UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Finančna matematika – 1. stopnja

Sabina Plemenitaš

TESTIRANJE NORMALNOSTI

Delo diplomskega seminarja

Mentor: doc. dr. Jaka Smrekar

Ljubljana, 2013

Kazalo

Slike

Tabele

Testiranje normalnosti

Povzetek

Pomembnost statistike je v vsakdanjosti zelo velika. Odločitve lažje sprejemamo, če imamo za to neko statistično podlago, ki podpira naša predvidevanja, pa naj gre za področje športa, biomedicine, psihologije, finančega sveta, ali kaj povsem drugega. Povsod lahko naletimo na statistično analiziranje podatkov.

Ob proučevanju določene lastnosti neke statistične populacije se srečujemo z vprašanjem, kako je slučajna spremenljivka, ki numerično kvalificira to lastnost, porazdeljena, saj je naše vsakršno nadaljnje analiziranje učinkovitejše ob primerni omejitvi nabora vseh možnih porazdelitev. Tipično se obrnemo na privzetke o normalnosti. Zaradi ugodnih lastnosti normalne porazdelitve lahko namreč nudimo kvalitetno ocenjevanje in proučevanje slučajne spremenljivke.

Seveda je potrebno vsakršen nedokazan privzetek o porazdelitvi slučajne spremenljivke ustrezno utemeljiti, za kar so nam na voljo statistični testi. V našem primeru se bomo srečevali s privzetkom normalnosti in eden izmed testov, ki testira, ali dani slučajni vzorec ustreza privzetku normalnosti, je d'Agostino-Pearsonov test. Le-ta temelji na testiranju asimetrije in sploščenosti, ki sta tipična kazalca normalnosti oziroma „nenormalnosti".

Kljub temu, da d'Agostino-Pearsonov test ni eksakten, vključuje dovolj dobre približke, na podlagi katerih lahko dokaj jasno sklepamo o privzetku normalnosti. Je močno orodje, ki na podlagi testiranja „življenjskih" realizacij zavrne hipotezo normalnosti ob vsakršnem dovolj velikem odstopanju od normalne porazdelitve, pa čeprav lahko grafična predstavitev podatkov nakazuje na spodobno obliko normalne krivulje.

Tests of normality

Abstract

Statistics is of great importance in everyday life. We can reach our decisions more easily if there is some statistical basis which supports our predictions, be it in the area of sports, biomedicine, psychology, finance or something else entirely. We can come across statistical analysis of data basically everywhere.

When examining a certain property of a statistical population, we are faced with the problem of how the random variable that numerically qualifies that property is distributed, since every further analysis we make is more efficient if the collection of all admissible distribution is suitably limited. Typically, we use the assumption of normality. Due to beneficial features of the normal distribution, the random variable can be assessed and examined appropriately.

Any unproven assumptions regarding the distribution of the random variable have to be substantiated accordingly with any of the statistical tests at hand. In our case, we are going to come across the assumption of normality; one of the tests used to figure out if the given random sample is normally distributed is the d'Agostino-Pearson test. It is based on testing of skewness and kurtosis, which are typical indicators of normality or non-normality.

The d'Agostino-Pearson test is not exact; however, it includes approximations which are good enough to make clear conclusions on the hypothesis of normality. It is a strong enough tool which can refute the hypothesis of normality by life realisation testing when the deviation from normal distribution is big enough, even though the graphic display of data indicates the appropriate form of a normal curve.

Math. Subj. Class. (2010): 60E05, 62F03, 62G10, 62F12, 62A01.

Ključne besede: asimetrija, sploščenost, cenilka asimetrije, cenilka sploščenosti, interval zaupanja, normalna porazdelitev, porazdelitev hi-kvadrat, testiranje hipotez, d'Agostino-Pearsonov K^2 test.

Keywords: skewness, kurtosis, tests of skewness, tests of kurtosis, confidence interval, normal distribution, chi-squared distribution, hypothesis testing, d'Agostino-Pearson K^2 test.

1. Inferenčna statistika

Statistika je veda, ki se ukvarja predvsem z analizo, interpretacijo in primerno predstavitvijo množic podatkov. Lahko se osredotočimo na analizo zbranih podatkov brez posploševanj, o čemer nam govori deskriptivna veja statistike, medtem ko je druga veja statistike inferenčna statistika, ki temelji na sklepanju iz lastnosti nekega vzorca populacije na lastnosti celotne statistične populacije. Pri tem s pojmom statistična populacija mislimo na tipično zelo veliko množico proučevanih podatkov oziroma elementov, ki jih imenujemo tudi statistične enote.

Kot že namiguje naslov poglavja, se bomo posvetili inferenci, saj je ta veja v vsakdanjosti pogostejša in ravno na tem področju je uporabnost testa normalnosti večja.

1.1. **Tipičen problem.**

Proučujemo določeno lastnost neke statistične populacije. Vsak njen element razumemo kot možen izid slučajnega eksperimenta, ki ga dosežemo z neko verjetnostjo. Proučevano lastnost numerično kvalificiramo v obliki slučajne spremenljivke. Slučajni vzorec je nato zaporedje neodvisnih replikacij slučajne spremenljivke X. Zanimajo nas parametri kot sta na primer matematično upanje ali disperzija. V praksi jih ne poznamo, saj imamo na voljo samo konkretne realizacije slučajnega vzorca, vendar pa zato iščemo čim bolj natančne ocene željenih parametrov. Pri tem upoštevamo dejstvo, da imamo realnoštevilsko slučajno spremenljivko. Denimo da ocenjujemo matematično upanje (normalno porazdeljene) slučajne spremenljivke X, ki ga označimo z μ.

Pri ocenjevanju se srečujemo s pojmom cenilke, ki ga bomo razložili še podrobneje v enem izmed prihodnjih poglavij, zaenkrat omenimo le, da je najboljša nepristranska cenilka za μ vzorčno povprečje, ki ga označimo z $\overline{X}$. Ta cenilka je preprosta ocena matematičnega upanja, s katero nimamo dovolj velikih možnosti, da zadanemo točno vrednost, zato si pomagamo z intervali zaupanja in tako zajamemo neko območje, na katerem naj bi se vrednost matematičnega upanja nahajala. Možnost, da je točna vrednost ocenjenega parametra v intervalu, je v tem primeru večja, zato je ta intervalska ocena sprejemljivejša.

Definicija 1.1. Naj bo n velikost vzorca in $\beta \in (0, 1)$ primerno dovolj veliko (dovolj blizu 1) število. *Interval zaupanja* za μ stopnje zaupanja β za vzorce velikosti n sestavljata taki merljivi funkciji vzorca $L : \mathbb{R}^n \to \mathbb{R}$ in $U : \mathbb{R}^n \to \mathbb{R}$, za kateri velja:

$$P(L(X_1, X_2, ..., X_n) \leq \mu \leq U(X_1, X_2, ..., X_n)) \geq \beta,$$

za vse možne porazdelitve slučajne spremenljivke X.

Opomba 1.2. S stopnjo zaupanja β navzdol omejimo verjetnost tistih vzorcev, za katere interval zaupanja vsebuje dejanski (iskani) parameter μ in temu primerno določimo funkciji L in U. Pazimo tudi, da funkciji določimo tako, da (v našem primeru) dobimo čim krajši interval zaupanja.

Za konstrukcijo natančnih intervalov zaupanja so potrebne dokaj stroge omejitve na množico dopustnih porazdelitev oziroma „model” za slučajno spremenljivko X. V primeru, ko imamo podan dovolj velik slučajni vzorec lahko z uporabo aproksimacije določimo približno porazdelitev slučajne spremenljivke X, pri tem pa dejanska porazdelitev ne igra ključne vloge. Če imamo na voljo majhen vzorec, takšnih sklepanj ne moremo uporabiti, zato se obrnemo na privzetke o dopustnih porazdelitvah, ki pa jih je potrebno še preveriti in potrditi. Takšnim problemom se bomo posvetili nekoliko kasneje, najprej navedimo vse potrebne obravnave pojmov in ustreznih definicij.

2. Zvezna slučajna spremenljivka

Iz verjetnosti in ostalih vej matematike vemo, da poznamo dva najpomembnejša tipa slučajnih spremenljivk - *diskretne* in *zvezne*. Zaradi narave problemov, ki se jim bomo v nadaljevanju posvečali, diskretnih slučajnih spremenljivk ne bomo obravnavali, navedli bomo le bistveno razliko med diskretnimi in zveznimi slučajnimi spremenljivkami, nato pa se bomo osredotočili le na slednje. Obravnavali bomo tudi za testiranje normalnosti najpomembnejša primera zveznih porazdelitev.

Razlika med omenjenimi slučajnimi spremenljivkami temelji na lastnosti zveznih slučajnih spremenljivk. Taka spremenljivka zavzame katerokoli številsko vrednost z verjetnostjo 0. Navedimo definicijo.

Definicija 2.1. Slučajna spremenljivka X je *zvezna*, če velja:

$$P(X = x) = 0, \forall x \in \mathbb{R}.$$

2.1. **Kumulativna porazdelitvena funkcija.**

Definicija 2.2. *Splošna kumulativna porazdelitvena funkcija* je funkcija, ki predstavlja verjetnost, da zvezna slučajna premenljivka X zavzame vrednost manjšo ali enako x. Zapišemo:

$$F_X(x) = P(X \leq x), \forall x \in \mathbb{R}.$$

Opomba 2.3. Zaradi zveznosti slučajne spremenljivke X veljajo enakosti:

$$F_X(x) = P(X < x) = P(X \leq x), \forall x \in \mathbb{R}.$$

Vsaka porazdelitvena funkcija je nepadajoča, zvezna z desne, njene vrednosti so vselej na intervalu $[0, 1]$, definirana pa je na množici realnih števil.

2.2. **Gostota porazdelitve.**

Definicija 2.4. Če za kumulativno porazdelitveno funkcijo slučajne spremenljivke X obstaja takšna integrabilna funkcija $f(t)$, da velja:

$$F_X(x) = \int_{-\infty}^{x} f(t)dt,$$

potem ji rečemo *gostota porazdelitve*.

Opomba 2.5. Če je slučajna spremenljivka X *absolutno zvezna*, potem za njeno kumulativno porazdelitveno funkcijo in gostoto velja:

$$f(x) = F'_X(x), \text{ za skoraj vse } x.$$

Gostota porazdelitve je vselej nenegativna, velja tudi normiranost, kar pomeni, da velja $\int_{-\infty}^{\infty} f(x)dx = 1$, s pomočjo gostote lahko dobimo torej tudi kumulativno porazdelitveno funkcijo slučajne spremenljivke.

2.3. **Normalna porazdelitev.**

Normalna porazdelitev, drugače poznana tudi kot Gaussova porazdelitev, je zahvaljujoč svojim lastnostim najpogosteje uporabljena porazdelitev na področju matematike, statistike in ekonometrije. Gre za zvezno verjetnostno porazdelitev vrednosti statističnih enot v statistični populaciji, ki je v grafični predstavitvi oblikovana kot tako imenovana normalna krivulja.

Normalna porazdelitev je natančno določena s parametrama $\mu \in \mathbb{R}$, ki predstavlja matematično upanje slučajne spremenljivke X, in $\sigma > 0$, ki predstavlja standardni odklon, le-ta je izračunan kot kvadratni koren disperzije ter določa razpršenost vrednosti slučajne spremenljivke X okoli matematičnega upanja. Prostor parametrov normalne porazdelitve je torej: $(\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$.

2.3.1. ***Kumulativna porazdelitvena funkcija.***

Definicija 2.6. Če velja $X \sim N(\mu, \sigma)$, potem ima slučajna spremenljivka X kumulativno porazdelitveno funkcijo:

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

2.3.2. ***Gostota porazdelitve.***

Definicija 2.7. Če velja $X \sim N(\mu, \sigma)$, potem ima slučajna spremenljivka X gostoto oblike:

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

2.3.3. ***Normalna krivulja.***

Graf gostote porazdelitvene funkcije normalne porazdelitve, prikazan na sliki 1, ima ugodne lastnosti, zanj veljajo naslednje trditve:

Trditev 2.8. *Graf je simetričen, zato imata matematično upanje in mediana enako vrednost, kar pomeni, da μ zavzame vrednost, od katere ima natanko polovica enot manjšo vrednost in natanko polovica enot večjo vrednost.*

Trditev 2.9. *Maksimum doseže v točki: $t_{max} = \mu$.*

Dokaz. Za iskanje ekstremov izračunamo prvi odvod gostote porazdelitvene funkcije normalne porazdelitve in ga enačimo z 0:

$$f'(t) = \frac{1}{\sigma\sqrt{2\pi}} \Big(-\frac{t}{\sigma^2} + \frac{\mu}{\sigma^2} \Big) e^{-\frac{(t-\mu)^2}{2\sigma^2}} = 0,$$

kar pomeni, da iščemo rešitev, pri kateri bo veljalo:

$$\frac{1}{\sigma\sqrt{2\pi}} \Big(\frac{t}{\sigma^2} \Big) e^{-\frac{(t-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} \Big(\frac{\mu}{\sigma^2} \Big) e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

Slednje je mogoče v točki $t = \mu$.
Ker velja $f''(\mu) < 0$, imamo v tej točki maksimum. Torej drži $t_{max} = \mu$. □

Trditev 2.10. *Ima dva prevoja v točkah: $t_{1,2} = \mu \pm \sigma$.*

Dokaz. Za določanje prevojev izračunamo drugi odvod gostote porazdelitvene funkcije in ga enačimo z 0:

$$f''(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \Big(-\frac{1}{\sigma^2} - \frac{t}{\sigma^2}\big(-\frac{t}{\sigma^2} + \frac{\mu}{\sigma^2}\big) + \frac{\mu}{\sigma^2}\big(-\frac{t}{\sigma^2} + \frac{\mu}{\sigma^2}\big)\Big) = 0.$$

Enačbo poenostavimo in dobimo:

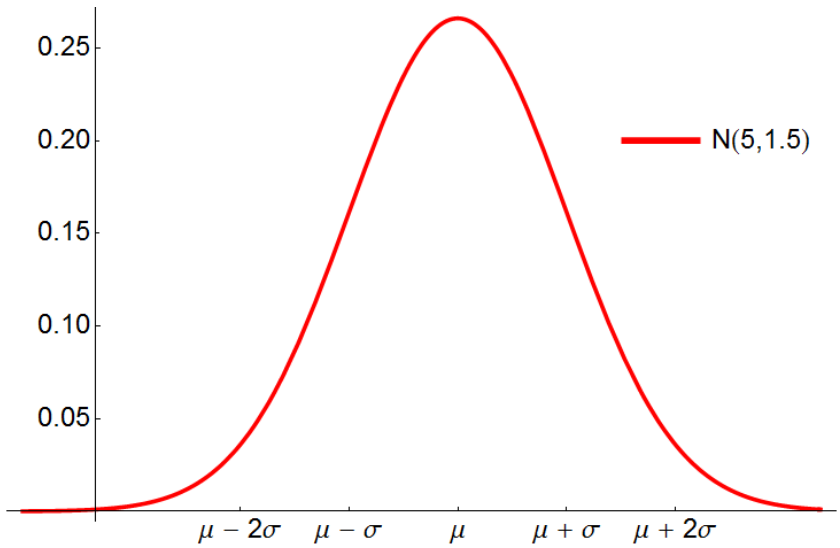
$$t^2 - 2t\mu + \mu^2 - \sigma^2 = 0,$$

torej iščemo rešitvi enačbe:

$$(t-\mu)^2 = \sigma^2$$

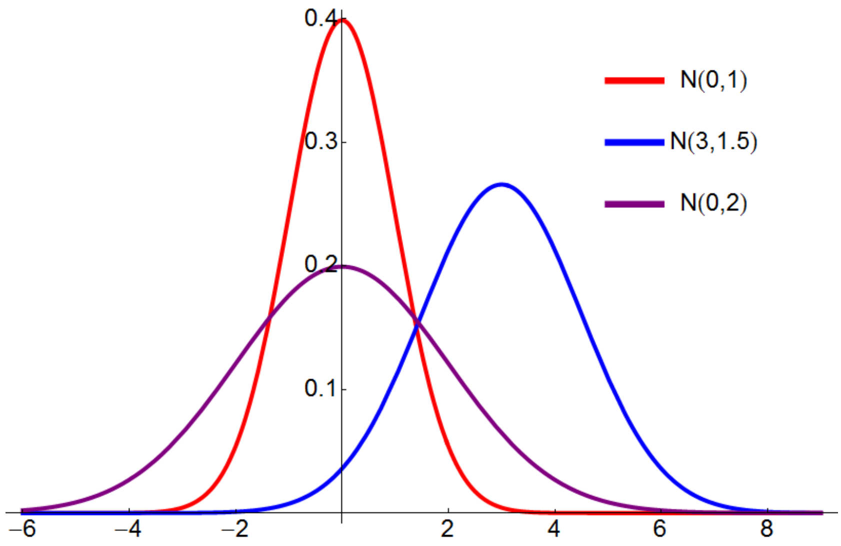
in dobimo prevoja $t_1 = \mu + \sigma$, $t_2 = \mu - \sigma$. □

Trditev 2.11. *Normalni krivulji ustreza tako imenovano empirično pravilo oziroma pravilo* $\pm 2\sigma$*, kar pomeni, da se znotraj intervala* $[\mu - 2\sigma, \mu + 2\sigma]$ *nahaja približno* 95% *vseh vrednosti.*



Slika 1. Normalna krivulja

V družino normalnih porazdelitev, katere primer vidimo na sliki 2, uvrščamo porazdelitve z različnima vrednostima parametrov μ in σ. Parameter μ določa vrednost, v kateri normalna krivulja doseže maksimum, medtem ko parameter σ opisuje razpršenost vrednosti okoli parametra μ. Ločimo visoko in nizko razpršenost.



SLIKA 2. Družina normalnih krivulj

2.3.4. ***Standardna normalna porazdelitev.***

Standardna normalna porazdelitev je natanko tista normalna porazdelitev, ki je točno določena z vrednostma parametrov $\mu = 0$ in $\sigma = 1$. S to porazdelitvijo je dandanes mogoče zelo enostavno in hitro računati, zato jo pogosto uporabljamo na področju statistike. Vsako normalno porazdelitev je možno tudi transformirati v standardno normalno porazdelitev, kar velja za še posebej pomembno lastnost normalnih porazdelitev. Formalno:

Trditev 2.12. *Če velja $X \sim N(\mu, \sigma)$, potem je $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$. Postopek vpeljave nove spremenljivke Z v normalno porazdelitev imenujemo standardizacija normalne porazdelitve, pri tem dobimo kumulativno porazdelitveno funkcijo:*

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt,$$

kjer je $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ gostota standardne normalne porazdelitve.

Dokaz. Z uporabo nastavka $Z = \frac{X-\mu}{\sigma}$ izračunamo:

$$\Phi(z) = P(Z \leq z) = P\Big(\frac{X-\mu}{\sigma} \leq z\Big) = P(X \leq z\sigma + \mu) =$$
$$\int_{-\infty}^{z\sigma+\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt.$$

Dobimo torej standardno porazdelitveno funkcijo, iz katere lahko razberemo tudi gostoto standardne porazdelitve. □

Verjetnost, da slučajna spremenljivka $Z \sim N(0, 1)$ zavzame vrednost manjšo od neke konkretne vrednosti z, zlahka odčitamo iz, v statistiki dobro poznane, tabele vrednosti standardne normalne porazdelitve, s pomočjo katere lahko izračunamo tudi naslednje verjetnosti:

- $P(Z < -a) = P(Z > a) = 1 - P(Z \leq a) = 1 - \Phi(a)$, kjer je $a > 0$,
- $P(a < Z \leq b) = P(Z \leq b) - P(Z \leq a) = \Phi(b) - \Phi(a)$.

Za nas pomemben del takšne tabele je prikazan v tabeli 1, označene vrednosti bomo namreč kasneje uporabili pri testiranju normalnosti, več o tem sledi v enem izmed prihodnjih poglavij.

	...	0.04	0.05	*0.06*	0.07	...
⋮		⋮	⋮	⋮	⋮	
1.6	...	0.9495	0.9505	0.9515	0.9525	...
1.7	...	0.9591	0.9599	0.9608	0.9616	...
1.8	...	0.9671	0.9678	0.9686	0.9693	...
1.9	...	0.9738	0.9744	*0.9750*	0.9756	...
2.0	...	0.9793	0.9798	0.9803	0.9808	...
⋮		⋮	⋮	⋮	⋮	

Tabela 1. Del tabele vrednosti standardne normalne porazdelitve

2.4. **Porazdelitev hi-kvadrat.**

Poradelitev hi-kvadrat je zvezna porazdelitev, ki jo na en način lahko predstavimo kot poseben primer Gama porazdelitve, in sicer velja $\chi^2_n = \text{Gama}(\frac{n}{2}, \frac{1}{2})$. Na drug način jo lahko predstavimo kot vsoto kvadratov n neodvisnih standardnih normalnih spremenljivk. Navedimo kar trditev.

Trditev 2.13. *Naj bodo $X_i, i = 1, 2, ..., n$, neodvisne in enako porazdeljene, vse po zakonu $N(0, 1)$. Potem je vsota kvadratov slučajnih spremenljivk X_i porazdeljena po zakonu hi-kvadrat z n prostostnimi stopnjami. Pišemo:*

$$X_1^2 + X_2^2 + ... + X_n^2 \sim \chi^2_n. \tag{1}$$

Porazdelitev hi-kvadrat zelo pogosto uporabljamo pri statističnem testiranju.

2.4.1. ***Kumulativna porazdelitvena funkcija.***

Definicija 2.14. Če velja $X \sim \chi^2_n$, potem ima slučajna spremenljivka X kumulativno porazdelitveno funkcijo:

$$F_X(x) = \frac{\gamma(\frac{n}{2}, \frac{x}{2})}{\Gamma(\frac{n}{2})},$$

kjer je $\gamma(\frac{n}{2}, \frac{x}{2}) = \int_0^{\frac{x}{2}} t^{\frac{n}{2}-1} e^{-t} dt$ in $\Gamma(\frac{n}{2}) = \int_0^{\infty} t^{\frac{n}{2}-1} e^{-t} dt$.

2.4.2. ***Gostota porazdelitve.***

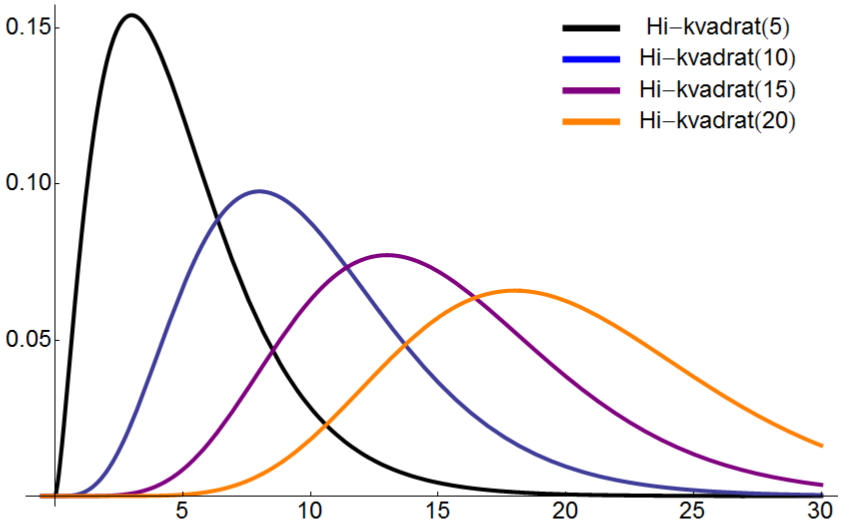
Definicija 2.15. Če velja $X \sim \chi^2_n$, potem ima slučajna spremenljivka X gostoto porazdelitve:

$$f(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}, \forall x > 0.$$

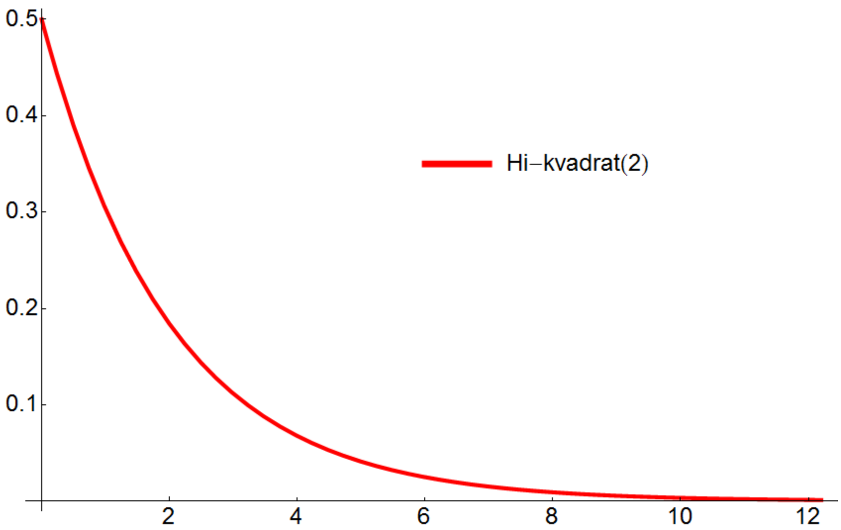
2.4.3. ***Krivulja gostote porazdelitve.***

Za krivuljo gostote porazdelitve hi-kvadrat je značilno, da z naraščanjem prostostnih stopenj graf dobiva podobno obliko, kot jo ima neka normalna krivulja. Slednjo lastnost lahko opazimo na primeru družine krivulj hi-kvadrat porazdelitve na sliki 3.

Sicer bo kasneje pri testiranju normalnosti pomembna porazdelitev hi-kvadrat z dvema prostostnima stopnjama, katere graf gostote je posebej izrisan na sliki 4.



Slika 3. Družina krivulj gostote porazdelitve hi-kvadrat



Slika 4. Krivulja gostote porazdelitve hi-kvadrat z dvema prostostnima stopnjama

V statistiki in verjetnosti je vpeljava pojma moment slučajne spremenljivke X pomembna, momenti različnih redov namreč opisujejo pričakovano vrednost, odklon, asimetrijo, sploščenost, itd., ki so nam na omenjenih področjih v veliko korist. Kot rečeno v realnosti pogosto nimamo podanih dovolj natančnih podatkov in hkrati pri proučevanju in analiziranju statističnih populacij ne moremo posegati po celoti, se pa srečujemo z vzorci, ki so izvzeti iz celotne populacije in zato ne poznamo natančnih vrednosti momentov, velja pa načelo, da vse, kar je mogoče izraziti z momenti, lahko tudi ocenimo s pomočjo momentov, ki jih izračunamo na podlagi realizacij oziroma slučajnih vzorcev in tako dobimo ocene za momente.

Definicija 3.1. *Necentralizirani moment slučajne spremenljivke X reda k*, kjer je $k \in \mathbb{N}$, je vpeljan kot:

$$E(X^k).$$

Le-ta obstaja, če obstaja matematično upanje $E(|X|^k)$.

Definicija 3.2. *Centralizirani moment slučajne spremenljivke X reda k*, kjer je $k \in \mathbb{N}$, je:

$$E((X - E(X))^k).$$

Le-ta obstaja, če obstaja matematično upanje $E(|X - E(X)|^k)$.

Opomba 3.3. Pri $k = 1$ dobimo *prvi* oziroma *začetni necentralizirani moment*, ki mu pravimo matematično upanje. Momente, pri katerih je $k \geq 2$, imenujemo *centralizirani momenti.*

Definicija 3.4. *Vzorčni k-ti centralizirani moment realizacije slučajnega vzorca* $(x_1, x_2, ..., x_n)$ je:

$$m_k = m_k(x_1, ..., x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k, \forall k > 1, \tag{2}$$

pri čemer je z x_i označenih n realizacij slučajne spremenljivke X, $\bar{x}$ pa predstavlja prvi necentralizirani vzorčni moment oziroma vzorčno povprečje, ki ustreza formuli $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Definicija 3.5. *Cenilka* je slučajna spremenljivka, ki na podlagi slučajnega vzorca ocenjuje neki neznan parameter, ki nas zanima.

Opomba 3.6. Cenilka je *nepristranska*, če je njeno matematično upanje enako parametru, ki ga ocenjujemo.

S pomočjo vzorčnih momentov med drugim ocenjujemo tudi asimetričnost in sploščenost slučajne spremenljivke X.

3.1. **Asimetrija.**

Asimetrija je eden od kazalecev, ki kaže na to, ali dana slučajna spremenljivka X ustreza normalni porazdelitvi ali ne. Natančneje asimetrija „meri” morebitno nesimetričnost verjetnostne porazdelitve slučajne spremenljivke.

Definicija 3.7. *Koeficient asimetričnosti* je normiran 3. centralizirani moment, če le-ta obstaja. Natančno vrednost koeficienta asimetričnosti slučajne spremenljivke X izračunamo po formuli:

$$A(X) = \frac{E((X-\mu)^3)}{\sigma(X)^3}.$$

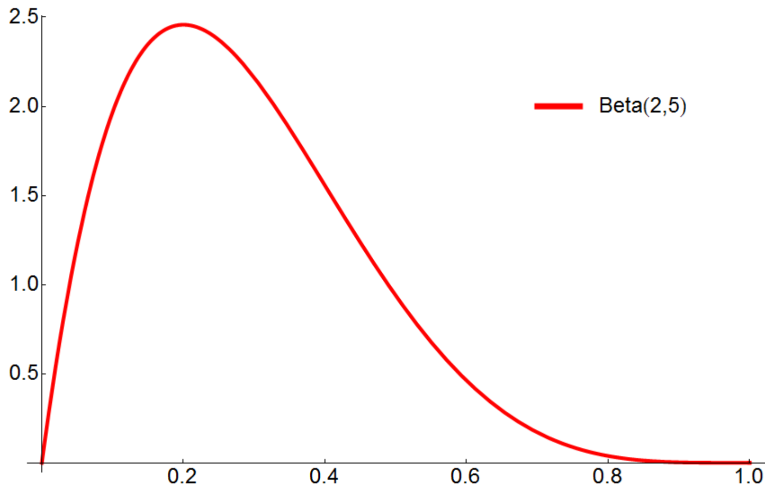
Za normalno porazdeljeno slučajno spremenljivko X je koeficient asimetričnosti enak 0. V kolikor se vrednosti koeficienta asimetričnosti občutno razlikujejo od 0, o normalnosti ne moremo govoriti.

Dejstvo, kako se vrednosti asimetrije razlikujejo od 0, nas v splošnem privede do dveh „šolskih" primerov asimetričnosti neke slučajne spremenljivke. Poglejmo oba primera.

3.1.1. ***Pozitivna asimetrija.***

Pri pozitivni asimetriji velja, da koeficient asimetričnosti zavzame pozitivno vrednost, torej $A(X) > 0$. Slednje se na krivulji gostote porazdelitve kaže tako, da je desni del krivulje nekoliko daljši, levo od parametra μ pa je ploščina med krivuljo ter abscisno osjo bistveno večja kot pri normalni porazdelitvi. Pri „življenjskih" porazdelitvah velja tudi, da matematično upanje zavzame večjo vrednost kot mediana. Torej $\mu > Me$.

Slika 5 prikazuje krivuljo gostote slučajne spremenljivke X, ki ima Beta porazdelitev. Velja $X \sim \mathrm{Beta}(2,5)$, ki je primer pozitivne asimetrije.

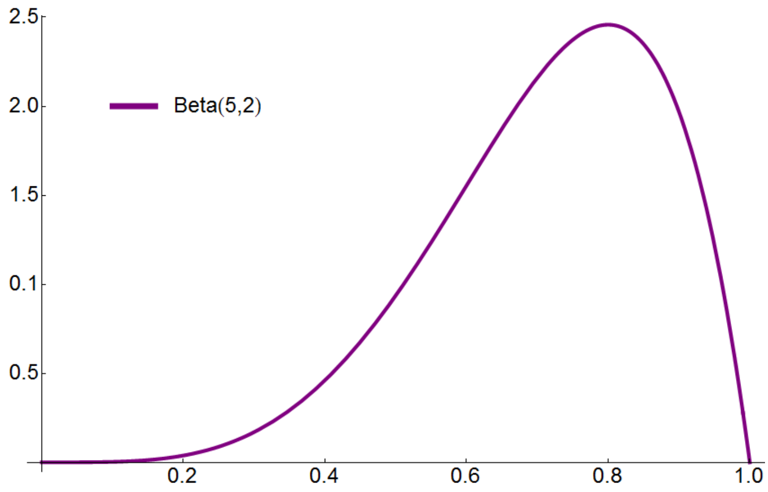


Slika 5. Pozitivna asimetrija

3.1.2. ***Negativna asimetrija.***

V primeru negativne asimetrije zavzame koeficient negativno vrednost, torej velja $A(X) < 0$. Krivulja je ravno „zrcalna" kot pri pozitivni asimetriji, kar pomeni, da je levi del asimetrične krivulje daljši, večina vrednosti porazdelitve pa se nahaja desno od μ. Pri „življenjskih" primerih velja $\mu < Me$.

Na sliki 6 je primer negativne asimetrije. Krivulja ustreza gostoti porazdelitve slučajne spremenljivke $X \sim \mathrm{Beta}(5,2)$.



SLIKA 6. Negativna asimetrija

3.2. **Sploščenost.**

Sploščenost, drugače tudi ukrivljenost ali s tujko kurtosis, je prav tako eden izmed kazalcev, ki kažejo na to, ali dana slučajna spremenljivka X ustreza normalni porazdelitvi. „Meri" koničastost oziroma ploščatost gostote verjetnostne porazdelitve slučajne spremenljivke.

Definicija 3.8. *Koeficient sploščenosti* je normiran 4. centralizirani moment, če le-ta obstaja. Natančno vrednost koeficienta sploščenosti slučajne spremenljivke X izračunamo po formuli:

$$K(X) = \frac{E((X-\mu)^4)}{\sigma(X)^4}. \tag{3}$$

V primeru, ko velja $X \sim N(\mu, \sigma)$, zavzame sploščenost vrednost 3. Če ocena vrednosti sploščenosti močno odstopa od teoretične, potem dvomimo o normalnosti.

Opomba 3.9. Pri obravnavanju sploščenosti lahko definiramo tudi tako imenovano *ekscesno sploščenost*:

$$EK(X) = \frac{E((X-\mu)^4)}{\sigma(X)^4} - 3. \tag{4}$$

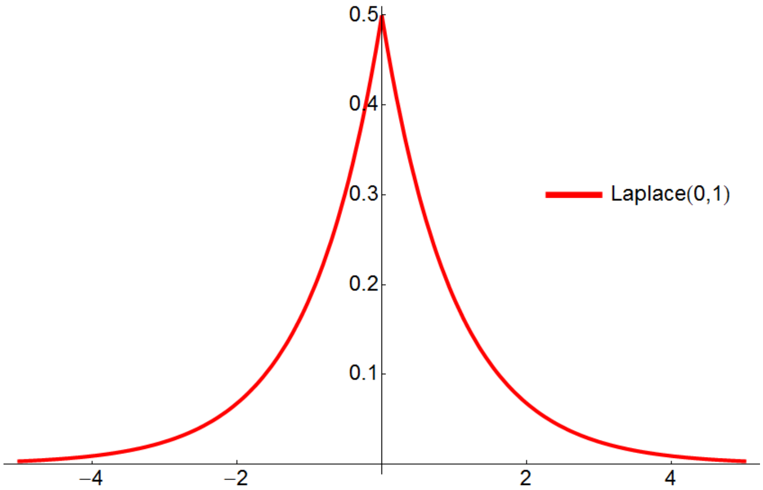
Razlika med formulama (3) in (4) je le v tem, da pri normalno porazdeljeni slučajni spremenljivki velja $EK(X) = 0$, medtem ko je $K(X) = 3$. Pri nadaljnjem obravnavanju sploščenosti se bomo držali formule (3).

Vrednosti različne od 3 torej nakazujejo na drugačno sploščenost od normalne porazdelitve in prav tako kot pri asimetriji ločimo dve možnosti sploščenosti.

3.2.1. ***Pozitivna sploščenost.***

Porazdelitev ima pozitivno sploščenost, velja $K(X) > 0$, kadar je krivulja gostote porazdelitve koničasta, kar pomeni, da ima ostrejši vrh ter daljši rep. Porazdelitev s pozitivno sploščenostjo v angleškem jeziku imenujemo *leptokurtic distribution*.

Primer porazdelitve, ki ima pozitivno sploščeno krivuljo gostote porazdelitve, je Laplaceova porazdelitev, ki jo prikazuje slika 7.

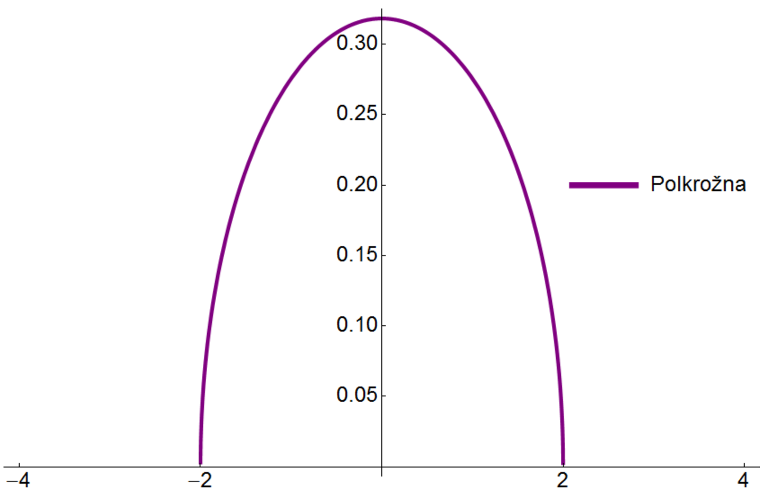


SLIKA 7. Pozitivna sploščenost

3.2.2. ***Negativna sploščenost.***

Porazdelitev z negativna sploščenostjo, v angleščini imenovana *platykurtic distribution*, je porazdelitev, katere krivulja gostote je sploščena. Ima torej nižji in zaobljen vrh ter krajši rep kot krivulja gostote normalne porazdelitve. V tem primeru je $K(X) < 0$.

Primer porazdelitve z negativno sploščenostjo je polkrožna porazdelitev, katere krivulja je na sliki 8.



SLIKA 8. Negativna sploščenost

3.3. **Cenilka asimetrije.**

Definicija 3.10. *Cenilka za asimetrijo* je oblike:

$$\hat{a}(x_1, x_2, ..., x_n) = \frac{m_3}{m_2^{\frac{3}{2}}}, \tag{5}$$

kjer m_k izračunamo po formuli (2).

Opomba 3.11. Če $X \sim N(\mu, \sigma)$, potem velja: $\lim_{n\to\infty} \hat{a}(X_1, X_2, ..., X_n) = 0$ (v verjetnosti).
Z besedami: če imamo normalno porazdeljeno slučajno spremenljivko X, potem cenilka asimetrije, izračanana na podlagi slučajnega vzorca velikosti n, konvergira verjetnostno k teoretični vrednosti 0, ko n pošljemo proti neskončno.

Trditev 3.12. *Naj velja $X \sim N(\mu, \sigma)$ ter naj bo dan slučajni vzorec velikosti n. Cenilka $\hat{a}$, izračunana na podlagi vzorca po formuli* (5), *je približno normalno porazdeljena z matematičnim upanjem* 0 *in standardnim odklonom $\sqrt{\frac{6}{n}}$. Velja še:*

$$\lim_{n\to\infty} A(\hat{a}) = 0,$$

$$\lim_{n\to\infty} K(\hat{a}) = 3.$$

Opomba 3.13. Da drži zgornja trditev, mora biti dan slučajni vzorec dovolj velik, sicer aproksimacije ne moremo izpeljati. Standardni odklon in sploščenost cenilke asimetrije za vzorec velikosti n sta podana z formulama (dokaz bomo izpustili):

$$\sigma(\hat{a}) = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} \tag{6}$$

$$K(\hat{a}) = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)} \tag{7}$$

Pričakujemo torej, da bodo momenti slučajne spremenljivke $\hat{a}$ z naraščanjem velikosti vzorca vedno bližje omenjenim asimptotičnim vrednostim. Slednje izračune, pridobljene s pomočjo formul (6) in (7), zabeležimo v spodnjo tabelo, ki predstavlja simulacijo vzorčenja iz normalne porazdelitve.

n	$\mu(\hat{a})$	$\sigma(\hat{a})$	$A(\hat{a})$	$K(\hat{a})$
25	0	0.435	0	3.578
50	0	0.326	0	3.452
100	0	0.238	0	3.284
250	0	0.153	0	3.131
500	0	0.109	0	3.069
1000	0	0.077	0	3.035
3000	0	0.045	0	3.012
5000	0	0.035	0	3.007
asimptotično	0	$\sqrt{\frac{6}{n}}$	0	3

Tabela 2. Simulacija vzorčenja za cenilko asimetrije iz normalne porazdelitve

Iz tabele 2 je moč opaziti, da potrebujemo zares velike vzorce podatkov, da dosegamo dovolj natančne vrednosti vzorčnih momentov. Velika odstopanja se občutijo tudi pri velikosti $n = 250$, medtem ko trdne približke, iz katerih bi morda lahko sklepali o normalnosti slučajne spremenljivke, dobimo šele pri $n = 5000$.

3.4. **Cenilka sploščenosti.**

Definicija 3.14. *Cenilka za sploščenost* je oblike:

$$\hat{k}(x_1, x_2, ..., x_n) = \frac{m_4}{m_2^2}, \tag{8}$$

kjer m_k izračunamo po formuli (2).

Opomba 3.15. Če $X \sim N(\mu, \sigma)$, potem velja: $\lim_{n\to\infty} \hat{k}(X_1, X_2, ..., X_n) = 3$ (v verjetnosti).

Trditev 3.16. *Naj bo $X \sim N(\mu, \sigma)$ in n velikost danega slučajnega vzorca. Cenilka sploščenosti $\hat{k}$, izračunana na podlagi vzorca po formuli* (8)*, je približno normalno porazdeljena z matematičnim upanjem* 3 *in standardnim odklonom $\sqrt{\frac{24}{n}}$. Velja tudi:*

$$\lim_{n\to\infty} A(\hat{k}) = 0,$$

$$\lim_{n\to\infty} K(\hat{k}) = 3.$$

Opomba 3.17. Tudi v tem primeru potrebujemo dovolj velik vzorec, da lahko izpeljemo aproksimacijo. Prvi štirje momenti cenilke sploščenosti za vzorec velikosti n se glasijo (dokaz bomo tudi tukaj izpustili):

$$\mu(\hat{k}) = \frac{3(n-1)}{(n+1)} \tag{9}$$

$$\sigma(\hat{k}) = \sqrt{\frac{24(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}} \tag{10}$$

$$A(\hat{k}) = \frac{n^2 - 5n + 2}{(n+7)(n+9)} \sqrt{\frac{216(n+3)(n+5)}{n(n-3)(n-2)}} \tag{11}$$

$$K(\hat{k}) = \frac{36(15n^6 - 36n^5 - 628n^4 + 982n^3 + 5777n^2 - 6402n + 900)}{n(n-3)(n-2)(n+7)(n+9)(n+11)(n+13)} + 3 \tag{12}$$

Izračune na podlagi zgornjih formul pri različnih velikostih n za slučajno spremenljivko $\hat{k}$ zapišemo v tabelo 3, kjer imamo simulacijo vzorčenja iz normalne porazdelitve. Iz tabele lahko razberemo, da pri majhnih vzorcih, recimo $n = 25$, vrednosti momentov, še posebej pa vrednost 4. momenta, v veliki meri odstopajo od asimptotičnih vrednosti, medtem se vzorec velikosti $n = 5000$, kot smo že ugotovili pri cenilki asimetrije, odreže nekoliko bolje.

n	$\mu(\hat{k})$	$\sigma(\hat{k})$	$A(\hat{k})$	$K(\hat{k})$
25	2.769	0.731	1.747	8.897
50	2.882	0.598	1.582	8.416
100	2.941	0455	1.277	6.774
250	2.976	0.301	0.878	4.865
500	2.988	0.216	0.639	4.003
1000	2.994	0.154	0.458	3.520
3000	2.998	0.089	0.267	3.178
5000	2.999	0.069	0.207	3.107
asimtotično	3	$\sqrt{\frac{24}{n}}$	0	3

TABELA 3. Simulacija vzorčenja za cenilko sploščenosti iz normalne porazdelitve

4. Testiranje hipotez

Testiranje hipotez je del statističnega sklepanja in je eno ključnih orodij testiranja. S statistično hipotezo namreč opisujemo naše domneve o porazdelitvi slučajne spremenljivke X, nato pa jih na podlagi rezultatov testiranj zavrnemo ali pa ne zavrnemo. Testiranje hipotez delimo na parametrično (možne porazdelitve so „lepo"parametrizirane s končno množico realnoštevilskih parametrov) in neparametrično (množica možnih porazdelitev je „velika").

4.1. **Določajne hipotez.**

Testiranje hipotez je sestavljano iz začetne - ničelne hipoteze in njej komplementarne hipoteze, ki jo imenujemo alternativna hipoteza. V alternativi zavzamemo vse preostale možnosti. Poglejmo najprej osnovne definicije.

Definicija 4.1. *Statistična hipoteza* ali *statistična domneva* je še nepreverjena izjava oblike:
Dejanska porazdelitev slučajne spremenljivke X pripada podmnožici množice dopustnih porazdelitev slučajne spremenljivke X. V tem smislu statistično hipotezo enačimo s to množico $H \subset \Theta$, tej izjavi pravimo ničelna hipoteza in jo označimo z H_0.
Podamo pa tudi njej komplementarno izjavo oblike:
Dejanska porazdelitev slučajne spremenljivke X pripada komplementarni podmnožici H^C. Kot prej izjavo enačimo s to množico, pravimo ji alternativna hipoteza in jo označimo s H_a.

Opomba 4.2. Množico dopustnih porazdelitev slučajne spremenljivk $X : \Omega \to \mathbb{R}$, enačimo z množico parametrov Θ, ki parametrizirajo dopustne porazdelitve.

Navedimo primer testiranja hipotez pri testiranju matematičnega upanja. V primeru „dvostranskega testiranja" so naše domneve oblike:

$$H_0 : \mu = \mu_0 \text{ proti } H_a : \mu \neq \mu_0,$$

če se odločimo za „enostransko testiranje", pa lahko imamo obliko:

$$H_0 : \mu \leq \mu_0 \text{ proti } H_a : \mu > \mu_0$$

ali

$$H_0 : \mu \geq \mu_0 \text{ proti } H_a : \mu < \mu_0.$$

4.2. **Statistični test.**

Definicija 4.3. *Statistični test* za vzorec velikosti n je odločitveno pravilo, ki na podlagi dejanskega vzorca velikosti n odloči, ali podpremo hipotezo H_0, ali H_a. Test je oblike:

- podpremo H_0, če velja *pogoj*,
- podpremo H_a, če velja *pogoj*.

Ali drugače:

- zavrnemo H_0, če velja *pogoj*,
- ne zavrnemo H_0, če velja *pogoj*.

4.3. **Napake.**

Pri odločitvah na podlagi vzorca lahko storimo napake. Poglejmo si spodnjo tabelo 4 možnih situacij glede na našo odločitev.

ODLOČITEV \ DEJANSKO STANJE	H_0 drži	H_0 ne drži
zavrnemo H_0	napaka	
ne zavrnemo H_0		napaka

Tabela 4. Možne napake pri testiranju hipotez

Seveda si želimo, da bi bili prikazani napaki čim manjši, vendar obeh hkrati ne moremo minimizirati, saj sta napaki komplementarni. Kar lahko storimo, je to, da si izberemo napako, ki je za naše testiranje pomembnejša ter le-to omejimo. Slednje pomeni, da bomo pomembnejšo napako ustrezno zmanjšali, na drugo napako pa v takšni meri ne bomo mogli vplivati, zato je tipično lahko velika. Izbrano minimizirano napako imenujemo *napaka prve vrste*, drugo pa *napaka druge vrste*.

Ob testiranju normalnosti bomo v prihajajoči obravnavi vedno privzeli, da je naša napaka prve vrste napaka, ki jo storimo v primeru, ko domnevo H_0 zavrnemo, čeprav drži.

4.4. **Stopnja značilnosti.**

Definicija 4.4. *Velikost* ali *stopnja testa* je največja možna verjetnost napake prve vrste.

Definicija 4.5. *Stopnja značilnosti*, ki jo označimo z α, je vnaprej določena zgornja meja za stopnjo testa. Drži torej:

$$P(\text{napaka prve vrste}) \leq \alpha.$$

Opomba 4.6. Na področju finančne matematike, statistične kemije in fizike najpogosteje izberemo $\alpha = 0.01$, na področju biomedicine večinoma velja izbira $\alpha = 0.05$, drugje lahko srečamo tudi izbiro $\alpha = 0.1$.

Pri testiranju hipotez je zelo pomembno, da stopnjo značilnosti določimo pred konkretnim testiranjem, saj s tem ohranimo objektivnost in realnost dobljenih rezultatov ter nadaljnjih odločitev. Velja tudi, da manjša kot je stopnja značilnosti, bolj natančne rezultate testiranja lahko pričakujemo, saj napako prve vrste z manjšo stopnjo značilnosti bolj omejimo in zato je verjetnost, da bi zavrnili pravilno ničelno hipotezo H_0 manjša.

4.5. **p-vrednost.**

Definicija 4.7. Če preizkušamo hipotezo H_0 proti H_a in imamo statistični test ter dejanski vzorec, potem je *p-vrednost* dejanskega vzorca najmanjši α, pri katerem naš test še ravno zavrne H_0.

S pomočjo p-vrednosti lahko testiranje hipotez zapišemo nekoliko drugače, test je pri tem oblike:

- zavrnemo H_0, če velja p-vrednost $\leq \alpha$,
- ne zavrnemo H_0, če velja p-vrednost $> \alpha$.

4.6. **Testiranje hipotez v normalnem modelu.**

V normalnem modelu $N(\mu, \sigma)$ sta koncepta interval zaupanja za μ in testiranje hipotez $H_0 : \mu = \mu_0$ proti $H_a : \mu \neq \mu_0$ (za vse možne μ_0) ekvivalentna, zato sklep o domnevah postavimo na podlagi naslednje definicije.

Definicija 4.8. Naj bo $X \sim N(\mu, \sigma)$ in $[L(X_1, X_2, ..., X_n), U(X_1, X_2, ..., X_n)]$ interval zaupanja za μ stopnje zaupanja β, pri čemer je n velikost vzorca. Potem ima dvostranski statistični test obliko:

- $H_0 : \mu = \mu_0$ zavrnemo, če $\mu \notin [L(X_1, X_2, ..., X_n), U(X_1, X_2, ..., X_n)]$ in
- H_0 ne zavrnemo, če $\mu \in [L(X_1, X_2, ..., X_n), U(X_1, X_2, ..., X_n)]$.

Opomba 4.9. Velja sledeča zveza med stopnjo značilnosti α in stopnjo zaupanja β:

$$\alpha = 1 - \beta.$$

Ob testiranju normalnosti se bomo držali izbire $\alpha = 0.05$ in $\beta = 0.95$.

5. Obravnava tipičnega problema

Vrnimo se sedaj k problemu, ki smo ga opisali že v 1. poglavju. Kot že omenjeno, nam je pri vsakršnem ocenjevanju parametrov slučajne spremenljivke X v pomoč vedenje o njeni verjetnostni porazdelitvi, saj nas to privede do natančnejših ocen ter lažjega analiziranja celotne statistične populacije. Omenili smo, da je določajne možnih porazdelitev odvisno tudi od velikosti vzorca, ki ga imamo na razpolago, zato poglejmo primere.

5.1. **Veliki vzorci.**

Ko imamo podan velik slučajni vzorec, npr. $n = 1000$, porazdelitev populacije ne igra odločilne vloge, saj v takšnem primeru uporabimo aproksimacijo s pomočjo centralnega limitnega izreka. Najprej navedimo ta izrek.

Izrek 5.1. Centralni limitni izrek
Naj bo $\{X_n\}_{n=1}^{\infty}$ zaporedje neodvisnih enako porazdeljenih slučajni spremenljivk iz L^2 in $S_n = \sum_{i=1}^{n} X_i$ zaporedje njihovih delnih vsot. Potem velja:

$$\frac{\sum_{i=1}^{n} X_i - nE(X)}{\sqrt{nD(X)}} \Rightarrow N(0,1),$$

kjer $\Rightarrow$ pomeni konvergenco po zakonu.

Opomba 5.2. $X \in L^2$ natanko takrat, ko obstaja $E(X^2) < \infty$.

Opomba 5.3. Velja $\sqrt{D(X)} = \sigma(X)$.

Slednje nam torej pove, da v primeru, ko imamo poljubno populacijo, ki je dovolj velika, standardizacija vzorčnega povprečja konvergira k standardni normalni porazdelitvi. Zato lahko privzamemo, da je standardizacija vzorčnega povprečja $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ približno standardno normalno porazdeljena, četudi same slučajne spremenljivke ne ustrezajo normalni porazdelitvi.

5.2. **Majhni vzorci.**

Pri majhnih vzorcih, npr. $n = 10$, je reševanje problema zahtevnejše, saj centralnega limitnega izreka ni moč uporabiti. Če se omejimo na zvezne slučajne spremenljivke, vemo, da imajo le-te gostoto porazdelitve, recimo ji $f(x)$ in tedaj velja:

$$P(X \in (a,b)) = \int_a^b f(x)dx.$$

Iz tega sledi:

$$P(\overline{X} \in (a,b)) = P\big(\frac{1}{n}\sum_{i=1}^{n} X_i \in (a,b)\big) = P\big(\sum_{i=1}^{n} X_i \in (na,nb)\big) = \int_{na}^{nb} g(t)dt,$$

kjer je $g(t)$ v splošnem zelo zapletena gostota porazdelitve, zato je slednji izraz težko izračunljiv.

Iz tega razloga se obrnemo na privzetke o normalnosti, ki nam nadaljnje računanje v veliki meri poenostavijo. Večinoma o teh privzetkih seveda nismo prepričani, zato nas slednje privede do tega, da s primernimi statističnimi testi preverimo, ali je vzorec konsistenten s takšnim privzetkom.

6. Testiranje normalnosti

Za testiranje normalnosti poznamo več statističnih testov. Najboljši temeljijo na vzorčnih momentih do $k = 4$, vendar jih le malo vključuje teste na podlagi asimetrije in sploščenosti. Eden takšnih, ki to lastnost imajo, pa je *d'Agostino-Pearsonov* K^2 *omnibus test*, ki se izkazuje za zanesljivega, čeprav ni eksakten. V končni fazi vključuje namreč zelo dobre približke, na podlagi katerih lahko dovolj trdno sklepamo o normalni predpostavki že za zelo majhne vzorce - test se obnese na vzorcih velikosti $n > 8$.

6.1. **Osnovne predpostavke in hipoteze.**

Testiramo slučajno spremenljivko $X : \Omega \to \mathbb{R}$, pri čemer je Ω prostor, v katerem z neko verjetnostjo črpamo podatke.

Predpostavimo, da imamo slučajni vzorec $(X_1, X_2, ...X_n)$ velikosti n, za katerega velja, da so slučajne spremenljivke X_i neodvisno enako porazdeljene, kar zapišemo kot:

$$X_i \sim X \text{ n.e.p.}$$

Sedaj preverjamo predpostavko o normalnosti, torej nas zanima, ali je slučajna spremenljivka X zares normalno porazdeljena, kar zapišemo pod osnovne hipoteze v obliki dvostranskega testa, ki se jih držimo skozi celotno testiranje, pri tem hipotezo H_0 imenujemo *hipoteza normalnosti.*

$$H_0 : X \sim N(\mu, \sigma) \textit{ za neka } \mu \in \mathbb{R} \textit{ in } \sigma \in (0, \infty),$$

$$H_a : X \textit{ ni normalno porazdeljena slučajna spremenljivka.}$$

Testiranje normalnosti se prične z dvema podtestoma, v katerih preverimo dva ključna kazalca nenormalnosti - asimetrijo in sploščenost slučajne spremenljivke X, le-ta pa se nato združita v končen K^2 omnibus test.

6.2. **Test asimetrije.**

Pri testiranju asimetrije preverjamo, ali je vrednost normiranega 3. centraliziranega momenta dovolj blizu teoretični vrednosti 0, pri pogoju, da osnovna ničelna hipoteza normalnosti drži. Če bi bila slučajna spremenljivka X zares normalno porazdeljena, bi morala držati tudi ničelna podhipoteza o vrednosti asimetrije. Pogojno na H_0 torej postavimo spodnji podhipotezi.

$$H_0' : A(X) = 0$$

$$\text{proti}$$

$$H_a' : A(X) \neq 0.$$

Na podlagi danega slučajnega vzorca, lahko s pomočjo cenilke (5) ocenimo vrednost asimetrije. Nadalje poglejmo definicije.

Definicija 6.1. Definiramo:

$$Y = \hat{a}\Big(\frac{(n+1)(n+3)}{6(n-2)}\Big)^{1/2}.$$

Izrek 6.2. *Pri pogoju, da drži H_0, velja:*

$$K(\hat{a}) = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}.$$

Definicija 6.3. Definiramo:

$$W^2 = -1 + (2(K(\hat{a})-1))^{1/2},$$

$$\delta = \frac{1}{\sqrt{lnW}},$$

$$\alpha = \Big(\frac{2}{W^2-1}\Big)^{1/2}.$$

Definicija 6.4. Končna transformacija je definirana kot

$$Z(\hat{a}) = \delta ln\Big(\frac{Y}{\alpha} + \big((\frac{Y}{\alpha})^2+1\big)^{1/2}\Big). \tag{13}$$

Srce testa predstavlja naslednja trditev (glej: *A Suggestion for Using Powerful and Informative Tests of Normality*):

Trditev 6.5. *Če je X normalno porazdeljena slučajna spremenljivka, kar pomeni, da je bil slučajni vzorec $(X_1, X_2, ..., X_n)$ generiran z neko normalno porazdelitvijo $N(\mu, \sigma)$, je transformacija $Z(\hat{a})$ približno standardno normalno porazdeljena že za $n > 8$.*

Definicija 6.6. *Statistični test za testiranje asimetrije*
Naj bo $[-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}]$ interval zaupanja za $Z(\hat{a})$ s stopnjo zaupanja $1-\alpha$, potem:

- ne zavrnemo H_0', če $Z(\hat{a}) \in [-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}]$,
- zavrnemo H_0' sicer.

Opomba 6.7. Navedeni interval zaupanja je simetričen in pri izbiri $\alpha = 0.05$ velja:

$$Z_{\frac{\alpha}{2}} = Z_{0.025} = \Phi^{-1}(0.975) \doteq 1.95996.$$[1]

Zgornjo približno vrednost lahko odčitamo tudi iz tabele 1, kjer so iskane vrednosti že označene. Velja

$$\Phi(1.96) \doteq 0.9750,$$

iz česar sledi

$$\Phi^{-1}(0.9750) \doteq 1.96.$$

V primeru, ko je vrednost slučajne spremenljivke $Z(\hat{a})$ zares vsebovana v intervalu zaupanja $[-1.96, 1.96]$, podhipoteze H_0' ne zavrnemo in lahko sklepamo, da velja $Z(\hat{a}) \approx N(0,1)$.

Trditev 6.8. *Ker velja $H_0' \subset H_0$, moramo ob zavrnitvi ničelne podhipoteze H_0', zavrniti tudi osnovno hipotezo normalnosti H_0. Obratno ne velja nujno.*

[1]Vrednost je izračunana s pomočjo programa Wolfram Mathematica.

6.3. **Test sploščenosti.**

Podobno kot pri testu asimetrije velja tudi za test sploščenosti. Pri slednjem testu preverjamo, ali je vrednost sploščenosti blizu vrednosti 3, pri tem predpostavimo, da velja hipoteza normalnosti. Pogojno na H_0 postavimo podhipotezi.

$$H_0'' : K(X) = 3$$

$$\text{proti}$$

$$H_a'' : K(X) \neq 3.$$

Vrednost sploščenosti ocenjujemo s pomočjo cenilke $\hat{k}$, ki ustreza formuli (8). Definirajmo še naslednje vrednosti.

Definicija 6.9. Standardizacijo cenilke $\hat{k}$ zapišemo:

$$x = \frac{(\hat{k} - E(\hat{k}))}{\sigma(\hat{k})}.$$

Izrek 6.10. *Pri pogoju, da drži H_0, velja:*

$$A(\hat{k}) = \frac{n^2 - 5n + 2}{(n+7)(n+9)} \sqrt{\frac{216(n+3)(n+5)}{n(n-3)(n-2)}}.$$

Definicija 6.11. Definirajmo še:

$$B = 6 + \frac{8}{A(\hat{k})}\Big(\frac{2}{A(\hat{k})} + \sqrt{1 + \frac{4}{A^2(\hat{k})}}\Big).$$

Definicija 6.12. Končna transformacija je definirana kot:

$$Z(\hat{k}) = \frac{\Big((1 - \frac{2}{9B}) - \big(\frac{1-\frac{2}{B}}{1+x\sqrt{\frac{2}{B-4}}}\big)^{1/3}\Big)}{\sqrt{\frac{2}{9B}}}. \tag{14}$$

Enako kot pri testiranju asimetrije, veljajo naslednje trditve in definicija (glej: *A Suggestion for Using Powerful and Informative Tests of Normality*):

Trditev 6.13. *Če je X normalno porazdeljena slučajna spremenljivka, je transformacija $Z(\hat{k})$ približno standardno normalno porazdeljena že za $n > 8$.*

Definicija 6.14. *Statistični test za testiranje sploščenosti*
Naj bo $[-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}]$ interval zaupanja za $Z(\hat{k})$ s stopnjo zaupanja $1 - \alpha$, potem:

- ne zavrnemo H_0'', če $Z(\hat{k}) \in [-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}]$,
- zavrnemo H_0'', če $Z(\hat{k}) \notin [-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}]$.

Ko je vrednost slučajne spremenljivke $Z(\hat{k})$ vsebovana v intervalu zaupanja, kjer velja $Z_{\frac{\alpha}{2}} \doteq 1.96$, potem podhipoteze H_0'' ne zavrnemo in obdržimo sklep $Z(\hat{k}) \approx N(0,1)$.

Trditev 6.15. *Velja $H_0'' \subset H_0$, zato moramo ob zavrnitvi podhipoteze H_0'', zavrniti tudi hipotezo H_0. Obratno prav tako ne velja nujno.*

Po testiranju asimetrije in sploščenosti tako pridemo do sklepov o vrednostih teh kazalcev, a nam to še vedno ne poda jasne slike za sklep o hipotezi normalnosti. Če do tega trenutka nismo zavrnili nobene izmed podhipotez H_0' in H_0'', se posvetimo še končnemu koraku testiranja normalnosti. Velja namreč, da lahko testa asimetrije in sploščenosti nakazujeta na normalno porazdelitev proučevane slučajne spremenljivke X, vendar se ob združitvi le-teh pokaže preveliko odstopanje od glavne hipoteze normalnosti, kar nas privede do sklepa, da naš privzetek normalnosti kljub vsemu ni dovolj trden in ga pri analiziranju podatkov ne moremo uporabiti.

6.4. K^2 **test.**

Korak, ki nas vodi do končnega sklepa o normalnosti slučajne spremenljivke X, je torej K^2 *omnibus test*. Pojem *omnibus test* nam pri tem nakazuje na dejstvo, da v končnem koraku združujemo več testov hkrati, v našem primeru združujemo dva - test asimetrije in test sploščenosti.

Najenostavnejša rešitev zadnjega dela testa bi sicer temeljila na neodvisnosti in normalnosti cenilk. V primeru, ko bi ti dve lastnosti držali, bi z lahkoto združili standardizaciji cenilk v tako imenovan K^2 test, za katerega bi nato, po veljavnosti (1), sledila porazdelitev χ_2^2. V tem primeru bi test izgledal sledeče:

$$\Big(\frac{\hat{a}-\mu(\hat{a})}{\sigma_1}\Big)^2 + \Big(\frac{\hat{k}-\mu(\hat{k})}{\sigma_2}\Big)^2 \sim \chi_2^2,$$

a bi za takšno zgradbo končnega testa potrebovali velike vzorce podatkov, da bi bila približna normalnost utemeljena in bi test upravičeno uporabili.

Test, ki se je izkazal za dovolj dobro rešitev že pri zelo majhnih vzorcih, sta prvič predstavila d'Agostino in Pearson leta 1974. Dokazala sta, da cenilki $\hat{a}$ in $\hat{k}$ sicer nista nujno neodvisni, sta pa nekorelirani, kar je zadosten pogoj, da se test izkaže za uporabnega. Navedimo najprej omenjeni test.

Definicija 6.16. K^2 *omnibus test* ima obliko

$$K^2 = Z^2(\hat{a}) + Z^2(\hat{k}), \tag{15}$$

kjer sta $Z(\hat{a})$ in $Z(\hat{k})$, izračunani po formulah (13) in (14), približno standardno normalno porazdeljeni slučajni spremenljivki, iz česar sledi, da je slučajna spremenljivka K^2 porazdeljena približno po zakonu χ_2^2.

Definicija 6.17. *Statistični test normalnosti*
Naj bo $[\chi^2_{2;-\frac{\alpha}{2}}, \chi^2_{2;\frac{\alpha}{2}}]$ interval zaupanja za K^2 s stopnjo zaupanja $1-\alpha$, potem:

- ne zavrnemo H_0, če $K^2 \in [\chi^2_{2;-\frac{\alpha}{2}}, \chi^2_{2;\frac{\alpha}{2}}]$,
- zavrnemo H_0, če $K^2 \notin [\chi^2_{2;-\frac{\alpha}{2}}, \chi^2_{2;\frac{\alpha}{2}}]$.

Opomba 6.18. Zgornji interval zaupanja ni simetričen. Pri izbiri $\alpha = 0.05$ velja

$$\chi^2_{2;-0.025} \doteq 0.0506356$$

in

$$\chi^2_{2;0.025} \doteq 7.37776.$$ [2]

[2] Obe vrednosti sta izračunani s pomočjo programa Wolfram Mathematica.

Hipotezo $H_0 : X \sim N(\mu, \sigma)$ za neka $\mu \in \mathbb{R}$ in $\sigma \in (0, \infty)$ zavrnemo ali obdržimo na podlagi zgornjega statističnega testa. Če hipotezo zavrnemo, potem končen sklep govori, da slučajni vzorec ni konsistenten z morebitnim privzetkom o normalnosti, v kolikor H_0 ne zavrnemo, sklepamo, da je privzetek o normalnosti slučajne spremenljivke X dovolj trden, zato ga obdržimo.

7. Testiranje konkretnih podatkov

V zadnjem poglavju se bomo soočili s potekom testiranja normalnosti na konkretnih podatkih. Kot že omenjeno, področje, kjer pogosto uporabljamo privzetke normalnosti, je biomedicina in ravno iz tega področja bomo analizirali izmerjeno količino holesterola LDL pri različnih in slučajno izbranih pacientih, ki so se odločili za zdravljenje. Omeniti velja, da kratica LDL označuje holesterol nizke gostote, ki mu pravimo tudi „nekoristen" oziroma „nezaželjen" holesterol, ki v telesu maši žile in s tem povzroča zdravstvene težave.

Na voljo imamo 980 meritev holesterola LDL pri različnih pacientih. Vsakemu pacientu je bil holesterol LDL izmerjen pred in po zdravljenju[3], sami pa izračunamo še razliko *pred* $-$ *po* zdravljenju. Zaradi ugodne velikosti vzorca, ki ga imamo na voljo, lahko postavimo sledeče predpostavke, potrebne za vsako izmed nadaljnjih testiranj.

- Imamo slučajni vzorec $(X_1, X_2, ...X_n)$, kjer je $n = 980$.
- Za slučajne spremenljivke X_i velja: $X_i \sim X$ n.e.p.
- $X \sim N(\mu, \sigma)$, za neka $\mu \in \mathbb{R}$ in $\sigma > 0$.

7.1. **Testiranje podatkov, izmerjenih pred zdravljenjem.**

Na sliki 9 se nahaja prikaz meritev, opravljenih pred zdravljenjem. Ob pogledu na histogram bi se lahko zazdelo, da kljub manjšim odstopanjem, meritve ustrezajo normalni porazdelitvi, a je brez izračunov o hipotezi normalnosti težko dovolj trdno sklepati.

7.1.1. ***Osnovni izračuni.***

Na osnovi zabeleženih meritev holesterola LDL pri pacientih izračunamo osnovne vzorčne momente, ki jih nato uporabimo pri testiranjih:

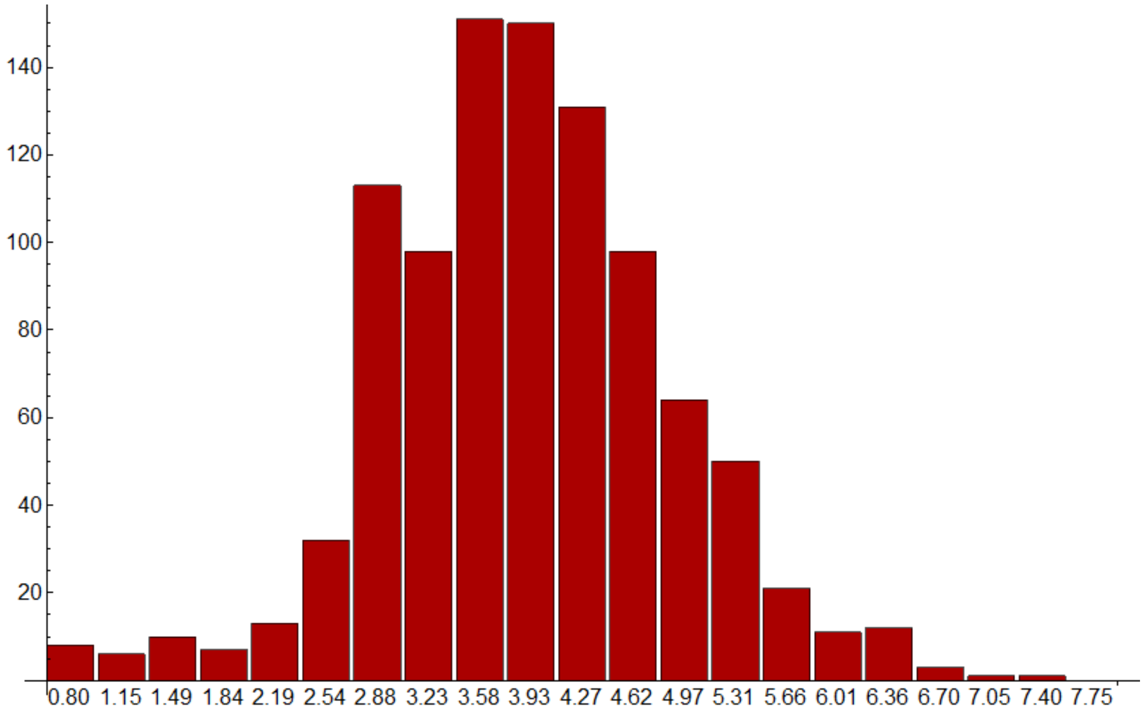
$$\overline{x} = 4.0676,$$

$$m_2 = 0.9760,$$

$$m_3 = -0.0635,$$

$$m_4 = 3.5831.$$

[3]Meritve, izmerjene pred in po zdravljenju, na podlagi katerih so poračunani sledeči izračuni, so zabeležene v prilogah na zadnjih straneh. Vsi izračuni in histogrami so narejeni s pomočjo programov Microsoft Excel in Wolfram Mathematica. Pri izrisu histogramov za izračun optimalnih razponov posameznih kategorij je uporabljena Scottova formula oblike $h = \frac{3.5\hat{\sigma}}{\sqrt[3]{n}}$, ki je še posebej primerna za prikaz morebitno normalno porazdeljenih podatkov.



SLIKA 9. Histogram podatkov, izmerjenih pred zdravljenjem

7.1.2. ***Test asimetrije.***

Postavimo domnevi $H_0' : A(X) = 0$ proti $H_a' : A(X) \neq 0$ ter pričnemo s testom asimetrije. Na podlagi slučajnega vzorca podatkov izračunamo:

$$\hat{a} = -0.0658,$$
$$Y = -0.8440,$$
$$\text{če drži } H_0, \text{ je } K(\hat{a}) = 3.0358,$$
$$W^2 = 1.0178,$$
$$\delta = 10.6338,$$
$$\alpha = 10.5868,$$
$$Z(\hat{a}) = -0.8468.$$

Ker $Z(\hat{a}) \in [-1.96, 1.96]$, podhipoteze H_0' ne zavračamo in nadaljujemo s testom sploščenosti.

7.1.3. ***Test sploščenosti.***

Testiramo $H_0'' : K(X) = 3$ proti $H_a'' : K(X) \neq 3$ ter izračunamo:

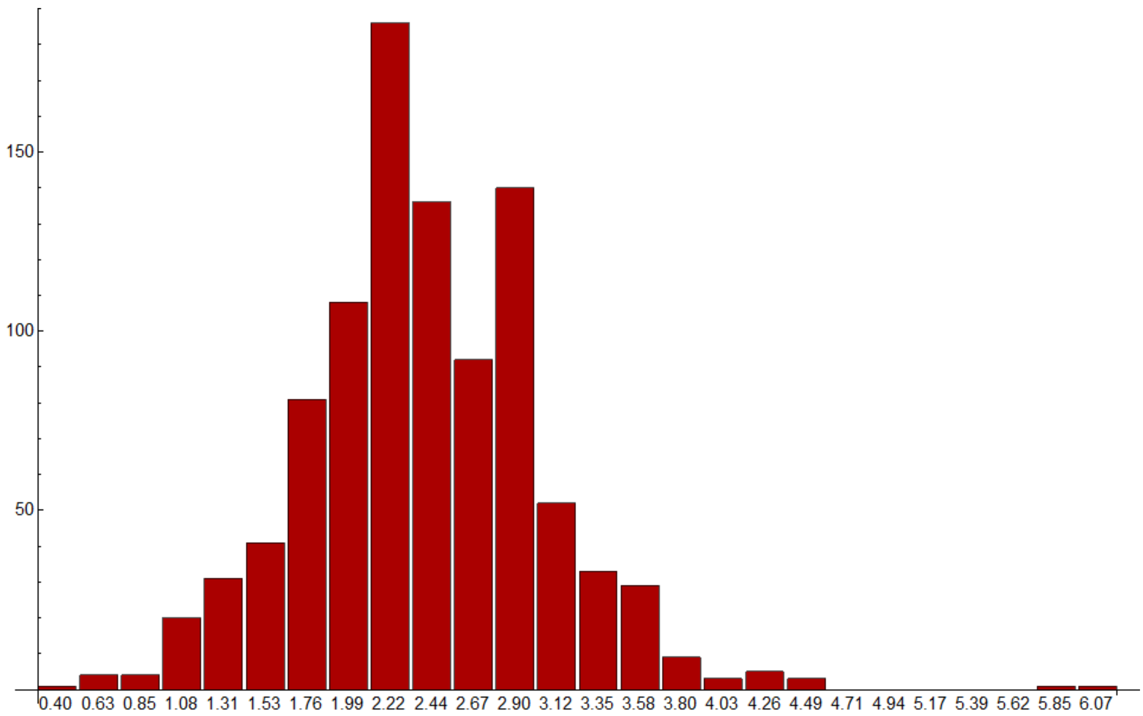
$$\hat{k} = 3.7615,$$
$$\mu(\hat{k}) = 2.9939,$$
$$\sigma(\hat{k}) = 0.1553,$$
$$x = 1.9477,$$
$$\text{če drži } H_0, \text{ je } A(\hat{k}) = 0.4626,$$
$$B = 157.5044,$$

$$Z(\hat{k}) = 3.7480.$$

Ker $Z(\hat{k}) \notin [-1.96, 1.96]$, podhipotezo H_0'' zavrnemo, iz česar sledi, da zavrnemo tudi ničelno hipotezo normalnosti v korist alternative. Končen sklep je, da podatki, izmerjeni pred zdravljenjem, niso porazdeljeni normalno.

7.2. **Testiranje podatkov, izmerjenih po zdravljenju.**

Slika 10 prikazuje podatke, izmerjene po zdravljenju. Iz histograma je razvidno odstopanje od normalnosti na desni strani histograma. Kljub temu ne moremo trditi, da podatki niso porazdeljeni normalno, zato poglejmo, kako se dani vzorec podatkov odreže pri testiranju normalnosti.



Slika 10. Histogram podatkov, izmerjenih po zdravljenju

7.2.1. ***Osnovni izračuni.***

Na osnovi danih podatkov izračunamo vzorčne momente:

$$\overline{x} = 2.5001,$$

$$m_2 = 0.4173,$$

$$m_3 = 0.1318,$$

$$m_4 = 0.8447.$$

7.2.2. ***Test asimetrije.***

Postavimo domnevi $H_0' : A(X) = 0$ proti $H_a' : A(X) \neq 0$ ter pričnemo s testom asimetrije. Na podlagi podatkov izračunamo:

$$\hat{a} = 0.4889,$$

$$Y = 6.2670,$$

$$\text{če drži } H_0, \text{ je } K(\hat{a}) = 3.0358,$$

$$W^2 = 1.0178,$$

$$\delta = 10.6338,$$

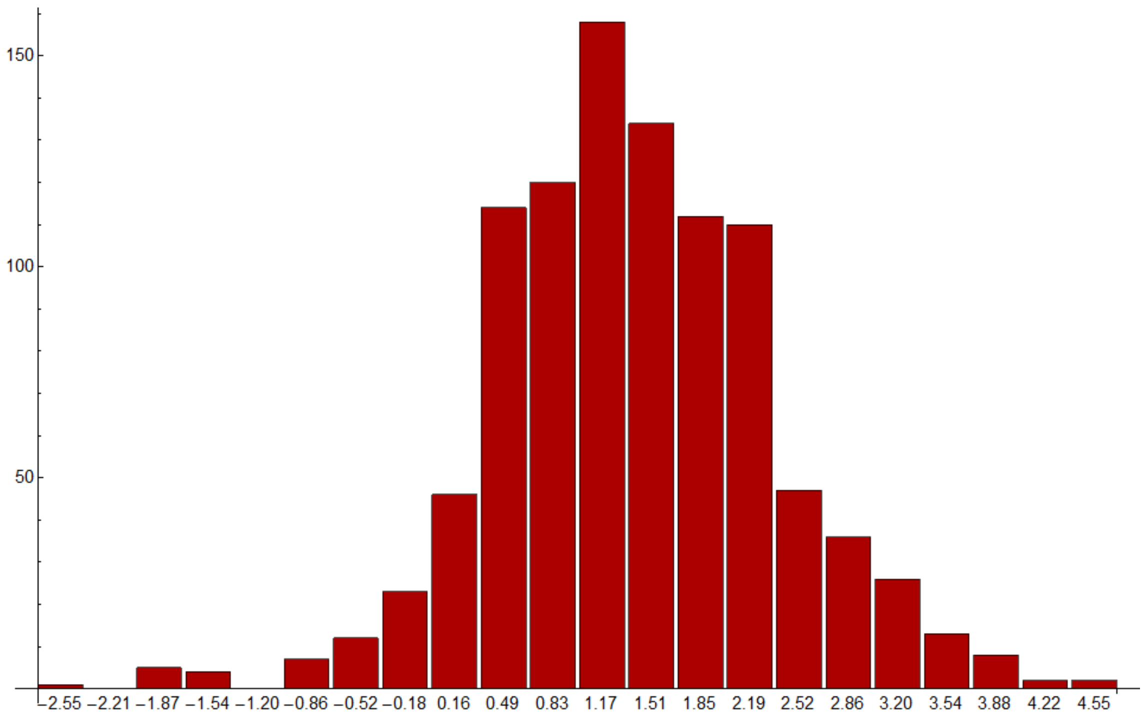
$$\alpha = 10.5868,$$

$$Z(\hat{a}) = 5.9753.$$

Ker $Z(\hat{a}) \notin [-1.96, 1.96]$, podhipotezo H_0' zavrnemo, zato moramo zavrniti tudi ničelno hipotezo H_0 ter zaključimo, da vzorec ni konsistenten s privzetkom normalnosti.

7.3. Testiranje razlike *pred* − *po* zdravljenju.

V večini primerov takšni tipi razlik podatkov še najbolj nakazujejo na lastnosti normalne porazdelitve, saj se zdi jasno, da bosta verjetnosti, da bi holesterol LDL z zdravljenjem ekstremno nadpovprečno ali ekstremno podpovprečno pozdravili, manjši. V kolikšni meri naša razlika podatkov, katere histogram je na sliki 11, dejansko ustreza normalni porazdelitvi, preverimo s testom normalnosti.



SLIKA 11. Histogram razlike *pred* − *po* zdravljenju

7.3.1. ***Osnovni izračuni.***

Na osnovi zabeleženih meritev holesterola LDL pri pacientih izračunamo:

$$\begin{aligned} \overline{x} &= 1.5675, \\ m_2 &= 0.9261, \\ m_3 &= -0.0476, \\ m_4 &= 3.4108. \end{aligned}$$

7.3.2. ***Test asimetrije.***

Postavimo domnevi $H_0' : A(X) = 0$ proti $H_a' : A(X) \neq 0$ ter pričnemo s testom asimetrije. Na podlagi vzorca podatkov izračunamo:

$$\begin{aligned} \hat{a} &= -0.0534, \\ Y &= -0.6850, \\ \text{če drži } H_0\text{, je } K(\hat{a}) &= 3.0358, \\ W^2 &= 1.0178, \\ \delta &= 10.6338, \\ \alpha &= 10.5868, \\ Z(\hat{a}) &= -0.6875. \end{aligned}$$

Ker $Z(\hat{a}) \in [-1.96, 1.96]$, podhipoteze H_0' ne zavrnemo in nadaljujemo s testom sploščenosti.

7.3.3. ***Test sploščenosti.***

Testiramo $H_0'' : K(X) = 3$ proti $H_a'' : K(X) \neq 3$ ter izračunamo:

$$\begin{aligned} \hat{k} &= 3.9771, \\ \mu(\hat{k}) &= 2.9939, \\ \sigma(\hat{k}) &= 0.1553, \\ x &= 2.4950, \\ \text{če drži } H_0\text{, je } A(\hat{k}) &= 0.4626, \\ B &= 157.5044, \\ Z(\hat{k}) &= 4.4711. \end{aligned}$$

Ker $Z(\hat{k}) \notin [-1.96, 1.96]$, podhipotezo H_0'' zavrnemo, iz česar sledi, da zavrnemo tudi ničelno hipotezo normalnosti v korist alternative. Končen sklep je, da razlika $pred - po$ zdravljenju ni porazdeljena normalno.

Iz končnih sklepov vseh treh primerov testiranja podatkov lahko potrdimo dejstvo, da četudi podatki, ki na prvo oko spominjajo na normalno porazdelitev, po ustreznem testu pokažejo povsem drugačne rezultate, zato sklepanja, ki ne temeljijo na testiranjih, tudi niso dovolj trdna. Statistični testi kot je na primer predstavljeni d'Agostino-Pearsonov K^2 omnibus test, so tako „nujno zlo" sodobne statistike, z njimi pa kljub določenim odstopanjem dobimo dovolj natančne zaključke o potrebnih privzetkih.

Literatura

[1] R. B. D'Agostino, A. Belanger in R. B. D'Agostino, Jr., *A Suggestion for Using Powerful and Informative Tests of Normality*, The American Statistician **44** (1990) 316–321.

[2] R. B. D'Agostino in E. S. Pearson, *Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$.*, Biometrika **60** (1973) 613–622.

[3] Henry C. Thode, Jr., *Testing for Normality*, Marcel Dekker, Inc., New York, 2002.

[4] M. Omladič, *Zapiski iz predavanj: Verjetnosti račun I.*, Ljubljana, 2012.

[5] J. Smrekar, *Zapiski iz predavanj: Statistika I., 2.del*, Ljubljana, 2013.

[6] *Normal distribution*, [ogled 15. 11. 2012], dostopno na `http://en.wikipedia.org/wiki/Normal_distribution`.

[7] *Chi-squared distribution*, [ogled 10. 7. 2013], dostopno na `http://en.wikipedia.org/wiki/Chi-squared_distribution`.

PRILOGA 1: Meritve holesterola LDL pred zdravljenjem

3.89	3.3	2.76	6.8	5.6	3.9	4.5	4.2	2.8	3.4	3
2.96	4.7	3.3	4	3.6	4.4	2.3	4.43	3.5	5.7	4.8
4.7	4.9	5.2	3.2	5.33	3.8	3.79	5.6	4.6	3.46	3.2
5.2	4	3.38	4.22	3.44	3.14	4.7	3.8	4	3.4	6.6
4.9	3.2	4.4	4.08	4.8	3.4	3.36	4.1	2.9	3.25	3.2
3.9	5.46	3.95	3.6	2.9	4	5.5	4	4.6	3.62	1.52
4.5	3	3.1	3.02	3.13	4.6	3.72	6.1	4.2	6.8	3.3
4.4	4.2	3.7	3.7	4	5.52	4.39	3	3.8	3.4	4
3.9	3.4	3.24	3.4	5.6	7.3	3.8	4.7	3	4.4	3
4.2	6.6	3.1	2.55	2.62	5.2	4.5	4.6	3.46	4.3	3.2
3.8	3.8	5.2	4.05	4.56	3.1355	4.1	3.9	4.3	4.6	2.98
3.5	3.41	4.8	3.5	3.8	3.8	4.1	4.4	3.7	5	3.14
4.5	5.4	4.2	3.2	4.6	3.42	4.1	2.2	3.2	4.79	3.5
4.1	4.1	3.4	5.6	4.41	3.12	4.6	4.7	4.1	3.63	4
3.1075	3.33	4.5	4.1	5	3.2	4.9	4	3.4	4.18	4.7
3.01	4.4	3.4	4	3.2	2.9	5.1	5	3.5	3.69	4.17
4.5	4.9	4.89	3.21	5.4	4.2	4.4	3.9	4	6.1	2.9
3.3	4.2	3.1	3.3	4.6	1.48	3.99	3	3.2	3	5.81
5.4	4.2	3.6	3.8	4.5	3.7	5.1	3.71	3.3	4.24	2.1
4.72	3.7	4.44	3.5	4.8	3.87	5.9	4.91	4.13	2.7	4.4
3.06	4.9	3.9	4.3	4.2	3.7	1.88	3.7	6	3.2	4.6
5.1	4.61	5.75	4.51	4.1	4.3	5	5	1.8	3.5	3.84
5.1	3.8	0.8	3.4	3.21	2.9	4.1	4.05	6.6	3.37	4.1
3.9	4.4	5.08	2.4	4	5.54	4.2	4.2	4.2	4.7	4.9
3.44	5.65	5.35	4.1	4	3.7	3.1	4.7	3.8	3.9	4.7
4.3595	1.02	5.54	3.71	4	4.1	4.9	4.1	4.85	6.3	4.29
5	2.7	4.9	4	5.7	4	3.2	5.3	4.4	3.2	4.26
4.5	3.7	5.4	3	4.4	3.87	2.3	4.5	3.2295	5.6	3.95
5.39	4.3	3.33	3.75	4.27	2.67	4.2	3.74	3.3	3.8	5.7
4.2	4	1.6	3.6	4.72	3	5.1	3	3.9	3.48	1.78
4.32	5.7	2.3	4.2	3.5	2.96	4.6	3.7	4.92	3.7	3.7
2.71	5.1	4.5	4.42	3.1	3.5	4.7	3.29	4.78	4.5	2.8
3.9	4.73	6.2	3.9	5.1	5.77	4	5.4	5.36	5.1	4.9
2.8	4.8	5.3	3.8	2.9	4.2	4.1	5	3.97	4	4.26
4.2	5.3	4.5	5.98	4.1	5.42	5	3.8	4.07	3.2	2.98
3.49	3.9	3.1	4.5	4.9	4.7	3	4.6	4.8	4.8	4.3
4.38	3.76	5.4	2.89	3.7	4.19	2.8	3.9	3.6	3.8	5.2
5.33	3.8	2.8	3.6	4.3	4.251	6.48	3.8	4.7	3.4	3.6
2.59	4.4	4.5	4.5	4.7	5.5	3.5	2.78	3.3	4.4	4.38
4.64	6.4325	3.9	0.964	4.4	4.5	6.9	4.4	5.9	5.3	3.3
4.7	3	4.1	3.7	4.44	3.01	1.4	4.1	4.15	5.44	4.9
3.9	4.7	2.9	5.89	2.96	5.4	3	3.9	5.4	3.9	3.1
7.5	4.7	3.2	3.9	4.77	3.75	4.49	4.4	4.37	3.1	4.57
1.5	2.9	3.7	2.86	4.61	4.28	3.77	4.563	3.3	4.02	4.6
4.8	4.02	4.09	3.28	3.5	3.3	3.8	4.06	4	3.7	3.7
3.88	5.3	4.9	4.9	4.1	5	2.9	5.5	5.29	4.4	4.72
3.3	4.2	3.8	3.7	4.47	2.8	3.8	4.3	6.51	3.9	4.43

3.6	4.9	5.4	3.1	2.83	4.6	3.1	3.7	4.58	3.97	4.9
5.7	3.64	4.1	3.8	4.6	2.8	4.4	3.7	4	6.6	3.6
6.37	4.2	3.7	3.1	2.98	5.31	5.4	5.9	5.9	3.65	4.73
2.59	4.66	4.38	5.2	4.6	4.1	4.95	4.9	4.8	5.3	4.3
3.12	3.44	5.6	6.54	3.23	4.91	5.17	3.2	2.9	1	4.89
5	3.8	4.5	3.7	3.54	3.6	5.27	2.7	5.5	2.5	3.4
3.9	4	5.1	2.9	3.2	4	4.19	4.5	6.1	3.6	3.122
3.3	3.74	2.9	3.9	4.17	3.24	3.5	3	1.84	3.3	3.6
4.2	2.9	3.43	5.16	4.16	2.9	4.1	3.3	4.01	4.3	4.2
3.7	3.2	3.1	3.68	1.53	3.3	3.2	2.7	4.6	4.67	3.7
2.99	4.35	5.55	3.8	4.39	3.6	4.3	3.5	4.48	5.4	5
1.4	1	4.74	4.8	3.8	4.3	2.3	4.26	4	4.9	3.3
3.1	4.7	3.4	4.17	4.01	3.4	3.2	4.9	4.6	2.7	4.5
3.4	3.97	4	4.1	6.6	3	4.2	3.2	2.7	4	4.1
6.1	3.9	3.64	3.3	3.9415	3.03	4.8	4.6	1.3	3.9	4.4
3.4395	5.85	2.7	4.08	4.98	5.93	4.4	4.1	5.6	4.56	4.7
4.6	2.23	4	4.72	4.6	4.6	5.05	3.64	4.2	4.3	4.8
3.9	2.25	5.6	5	6.33	4.1	4.68	3.7	3.06	2.8	3.9
3.2	3.08	4.51	2.4	4.72	4.2	5.62	5.6	5.6	4.8	4.63
3.9	3.26	4.1	3.1	4.2	3.56	3.56	3.07	2.41	4.69	3.96
5.1	1.87	2.9	3.9	3.7	4	4.02	2.7	4.9	4.73	4.71
3.4	3.41	3.3	5.5	5	4.4	2	3.6	3.8	3.81	3.7
3.7	5.49	3.9	3.1	1.08	4.81	4.93	3.5	5.7	4.3	4.6
4.2	4.16	5.4	5.7	4.69	3.7	4.9	3.5	4.82	4.8	3.6
3.8	4.65	5.7	4.8	6.1	3.1	5	3.5	4.6	3.6	6.33
3.3	3.2	5.08	3.9	4.7	5.9	4.1	3.4	5	3.8	4.26
3.9	2.8	4.4	3.5	5.4	4	3.9	4.1	0.9	3.3	4.9
4.07	3.49	3.71	1.9	3.53	3.6	2.1335	4	4.3	6.2	2.9
5	4.8	3.94	3.4	3	3.4	5.3	3.1	3.6	1.33	5.3
0.9	4.4	3.19	4.8	4.1	4.2	3.04	4.6	4.1	5.1	4
5.4	3.2	4.4	2.4	4.4	5	5.1	6.2	3.3	3.6	5.2
4.25	3	4.24	4.7	3.48	4	3.9	4.5	3.2	3.52	5
5.57	4.5	5.4	5.14	3.87	3.1	3.3	4.1	3.5	4.2	3.2
2.88	4.26	5.1	2.63	4.5	4.78	3.9	4.6	5.1	3.9	
4.15	4.2	4.13	4.65	3	4.3	4.3	3	4.08	6.41	
6.44	5.5	3.1	5.03	4.3	4.6	5.5	4.59	3.6	4.6	
3.7	3.6	4.7	3.4	3.9	3.8	1.75	4.5	4.95	3.2	
3.2	3.69	2.76	4.27	5.2	4.7	4.2	4.15	4.3	4.16	
3.99	4.02	5.4	3.9	2.8	3.72	3.84	5.44	3.6	3.7	
3.7	1.4	3.3	5.06	3.5	3.3	2.4	4.5	4	2.8	
4	2.9	3.87	3.5	4.12	4.4	4.8	1.9	3.5	4.27	
4.6	4.9	3.73	5	5.26	3.4	1.52	4.8	5	3.8	
4.1	5	4.3	1.657	3.3	3.8	4.96	4.35	4.59	4.4	

PRILOGA 2: Meritve holesterola LDL po zdravljenju

3	2	2.65	3.2	4.5	1.8	1.5	2.9	2	2.3	2.9
1.89	3	1.9	3.2	2.3	3.5	2	3.37	2.16	3.1	2.3
2.2	2.5	2.3	2.05	2.16	3.2	2.9	3	2.25	1.8	2.1
3.7	3	1.94	3	2.55	2.25	2.7	2.7	3	2.26	3
3.1	1.4	2.9	2.13	2.6	2	1.7	3	1.8	2.25	2.9
1.42	2.5	2.8	2.38	2.755	3.27	3.1	2.6	2.4	2.1	3.3
1.53	2.3	1.7	2.1	2.93	2.9	1.19	3.9	1.92	3.05	2.4
2.25	1.8	1.7	1.96	2.1	3	2.71	2.9	2.3	1.8	2
2.4	2.7	2.49	3	2.9	3.5	3.3	2.8	1.17	2.6	2
3.1	3.3	2.9	1.43	2.4	3.1	2.1	3.1	3.8	2.6	2.8
2.49	2	2.2	2.5	3.47	1.8	1.6	2.78	2.445	2.7	2.4
2.9	2.52	2.7	2.3	2.5	2.8	3.5	3.05	2.4	3.4	2
3.2	3.05	3.6	2.1	2.17	2.85	3.3	1.8	2.4	1.79	2.2
2.9	3.3	2.7	2.7	2.4	1.63	1.1	2.1	2.6	2.82	2.7
2.99	1.97	1.8	3.1	2.5	2.5	3.4	2.9	2.6	2.31	2.6
2.56	2.1	2.3	2	1.3	2.8	2.6	2	2.7	2.36	3.05
3.4	1.9	2.5	2.41	2	3.2	2.5	2.3	2.9	1.94	2.1
2.8	2.9	2.3515	1.4	3.8	1.52	3.17	2	2.9	2.4	2.6
3.2	2.2	2.4	2.31	3	2.9	1.4	2.38	2.6	2.21	1.95
4.59	2.27	2.43	2.8	2.2	2.8	3.15	2.69	1.95	1.4	2.4
2.21	2.4	2.71	2.5	3.1	1.8	2.21	2.23	2.1	1.9	2.3
2.6	2.1	2.63	2.4	2.54	2.6	3.3	3.5	2.3	2.17	1.98
2.4	2.7	0.9	2	1.8	2.63	1.2	2.21	4.2	2.56	3.5
2.2	2.8	0.89	1.9	2.9	3.7	2	2.5	2.6	3	3.2
2.4	3.49	3.6	1.7	2.3	2.3	2.5	3	1.9	1.9	2.5
5.8585	1.67	3.27	2	2.6	1.9	2.4	1.8	2.4	3.9	2.42
4	3.2	2.7	2.4	2.8	2.1	2.5	3.4	3.4	2.7	1.89
2.2	2.2	2.8	2.5	1.3	2.1	2.2	2.9	3.7	2.9	2
3.3	1.76	2.05	2.23	1.9	2.1	3	3.03	2.3	1.8	2.9
1.6	2.9	1.34	1.8	2.5	1.8	3.3	1.8	2.5	2.48	2.23
2.5	3.6	2.2	2.3	2.7	0.83	2.9	2.2	2.5	1.73	2.7
3.31	2.1	3	3.46	2.2	2.31	2.5	1.8	2.3	3.1	2.1
3	3.21	3.4	2.6	1.8	3	3	3.9	3.1	3.3	2
0.4	3	4.1	1.66	2.8	2.4	2	2.89	2.86	3	2.82
2.1	3	3.4	2.3	2.9	3.71	3.7	2.4	2.42	2.8	2.1
2.1	2.5	2	3.4	1.7	2.6	2.1	3.23	4.7	1.9	1.8
2.41	1.48	3	2.12	2	3.2	2.35	1.7	2.5	2.8	2.9
2.32	2.8	2	2.3	2.5	3.04	4.29	1.9	3.11	2	1.3
2.02	3.1	3.1	2.1	2.36	1.46	2.1	2.37	1.6	1.46	2.66
1.8	2.2	1.9	2.2	0.7	1.98	3.4	3.1	2.5	3.1	2.3
4.3	2.4	2.2	2	3.22	2.06	1.64	1.6	2.78	2.395	1.9
2.3	2.6	2.4	1.18	1.45	3.8	2.2	2.5	3.8	1.7	2.1
4.1	2.7	2.5	2.5	3.9	1.86	2.33	2.8	1.3	2.5	2.9
2.2	2.2	2.65	2.53	1.93	2.1	1.9	2.1	3.1	3.605	3.8
2.6	2.03	2.81	2.9	2.7	2.7	2.7	2.67	2.5	2.2	2
2.98	3	2.4	2.4	2.2	3.5	1.91	3.5	2.74	2.3	2.39
2.69	2.4	1.9	2.9	2.07	2.1	3.02	2.5	2.02	3.1	2.39

1.56	2.49	2.9	2.9	3.5	2.7	2.2	2.7	3.8	3.25	2.3
2	2.9	2.5	2.1	2.5	2	2.2	3.01	2.3	2.8	2.4
3.77	3	1.8	2.8	2.62	3.63	3	4.3	3	3.1	2.04
1.9	3.24	2.5	2.5	3.16	1.3	3.78	2.7	3.5	2.5	3.15
2.6	2	3	2.41	1.61	2.61	3	2.3	2	1.3	3.24
3.2	2.3	1.8	2.4	2.23	2.9	3.4	2.1	2.8	2	2.41
2.2	2.1	2.6	1.9	2.9	2.2	3.14	1.3	2.6	2.6	0.84
2.4	1.46	1.6	2.4	2.65	2.1	2.82	2.2	1.32	2.6	2.4
3.1	3.02	2.18	1.89	1.44	1.83	2.6	1.5	3.01	2.4	2.8
6.25	2.7	2.3	2.35	3.3	1.72	2.2	2	1.9	2.42	2.1
1.8	2.17	3.5	3.3	2.69	2.1	2.24	1.25	1.11	2.5	3.5
2.8	1	3	2.4	3.2	2.4	1.64	2.6	2.35	2.4	4
2.1	3.1	3.2	1.98	2.9	2.4	2.4	1.8	3.1	1.4	3
2.6	1.6	2.7	2.4	4.4	1.19	2.8	2.5	2.8	2.6	2.43
2.4	1.21	1.8	2.5	2.14	2.28	1.4	3	2.77	2.9	2.1
1.71	2.65	2	1.57	2.97	3.64	2.5	2.2	2.6	1.8	3
2.6	2.56	2.48	2.58	2.9	1.41	3.8	2.08	2.5	2.62	2.8
3.2	1.8	2.9	1	2.68	3.7	3.1	2.15	1.83	2.1	2.5
2.4	1.67	3.7	2.1	2.45	2.1	3	2.9	2.6	2.6	1.8
1.8	2.84	3.5	2.8	2.3	1.68	2.49	2.95	2.41	3	1.51
2.71	1.7	1.7	2.44	2.4	3	1.93	2	2.9	2.7	3
2.3	3.08	2.4	3	1.52	1.67	1.8	2.7	2.4	2.31	2.69
2	2.19	2.4	2.9	2.84	3	2.93	3	2.2	2.1	3
2.9	2.09	2.7	2.1	3.28	2.9	2.6	2.3	1.6	1.15	2.3
1.66	3.9	2.8	3.8	2.8	1.8	2.5	2.6	2.5	2.9	2.84
1.85	2.15	3.7	2.9	1.9	2	3.55	3	1.7	2.2	4.02
2.6	2.8	1.7	2.3	2.59	1.95	3	1.9	1.55	1.8	2.5
2.3	1.88	2.44	1.5	3.1	2.75	2.09	2.4	2.5	2.8	2.5
1.5	3.07	2.47	2.2	2	2.2	3.2	1.9	2.1	1.5	4.3
2.6	2.8	1.92	2.6	2.5	3.5	2.45	2.7	2.2	3.9	1.5
3.6	2.5	3.19	2.1	2.3	2.3	2.7	2.9	1.75	2.2	2.4
3.06	2.5	3.2	2.6	3.48	2.8	2.9	2.7	1.2	3.14	2.6
3.65	2.4	2.6	1.69	2.5	2.5	1.3	1.43	2.9	2.2	2.5
2.1	2.81	1.93	2.2	2.1	2.89	2.3	2.6	1.8	2.5	
3.15	2.1	2.09	2.6	2.02	3.6	2.3	2.4	2.45	1.72	
2.52	2.03	2.5	3.2	1.8	2.5	3.2	1.7	2.6	3.2	
3.4	1.5	2.9	2.2	2.3	2.5	2.21	1.2	3.12	1.6	
2.1	2.12	1.89	2.7	2.9	2.01	2.5	3.33	3	2.41	
2.3	2.2	2.4	2.5	3.1	2.53	2.4	3.475	2.6	2	
2.1	3.2	3	1.91	2.1	2.4	2.7	2	2	1.75	
3.1	2.2	2.45	3.1	2.92	2.8	3.1	0.69	2.3	3.5	
3.8	2.4	1.46	3.3	2.4	2.35	2.1	3.8	2.7	2.6	
3.57	3.3	2.82	1.35	2.5	2.1	3.3	2.54	2.38	1.9	