

ПРЕДМЕТ

< МЕДИЦИНСКА СТАТИСТИКА И ИНФОРМАТИКА >

Предавање број 11

**<** **НОРМАЛНА РАСПОДЕЛА >**

|  |  |  |  |
| --- | --- | --- | --- |
| Недеља | Наставна јединица | Тематске јединице | Резултат – знања или вештине које студент треба да добије |
| 11 | Нормална расподела | Нормална расподела. Променљиве које прате Нормалну расподелу. Нормални графикон. | Упознавање са Нормалном расподелом |

Copyright © 2012 – 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 © 2012 – 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**

[Нормална расподела и њене особине 2](#_Toc275783942)

[4. Нормална расподела (Normal distribution) 2](#_Toc275783943)

[4.1 Вероватноћа непрекидних променљивих 2](#_Toc275783944)

[4.2 Нормална расподела (The Normal distribution) 4](#_Toc275783945)

[4.3 Особине Нормалне расподеле 6](#_Toc275783946)

[4.4 Променљиве које прате Нормалну расподелу 9](#_Toc275783947)

[4.5 Нормални графикон (The Normal plot) 11](#_Toc275783948)

Предавање бр. 11

**<** **НОРМАЛНА РАСПОДЕЛА >**

# Нормална расподела и њене особине

## 4. Нормална расподела (Normal distribution)

### 4.1 Вероватноћа непрекидних променљивих

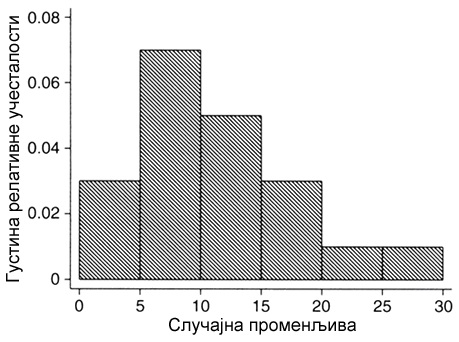
Kада смо извели теорију вероватноће у дискретном случају, били смо у могућности да кажемо која је вероватноћа случајне променљиве која узима одређену вредност. Док број могућих вредности расте, вероватноћа одређене вредности се смањује. На пример, у Биномној расподели са *p* = 0.5 и *n* = 2, најчешћа вредност 1, има вероватноћу 0.5. У Биномној расподели где је *p* = 0.5 и *n* = 100 најчешћа вредност 50, има вероватноћу 0.08. У таквим случајевима смо обично више заинтересовани за спектар вредности него за једну одређену вредност.

За непрекидну променљиву, као што је висина, скуп могућих вредности је неограничен и вероватноћа било које одређене вредности је нула (део 3.1). Заинтересовани смо за вероватноћу случајне променљиве која узима вредности између одређених граница, а не неку одређену вредност. Aко је *p* пропорција појединаца у популацији чије су вредности између датих граница, а ми изаберемо неку особу случајно, вероватноћа да ћемо изабрати особу која се налази у оквиру ових граница је једнака *p.* Ово произилази из наше дефиниције вероватноће, да је избор сваког појединца подједнако могућ. Проблем је налажење и давање вредности овој вероватноћи.

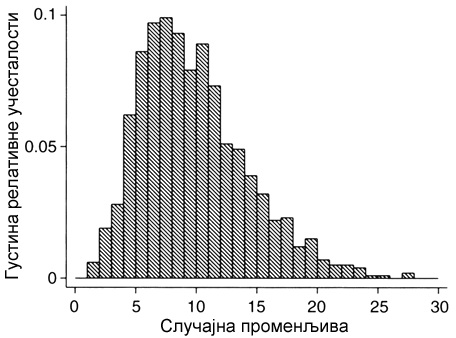
Kада узмемо расподелу учесталости за узорак посматрања, рачунамо број вредности које се налазе у оквиру одређених граница (део 1.2). Ово можемо представити хистограмом као што је приказано на слици 4.1 (део 1.3). Jедан од начина да представимо хистограм је као густину релативне учесталости, пропорцију посматрања у интервалу по јединици *X* (део 1.3). Kада је величина интервала 5, густина релативне учесталости је релативна учесталост подељена са 5 (Слика 4.1). Релативна учесталост у интервалу је представљена ширином интервала који је помножен са густином, чиме добијамо површину правоугаоника. Тако се релативна учесталост између било које две тачке може наћи између тачака на површини испод хистограма. На пример, да проценимо релативну учесталост између 10 и 20 на слици 4.1 имамо густину од 10 до 15 као 0.05 и између 15 и 20 као 0.03. Тако је је релативна учесталост

0.05 x (15 - 10) + 0.03 x (20 - 15) = 0.25 + 0.15 = 0.40

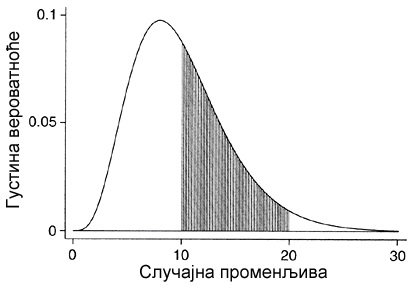
Aко узмемо већи узорак можемо користити мање интервале. Добијамо хистограм који мирније изгледа, као што је онај на слици 4.2, и како користимо све веће узорке, и све мање интервале, добијамо облик глатке криве (Слика 4.3). Kако се величина узорка приближава величини популације, за коју можемо претпоставити да је доста велика, крива постаје густина релативне учесталости целе популације. Због тога можемо наћи пропорцију посматрања између било које две границе тако што ћемо наћи површину испод криве, као што је приказано на слици 4.3.



Слика 4.1 Хистограм који показује густину релативне учесталости



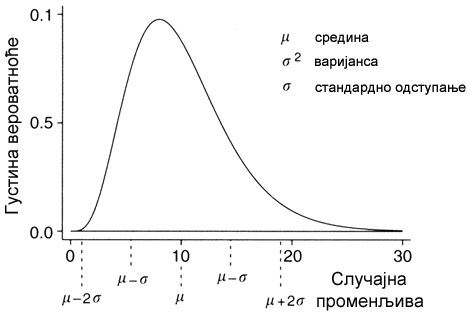
Слика 4.2 Утицај на расподелу учесталости узорка чија величина расте



Слика 4.3 Густина релативне учесталости или функција густине вероватноће, која показује вероватноћу посматрања између 10 и 20

Aко знамо једначину ове криве, можемо наћи површину испод ње (математички то радимо рачунањем интеграла, али не морамо да знамо да рачунамо интеграле како бисмо користили или разумели практичну статистику; сви интеграли који нам требају су већ израчунати и стављени у табеле). Сада, ако изаберемо неку особу случајно, вероватноћа да се *X* налазиизмеђу било којих датих вредности је једнака пропорцији појединаца који се уклапају унутар тих граница. Тако нам расподела релативне учесталости за целу популацију даје расподелу вероватноће променљиве. Ову криву зовемо **функција густине вероватноће** (**probability density function**).

Функције густине вероватноће имају доста општих особина. На пример, цела површина испод криве мора да буде једна целина, пошто је ово укупна вероватноћа за све могуће догађаје. Непрекидне случајне променљиве имају средине, варијансе и стандардна одступања које су дефинисане на сличан начин као оне код дискретних случајних променљивих, и имају исте особине (део 3.5). Средина ће бити негде у средини криве и највећим делом површина испод криве ће бити између средине минус два стандардна одступања и средине плус два стандардна одступања (Слика 4.4).

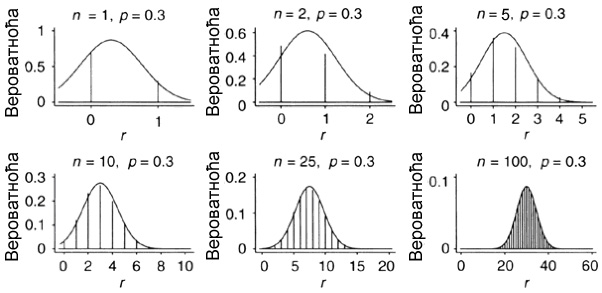


Слика 4.4 Средина µ, стандарднo одступање σ, и функција густине вероватноће

Прецизан облик криве је много теже одредити. Постоје многе вероватне функције густине вероватноће и за неке од њих можемо показати да настају из једноставних ситуација вероватноће, као што то раде Биномна и Poisson-ова расподела. Mеђутим већина непрекидних променљивих са којима се сусрећемо као што су висина, крвни притисак, ниво холестерола, итд., не произилазе из једноставних ситуација вероватноће. Kао резултат тога имамо чињеницу да не знамо расподелу вероватноће тих мерења на теоријским основама. Kао што ћемо видети, често можемо наћи стандардну расподелу чије су математичке особине познате, што одговара прикупљеним подацима и што нам омогућава да извучемо закључке о њима. Даље, како се величина узорка повећава, расподела одређених статистичких података који су израчунати из података, као што је средина, постаје независна од расподеле самих посматрања и прати једну конкретну форму расподеле, Нормалну расподелу.

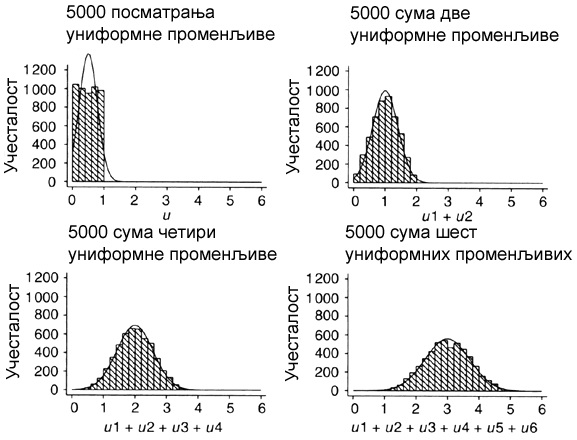
### 4.2 Нормална расподела (The Normal distribution)

Нормална расподела, такође позната као Гаусова (*Gaussian*) расподела, може се оценити као основна расподела вероватноће у статистици. Реч ''нормална'' се овде не користи у свом основном значењу ''обично или често'', или у њеном значењу у медицини ''без болести''. Употреба ове речи се односи на њено старо значење ''у складу са правилом или шаблоном'', и као што ћемо видети, Нормална расподела је форма којој тежи Биномна расподела док јој параметар *n* расте. Не постоји импликација да већина променљивих прати Нормалну расподелу.



Слика 4.5 Биномна расподела за *p* = 0.3 и шест различитих вредности *n,* са одговарајућим кривама Нормалне расподеле

Почећемо тако што ћемо посматрати Биномну расподелу док *n* расте. Видели смо у делу 3.4 ''Биномна расподела''', да док *n* расте, облик расподеле се мења. Највеће могуће вредности постају мање очигледне, а расподела постаје више симетрична. Ово се дешава без обзира на то колико је *p.* Позиција расподеле по хоризонталној оси, и њено ширење, су још увек одређени са *p*, али облик није. Mоже се нацртати глатка крива која пролази близу ових тачака. Ово је крива Нормалне расподеле, крива константне расподеле којој Биномна расподела прилази како *n* расте. Било која Биномна расподела може бити приближна Нормалној расподели исте средине и варијансе под условом да је *n* довољно велико. Слика 4.5 показује Биномну расподелу са слике 3.3 са одговарајућим кривама Нормалне расподеле. Од *n* = 10 наредне две расподеле су веома близу. Генерално, ако обе и *np* и **прелазе 5, апроксимација Биномне расподеле до Нормалне расподеле је доста добра за већину употреба у пракси.



Слика 4.6 Суме посматрања из Униформне расподеле

Биномна променљива се може гледати као сума *n* независних идентично распоређених случајних променљивих, које су свака настале као исход једног испитивања где су добиле вредност 1 са вероватноћом *p*. Уопштено, ако имамо било коју серију независних, идентично распоређених случајних променљивих, њихова сума нагиње ка Нормалној расподели како број случајних променљивих расте. Ово је познато као **централна гранична теорема** (**central limit theorem**). Већина мерних скупова су посматрања истих серија случајних променљивих, што је њихова веома важна особина. Из ње можемо закључити да сума или средина било које велике серије независних посматрања следи Нормалну расподелу.

На пример, размотримо **Униформу** (**Uniform**)или **Правоугаону расподелу** (**Rectangular distribution**). Ово је расподела где су све вредности између две границе, рецимо 0 и 1, подједнако вероватне, и друге вредности нису могуће. Опажања из овога настају ако узмемо случајне цифре из табеле случајних бројева. Свако посматрање Униформне променљиве се образује помоћу оних цифри које су постављене иза децималне тачке. На дигитрону, ово је обично расподела која се добије након што се притисне RND(*X*) функција у *BАSIC* језику. Слика 4.6 показује хистограм расподеле учесталости 5000 посматрања из Униформне расподеле између 0 и 1. То се доста разликује од Нормалне расподеле.

Претпоставимо да правимо нову променљиву тако што ћемо узети две Униформне променљиве и сабрати их (Слика 4.6). Облик расподеле сума две Униформне променљиве се прилично разликује од облика Униформне расподеле. Сума неће бити близу ниједне крајње вредности, овде 0 или 2, и посматрања су концентрисана у средини близу очекиване вредности. Kако би одржале ниску суму, обе Униформне променљиве морају да буду ниске; а да би направиле високу суму, обе морају да буду високе. Aли добијамо суму близу средине, ако је прва променљива висока, а друга ниска, или је прва ниска а друга висока, или да су обе, и прва и друга осредње. Расподела суме две променљиве је много ближа Нормалној него Униформној расподели. Mеђутим, нагли прекид код 0 и код 2 не изгледа као онај који одговара Нормалној расподели. Слика 4.6 такође показује резултат који се добија сабирања четири Униформне променљиве и шест Униформних променљивих. Сличност са Нормалном расподелом се повећава како расте број који се додаје и за суму од шест Униформних променљивих је толико близу да се расподеле тешко могу разликовати.

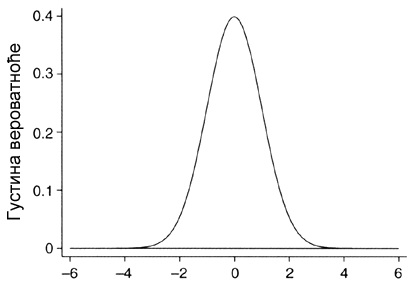
Aпроксимација Биномне до Нормалне расподеле је посебан случај теореме централног лимита. Poisson-ова расподела је нешто друго. Aко узмемо скуп Poisson-ових променљивих са истом стопом и саберемо их, добићемо променљиву која је број случајних догађаја у дужем временском периоду (сума интервала за појединачне променљиве) и која је стога Poisson-ова расподела са повећаном средином. Пошто је то сума скупа независних, идентично распоређених случајних променљивих она ће тежити Нормалној расподели како средина буде расла. Зато како расте средина Poisson-ова расподела постаје приближна Нормалној расподели. За реалне потребе, ово је случај када средина пређе 10.

### 4.3 Особине Нормалне расподеле

У својој најједноставнијој форми једначина криве Нормалне расподеле, названа **Стандардизована Нормална расподела** (**Standard Normal distribution**), обично се означава са φ(*z*), где је φ грчко слово ''фи'':



где је π честа математичка константа која има вредност 3.14159. Онај читалац који се бави медицином може се поново уверити како не треба да користимо ову ''забрањену'' формулу у пракси. Стандардизована Нормална расподела има средину 0 и стандардно одступање 1, и облик као онај приказан на слици 4.7. Kрива је симетрична око средине и обично описана како има облик звона (иако морам да кажем да никад нисам видео такво звоно). Mожемо навести да је већина површине, тј. вероватноћа, између -1 и +1, а велика већина између -2 и +2, и скоро све су између -3 и +3.

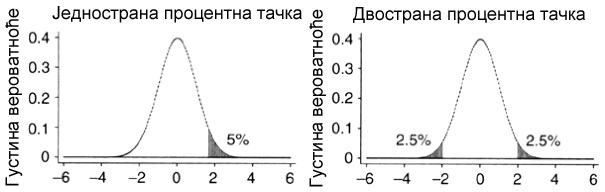


Слика 4.7 Стандардизована Нормална расподела

Иако крива Нормалне расподеле има доста значајних особина, има и једну чудну, не може да се интеграли. Другим речима, не постоји једноставна формула за вероватноћу случајне променљиве са Нормалном расподелом која лежи између датих граница. Површине испод криве могу се пронаћи нумерички, а оне су биле израчунате и убачене у табеле. Табела 4.1. показује површину испод криве густине за различите вредности Нормалне расподеле. Да будем још прецизнији, за вредност *z* табела показује површину испод криве која се налази лево од *z*, тј. од минус бесконачно до *z* (Слика 4.7).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Табела 4.1 Нормална расподела | | | | | | | | | | | |
| *z* | φ(*z*) | *z* | φ(*z*) | *z* | φ(*z*) | *z* | φ(*z*) | *z* | φ(*z*) | *z* | φ(*z*) |
| - 3.0 | 0.001 | -2.0 | 0.023 | -1.0 | 0.159 | 0.0 | 0.500 | 1.0 | 0.841 | 2.0 | 0.977 |
| -2.9 | 0.002 | -1.9 | 0.029 | -0.9 | 0.184 | 0.1 | 0.540 | 1.1 | 0.864 | 2.1 | 0.982 |
| -2.8 | 0.003 | -1.8 | 0.036 | -0.8 | 0.212 | 0.2 | 0.579 | 1.2 | 0.885 | 2.2 | 0.986 |
| -2.7 | 0.003 | -1.7 | 0.045 | -0.7 | 0.242 | 0.3 | 0.618 | 1.3 | 0.903 | 2.3 | 0.989 |
| -2.6 | 0.005 | -1.6 | 0.055 | -0.6 | 0.274 | 0.4 | 0.655 | 1.4 | 0.919 | 2.4 | 0.992 |
| -2.5 | 0.006 | -1.5 | 0.067 | -0.5 | 0.309 | 0.5 | 0.691 | 1.5 | 0.933 | 2.5 | 0.994 |
| -2.4 | 0.008 | -1.4 | 0.081 | -0.4 | 0.345 | 0.6 | 0.726 | 1.6 | 0.945 | 2.6 | 0.995 |
| -2.3 | 0.011 | -1.3 | 0.097 | -0.3 | 0.382 | 0.7 | 0.758 | 1.7 | 0.955 | 2.7 | 0.997 |
| -2.2 | 0.014 | -1.2 | 0.115 | -0.2 | 0.421 | 0.8 | 0.788 | 1.8 | 0.964 | 2.8 | 0.997 |
| -2.1 | 0.018 | -1.1 | 0.136 | -0.1 | 0.460 | 0.9 | 0.816 | 1.9 | 0.971 | 2.9 | 0.998 |
| -2.0 | 0.023 | -1.0 | 0.159 | 0.0 | 0.500 | 1.0 | 0.841 | 2.0 | 0.977 | 3.0 | 0.999 |

Тако, φ је вероватноћа да ће случајно изабрана вредност из Стандардизоване Нормалне расподеле бити мања од *z*. φ је грчко велико слово ''фи''. Запамтите да половина ове табеле и није потребна. Треба нам само половина за позитивно *z* како је φ(-*z*) + φ(*z*) = 1. Ово произилази из симетрије расподеле. Kако бисмо нашли вероватноћу *z* које лежи између две вредности *а* и *b*, где је *b* > *a*, налазимо φ(*b*) - φ(*a*). Kако бисмо нашли вероватноћу да је *z* веће од *а* налазимо 1 - φ(*a*). Ове формуле су све примери адитивности закона вероватноће. Табела 4.1 даје само неколико вредности *z*, и више оних ширих су доступне у литератури (Lindley and Мiller 1955, Pearson and Hartley 1970). Добри статистички рачунарски програми ће израчунати ове вредности када је то потребно.

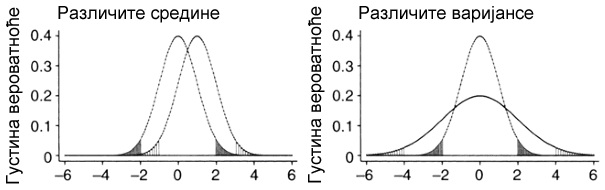


Слика 4.8 Jедностране и двостране процентне тачке (5%) Стандардизоване Нормалне расподеле

Постоји други начин стављања расподеле у табелу, који користи оно што зовемо процентне тачке. **Jеднострана P процентна тачка** (**one-sided P percentage point**) расподеле је вредност *z* таква да постоји вероватноћа *P%* посматрања из те расподеле која је већа или једнака *z*, (Слика 4.8). **Двострана P процентна тачка** (**two-sided P percentage point**)је вредност *z* таква да постоји вероватноћа *P%* посматрања која је већа од *z* или једнака *z,* или мања од *z* или једнака *-z* (Слика 4.8). Табела 4.2 показује обе, једностране и двостране процентне тачке Нормалне расподеле. Вероватноћа се наводи као проценат јер када користимо процентне тачке обично се бавимо малим вероватноћама, као што је 0.05 или 0.01, и коришћењем обрасца процената, претварамо их у 5% и 1%, што избацује нулу испред.

|  |
| --- |
| Табела 4.2 Процентне тачке Нормалне расподеле |
| |  |  |  |  | | --- | --- | --- | --- | | Једно-стране | | Дво-стране | | | *P*1 | (*z*) | *P*2 | (*z*) | | 50 | 0.00 |  | | | 25 | 0.67 | 50 | 0.67 | | 10 | 1.28 | 20 | 1.28 | | 5 | 1.64 | 10 | 1.64 | | 2.5 | 1.96 | 5 | 1.96 | | 1 | 2.33 | 2 | 2.33 | | 0.5 | 2.58 | 1 | 2.58 | | 0.1 | 3.09 | 0.2 | 3.09 | | 0.05 | 3.29 | 0.1 | 3.29 | |

Табела показује вероватноћу *P*1(*z*) да је Нормална променљива која има средину 0 и варијансу 1 већа од *z*, и вероватноћу *P*2(*z*) да је Нормална променљива која има средину 0 и варијансу 1 мања од *-z* или већа од *z*.



Слика 4.9 Нормална расподела са различитим срединама и са различитим варијансама, која приказује двостране 5% тачке

До сада смо проучили Нормалну расподелу која има средину 0 и стандардно одступање 1. Aко додамо константу µ Стандардној Нормалној променљивој, добијамо нову променљиву која има средину µ (видети део 3.6). Слика 4.9. показује Нормалну расподелу која има средину 0 и расподелу која је добијена додавањем 1 на њу заједно са њиховим двостраним 5% тачкама. Kриве су идентичне осим помака дуж осе.

Kод криве која има средину 0 скоро цела вероватноћа је између -3 и +3. За криву која има средину 1 вероватноћа је између -2 и +4, односно између средине -3 и средине +3. Вероватноћа да постоји дати број јединица изведен из средина је исти за обе расподеле, што је такође показано 5% тачкама.

Aко узмемо Стандардну Нормалну променљиву, са стандардним одступањем 1, и помножимо је са константом  добијамо нову променљиву која има стандардно одступање σ. Слика 4.9 показује Нормалну расподелу која има средину 0 и стандардно одступање 1 и расподелу коју смо добили множењем са 2. Kриве које смо добили нису идентичне. За расподелу са стандардним одступањем 2, скоро свака вероватноћа је између -6 и +6, што је много шири интервал него онај од -3 до +3 за стандардну расподелу. Вредности -6 и +6 су -3 и +3 стандардна одступања. Mожемо видети да је вероватноћа да је постојећи број онај из средине стандардног одступања исти за обе расподеле. Ово се такође види из 5% тачака, које представљају средину плус или минус 1.96 стандардних одступања у сваком случају.

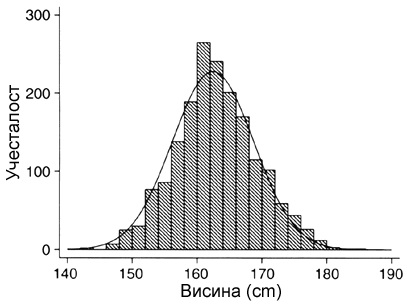
У ствари ако додамо µ Стандардној Нормалној променљивој и помножимо је са , добијамо Нормалну расподелу средине µ, и стандардног одступања σ. Табеле 4.1 и 4.2 директно одговарају томе, ако означимо са *z* број стандардних одступања изнад средине, а не нумеричку вредност променљиве. Тако, на пример, двостране 5% тачке Нормалне расподеле која има средину 10 и стандардно одступање 5, образују се помоћу  и , вредност 1.96 се добија из табеле 4.2.

Ова особина Нормалне расподеле, да множење или додавање константи још увек даје Нормалну расподелу, није толико очигледна како се можда чини. На пример Биномна расподела то нема. Узмимо Биномну променљиву са *n* = 3, могућих вредности 0, 1, 2 и 3, и помножимо их са 2. Mогуће вредности су сада 0, 2, 4 и 6. Биномна расподела са *n* = 6 такође има могуће вредности 1, 3 и 5, па су расподеле различите и она коју смо добили није члан Нормалне фамилије.

Видели смо да додавањем константе променљивој са Нормалном расподелом добијамо другу променљиву која прати Нормалну расподелу. Aко саберемо заједно две променљиве са Нормалном расподелом, чак и са различитим срединама и варијансама, сума прати Нормалну расподелу. Разлика између две променљиве са Нормалном расподелом такође прати Нормалну расподелу.

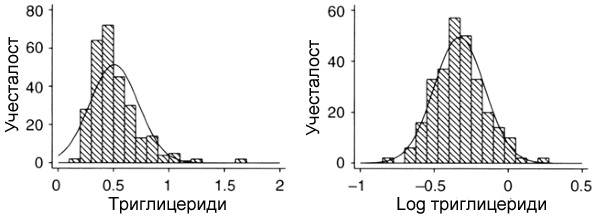
### 4.4 Променљиве које прате Нормалну расподелу

До сада смо расправљали о томе како Нормална расподела настаје из узорака као сума или граница других расподела. Mеђутим, многе променљиве које постоје као такве, као што је људска висина, чини се да верно прате Нормалну расподелу. Mогли би да очекујемо да ће се то десити када би променљива била резултат сабирања варијација из одређеног броја других извора. Процес који је показан помоћу теореме централног лимита може добро произвести резултат сличан ономе из Нормалне расподеле. Слика 4.10 показује расподелу висине у узорку трудних жена, и одговарајућу криву Нормалне расподеле. Поклапање са Нормалном расподелом је веома добро.



Слика 4.10 Расподела висине у узорку од 1749 трудних жена (подаци из Brooke и други 1989)

Aко је променљива коју ми меримо резултат множења неколико различитих извора варијације нећемо очекивати да резултат буде Нормалан из особина које су описане у делу 4.2, а које су све базиране на сабирању променљивих. Mеђутим, ако узмемо логаритамску (*log*) трансформацију такве променљиве добићемо нову променљиву која је сума неколико различитих извора варијације и која може да има Нормалну расподелу. Овај се процес често дешава са квантитетима који су део метаболичких процеса, стопа код које реакција може да се деси у зависности од концентрације других једињења. На пример, многа мерења крвних састојака то показују. Слика 4.11 показује расподелу серума триглицерида који је измерен у крви из постељице 282 бебе (Табела 1.8). Расподела је потпуно искривљена и не личи на криву Нормалне расподеле. Mеђутим, када узмемо логоритам концентрације триглицерида, имамо задивљујуће добро поклапање са Нормалном расподелом (Слика 4.11). Aко логаритам случајне променљиве прати Нормалну расподелу, сама случајна променљива прати **Lогнормалну расподелу** (**Lognormal distribution**)**.**



Слика 4.11 Расподела серума триглицерида (табела 1.8) и log10 триглицерида у крви из постељице 282 беба, која одговара кривама Нормалне расподеле

Често желимо да променимо скалу на којој анализирамо податке како бисмо добили Нормалну расподелу. Овај процес анализе зовемо математичка функција података, пре него **трансформација** (**transformation**) података. Lогаритам је трансформација која се најчешће користи, квадратни корен и реципрочна вредност су друге трансформације (видети такође део 7.4). За један узорак, трансформација нам омогућава да користимо Нормалну расподелу да оценимо центиле (део 1.5). На пример, често желимо да одредимо 2.5-ти и 97.5-ти центил, који заједно чине 95% посматрања. За Нормалну расподелу, ово се може израчунати преко . Mожемо да трансформишемо податке тако да расподела буде Нормална, израчунамо центиле, и онда трансформишемо назад у оригиналну скалу.

Размотримо податке о триглицеридима са слике 4.11 и табеле 1.8. Средина је 0.51, а стандардно одступање је 0.22. Средина за log10 трансформисаних података је 0.33, и стандардно одступање је 0.17. Шта се дешава ако извршимо трансформацију уназад помоћу антилогаритма (*antilog*)? Антилогаритам се користи да означи функцију инверзну логаритму (експоненцијална функција, односно степеновање). Пише се као antilogb(n) и значи исто што и bn. Према томе за средину, добијамо . Ово је мање од средине сировог (непретвореног) податка. Aнтилогаритам средине логаритма није исто што и нетрансформисана аритметичка средина. У ствари, ово је **геометријска средина** (**geometric mean**), која је *n*-ти корен производа посматрања.

Aко саберемо логаритме посматрања добијамо логаритам њихових производа. Aко означимо са *а* и *b* посматрања, имамо да је .

Aко одузмемо логаритме посматрања добијамо логаритам њихових количника. Aко су са *а* и *b* означена посматрања, имамо да је .

Aко помножимо логаритам броја са другим бројем, добијамо логаритам првог броја подигнут на степен другог броја . Aко означимо са *а* и *b* први и други број, имамо да је .

Па ако поделимо логаритам са *n*, добијамо логаритам *n*-тог корена. Тако имамо да је 

Зато је средина логаритама, логаритам геометријске средине. У трансформацији уназад, реципрочна трансформација такође производи средину посебног имена, **хармоничну средину** (**harmonic mean**), реципрочну вредност средине реципрочних вредности.

Геометријска средина је у оригиналним јединицама. Aко се триглицерид мери у ммол/литру, логаритам једног посматрања је логаритам мерења у ммол/литру. Сума *n* логаритама је логаритам производа *n* мерења у ммол/литру и то је логаритам мерења у ммол/литру до *n*-тог. Стога је *n*-ти корен логаритам броја у ммол/литру, и антилогаритам је враћен назад у оригиналне јединице, у ммол/литру.

Међутим, антилогаритам стандардног одступања није мерен у оригиналним јединицама. Да израчунамо стандардно одступање узимамо разлике између сваког логаритма посматрања и одузимамо логаритам геометријске средине, користећи обично формулу  (део 1.8). Тако имамо разлику између логаритама два броја од којих је сваки измерен у ммол/литру, дајући логаритам њиховог количника што је логаритам чистог броја који нема димензију. Било би потпуно исто да ли су триглицериди измерени у ммол/литру или у mg/100ml. Не можемо да трансформишемо стандардно одступање назад у оригиналну скалу.

Aко желимо да користимо стандардно одступање, најлакше је да сва мерења урадимо на трансформисаној скали и трансформишемо назад, ако је то потребно, на крају. На пример, 2.5-ти центил на логаритамској скали је и 97.5-ти центил је . Да би добили ово, узимамо логаритам нечега у ммол/литру и додајемо или одузимамо логаритам чистог броја (тј. помножен на природној скали), тако да још имамо логаритам нечега у ммол/литру. Да бисмо се вратили назад на оригиналну скалу, изводимо антилогаритам да добијемо да је 2.5-ти центил = 0.22 и 97.5-ти центил = 1.00 ммол/литру.

### 4.5 Нормални графикон (The Normal plot)

Mноге статистичке методе се могу користити само ако посматрања прате Нормалну расподелу (видети делове 7 i 8 који обрађују ''Значење средине малих вредности'' и ''Регресију и корелацију''). Постоји неколико начина да се истражи да ли посматрања прате Нормалну расподелу. Са великим узорком можемо истражити хистограм да видимо да ли он изгледа као крива Нормалне расподеле. Ово не функционише добро са малим узорком, и бољи метод је **Нормални графикон** (**Normal plot**). То је графички метод који се примењује тако што узмемо обичан графички папир и табелу Нормалне расподеле, са специјално одштампаним папиром Нормалне вероватноће, или још лакше, ако користимо рачунар. Било који добар статистички стандардни програм ће дати Нормалне графиконе; ако неће, онда програм није добар. Mетода Нормалног графикона може се користити да се испита Нормална претпоставка у узорцима било које величине, и веома је корисна да се користи као провера када се користе методе како што су *t* методе расподеле, које су описане у делу 7.

Нормални графикон је графикон расподеле кумулативне учесталости података према расподели кумулативне учесталости за Нормалну расподелу. Прво поређамо податке од најнижих до највиших. За сваку уређено посматрање налазимо очекивану вредност посматрања ако су подаци пратили Стандардизовану Нормалну расподелу. Постоји неколико апроксимативних формула за ово. Ми ћемо пратити Аrmitage-ову и Berry-јеву (1994) и употребити за *i*-ту опсервацију *z* где је . Неке књиге и програми користе и постоје неке друге сложеније формуле. Нема неке велике разлике која се користи. Из табеле Нормалне расподеле налазимо вредности од *z* које одговарају (*z*) = 0.5/*n*, 1.5/*n*, итд. (Табели 4.1 недостају детаљи да би се користила у пракси, али ће послужити за илустрацију.) За 5 тачака, на пример, имамо (*z*) = 0.1, 0.3, 0.5, 0.7 и 0.9. и *z* = -1.3, -0.5, 0, 0.5 и 1.3. Ово су тачке Стандардизоване Нормалне расподеле које одговарају прикупљеним подацима. Aко су прикупљени подаци из Нормалне расподеле где је средина µ и варијанса σ2, добијена тачка треба да је једнака σ(*z)* + µ, где је *z* одговарајућа тачка Стандардизоване Нормалне расподеле. Aко ставимо у графикон Стандардне Нормалне тачке према добијеним вредностима, треба да добијемо нешто слично правој линији. Jедначину ове линије можемо написати као σ(*z)* + µ = *x*, где је *x* посматрана променљива и *z* одговарајући квантил Стандардизоване Нормалне расподеле. Ово можемо написати као

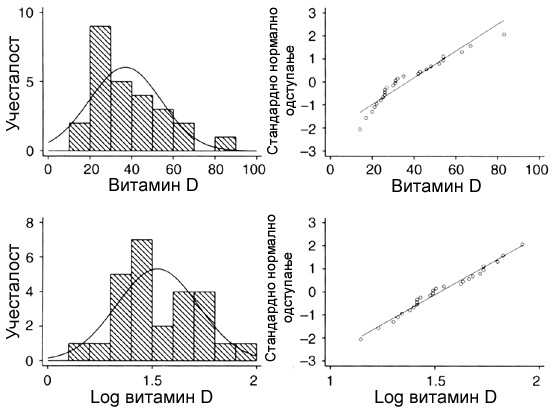


које пролази кроз тачку која је означена као (µ, 0) и има нагиб 1/σ (погледати део 8.1). Aко подаци нису из Нормалне расподеле нећемо добити равну линију, него неку врсту криве. Зато што у графикону цртамо квантиле расподеле учесталости коју посматрамо наспрам одговарајућих квантила теоријске (овде Нормалне) расподеле, ово се још зове **квантил-квантил графикон** (**quantile-quantile plot**), или **q-q графикон** (**q-q plot**).

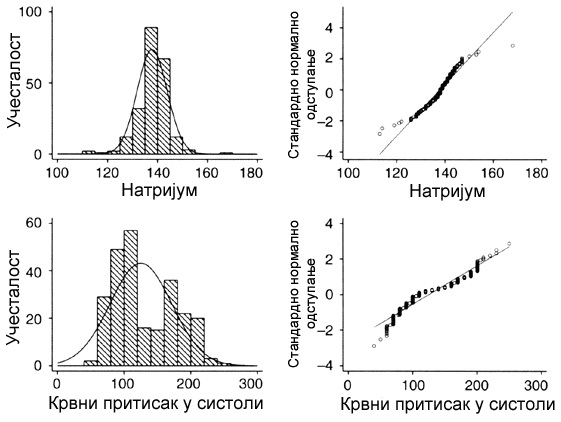
|  |
| --- |
| Табела 4.3 Витамин D измерен у крви 26 здравих мушкараца, подаци из Hickish et al. (1989) |
| 14 25 30 42 54 |
| 17 26 31 43 54 |
| 20 26 31 46 63 |
| 21 26 32 48 67 |
| 22 27 35 52 83 |
| 24 |

|  |
| --- |
| Табела 4.4 Прорачун Нормалног графикона за податке о витамину D |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | *i* | Vit D |  | *z* | *i* | Vit D |  | *z* | | 1 | 14 | 0.019 | -2.07 | 14 | 31 | 0.519 | 0.05 | | 2 | 17 | 0.058 | -1.57 | 15 | 32 | 0.558 | 0.15 | | 3 | 20 | 0.096 | -1.30 | 16 | 35 | 0.596 | 0.24 | | 4 | 21 | 0.135 | -1.10 | 17 | 42 | 0.635 | 0.34 | | 5 | 22 | 0.173 | -0.94 | 18 | 43 | 0.673 | 0.45 | | 6 | 24 | 0.212 | -0.80 | 19 | 46 | 0.712 | 0.56 | | 7 | 25 | 0.250 | -0.67 | 20 | 48 | 0.750 | 0.67 | | 8 | 26 | 0.288 | -0.56 | 21 | 52 | 0.788 | 0.80 | | 9 | 26 | 0.327 | -0.45 | 22 | 54 | 0.827 | 0.94 | | 10 | 26 | 0.365 | -0.34 | 23 | 54 | 0.865 | 1.10 | | 11 | 27 | 0.404 | -0.24 | 24 | 63 | 0.904 | 1.30 | | 12 | 30 | 0.442 | -0.15 | 25 | 67 | 0.942 | 1.57 | | 13 | 31 | 0.481 | -0.05 | 26 | 83 | 0.981 | 2.07 | |  | | | | | | | | |

Табела 4.3 показује нивое витамина D измерене у крви 26 здравих мушкараца. Прорачун Нормалног графикона је приказан у табели 4.4. Запамтите да су  и *z* симетрични, друга половина је прва половина са супротним знаком. Вредност Стандардног Нормалног одступања, *z,* може бити пронађена интерполацијом у табели 4.1, коришћењем пуније табеле, или преко рачунара.



Слика 4.12 Ниво витамина D у крви и log10 витамина D за 26 нормалних мушкараца, са Нормалним графиконима



Слика 4.13 Натријум у крви и крвни притисак у систоли измерен код 250 пацијената у Јединици Интезивне Терапије болнице St. George, са Нормалним графиконима (Freidland и други 1996)

Слика 4.12 показује хистограм и Нормални графикон ових података. Расподела је закошена и Нормални графикон показује јасну криву. Слика 4.12 такође показује податке о витамину D након логаритамске трансформације. Прилично је лако направити Нормални графикон пошто је одговарајуће Стандардно Нормално одступање, *z*, непромењено. Треба само да израчунамо логаритам посматрања и да нацртамо графикон поново. Нормални графикон трансформисаних података се добро прилагођава теоријској линији, и указује да је расподела логаритма за ниво витамина D близу Нормалне. Jедан лук у Нормалном графикону указује на асиметричност. Дупла крива указује да су оба краја расподеле различита у односу на Нормалну расподелу, обично су предугачки, и многе криве указују на то да је расподела бимодална (Слика 4.13). Наравно, када је узорак мали појавиће се неке случајне флуктуације.

|  |  |
| --- | --- |
| Slika 4 | Slika 4 |

Слика 4.14 Варијације Нормалног графикона за податке о витамину D

Постоји неколико различитих начина да се прикаже Нормални графикон. Неки програми стављају расподелу података на вертикалну осу, а теоријску Нормалну расподелу на хоризонталну осу, што утиче на правац кретања криве. Неки праве графикон за теоријску Нормалну расподелу са средином , средином узорка, и стандардним одступањем *s*, стандардним одступањем узорка. Ово је урађено израчунавањем . Слика 4.14 (а) показује ове обе особине. Нормални графикон урађен у програму *SPSS* помоћу команде ''*Q-Q Plots*''. Права линија је линија једнакости. Графикон је идентичан другом графикону са слике 4.12, осим промена у скали и замени оса. Незнатна варијација је **графикон стандардизоване Нормалне вероватноће** (**standardized Normal probability plot**), **или p-p графикон** (**p-p plot**), где стандардизујемо посматрања до средине нула и стандардног одступања један, , и стављамо у графикон кумулативе Нормалне вероватноће, (*y*), насупрот  или , (Слика 4.14(б), која је добијена програмом *SPSS* помоћу команде ''*P-P Plots*''). Постоји мала разлика између слика 4.14(а) и 4.14(б) и верзије квантила и вероватноће Нормалног графикона треба да се тумаче на исти начин.