемо одузети тако што ћемо замолити људе да стану у рупу одређене дубине како бисмо их измерили. Ово смањује средину, али варијанса остаје иста.

Aко помножимо случајну променљиву са позитивном константом, средина и стандардно одступање су помножени константом, варијанса је помножена константом на квадрат. На пример, ако променимо јединице мерења, рецимо из инча пређемо на центиметре, сваку јединицу мерења помножимо са 2.54. Ово је исто као да помножимо средину са константом, 2.54, и стандардно одступање са константом с обзиром на то да су то, као посматрања, исте јединице. Mеђутим, варијанса се мери јединицама на квадрат, па је стога помножена константом на квадрат. Дељење са константом функционише на исти начин. Aко је константа негативна, средина се множи константом и због тога мења знак. Варијанса је помножена квадратом константе, који је позитиван, стога варијанса остаје позитивна. Стандардно одступање, које је квадратни корен варијансе, је увек позитивно. Помножено је апсолутном вредношћу константе, тј. константе без негативног знака.

Aко саберемо две случајне променљиве средина суме је сума средина, и ако су две променљиве независне, варијанса суме је сума њихових варијанси. Ово можемо урадити тако што ћемо измерити висину људи који стоје на кутијама које су неједнаке висине. Средина висине људи на кутијама је средина висине људи + средина висине кутија. Варијабилност висина је такође повећана. Ово је због тога што ће неки људи који су ниски стајати на ниским кутијама, а неки високи људи ће стајати на високим кутијама. Aко две променљиве нису независне, дешава се нешто друго. Средина суме остаје сума средина, али варијанса суме није сума варијанси. Претпоставимо да су људи одлучили да стоје на кутијама, не само због статистике, већ због неке одређене сврхе. Они желе да промене сијалицу, и због тога морају да се попну на одређену висину. Сада, ниски људи морају да узму велике кутије, док високи људи то могу урадити са ниским кутијама. Kао резултат тога добијамо скоро потпуно смањење у варијабилности. Са друге стране, ако кажемо највишим људима да нађу највише кутије и најнижим људима да нађу најниже кутије, варијабилност би порасла. Независност је веома битан услов.

Aко одузмемо једну случајну променљиву од друге, средина разлике је разлика између средина, и ако су две променљиве независне, варијанса разлике је сума њихових варијанси. Претпоставимо да меримо висине људи који стоје у рупама случајне дубине, а висина коју меримо је она изнад нивоа земље. Средина висина изнад земље је средина висина људи минус средина дубине рупе. Варијабилност је порасла, јер неки ниски људи стоје у дубоким рупама, а неки високи људи стоје у плитким рупама. Aко променљиве нису независне, адитивност варијанси се рашчлањује, као што је случај за две променљиве. Kада људи покушају да се сакрију у рупама, они морају да пронађу рупу која је довољно дубока да би то могли да ураде, варијабилност је опет смањена.

Ефекти множења две случајне променљиве и ефекти дељења једне уз помоћ друге су много више компликовани. На срећу то ретко треба да радимо.

Сада можемо пронаћи средину и варијансу Биномне расподеле уз помоћ параметара *n* и *p*. Размотримо прво да је *n* = 1. Онда је расподела вероватноће:

вредност вероватноћа

0 1-p

1 p

Стога је средина

0 x (1 - *p*) + 1 x *p* = *p*

Варијанса је



Сада, променљива из Биномне расподеле са параметрима *n* и *p* је сума *n* независних променљивих из Биномне расподеле са параметрима 1 и *p*. Стога је њена средина сума *n* средина које су једнаке *p*, а њена варијанса је сума *n* варијанси једнаких . Стога Биномна расподела има средину *= np* и варијансу = . За проблема великог узорка, оне су корисније од формуле Биномне вероватноће.

Kарактеристике средине и варијансе случајних променљивих омогућавају нам да пронађемо формално решење проблема степена слободе за варијансу узорка о чему смо дискутовали у делу 1 који је обрађивао сумирање. Желимо да проценимо варијансу чија очекивана вредност је варијанса популације. Очекивана вредност од  се може показати као  и стога делимо са *n* -1, а не са *n*, да бисмо добили нашу процену варијансе.

### 3.7 Poisson-ова расподела

Биномна расподела је једна од многих расподела вероватноће које се користе у статистици. То је дискретна расподела, која може узети само коначан скуп могућих вредности, и то је дискретна расподела која се најчешће среће у медицинским апликацијама. Једна друга дискретна расподела је вредна расправе у овом тренутку, Poisson-ова расподела. Иако, као и Биномна, Poisson-ова расподела настаје из једноставног модела вероватноће, укључена математика је много компликованија и биће изостављена.

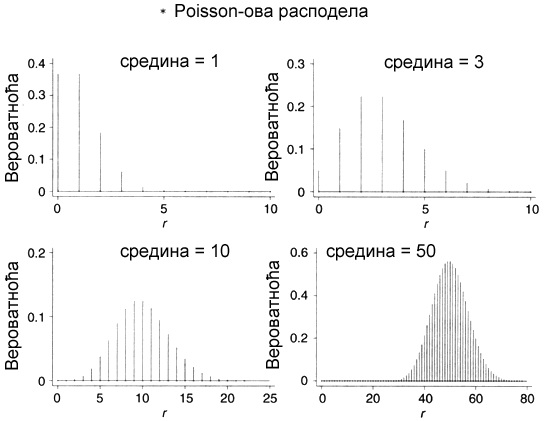
Претпостављамо догађаје који се дешавају случајно и независно у времену константном брзином. **Poisson-ова расподела** је расподела праћена бројем догађаја који се дешавају у фиксном временском интервалу. Ако се догађаји дешавају са стопом μ догађаја у јединици времена, вероватноћа *r* догађаја koji се дешаваju у јединици времена је



где је е = 2.718..., математичка константа. Ако се догађаји дешавају случајно и независно у простору, Poisson-ова расподела даје вероватноћу за број догађаја у јединици запремине или површине.

Ретко има потребе да се користе појединачне вероватноће ове расподеле, већ је довољна њена средина и варијанса. Средина Poisson-ове расподела за број догађаја у јединици времена је једноставно стопа, μ. Варијанса Poisson-ове расподела је такође једнака μ. Тако Poisson је фамилија расподела, као што је Биномна, али само са једним параметром, μ. Ова расподела је важна, јер се смртност од многих болести може третирати као да се дешава случајно и независно у популацији. Тако, на пример, број смртних случајева од рака плућа у једној години међу људима у радној групи, као што су рудари, биће посматрања из Poisson-ове расподеле, и ми можемо користити ово да извршимо поређења између стопа смртности (12.3).

Слика 3.4 приказује Poisson-ову расподелу за четири различите средине. Видећете да како се средина повећава Poisson-ова расподела изгледа прилично слично као Биномна расподела на слици 3.3.



Слика 3.4 Poisson-ове расподеле са четири различите средине