Statistički testovi za lokacijski parametar

Bilandžić, Lana Lucia

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Department of Mathematics / Sveučilište Josipa Jurja Strossmayera u Osijeku, Odjel za matematiku**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:126:010014

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-04-26**

Repository / Repozitorij:

Repository of School of Applied Mathematics and Computer Science

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku
Diplomski studij matematike
Financijska matematika i statistika

Lana Lucia Bilandžić

Statistički testovi za lokacijski parametar

Diplomski rad

Osijek, 2021.

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku
Diplomski studij matematike
Financijska matematika i statistika

Lana Lucia Bilandžić

Statistički testovi za lokacijski parametar

Diplomski rad

Mentor: doc. dr. sc. Danijel Grahovac

Osijek, 2021.

Sadržaj

1 Uvod **2**

2 Testiranje statističkih hipoteza **3**

3 Statistički testovi za lokacijski parametar za jedan uzorak **8**
3.1 Z-test 8
3.2 Studentov t-test 10
3.2.1 Veliki uzorci 12
3.3 Wilcoxonov test ranga i predznaka 16
3.3.1 Normalna aproksimacija za velike uzorke 17
3.4 Test predznaka 18

4 Statistički testovi za lokacijski parametar za dva uzorka **19**
4.1 Z-test 19
4.2 t-test 19
4.3 Behrens-Fisherov problem 22
4.3.1 Welchov t-test 22

5 Mann-Whitney-Wilcoxonov test **24**
5.1 Veliki uzorci 26

6 Simulacije **27**

Literatura **35**

Sažetak **36**

Summary **36**

Životopis **37**

1 Uvod

Testiranje usporedbe lokacija ili centralne tendencije, odnosno 'sredine', dvaju nezavisnih uzoraka u svrhu znanstvenih istraživanja sve je češća pojava. Tako, na primjer, u biomedicinskom istraživanju, istraživač je zainteresiran o efektu nekog tretmana u smislu mjere lokacijskog pomaka. Zimmerman (1987., 1992.) i Gibbons i Chakraborti (1991.,1992.) postavljaju pitanje o prigodnom testu za analiziranje lokacijskog pomaka. Radi podložnosti ovisnosti o distribuciji uzoraka, postoje razni statistički testovi korišteni u tu svrhu, a interpretiraju se kao dokaz o jednakosti ili različitosti očekivanja ili medijana.

Najčešće korišten pristup pri testiranju jednakosti očekivanja podrazumijeva *Studentov* t-test na dvama uzorcima. Kako je isti određen pretpostavkama o normalnosti podataka i jednakošću varijanci populacija, koje su nerijetko narušene u primjeni, stvara se potreba za novim testovima. Dizajniran za različite varijance, uz očuvanu pretpostavku o normalnosti, *Welchov* t-test nastaje kao modifikacija t-testa. Uz narušenost pretpostavke o normalnosti distribucija, dimenzija uzorka bitan je čimbenik. Naime, ako su uzorci dovoljno veliki, koristeći asimptotske rezultate, navedeni testovi i dalje su prihvatljivi. *Welchov* t-test suzbija problem nejednakosti varijanci, ali ne i pretpostavku o normalnosti. Postavlja se pitanje koji test koristiti za male uzorke koji nisu normalno distribuirani. Najčešća neparametarska alternativa je MWW test koji je egzaktan, a čiji se rezultati interpretiraju kao test o jednakosti medijana jedino ako su distribucije uzoraka jednake osim za mogući lokacijski pomak. Stoga je MWW test dobar izbor za male uzorke, no osjetljiv je na pretpostavke o nejednakosti varijanci u kombinaciji s velikom razlikom dimenzija uzoraka.

Za početak ćemo definirati neke osnovne pojmove vezane za testiranje statističkih hipoteza. Objasnit ćemo testove za lokacijski parametar za jedan uzorak; parametarske testove o očekivanju, z test i *Studentov* t-test te neparametarske testove o medijanu, *test predznaka* i *Wilcoxonov test ranga i predznaka*. Nakon toga ćemo objasniti problematiku o testiranju lokacijskog parametra za dva uzorka. Jakosti predloženih statističkih testova ovisit će o različitim čimbenicima i pretpostavkama distribucije uzoraka, kao što su veličine uzoraka te njihova razlika, nejednakost varijanci, asimetričnost distribucija te težina u repovima distribucija. Obzirom na navedeni teorijski dio, simulacijama ćemo pokušati odrediti prikladan test u ovisnosti o navedenim čimbenicima.

2 Testiranje statističkih hipoteza

Statističko zaključivanje o nekom obilježju populacije provodimo na temelju prikupljenih podataka dijela populacije. Pri tome, podatke kojima raspolažemo smatramo realizacijama slučajnog vektora $\mathbf{X} = (X_1, \ldots, X_n)$. Slučajni vektor $\mathbf{X}$ naziva se **slučajni uzorak**.

Definicija 2.1. Statistički model $\mathcal{P}$ *je familija dozvoljenih funkcija distribucije slučajnog vektora za koji baza podataka čini jednu realizaciju.*

Definicija 2.2. *Slučajni vektor* $\mathbf{X} = (X_1, \ldots, X_n)$ *je* **jednostavni slučajni uzorak** *iz distribucije* F *ako vrijedi:*

1. $X_1, \ldots, X_n$ *su nezavisne,*
2. $X_1, \ldots, X_n$ *imaju funkciju distribucije* F.

Mnoge praktične situacije zahtijevaju donošenje odluke o istinitosti ili neistinosti određene slutnje. Slutnja koju želimo testirati formulira se u terminima distribucije slučajne varijable te se naziva *statistička hipoteza*, $\mathcal{H}$. Obzirom da je statistički model $\mathcal{P}$ familija dozvoljenih funkcija distribucija prilikom zaključivanja o zadanom problemu, svaka statistička hipoteza $\mathcal{H}$ podskup je statističkog modela $\mathcal{P}$. Općenito se problem testiranja statističke hipoteze na temelju realizacija $(x_1, \ldots, x_n)$ slučajnog uzorka $(X_1, \ldots, X_n)$ svodi na donošenje odluke o prihvaćanju ili odbacivanju hipoteze.
Testiranje se provodi formuliranjem dviju hipoteza, $\mathcal{H}_0$ koju nazivamo **nul − hipoteza** te $\mathcal{H}_1$, **alternativna hipoteza**. Nul-hipoteza i alternativna hipoteza dijele statistički model $\mathcal{P}$ na dva disjunktna podskupa

$$\mathcal{P} = \mathcal{H}_0 \cup \mathcal{H}_1.$$

Pravilo temeljeno na realizaciji slučajnog uzorka na temelju kojeg donosimo odluku o odbacivanju ili neodbacivanju nul-hipoteze naziva se *statistički test*. Statistički test dijeli skup svih mogućih realizacija na dva disjunktna skupa C_r i C_r^c. C_r je područje odbacivanja nul-hipoteze $\mathcal{H}_0$ te se naziva **kritično područje**. Odluke o odbacivanju ili neodbacivanju nul-hipoteze u korist alternativne hipoteze donesene statističkim testom temeljene na uzorcima iz populacije mogu biti ispravne ili pogrešne.
Dvije su mogućnosti pogrešne odluke:

pogreška I. tipa : *odbaciti* $\mathcal{H}_0$ *ako je ona istinita*

pogreška II. tipa : *ne odbaciti* $\mathcal{H}_0$ *ako je* $\mathcal{H}_1$ *istinita.*

Cilj je kreirati test koji će imati što je moguće manje vjerojatnosti pogreške oba tipa, a one ovise o distribuciji varijable na koju se odnose hipoteze. Stoga, za svaku distribuciju F iz statističkog modela $\mathcal{P}$ moramo razmatrati vjerojatnost pogreške. U tu svrhu definiramo funkciju jakosti testa.

Definicija 2.3. *Neka je* $\mathbf{X} = (X_1, \ldots, X_n)$ *slučajni uzorak statističkog modela* $\mathcal{P}$ *i* C_r *kritično područje statističkog testa.* **Funkcija jakosti testa** *je funkcija* $\pi : \mathcal{P} \to [0, 1]$ *definirana izrazom*

$$\pi(F) = P_F(\mathbf{X} \in C_r).$$

Ako je $F \in \mathcal{H}_0$, onda je $\pi(F)$ vjerojatnost pogrešnog odbacivanja $\mathcal{H}_0$, tj. vjerojatnost pogreške prvog tipa.
Ako je $F \in \mathcal{H}_1$, onda je $\pi(F)$ vjerojatnost dobre odluke jer je $\mathcal{H}_0$ odbačena u korist $\mathcal{H}_1$.
Vjerojatnost pogreške drugog tipa je $P_F(\mathbf{X} \in {C_r}^c) = 1 - P_F(\mathbf{X} \in C_r) = 1 - \pi(F)$.
Vjerojatnosti pogreške prvog i drugog tipa često se označavaju s $\alpha(F)$ i $\beta(F)$, pri čemu su α i β preslikavanja za koje vrijedi

$$\alpha : \mathcal{H}_0 \to [0,1],\ \alpha(F) = \pi(F),$$

$$\beta : \mathcal{H}_1 \to [0,1],\ \beta(F) = 1 - \pi(F).$$

Odabir kritičnog područja temelji se na izboru maksimalne vjerojatnosti pogreške prvog tipa koja se želi prihvatiti. Odabrana maksimalna vjerojatnost pogreške prvog tipa naziva se **razina značajnosti testa** i označava s α. Uglavnom se za razinu značajnosti testa uzimaju brojevi 0.01, 0.05 ili 0.1. Najmanja razina značajnosti uz koju bi $\mathcal{H}_0$ bila odbačena naziva se $p - vrijednost$.
Kritično se područje obično izražava u terminima funkcije slučajnog uzorka $T = t(X_1, \ldots, X_n)$, a funkcija $T = t(\mathbf{X}) : \mathbb{R}^n \to \mathbb{R}$ naziva se *test statistika*.
Radi zaključivanja o lokacijskom parametru distribucije, pri kreiranju statističkih testova fokusirat ćemo se na parametarski statistički model $\mathcal{P} = \{F_\theta : \theta \in \Theta\}$, pri čemu je Θ prostor parametara. Postupak kreiranja testa započinje proučavanjem problema procjene za parametar.

Intuitivni pristup kreiranja statističkog testa podrazumijeva:

1. Pronaći test statistiku $T = t(\mathbf{X})$ čija se distribucija razlikuje u uvjetima $\mathcal{H}_0$ i $\mathcal{H}_1$. Znati distribuciju test statistike za svaki parametar iz $\mathcal{H}_0$.

2. Podijeliti skup svih mogućih realizacija test statistike na C_r i C_r^c na način da je

$$P_\theta(t(\mathbf{X}) \in C_r) \leqslant \alpha, \quad \forall\ \theta \in \mathcal{H}_0.$$

   Realizacije iz odabranog kritičnog područja idu u prilog alternativnoj hipotezi, pri čemu je α maksimalna vjerojatnost pogreške prve vrste.

Jedna od metoda procjene parametara je metoda maksimalne vjerodostojnosti, a ista se temelji na procjeni maksimizacije funkcije vjerodostojnosti.

Definicija 2.4. *Za danu realizaciju* $\mathbf{x} = (x_1, \ldots, x_n)$ *slučajnog uzorka* $(X_1, \ldots, X_n)$ *iz gustoće* $f(\mathbf{x}; \boldsymbol{\theta})$ *funkcija vjerodostojnosti je* $L(\boldsymbol{\theta}; \mathbf{x}) = f(\mathbf{x}; \boldsymbol{\theta})$, *tj. gustoća za fiksan* $\mathbf{x}$ *razmatrana kao funkcija parametra.*

Za jednostavni slučajni uzorak $\mathbf{X} = (X_1, \ldots, X_n)$ iz gustoće $f(\mathbf{x}; \boldsymbol{\theta})$, funkcija vjerodostojnosti je oblika

$$L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}).$$

Definicija 2.5. *Neka je* $\mathbf{X}$ *slučajni uzorak iz funkcije gustoće* $f(\mathbf{x};\boldsymbol{\theta})$, $\boldsymbol{\theta}\in\Theta$. *Uz danu realizaciju tog uzorka* $\mathbf{x}$, *neka funkcija vjerodostojnosti* $L(\boldsymbol{\theta};\mathbf{x})$ *postiže svoj maksimum po* $\boldsymbol{\theta}$ *u* $s(\mathbf{x})$, *tj.*

$$\max_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta};\mathbf{x}) = L(s(\mathbf{x});\mathbf{x}), \qquad s(\mathbf{x})\in\Theta.$$

Tada statistiku $S = s(\mathbf{X})$ *zovemo procjenitelj maksimalne vjerodostojnosti (eng. Maximum Likelihood Estimator).*

Za procjenitelje maksimalne vjerodostojnosti kraće ćemo pisati ML procjenitelji.

Primjer 2.6. *Neka je* $\mathbf{X} = (X_1,\ldots,X_n)$ *jednostavni slučajni uzorak iz* $\mathcal{N}(\mu,\sigma^2)$. *Procijenimo parametre* μ *i* σ^2 *metodom maksimalne vjerodostojnosti.*
Funkcija vjerodostojnosti je oblika

$$L(\mu,\sigma^2;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}.$$

Kako je funkcija prirodnog logaritma strogo monotono rastuća, maksimizacija funkcije vjerodostojnosti odgovara maksimizaciji

$$\ln L(\mu,\sigma^2;\mathbf{x}).$$

Tražimo vrijednosti parametara μ *i* σ^2, *označimo ih s* μ_{ML} *i* σ^2_{ML}, *koji maksimiziraju*

$$\begin{aligned}
\ln L(\mu,\sigma^2;\mathbf{x}) &= \sum_{i=1}^{n} \ln\Big(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\Big) \\
&= \sum_{i=1}^{n} \Big(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x_i-\mu)^2\Big) \\
&= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.
\end{aligned}$$

Izjednačavanjem parcijalnih derivacija pripadne funkcije s nulom, dobivamo kandidate za ekstreme:

$$\frac{\partial(\ln L(\mu,\sigma^2;\mathbf{x}))}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\Big(\sum_{i=1}^{n} x_i - n\mu\Big) = 0$$

$\Longleftrightarrow$

$$\mu_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n;$$

$$\frac{\partial(\ln L(\mu,\sigma^2;\mathbf{x}))}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$\Longleftrightarrow$

$$\sum_{i=1}^{n}(x_i-\mu)^2 = n\sigma^2,$$

odnosno, uvrštavanjem μ_{ML},

$$\sigma^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2.$$

Da su to zaista vrijednosti parametara za koje se postiže maksimum logaritma funkcije vjerodostojnosti, potvrđuje determinanta pripadne Hesseove matrice (vidi [9]).

Testovi kreirani intuitivnim pristupom baziraju se na razumnoj test statistici, no ne znači da su upravo takvi testovi najbolji u nekom smislu. Strategija u odabiru najboljeg testa je odabrati onaj s najmanjom vjerojatnošću pogreške druge vrste.

Definicija 2.7. *Neka je* $\mathbf{X} = (X_1, \ldots, X_n)$ *slučajni uzorak iz funkcije gustoće* $f(\mathbf{x}; \theta)$. *Test za testiranje hipoteza*

$$\mathcal{H}_0 : \theta = \theta_0$$

$$\mathcal{H}_1 : \theta = \theta_1$$

s pripadnim kritičnim područjem C_r^* *je najjači test nivoa značajnosti* α *ako vrijedi:*

1. $P_{\theta_0}(\mathbf{X} \in C_r^*) = \alpha$
2. $P_{\theta_1}(\mathbf{X} \in C_r^*) \geqslant P_{\theta_0}(\mathbf{X} \in C_r)$

za svako drugo kritično područje C_r *nivoa značajnosti* α. *Takvo se kritično područje* C_r^* *naziva najjače kritično područje nivoa značajnosti* α.

Teorem 2.8. ***Neyman-Pearsonova lema.***[1]
Neka je $\mathbf{X} = (X_1, \ldots, X_n)$ *slučajni uzorak iz funkcije gustoće* $f(\mathbf{x}; \theta)$. *Označimo s*

$$\lambda(\mathbf{x}) = \frac{f(\mathbf{x}; \theta_0)}{f(\mathbf{x}; \theta_1)}$$

i

$$C_r^* = \{\mathbf{x} : \lambda(\mathbf{x}) \leqslant k\},$$

pri čemu je k *konstanta za koju vrijedi* $P_{\theta_0}(\mathbf{X} \in C_r^*) = \alpha$. *Tada je* C_r^* *najjače kritično područje nivoa značajnosti* α *za testiranje hipoteza*

$$\mathcal{H}_0 : \theta = \theta_0$$

$$\mathcal{H}_1 : \theta = \theta_1.$$

Neyman-Pearsonov pristup kreiranju statističkog testa odnosi se na testiranje jednostavnih hipoteza, [1] no može se ponekad proširiti i na složene.
Test za testiranje složenih hipoteza, koje nisu nužno jednostrane, može se kreirati *metodom generaliziranog kvocijenta vjerodostojnosti* koji nastaje kao generalizacija Neyman-Pearsonove metode, [1].

[1] Jednostavna hipoteza je ona koja jednoznačno određuje distribuciju, npr. $\mathcal{H} : \theta = 5$. U suprotnome, kažemo da je hipoteza složena, npr. $\mathcal{H} : \theta > 5$.

Definicija 2.9. *Neka je dan* $\mathbf{X} = (X_1, \ldots, X_n)$ *slučajni uzorak iz modela* $\{f(\mathbf{x}; \boldsymbol{\theta}) : \theta \in \Theta\}$ *i hipoteze*

$$\mathcal{H}_0 : \boldsymbol{\theta} \in \Theta_0$$

$$\mathcal{H}_1 : \boldsymbol{\theta} \in \Theta_1 = \Theta\Theta_0.$$

Generalizirani kvocijent vjerodostojnosti definiran je kao

$$\lambda(\mathbf{x}) = \frac{\underset{\theta\in\Theta_0}{max} f(\mathbf{x}; \boldsymbol{\theta})}{\underset{\theta\in\Theta}{max} f(\mathbf{x}; \boldsymbol{\theta})} = \frac{f(\mathbf{x}; \boldsymbol{\theta}_{0ML})}{f(\mathbf{x}; \boldsymbol{\theta}_{ML})}, \tag{1}$$

pri čemu je $\boldsymbol{\theta}_{ML}$ *ML procjenitelj na cijelom skupu dozvoljenih parametara* Θ*, dok je* $\boldsymbol{\theta}_{0ML}$ *ML procjenitelj za* $\boldsymbol{\theta}$ *pod restrikcijom da je* $\mathcal{H}_0$ *istinita.*

Test se kreira tako da male vrijednosti $\lambda(\mathbf{x})$ idu u prilog alternativnoj hipotezi, odnosno

$$C_r = \{\mathbf{x} : \lambda(\mathbf{x}) \leqslant k\},$$

pri čemu je k određen odabranim nivoom značajnosti α.

3 Statistički testovi za lokacijski parametar za jedan uzorak

Testovi za lokacijski parametar distribucije podrazumijevat će testiranje hipoteza o očekivanju μ ili medijanu M kao mjeri "sredine" distribucije. Testovima o očekivanju želimo testirati hipotezu

$$\mathcal{H}_0 : \mu = \mu_0,$$

pri čemu je μ nepoznato očekivanje, a μ_0 vrijednost s kojom uspoređujemo μ. Usredotočit ćemo se na dvostrane statističke testove, ondosno testove čiji je oblik alternativne hipoteze

$$\mathcal{H}_1 : \mu \neq \mu_0.$$

3.1 Z-test

Pretpostavke koje želimo ocijeniti su dolazi li uzorak od n ispitanika iz populacije u kojoj je očekivanje jednako određenoj vrijednosti.
Z test upotrebljava se isključivo u slučaju ako je populacijska varijanca *poznata*.

Neka je dan statistički model jednostavnog slučajnog uzorka iz *normalne* distribucije s očekivanjem μ i varijancom σ^2, tj. $\mathbf{X} = (X_1, \ldots, X_n)$ je jednostavni slučajni uzorak iz $\mathcal{N}(\mu, \sigma^2)$, pri čemu je σ^2 *poznata*. Važnost pretpostavke o normalnosti smanjuje se povećanjem veličine uzorka.

Test za testiranje hipoteza

$$\mathcal{H}_0 : \mu = \mu_0 \tag{2}$$
$$\mathcal{H}_1 : \mu \neq \mu_0, \tag{3}$$

pri čemu je σ^2 poznato, kreirat će se korištenjem generaliziranog kvocijenta vjerodostojnosti (1),

$$\lambda(\mathbf{x}) = \frac{f(\mathbf{x}; \mu_0, \sigma^2)}{\max\limits_{\mu \in \mathbb{R}} f(\mathbf{x}; \mu, \sigma^2)} = \frac{f(\mathbf{x}; \mu_0, \sigma^2)}{f(\mathbf{x}; \mu_{ML}, \sigma^2)}. \tag{4}$$

Funkcija gustoće u ovom modelu je oblika:

$$f(\mathbf{x}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

a ML procjenitelj očekivanja μ, dobiven maksimizacijom funkcije vjerodostojnosti u (2.6), je aritmetička sredina uzorka:

$$\mu_{ML} = \overline{x}_n.$$

Uvrštavanjem u (4), generalizirani kvocijent vjerodostojnosti za realizaciju $\mathbf{x}$ dan je s:

$$\lambda(\mathbf{x}) = e^{-\frac{1}{2\sigma^2}(-2n\mu_0\overline{x}_n + n\mu_0^2 + n\overline{x}_n^2)}.$$

Kritično je područje oblika:

$$C_r = \Big\{\mathbf{x} : \lambda(\mathbf{x}) \leqslant k\Big\},$$

pri čemu je k određen odabranim nivoom značajnosti α.
Uvjet $\lambda(\mathbf{x}) \leqslant k$ ekvivalentan je uvjetu $\ln\lambda(\mathbf{x}) \leqslant k_1$, tj.

$$\ln\lambda(\mathbf{x}) = -\frac{n}{2\sigma^2}(\overline{x}_n - \mu_0)^2 \leqslant k_1$$

odgovara

$$\Big(\frac{\overline{x}_n - \mu_0}{\sigma}\sqrt{n}\Big)^2 \geqslant k_2,$$

odnosno,

$$\left|\frac{\sqrt{n}(\overline{x}_n - \mu_0)}{\sigma}\right| \geqslant k_3.$$

Konačno, kritično je područje oblika:

$$C_r = \left\{\mathbf{x} : \left|\frac{\sqrt{n}(\overline{x}_n - \mu_0)}{\sigma}\right| \geqslant k\right\},$$

a pripadna test-statistika, nastala uvjetom kritičnog područja uz istinitost nul-hipoteze,

$$Z = \left|\frac{\sqrt{n}(\overline{x}_n - \mu_0)}{\sigma}\right|$$

ima standardnu normalnu distribciju, dok je k $(1 - \frac{\alpha}{2})$ kvantil pripadne distribucije.

Dakle, z-test o očekivanju normalne distribucije, uz poznatu varijancu, za testiranje hipoteza (2) i (3) koristi test-statistiku

$$Z = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma},$$

za koju je, u uvjetima istinitosti $\mathcal{H}_0$, $Z \sim \mathcal{N}(0,1)$. Pripadno kritično područje nivoa značajnosti α je oblika $C_r = \{\mathbf{x} : |Z| \geqslant z_{1-\frac{\alpha}{2}}\}$, pri čemu je $z_{1-\frac{\alpha}{2}}$ $(1 - \frac{\alpha}{2})$ kvantil standardne normalne distribucije.

3.2 Studentov t-test

U slučaju da je varijanca uzorka *nepoznata*, za testiranje hipoteza o vrijednosti očekivanja, koristi se *Studentov* t-test.
Pretpostavimo ponovno model jednostavnog slučajnog uzorka iz $\mathcal{N}(\mu, \sigma^2)$, ali uz nepoznatu varijancu uzorka σ^2. Sada se koristi uzoračka varijanca kao procjenitelj iste.
T-test jedan je od mnogih statističkih testova čija se test statistika bazira na Studentovoj t distribuciji. Poput normalne distribucije, t-distribucija je također neprekidna čija je funkcija gustoće simetrična, u obliku zvona ("zvonasta"). Razlika je u tome što t distribucija ima teže repove i t-interval je širi, u smislu procjene očekivanja, nego kad se koriste kvantili standardne normalne distribucije.

Neka je dan statistički model jednostavnog slučajnog uzorka iz normalne distribucije s očekivanjem μ i varijancom σ^2. Test za testiranje hipoteza:

$$\mathcal{H}_0 : \mu = \mu_0 \tag{5}$$
$$\mathcal{H}_1 : \mu \neq \mu_0, \tag{6}$$

pri čemu je σ^2 nepoznato, kreirat će se korištenjem generaliziranog kvocijenta vjerodostojnosti

$$\lambda(\mathbf{x}) = \frac{\underset{\sigma^2>0}{max} f(\mathbf{x}; \mu_0, \sigma^2)}{\underset{\mu\in\mathbb{R},\sigma^2>0}{max} f(\mathbf{x}; \mu, \sigma^2)} = \frac{f(\mathbf{x}; \mu_0, \sigma^2_{0ML})}{f(\mathbf{x}; \mu_{ML}, \sigma^2_{ML})}. \tag{7}$$

Funkcija gustoće u ovom modelu je oblika:

$$f(\mathbf{x}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

a ML procjenitelji za parametre, dobiveni maksimizacijom funkcije vjerodostojnosti u 2.6, su sljedeći:

$$\mu_{ML} = \overline{x}_n$$
$$\sigma^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2$$
$$\sigma^2_{0ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2.$$

Pri tome su μ_{ML} i σ^2_{ML} ML procjenitelji za μ i σ^2 na cijelom skupu dozvoljenih vrijednosti parametara, dok je σ^2_{0ML} ML procjenitelj za σ^2 u uvjetima istinitosti nul-hipoteze, tj. za $\mu = \mu_0$. Uvrštavanjem u (7), generalizirani kvocijent vjerodostojnosti za realizaciju $\mathbf{x}$ dan je s:

$$\lambda(\mathbf{x}) = \frac{\sigma^2_{ML}}{\sigma^2_{0ML}}.$$

U skladu s Neyman-Pearsonovom lemom, kritično je područje oblika:

$$C_r = \Big\{ \mathbf{x} : \lambda(\mathbf{x}) \leqslant k \Big\},$$

pri čemu je k određen odabranim nivoom značajnosti α.
Kako je

$$\frac{\sigma_{0ML}^2}{\sigma_{ML}^2} = \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sum_{i=1}^n (x_i - \overline{x}_n)^2} = \frac{\sum_{i=1}^n (x_i - \overline{x}_n + \overline{x}_n - \mu_0)^2}{\sum_{i=1}^n (x_i - \overline{x}_n)^2} = 1 + \frac{n(\overline{x}_n - \mu_o)^2}{\sum_{i=1}^n (x_i - \overline{x}_n)^2},$$

uvjet $\lambda(\mathbf{x}) \leqslant k$ odgovara:

$$\frac{\sigma_{0ML}^2}{\sigma_{ML}^2} - 1 \geqslant k_1,$$

odnosno,

$$\left| \frac{\sqrt{n}(\overline{x}_n - \mu_0)}{\sqrt{\sum_{i=1}^n (x_i - \overline{x}_n)^2}} \right| \geqslant k_2.$$

Iskoristivši izraz za uzoračku funkciju varijance,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x}_n)^2,$$

dani je uvjet

$$\left| \frac{\sqrt{n}(\overline{x}_n - \mu_0)}{s} \right| \geqslant k_3.$$

Kritično je područje oblika:

$$C_r = \left\{ \mathbf{x} : \left| \frac{\sqrt{n}(\overline{x}_n - \mu_0)}{s} \right| \geqslant k \right\},$$

a pripadna test-statistika, nastala uvjetom kritičnog područja uz istinitost nul-hipoteze,

$$T = \left| \frac{\sqrt{n}(\overline{x}_n - \mu_0)}{s} \right|$$

ima Studentovu distribuciju s $n-1$ stupnjeva slobode, dok je k $(1 - \frac{\alpha}{2})$ kvantil pripadne distribucije.

Dakle, t-test o očekivanju normalne distribucije s nepoznatom varijancom za testiranje hipoteza (5) i (6) koristi test-statistiku

$$T = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S},$$

za koju je, u uvjetima istinitosti $\mathcal{H}_0$, $T \sim t(n-1)$. Pripadno kritično područje nivoa značajnosti α je oblika $C_r = \{\mathbf{x} : |t(\mathbf{x})| \geqslant t(n-1)_{1-\frac{\alpha}{2}}\}$ obzirom da velike apsolutne vrijednosti ove statistike idu u prilog alternativnoj hipotezi.

3.2.1 Veliki uzorci

Ukoliko populacija nema normalnu distribuciju, a uzorak je dovoljno velik, centralni granični teorem dozvoljava korištenje z testa u svrhu testiranja hipoteza o očekivanju.

Teorem 3.1. *(Levyjev centralni granični teorem, [8].) Neka je* $(X_n,\ n \in \mathbb{N})$ *niz nezavisnih jednako distribuiranih slučajnih varijabl s očekivanjem* μ *i varijancom* $\sigma^2 < \infty$*. Tada niz slučajnih varijabli*

$$\left(\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n},\ n \in \mathbb{N} \right)$$

konvergira po distribuciji prema standardnoj normalnoj slučajnoj varijabli, tj.

$$\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \ \xrightarrow{D}\ \mathcal{N}(0,1),\ n \to \infty.$$

Dakle, prema centralnom graničnom teoremu distribucija test statistike z testa,

$$Z = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma},$$

konvergira prema standardnoj normalnoj distribuciji. Uz korištenje teorema Slutsky [2], konzistentnost procjenitelja S_n za parametar σ, tj.

$$S_n \xrightarrow{\mathrm{P}}\ \sigma,\ n \to \infty,$$

odnosno

$$\frac{\sigma}{S_n} \xrightarrow{\mathrm{P}}\ 1,\ n \to \infty$$

podrazumijeva

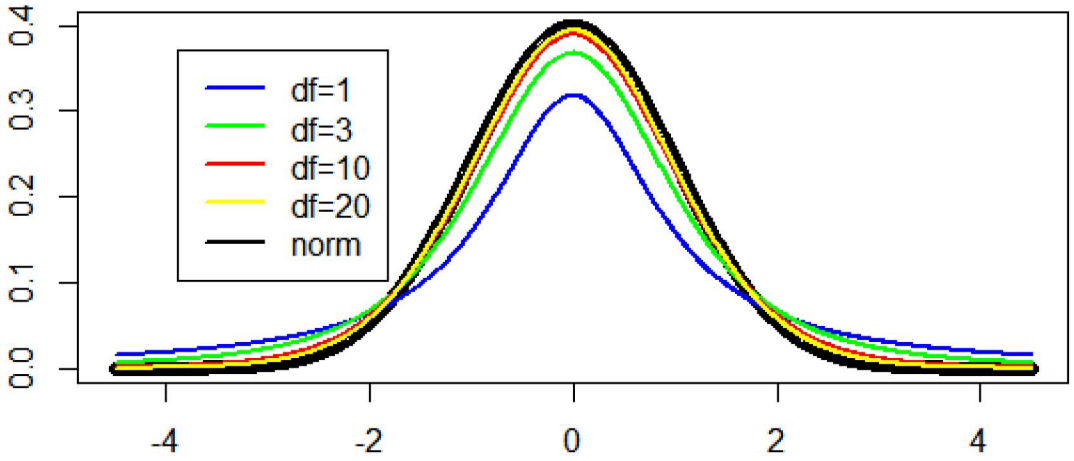
$$T = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} = \frac{\overline{X}_n - \mu}{S_n}\sqrt{n} = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \cdot \frac{\sigma}{S_n} \ \xrightarrow{\mathrm{D}}\ Z \sim \mathcal{N}(0,1),\ n \to \infty.$$

Dakle, za velike uzorke može se koristi i test statistika t-testa,

$$T = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n}.$$

Kako testovi vrijede aproksimativno za velike uzorke, radi korištenja procjene za σ, preporučeno je ipak za velike uzorke koristiti t-test. Na taj se način osigurava dodatna korekcija za traženu razinu značajnosti statističkog testa. Također, za velike n, distribucija $t(n-1)$ bliska je standardnoj normalnoj distribuciji što ide u prilog korištenju t testa [10]. Na sljedećoj se slici može vidjeti usporedba funkcija gustoće Studentove distribucije sa standardnom normalnom distribucijom.

[2] Pogledati [1], str.248.



Slika 1: Aproksimacija Studentove distribucije normalnom

Primjer 3.2. *Usporedba z testa i t testa za normalno distribuirane uzorke. Testiranje hipoteza*

$$\mathcal{H}_0 : \mu = 0$$
$$\mathcal{H}_1 : \mu \neq 0,$$

za generirane slučajne uzorke iz $\mathcal{N}(0,1)$ *različitih veličina korištenjem z testa i t testa u* 1000 *simulacija. Podaci iz normalne distribucije generirani su ugrađenom funkcijom rnorm(). Za svaki se test procjenjuje vjerojatnost pogreške prve vrste, odnosno udio pogrešno odbačenih* $\mathcal{H}_0$ *u ukupno* 1000 *simulacija, pri čemu se boljim testom smatra test čija je vjerojatnost pogreške prve vrste manja, odnosno bliža nominalnoj vrijednosti* $\alpha = 0.05$*. Vjerojatnosti pogreške prve vrste za svaki test, u ovisnosti o dimenziji uzorka, prikazane su u sljedećoj tablici.*

n	z-test	t-test
10	0,055	0,058
15	0,056	0,055
30	0,054	0,051
50	0,036	0,041
100	0,055	0,055
150	0,04	0,04
175	0,058	0,061
200	0,042	0,039

Tablica 1: Usporedba vjerojatnosti pogrešaka 1. vrste koristeći z test i t test u ovisnosti o različitoj dimenziji uzorka iz normalne distribucije

Iz tablice je vidljivo da su za normalne podatke pripadne vjerojatnosti poprilično jednake i da nema nekih preferencija u odabiru testa, što je i u skladu s navedenom teorijom da se Studentova distribucija $t(n-1)$ dobro aproksimira standardnom normalnom za velike n $(n > 30)$.

U prethodnom su primjeru simulacije provedene za normalne podatke, što i je osnovna teorijska pretpostavka koja mora biti zadovoljena da bi se koristili z i t test. Iz tog su razloga vjerojatnosti pogreške prve vrste za male uzorke također zadovoljavajuće, no problem se pojavljuje ukoliko podaci nisu normalno distribuirani. Spomenuto je da, iako uzorak nije normalno distribuiran, ako je veličina uzorka dovoljno velika, korištenje t testa pokazat će se uspješnim.

Primjer 3.3. *Usporedba z testa i t testa za uzorke iz eksponencijalne distribucije.*
Za generirane slučajne uzorke iz $\mathcal{E}(1)$ korištenjem z testa i t testa testirane su hipoteze

$$\mathcal{H}_0 : \mu = 1$$
$$\mathcal{H}_1 : \mu \neq 1.$$

U 1000 *simulacija, za različite dimenzije uzorka, procijenjene su vjerojatnosti pogreške prve vrste te su prikazane u sljedećoj tablici.*

n	**z-test**	**t-test**
7	0,038	0,109
10	0,04	0,103
15	0,048	0,083
30	0,053	0,083
100	0,05	0,059
150	0,042	0,048
175	0,045	0,057
200	0,048	0,052

Tablica 2: Usporedba vjerojatnosti pogrešaka 1. vrste koristeći z test i t-test u ovisnosti o različitoj dimenziji uzorka iz eksponencijalne distribucije

Sada je vidljiva razlika dobivenih vjerojatnosti pogrešaka prve vrste u ovisnosti o dimenziji uzorka. Naime, vjerojatnosti pogrešaka prve vrste dobivene t-testom poprilično su velike u odnosu na nominalnu razinu značajnosti $\alpha = 0.05$. *Porastom dimenzije uzorka, za* $n > 100$, *ta se vjerojatnost smanjuje i bliža je nominalnoj. Z test pokazuje zadovoljavajuće rezultate čak i za male uzorke.*

Ako su pretpostavke populacijske distribucije upitne ili određene pretpostavke nisu zadovoljene, tada se umjesto parametarskih metoda koriste neparametarske jer ne pretpostavljaju posebnu distribuciju. Ako je parametarska metoda primjenjiva, zaključivanje o lokacijskom parametru podrazumijeva, između ostalog, testiranje hipoteza o populacijskom očekivanju kao mjeri centralne tendencije. Tako je za provođenje z-testa i *Studentovog* t-testa, u svrhu testiranja očekivanja, potrebna pretpostavka da podaci dolaze iz normalne distribucije. Ako je pretpostavka o normalnosti distribucije narušena, jedna od alternativa za zaključivanje o lokacijskom parametru je korištenje neparametarskih procedura. U neparametarskoj proceduri za lokacijski parametar uvijek se referiramo na populacijski *medijan* radije nego populacijsko očekivanje. Poznati neprametarski testovi o populacijskom medijanu su *test predznaka* i *Wilcoxonov test ranga i predznaka*. Prednosti testova su upravo neovisnost o tipu distribucije, a izračuni su brzi i egzaktni za male uzorke.

3.3 Wilcoxonov test ranga i predznaka

Wilcoxonov test ranga i predznaka poznati je neparametarski test o populacijskom medijanu. Budući da ne zahtijeva pretpostavku o normalnosti populacijske distribucije, koristi se umjesto z-testa i t-testa kada je pretpostavka o normalnosti narušena. Test statistika bazira se na predznacima suma rangova razlike podataka te pretpostavljene vrijednosti podataka, a pretpostavke koje moraju biti zadovoljene su:

1. neprekidnost populacijske distribucije
2. simetričnost pripadne funkcije gustoće oko medijana, [10], [1].

Neka je $\mathbf{X} = (X_1, \ldots, X_n)$ jednostavni slučajni uzorak iz neprekidne distribucije s funkcijom gustoće $f_{\mathbf{x}}$ koja je simetrična s obzirom na medijan M.
Test za testiranje hipoteza

$$\mathcal{H}_0 : M = M_0 \tag{8}$$
$$\mathcal{H}_1 : M \neq M_0, \tag{9}$$

koristi rangove razlika podataka i pretpostavljenog medijana M_0. Označimo s

$$d_i = x_i - M_0, \quad i = 1, \ldots, n,$$

razlike podataka x_i i pretpostavljenog medijana M_0. Rangiramo apsolutne razlike $|d_i|$ na način da je najmanja apsolutna razlika ranga 1, sljedeća ranga 2 i tako do najveće uz uvjet da se apsolutna razlika za koju vrijedi

$$|d_i| = 0$$

ne rangira. Jednakim se vrijednostima $|d_i|$ dodjeljuje pripadna aritmetička sredina. Nakon što smo apsolutnim razlikama $|d_i|$ dodjelili rang, svakom se rangu dodjeljuje predznak + ili - jednak predznaku pripadne razlike d_i. Zasebno se sumiraju rangovi pozitivnih razlika, u oznaci R_+, i rangovi negativnih razlika, u oznaci R_-. Mora vrijediti da je

$$R_+ + R_- = \frac{n(n+1)}{2},$$

pri čemu je n broj razlika koje su rangirane. Pod uvjetima nul-hipoteze očekuje se da distribucija razlika bude aproksimativno simetrična oko nule, odnosno da je

$$R_+ = R_-$$

što je jednako

$$\frac{n(n+1)}{4}.$$

Za Wilcoxonovu test statistiku uzima se

$$R = \min\{R_+, R_-\},$$

pri čemu je očekivana vrijednost test statistike u uvjetima nul-hipoteze $\frac{n(n+1)}{4}$.
Za odabrani nivo značajnosti α, $R_+ > R_-$ odgovara alternativnoj hipotezi $M > M_0$, dok $R_+ < R_-$ odgovara alternativnoj hipotezi $M < M_0$. U uvjetima istinitosti nul-hipoteze (8), test statistika R je diskretna slučajna varijabla sa skupom realizacija $\left\{0, 1, \ldots, \frac{n(n+1)}{2}\right\}$ čija je distribucija egzaktna.

3.3.1 Normalna aproksimacija za velike uzorke

Pretpostavimo da je $R_+ < R_-$, odnosno da je Wilcoxonova test statistika R suma pozitivnih rangova. Za velike uzorke distribuciju od R nije jednostavno egzaktno izračunati pa se koristi asimptotska verzija testa temeljena na centalnom graničnom teoremu, [10].
Označimo s

$$R = \sum_{i=1}^{n} i\ U_i, \quad U_i = \begin{cases} 1, & d_i = x_i - M_0 > 0 \\ 0, & d_i = x_i - M_0 < 0. \end{cases}$$

U uvjetima $\mathcal{H}_0$, $U_1, \dots U_n$ je n nezavisnih jednako distribuiranih Bernoullijevih slučajnih varijabli

$$U_i \sim \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix},\ i = 1, \dots, n.$$

Sada je

$$E[R] = \sum_{i=1}^{n} i\ E[U_i] = \sum_{i=1}^{n} \frac{i}{2} = \frac{1}{2}\frac{n(n+1)}{2} = \frac{n(n+1)}{4},$$

$$Var(R) = \sum_{i=1}^{n} i^2\ Var(U_i) = \sum_{i=1}^{n} \frac{i^2}{4} = \frac{1}{4}\frac{n(n+1)(2n+1)}{6} = \frac{n(n+1)(2n+1)}{24}.$$

Za velike se uzorke stoga koristi test statistika

$$T = \frac{R - \mu_R}{\sigma_R} \xrightarrow{\text{D}} \mathcal{N}(0,1),\ n \to \infty,$$

gdje je

$$\mu_R = \frac{n(n+1)}{4}, \quad \sigma_R^2 = \frac{n(n+1)(2n+1)}{24}.$$

Ukoliko u podacima postoje jednake vrijednosti, *eng. ties*, asimptotskoj se test statistici T dodaje dodatna korekcija. Naime, kako se jednakim vrijednostima d_i dodaje njihov prosječni rang, varijanca od T reducira se s

$$\frac{t^3 - t}{48},$$

za svaku grupu jednakih razlika. Nova test statistika je

$$T' = \frac{R - \mu_R}{\sigma_R'} \xrightarrow{\text{D}} \mathcal{N}(0,1),\ n \to \infty,$$

pri čemu je

$$\sigma_R'^2 = \frac{n(n+1)(2n+1)}{24} - \frac{\sum t_j^3 - \sum t_j}{48},$$

za $j = 1, \dots, m$ grupa jednakih razlika.

3.4 Test predznaka

Test predznaka najstariji je neparametarski test za lokacijski parametar. Naziva se testom predznaka jer u test statistici koristi predznake podataka, odnosno broji pozitivne ili negativne realizacije u odnosu na pretpostavljenu vrijednost medijana.

Neka je $\mathbf{X} = (X_1, \dots X_n)$ jednostavni slučajni uzorak iz *neprekidne* distribucije F slučajne varijable X. Želimo testirati hipotezu da je p_0 kvantil $Q(p_0)$ jednak nekoj zadanoj vrijednosti x_0

$$\mathcal{H}_0 : Q(p_0) = x_0.$$

Funkcija kvantila definirana je s

$$Q(p) = \inf \big\{x \in \mathbb{R} : F(x) \geqslant p\big\},\ p \in (0,1).$$

Za neprekidne distribucije kvantili su jedinstveno određeni inverzom funkcije distribucije. Sada se H_0 može ekvivalentno zapisati kao:

$$\mathcal{H}_0 : F(x_0) = P(X \leqslant x_0) = p_0.$$

Odgovarajući uzorak je

$$\big(\mathbb{1}_{\{X_1 \leqslant x_0\}}, \dots, \mathbb{1}_{\{X_n \leqslant x_0\}}\big),$$

a pripadna test statistika koja broji negativne predznake u realizacijama $\{x_i - x_0\}$ je

$$T = \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leqslant x_0\}} \sim \mathcal{B}(n, p_0).$$

Dakle, ukoliko je $x_0 = Q(p_0)$, onda frekvencija negativnih vrijednosti razlika $\{x_i - x_0\}$ ima $\mathcal{B}(n, p_0)$. Za $p_0 = \frac{1}{2}$, $Q(p_0)$ je upravo medijan od X, što je mjera "sredine"distribucije.

Test predznaka egzaktan je te je primjenjiv na male uzorke. Iz tog se razloga koristi pri testiranju centralne tendencije distribucije u slučaju kada pretpostavke t-testa nisu zadovoljene. *Wilcoxonov test ranga i predznaka* jači je test u odnosu na *test predznaka* jer, osim što koristi predznake razlika između realizacija i pretpostavljenog medijana kao *test predznaka*, koristi i veličine tih razlika na način da ih rangira po veličini. Međutim, *Wilcoxonov test ranga i predznaka* ne preporuča se kada ima razloga vjerovati da je populacijska distribucija asimetrična. U tom slučaju preferira se *test predznaka*, [10].

4 Statistički testovi za lokacijski parametar za dva uzorka

Uspoređivanje lokacijskih parametara, "sredina", dvaju nezavisnih uzoraka u parametarskim se procedurama svodi na zaključivanje o populacijskim očekivanjima. Odabir prikladnog testa u tu svrhu ovisit će o pretpostavkama koje uzorci zadovoljavaju. Na osnovu dva nezavisna jednostavna slučajna uzorka $\mathbf{X} = (X_1, \ldots, X_n)$ i $\mathbf{Y} = (Y_1, \ldots, Y_m)$, želi se testirati

$$\mathcal{H}_0 : \mu_1 = \mu_2,$$

pri čemu je $\mu_1 = E[X_i]$ te $\mu_2 = E[Y_i]$. U primjeni se zaključivanje o lokacijskom pomaku svodi na testiranje o razlici populacijskih očekivanja $\mu_1 - \mu_2$. Najpoznatiji parametarski test o jednakosti populacijskih očekivanja je *Studentov t-* test, a njegovo korištenje zahtijeva zadovoljavanje određenih pretpostavki distribucije uzorka, kao što je, već spomenuta, normalnost. Osim nezavisnosti i normalne distribuiranosti uzoraka, bitna je pretpostavka o jednakosti populacijskih varijanci, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Kao i u slučaju *Studentovog* testa na jednom uzorku, populacijske varijance su nepoznate te se u test-statistici procjenjuju tzv. *zajedničkom varijancom*. Z test je statistički test koji se koristi za testiranje populacijskih očekivanja za normalne uzorke u kojima su varijance poznate, i ne nužno jednake.

4.1 Z-test

Neka je $\mathbf{X} = (X_1, \ldots, X_n)$ jednostavni slučajni uzorak iz $\mathcal{N}(\mu_1, \sigma_1^2)$, $\mathbf{Y} = (Y_1, \ldots, Y_m)$ jednostavni slučajni uzorak iz $\mathcal{N}(\mu_2, \sigma_2^2)$, te neka su uzorci međusobno nezavisni, a σ_1^2 i σ_2^2 poznati. Intuitivnim pristupom kreiramo test za testiranje hipoteza

$$\mathcal{H}_0 : \mu_1 = \mu_2$$
$$\mathcal{H}_1 : \mu_1 \neq \mu_2.$$

ML procjenitelj za $\mu_1 - \mu_2$ u ovom modelu je $\overline{X}_n - \overline{Y}_m$, a obzirom da je ta razlika linearna kombinacija nezavisnih i normalnih distribucija,

$$\overline{X}_n - \overline{Y}_m \sim \mathcal{N}\Big(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\Big).$$

Sada zaključivanje o $\mu_1 - \mu_2$ možemo temeljiti na test statistici

$$T(\mathbf{X}, \mathbf{Y}) = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}.$$

U uvjetima istinitosti $\mathcal{H}_0$ $T(\mathbf{X}, \mathbf{Y}) \sim \mathcal{N}(0, 1)$.

4.2 t-test

Neka je $\mathbf{X} = (X_1, \ldots, X_n)$ jednostavni slučajni uzorak iz $\mathcal{N}(\mu_1, \sigma_1^2)$, $\mathbf{Y} = (Y_1, \ldots, Y_m)$ jednostavni slučajni uzorak iz $\mathcal{N}(\mu_2, \sigma_2^2)$, te neka su uzorci međusobno nezavisni. Klasični t-test pretpostavlja da su varijance σ_1 i σ_2 nepoznate, ali i jednake, tj. $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

U svrhu procjene varijance σ^2 koristi se tzv. *zajednička varijanca*

$$S_p^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 + \sum_{j=1}^{m}(Y_j - \overline{Y}_m)^2}{n+m-2} = \frac{(n-1)S_n^2 + (m-1)S_m^2}{n+m-2}$$

nastala težinskom kombinacijom procjenitelja S_n^2 i S_m^2, pri čemu su

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

$$S_m^2 = \frac{1}{n-1}\sum_{j=1}^{m}(Y_j - \overline{Y}_m)^2.$$

Teorem 4.1. *Neka su $X_1, \ldots, X_n$ nezavisne i jednako distribuirane normalne slučajne varijable s očekivanjem μ_1 i varijancom σ^2, a $Y_1, \ldots, Y_m$ nezavisne i jednako distribuirane normalne slučajne varijable s očekivanjem μ_2 i tom istom varijancom σ^2. Nadalje, neka su $\mathbf{X} = (X_1, \ldots, X_n)$ i $\mathbf{Y} = (Y_1, \ldots, Y_m)$ međusobno nezavisni. Tada statistika*

$$\mathcal{T} = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

ima Studentovu distribuciju s n+m-2 stupnjeva slobode, tj.

$$\mathcal{T} \sim t(n+m-2).$$

Dokaz. Treba prikazati $\mathcal{T}$ kao kvocijent standardne normalne i korijena χ^2 slučajne varijable podijeljene svojim stupnjem slobode, nezavisne od varijable brojnika. Kako je

$$E\Big[\overline{X}_n\Big] = E\Big[\frac{1}{n}\sum_{i=1}^{n} X_i\Big] = \frac{1}{n}n\mu_1 = \mu_1,$$

$$E\Big[\overline{Y}_m\Big] = E\Big[\frac{1}{m}\sum_{j=1}^{m} Y_j\Big] = \frac{1}{m}m\mu_2 = \mu_2,$$

$$Var\Big(\overline{X}_n\Big) = Var\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n},$$

$$Var\Big(\overline{Y}_m\Big) = Var\Big(\frac{1}{m}\sum_{j=1}^{m} Y_j\Big) = \frac{1}{m^2}m\sigma^2 = \frac{\sigma^2}{m},$$

a $\overline{X}_n$ i $\overline{Y}_m$ međusobno su nezavisni, slučajna varijabla $\overline{X}_n - \overline{Y}_m$ ima distribuciju s očekivanjem

$$E\Big[\overline{X}_n - \overline{Y}_m\Big] = \mu_1 - \mu_2$$

i varijancom

$$Var\Big(\overline{X}_n - \overline{Y}_m\Big) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m} = \sigma^2\Big(\frac{1}{n} + \frac{1}{m}\Big).$$

Standardizacijom iste dobivamo slučajnu varijablu

$$Z = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0,1).$$

Znamo da je

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

i nezavisno od $\overline{X}_n$,

$$\frac{m-1}{\sigma^2} S_m^2 \sim \chi^2(m-1)$$

i nezavisno od $\overline{Y}_m$. Također, one su međusobno nezavisne pa i njihova suma ima χ^2 distribuciju sa sumom stupnjeva slobode, odnosno

$$\frac{n-1}{\sigma^2} S_n^2 + \frac{m-1}{\sigma^2} S_m^2 = \frac{1}{\sigma^2}\Big((n-1)S_n^2 + (m-1)S_m^2\Big) = \frac{n+m-2}{\sigma^2} S_p^2 \sim \chi^2(n+m-2).$$

Radi pretpostavke o nezavisnosti u uzorcima i oblika S_p^2, vidimo da je S_p^2 nezavisno od $\overline{X}_n - \overline{Y}_m$. Konačno,

$$\mathcal{T} = \frac{Z}{\sqrt{\frac{\frac{n+m-2}{\sigma^2} S_p^2}{n+m-2}}} = \frac{\frac{(\overline{X}_n - \overline{Y}_m) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}}{\frac{1}{\sigma} S_p} = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

□

Na temelju prethodnog rezultata kreiramo test za testiranje hipoteza

$$\mathcal{H}_0 : \mu_1 = \mu_2$$

$$\mathcal{H}_0 : \mu_1 \neq \mu_2.$$

U slučaju istinistosti hipoteze $\mathcal{H}_0 : \mu_1 = \mu_2$, test statistika

$$T = \mathcal{T}(\mathbf{X}, \mathbf{Y}) = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\overline{X}_n - \overline{Y}_m}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

ima Studentovu distribuciju s $n+m-2$ stupnjeva slobode, $\mathcal{T} \sim t(n+m-2)$.
Velike vrijednosti test statistike idu u prilog alternativnoj hipotezi pa je pripadno kritično područje, razine značajnosti α, $C_r = \{|\mathcal{T}(\mathbf{x}, \mathbf{y})| \geqslant t_{1-\frac{\alpha}{2}}(n+m-2)\}$, pri čemu je $t_{1-\frac{\alpha}{2}}(n+m-2)$ pripadni kvantil $t(n+m-2)$ distribucije.

4.3 Behrens-Fisherov problem

Kako je *Studentov* t-test određen pretpostavkama o normalnosti podataka i jednakosti varijanci koje su u primjeni uglavnom narušene, stvara se potreba za novim testovima prikladnim za testiranje o populacijskim očekivanjima. Problem testiranja hipoteza o razlici očekivanja iz dviju nezavisnih normalno distribuiranih populacija, pri čemu su pripadne varijance nepoznate i različite, poznat je pod nazivom $Behrens-Fisherov\ problem$. Najpoznatiji statistički testovi korišteni u tu svrhu su $Welchov$ t-test i neparametarski $Brunner\ Munzelov$ test, [5].

4.3.1 Welchov t-test

S obzirom da klasični t-test na dvama uzorcima pretpostavlja da su varijance σ_1^2 i σ_2^2 nepoznate a jednake, pretpostavka o jednakosti dviju varijanci unosi dodatnu nesigurnost. $Welchov$ t-test nastao je kao modifikacija *Studentovog* t testa u svrhu suzbijanja problema nejednakosti varijanci kao solucija $Behrens-Fisherova\ problema$, [4]. Ako su populacijske varijance različite, prirodni procjenitelj varijance od $\overline{X}_n - \overline{Y}_m$ je

$$\frac{S_n^2}{n} + \frac{S_m^2}{m}.$$

Pretpostavimo sada da su $\mathbf{X} = (X_1, \ldots, X_n)$ jednostavni slučajni uzorak iz $\mathcal{N}(\mu_1, \sigma_1^2)$ i $\mathbf{Y} = (Y_1, \ldots, Y_m)$ jednostavni slučajni uzorak iz $\mathcal{N}(\mu_2, \sigma_2^2)$ nezavisni, pri čemu su σ_1^2 i σ_2^2 nepoznati. Test statistika koju koristimo za testiranje hipoteza

$$\mathcal{H}_0 : \mu_1 = \mu_2$$
$$\mathcal{H}_1 : \mu_1 \neq \mu_2.$$

je

$$T(\mathbf{X}, \mathbf{Y}) = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_n^2}{n} + \frac{S_m^2}{m}}}$$

koja u uvjetima $\mathcal{H}_0$ ima Studentovu distribuciju $\mathcal{T}_\nu$, gdje je ν aproksimativna vrijednost broja stupnjeva slobode koja se dobije tzv. Welch-Satterthwaitovom formulom

$$\nu = \frac{\left(\frac{s_n^2}{n} + \frac{s_m^2}{m}\right)^2}{\frac{s_n^4}{n^2(n-1)} + \frac{s_m^4}{m^2(m-1)}}.$$

Kritično područje određuje se analogno kao kod klasičnog t-testa na dvama uzorcima. S obzirom na dodatnu pretpostavku o jednakosti varijanci kod klasičnog t-testa, preporuka je stoga koristiti Welchov t-test.

U primjeni, npr. u medicinskom istraživanju, pretpostavke o normalnosti distribucije uzoraka i jednakosti varijanci uglavnom su narušene. Podaci su realizacije asimetričnih distribucija (eng.

skewed) uz prisutne stršeće vrijednosti, realizacije koje nisu konzistentne s ostalima i svojom prisutnošću znatno utječu na povećanje očekivanja i varijance distribucije, [4].

Za izrazito asimetrične distribucije očekivanje stoga može biti loša mjera centralne tendencije. Očekivanje je dobra mjera "sredine" za distribucije koje nisu zakrivljene, tj. simetrične su, za distribucije koje nemaju teških repova i za koje nisu prisutne stršeće vrijednosti.
Problem utjecaja stršećih vrijednosti može se riješiti različitim procedurama, kao što su tzv. "*rezanje*" *podataka*, (eng. *trimming*), ili transformacija podataka nekom funkcijom. "*Rezanje*" *podataka* dobra je transformacija ukoliko je distribucija uzorka iz koje dolaze podaci ona s teškim repovima, a podrazumijeva uklanjanje ekstremnih realizacija iz repova distribucije, [10].

Yuen Welchov test nastaje kao modifikacija *Welchovog* t-testa, a testira hipoteze o "odrezanim" (ili "poboljšanim") očekivanjima, nastalim uklanjanjem vrijednosti sa svakog repa uz fiksiran *iznos poboljšanja* γ, [4]. Uz 0% poboljšanja, *Yuen Welchov* test ekvivalentan je *Welchovom* testu. Općenito se uzima $\gamma = 0.2$, što odgovara odbacivanju 20% najmanjih i najvećih podataka iz svakog uzorka, a za testiranje hipoteza o "poboljšanim" očekivanjima korištena test statistika ponovno je iz *Studentove* t distribucije, ali s korigiranim stupnjem slobode.

Uklanjanje podataka iz neke distribucije znači smanjenje veličine uzorka što rezultira smanjenjem jakosti statističkog testa. Transformacija podataka funkcijom drugog korijena ili logaritmom, često smanjuje utjecaj stršećih vrijednosti što rezultira homogenošću varijanci te normalizira distribuciju. Tako je transformacija drugim korijenom korisna u slučaju da podaci dolaze iz pozitivno zakrivljene distribucije. Smanjit će veličinu razlike podataka između repova distribucije, na način da će se desna strana distribucije "pomaknuti" prema sredini te će se time stabilizirati varijanca. Logaritamska će transformacija također suzbiti problem heterogenosti varijanci te normalizirati podatke za pozitivno zakrivljene distribucije. U slučaju negativne zakrivljenosti distribucije, prikladna je kvadratna transformacija, [10].

Osim različitih transformacija za suzbijanje nenormalnosti pri korištenju parametarskih testova, koriste se testovi koji se ne oslanjaju na pretpostavke o tipu distribucije - *neparametarski testovi*. Medijan, kao mjera centralne tendencije, manje je osjetljiv na zakrivljenost distribucije, te se u takvim slučajevim za testiranje hipoteza o "sredini" provodi *neparametarskim procedurama* uspoređujući populacijske medijane. Ako podaci ne dolaze iz normalne distribucije, najčešća neparametarska alternativa *Studentovom* testu je $Mann-Whitney-Wilcoxonov$ *test*, [7].

5 Mann-Whitney-Wilcoxonov test

Mann-Whitney-Wilcoxonov test, poznat još kao Wilcoxonov test sume rangova ili Mann-Whitneyev U test, je neparametarski test jer se ne oslanja na pretpostavke o tipu distribucije.
Neka su $\mathbf{X} = (X_1, \ldots, X_n)$ i $\mathbf{Y} = (Y_1, \ldots, Y_m)$ nezavisni jednostavni slučajni uzorci iz neprekidnih distribucija F_X i F_Y. U MWW testu pretpostavljamo da F_X i F_Y imaju isti "oblik", ali se mogu razlikovati za lokacijski pomak, tj. za neki $\delta \in \mathbb{R}$

$$F_X(x) = F_Y(x - \delta).$$

Uočimo da to znači da za svaki $x \in \mathbb{R}$

$$P(X \leqslant x) = F_X(x) = F_Y(x - \delta) = P(Y \leqslant x - \delta) = P(Y + \delta \leqslant x),$$

odnosno,

$$X \stackrel{d}{=} Y + \delta \iff X - \delta \stackrel{d}{=} Y.$$

Sada testiramo nul-hipotezu

$$\mathcal{H}_0 : F_X(x) = F_Y(x), \forall x \in \mathbb{R},$$

koju možemo iskazati kao

$$\mathcal{H}_0 : \delta = 0.$$

Alternativne hipoteze mogu biti:

1. dvostrana hipoteza - postoji lokacijski pomak $\delta \in \mathbb{R}$

   $$\mathcal{H}_1 : \text{postoji } \delta \in \mathbb{R} \text{ takav da je } F_X(x) = F_Y(x - \delta), \text{ za svaki } x \in \mathbb{R},$$

   posebno, i distribucije s razlikuju

   $$F_X(x) < F_Y(x) \text{ ili } F_X(x) > F_Y(x), \text{ za svaki } x \in \mathbb{R}$$

2. jednostrana hipoteza - postoji pozitivan lokacijski pomak ($\delta > 0$) X u odnosu na Y

   $$\mathcal{H}_1 : \text{postoji } \delta > 0 \text{ takav da je } F_X(x) = F_Y(x - \delta), \text{ za svaki } x \in \mathbb{R},$$

   odnosno,

   $$\mathcal{H}_1 : X \stackrel{d}{=} Y + \delta \text{ za neki } \delta > 0.$$

   Tu hipotezu interpretiramo kao "X je veći od Y",

   Primijetimo da je u situaciji u kojoj je $X > Y$, zapravo $F_X(x) \leqslant F_Y(x)$, pri čemu se stroga nejednakost pojavljuje za barem jedan $x \in \mathbb{R}$.

3. jednostrana hipoteza - postoji negativan lokacijski pomak ($\delta < 0$) X u odnosu na Y

$$\mathcal{H}_1 : \text{postoji } \delta < 0 \text{ takav da je } F_X(x) = F_Y(x - \delta), \text{ za svaki } x \in \mathbb{R},$$

odnosno,

$$\mathcal{H}_1 : X \stackrel{d}{=} Y + \delta \text{ za neki } \delta < 0.$$

Tu hipotezu interpretiramo kao "X je manji od Y",

Sada je za $X < Y$, $F_X(x) \geqslant F_Y(x)$, pri čemu se stroga nejednakost pojavljuje za barem jedan $x \in \mathbb{R}$.

Test statistika kreira se na temelju ranga združenog skupa podataka. Neka je $x_1, \ldots, x_n$ realizacija jednostavnog slučajnog uzorka $\mathbf{X} = (X_1, \ldots, X_n)$ iz neprekidne distribucije F_X, a $y_1, \ldots, y_m$ realizacija jednostavnog slučajnog uzorka $\mathbf{Y} = (Y_1, \ldots, Y_m)$ iz neprekidne distribucije F_Y. Formiramo vektor svih podataka $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ te ih sortiramo u rastućem poretku. Temelj za kreiranje statistike čini uređeni niz svih podataka $z_{(1)}, z_{(2)}, \ldots, z_{(n+m)}$. Pretpostavimo bez smanjenja općenitosti da je $n < m$, test statistika MWW testa koristit će sumu rangova kraćeg niza, u ovom slučaju $\mathbf{x}$. Da bi se odredili rangovi podataka, potrebno je utvrditi pozicije podataka. Svakom podatku iz $\mathbf{Z}$ pridružujemo njegov **rang** na način da:

1. ako u $\mathbf{Z}$ nema jednakih elemenata, rang podatka bit će njegov redni broj u $\mathbf{Z}$
2. ako u $\mathbf{Z}$ ima jendakih elemenata, svim se jednakim podacima pridružuje rang koji će biti jednak aritmetičkoj sredini njihovih rednih brojeva, tzv. korigirani rang.

Pretpostavimo da u $\mathbf{Z}$ nema jednakih elemenata, tj. da su rangovi podataka jednoznačno definirani, te u svrhu kreiranja test statistike označimo rangove podataka iz prvog uzorka s $\mathbf{r} = (r_X(1), \ldots, r_X(n))$ čiju sumu koristimo za kreiranje test statistike, tj.

$$w' = \sum_{i=1}^{n} r_X(i)$$

je realizacija test statistike W'. Uočimo da minimalna i maksimalna suma rangova ovise samo o dimenzijama uzoraka n i m. Naime,

$$w'_{min} = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

$$w'_{max} = m + 1 + m + 2 + \cdots + m + n = nm + \frac{n(n+1)}{2},$$

pri čemu će se w'_{min} realizirati ako su svi podaci iz $\mathbf{x}$ manji od podataka iz $\mathbf{y}$, a w'_{max} realizirat će se ako su svi podaci iz $\mathbf{y}$ manji od podataka iz $\mathbf{x}$. Zanima nas distribucija statistike W u uvjetima istinitosti hipoteze $\mathcal{H}_0$ za koju ne postoji lokacijski pomak. Ako su obje distribucije jednake, $F_X(x) = F_Y(x)$, i dolaze iz modela nezavisnih jednostavnih slučajnih uzoraka, distribucija rangova podataka iz $\mathbf{x}$ trebala bi odgovarati potpunoj slučajnosti, odnosno odabiru n pozicija

od ukupnih $n+m$ koje se računaju primjenom osnovnih principa prebrojavnja. Dakle, suma rangova iz $\mathbf{x}$ W' je diskretna slučajna varijabla sa slikom

$$\{w'_{min}, w'_{min}+1, \ldots, w'_{max}\} = \left\{ \frac{n(n+1)}{2}, \frac{n(n+1)}{2}+1, \ldots, nm+\frac{n(n+1)}{2} \right\}$$

čija se distribucija može egzaktno izračunati. Obzirom na skup vrijednosti statistike W' koja sadrži $nm+1$ elemenata, koristi se distribucija statistike

$$W = W' - \frac{n(n+1)}{2}$$

sa skupom vrijednosti $\{0, 1, \ldots, nm\}$.

5.1 Veliki uzorci

Za velike uzorke distribuciju od W' nije jednostavno egzaktno izračunati pa se koristi asimptotska verzija testa temeljena na centalnom graničnom teoremu:

$$\frac{W' - \mu_{w'}}{\sigma_{w'}} \xrightarrow{\mathrm{D}} \mathcal{N}(0,1),$$

gdje je

$$\mu_{w'} = \frac{n(n+m+1)}{2}, \quad \sigma^2_{w'} = \frac{nm(n+m+1)}{12}.$$

Uočimo da se alternativne hipoteze mogu izraziti i u terminima očekivanja. Naime, ako pretpostavimo da je za neki $\delta \in \mathbb{R}$

$$F_X(x) = F_Y(x-\delta), \text{ za svaki } x \in \mathbb{R},$$

odnosno,

$$X \stackrel{d}{=} Y + \delta,$$

onda je i

$$E[X] = E[Y] + \delta \iff \mu_1 - \mu_2 = \delta.$$

MWW test koristi se stoga kao alternativa klasičnom t-testu kada pretpostavka o normalnosti t-testa nije zadovoljena jer nema pretpostavki o tipu distribucije, a i egzaktan je za male uzorke. Ako su varijance jednake, porastom zakrivljenosti distribucije uzoraka, jakost *MWW* testa prestiže jakost t-testa bez obzira na veličinu uzoraka, [7].

6 Simulacije

U sljedećim će se simulacijama pokušati odabrati prikladan statistički test za usporedbu centralnih tendencija, očekivanja ili medijana, dviju nezavisnih grupa u ovisnosti o tipu distribucije iz koje dolaze podaci, veličini uzoraka te različitosti varijanci. Testiranje o lokacijskim parametrima za dva uzorka često u primjeni podrazumijeva različitost veličina uzoraka te različitost varijanci. Takvi su slučajevi razmatrani u sljedećim simulacijama te se, na osnovu dobivenih procijenjenih vjerojatnosti pogrešaka prve vrste, odabire prikladan test. Prikladan je test onaj čija je vjerojatnost pogreške prve vrste bliska nominalnoj, odnosno maksimalnoj vjerojatnosti pogreške prve vrste $\alpha = 0.05$.

Ukoliko je vjerojatnost pogreške prve vrste nekog testa veća od nominalne vrijednosti, test se smatra *liberalnim*. U slučaju da je vjerojatnost pogreške prve vrste manja od nominalne vrijednosti, test se smatra *konzervativnim*, [7].

Primjer 6.1. *Usporedba t-testa, Welchovog t-testa i MWW testa za uzorke iz normalne distribucije.*
Za generirane slučajne uzorke iz normalne distribucije, u 1000 *ponavljanja, procijenjene vjerojatnosti pogreške prve vrste prikazane su u sljedećoj tablici.*

n	**m**	$\frac{\sigma_1}{\sigma_2}$	**Studentov t-test**	**MWW test**	**Welchov t-test**
5	5	1	0,053	0,034	0,046
10	10	1	0,053	0,044	0,05
20	20	1	0,046	0,048	0,043
5	10	1	0,048	0,037	0,04
5	15	1	0,035	0,033	0,044
10	20	1	0,055	0,061	0,058
5	10	$\frac{1}{2}$	0,015	0,017	0,044
5	15	$\frac{1}{2}$	0,01	0,02	0,051
10	20	$\frac{1}{2}$	0,017	0,027	0,046
10	5	$\frac{1}{2}$	0,119	0,074	0,053
15	5	$\frac{1}{2}$	0,155	0,091	0,052
20	10	$\frac{1}{2}$	0,116	0,082	0,056
10	10	$\frac{1}{2}$	0,068	0,066	0,056
20	20	$\frac{1}{2}$	0,047	0,06	0,045
5	10	$\frac{1}{4}$	0,021	0,027	0,053
5	15	$\frac{1}{4}$	0,002	0,01	0,045
10	20	$\frac{1}{4}$	0,01	0,032	0,043
10	5	$\frac{1}{4}$	0,164	0,069	0,048
15	5	$\frac{1}{4}$	0,247	0,114	0,051
20	10	$\frac{1}{4}$	0,147	0,107	0,041
10	10	$\frac{1}{4}$	0,057	0,068	0,047
20	20	$\frac{1}{4}$	0,056	0,062	0,052

Tablica 3: Usporedba vjerojatnosti pogreške 1. vrste koristeći *Studentov* t-test, MWW test i $Welchov$ t-test u ovisnosti o različitim dimenzijama uzoraka i različitim omjerima standardnih devijacija uzoraka iz normalne distribucije

Ako su varijance uzoraka jednake, *Studentov* i $Welchov$ t-test ne pokazuju značajna odstupanja bez obzira na veličinu uzoraka. U slučaju različitih varijanci, stvari se bitno mijenjaju. Naime, ako uzorak manje dimenzije ima ujedno i veću varijancu, onda *Studentov* t test i MWW test imaju znatno veću vjerojatnost pogreške prve vrste od nominalne - *preliberalni* su. Ako uzorak veće dimenzije ima ujedno i veću varijancu, onda *Studentov* t test i MWW test imaju znatno manju vjerojatnost pogreške prve vrste od nominalne - *prekonzervativni* su, [7]. U oba se slučaja $Welchov$ test pokazuje kao prikladan, što i nalaže teorija da se isti koristi u slučaju različitih varijanci uzoraka kao rješenje $Behrens - Fisherova\ problema$. MWW test osjetljiv je na različitost varijanci te se u takvim slučajevima ne preporučuje bez obzira na dimenziju uzoraka, dok je *Studentov* t-test prikladan u slučajevima različitih varijanci ako su dimenzije uzoraka međusobno jednake.

U sljedećem ćemo primjeru simulirati uzorke iz *Laplaceove* distribucije. Slučajna varijabla X ima *Laplaceovu* distribuciju s realnim parametrima μ i $b > 0$ ako je njena funkcija gustoće dana izrazom

$$f_X(x) = \frac{1}{2b}\, e^{-\dfrac{|x-\mu|}{b}}, \quad x \in \mathbb{R}.$$

Pišemo $X \sim \mathcal{L}ap(\mu, b)$. Parametar μ je lokacijski parametar i odgovara očekivanju, dok je b parametar skale. Očekivanje i varijanca dani su s $E[X] = \mu,\ Var(X) = 2b^2$.

Primjer 6.2. *Usporedba t-testa, Welchovog t-testa i MWW testa za uzorke iz Laplaceove distribucije.*

Ugrađenom funkcijom rlaplace() generirane su slučajne varijable

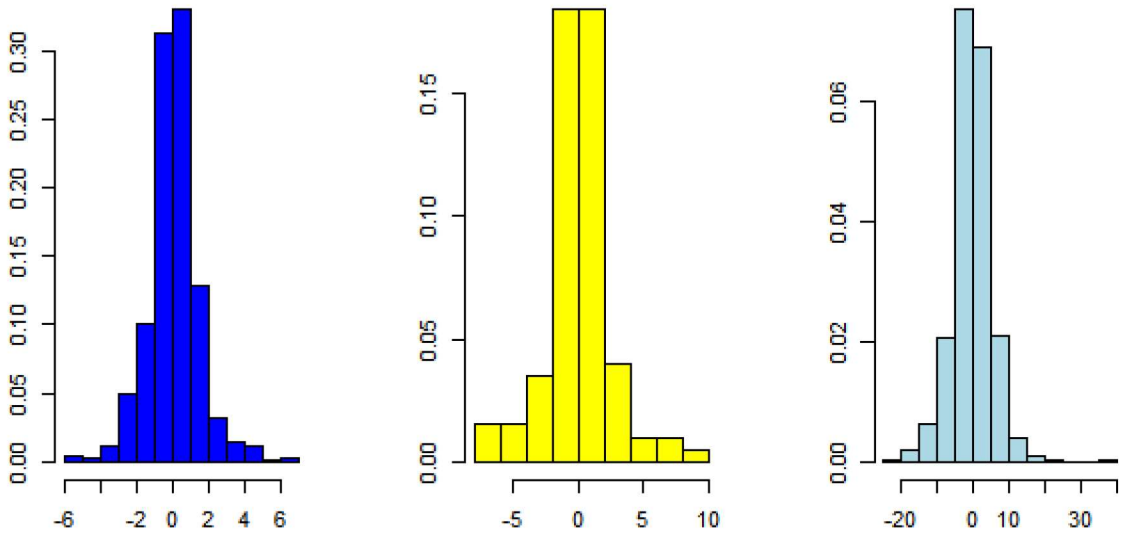
$$X \sim \mathcal{L}ap(0,1),\ Y \sim \mathcal{L}ap(0,2),\ Z \sim \mathcal{L}ap(0,4)$$

čija su očekivanja 0, *a varijance* $2, 8, 32$ *redom. Standardne se devijacije odnose kao* $1:2:4$, *te ćemo s obzirom na različite veličine uzoraka i omjere standardnih devijacija provjeravati odstupanje vjerojatnosti pogrešaka prve vrste navedenih testova u odnosu na nominalnu veličinu* 0.05. *Za generirane slučajne uzorke iz Laplaceove distribucije, u* 1000 *ponavljanja, procijenjene vjerojatnosti pogrešaka prve vrste prikazane su u sljedećoj tablici.*

n	m	$\frac{\sigma_1}{\sigma_2}$	Studentov t-test	MWW test	Welchov t-test
5	5	1	0,05	0,034	0,044
10	10	1	0,042	0,04	0,039
20	20	1	0,046	0,044	0,047
5	10	1	0,048	0,044	0,048
5	15	1	0,048	0,047	0,04
10	20	1	0,054	0,052	0,044
5	10	$\frac{1}{2}$	0,017	0,021	0,044
5	15	$\frac{1}{2}$	0,016	0,025	0,045
10	20	$\frac{1}{2}$	0,018	0,038	0,046
10	5	$\frac{1}{2}$	0,09	0,053	0,038
15	5	$\frac{1}{2}$	0,154	0,079	0,051
20	10	$\frac{1}{2}$	0,105	0,084	0,048
10	10	$\frac{1}{2}$	0,053	0,05	0,049
20	20	$\frac{1}{2}$	0,043	0,045	0,048
5	10	$\frac{1}{4}$	0,009	0,022	0,042
5	15	$\frac{1}{4}$	0,004	0,016	0,043
10	20	$\frac{1}{4}$	0,011	0,037	0,05
10	5	$\frac{1}{4}$	0,151	0,082	0,046
15	5	$\frac{1}{4}$	0,253	0,110	0,048
20	10	$\frac{1}{4}$	0,165	0,112	0,053
10	10	$\frac{1}{4}$	0,061	0,072	0,051
20	20	$\frac{1}{4}$	0,045	0,058	0,043
50	50	$\frac{1}{4}$	0,048	0,068	0,048
100	100	$\frac{1}{4}$	0,051	0,073	0,051

Tablica 4: Usporedba vjerojatnosti pogrešaka 1. vrste koristeći *Studentov* t-test, MWW test i $Welchov$ t-test u ovisnosti o različitim dimenzijama uzoraka i različitim omjerima standardnih devijacija uzoraka iz Laplaceove distribucije

Iz tablice je vidljivo da, iako pretpostavka o normalnosti distribucije nije zadovoljena, $Welchov$ t-test najbolje kontrolira željenu vjerojatnost pogreške prve vrste. Ponovno su *Studentov* t-test i MWW test prekonzervativni u slučaju pozitivne korelacije dimenzije uzorka i standardne devijacije, dok su u suprotnom preliberalni. MWW test sve je lošiji porastom dimenzija uzoraka i omjera standardnih devijacija. U slučaju različitih varijanci, prikladan je i *Studentov* t-test ako su dimenzije uzoraka jednake i dovoljno velike.



Slika 2: Histogrami generiranih Laplaceovih distribucija

Dakle, iako generirani uzorci nisu normalno distribuirani, *Welchov* t-test pokazuje dobre rezultate, što opravdavamo svojstvom simetričnosti Laplaceove distribucije.

U sljedećem primjeru pogledajmo kako simetričnost distribucije uzorka utječe na navedene testove. U tu svrhu, generirajmo uzorke iz *Gamma* distribucije. Slučajna varijabla X ima *Gamma* distribuciju s parametrima $\alpha > 0$ i $\beta > 0$ ako je njena funkcija gustoće dana izrazom

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\beta x},\ x > 0,$$

pri čemu je $\Gamma(\alpha)$ *Gamma* funkcija. [3]
Pišemo $X \sim Gamaa(\alpha, \beta) = \Gamma(\alpha, \beta)$. Parametar α naziva se parametrom oblika, dok je β parametar stope. Očekivanje i varijanca dani su s $E[X] = \frac{\alpha}{\beta}$, $Var(X) = \frac{\alpha}{\beta^2}$.

Primjer 6.3. *Usporedba t-testa, Welchovog t-testa i MWW testa za uzorke iz Gamma distribucije.*

Ugrađenom funkcijom rgamma() generirane su slučajne varijable

$$X \sim \mathcal{G}amma(4, 2),\ Y \sim \mathcal{G}amma(2, 1),\ Z \sim \mathcal{G}amma(1, 1/2)$$

čija su očekivanja 2, *a varijance* $1, 2, 4$ *redom. Obzirom na različite veličine uzoraka i omjere varijanci uzoraka, provjeravat će se odstupanje vjerojatnosti pogrešaka prve vrste navedenih testova*

[3] Gamma funkcija definirana je kao $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t},\ x > 0.$

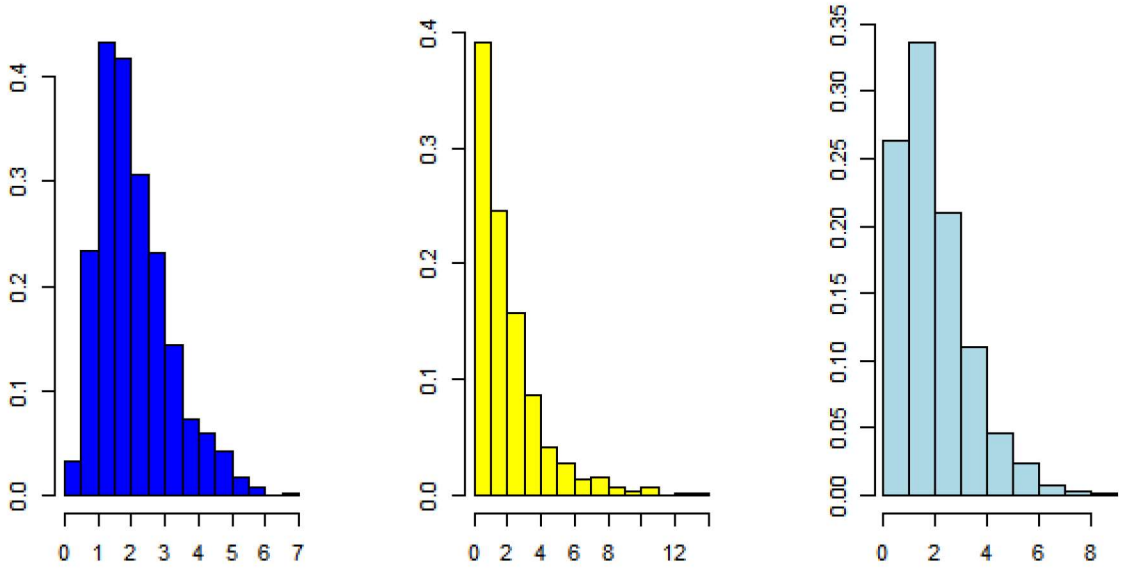
u odnosu na nominalnu veličinu 0.05. *Za generirane slučajne uzorke iz Gamma distribucije, u* 1000 *ponavljanja, procijenjene vjerojatnosti pogrešaka prve vrste prikazane su u sljedećoj tablici.*

n	**m**	$\frac{\sigma_1^2}{\sigma_2^2}$	**Studentov t-test**	**MWW test**	**Welchov t-test**
5	5	1	0,051	0,037	0,047
10	10	1	0,049	0,05	0,048
20	20	1	0,052	0,05	0,052
5	10	1	0,06	0,05	0,059
5	15	1	0,044	0,042	0,063
10	20	1	0,034	0,043	0,044
5	10	$\frac{1}{2}$	0,029	0,029	0,042
5	15	$\frac{1}{2}$	0,021	0,031	0,055
10	20	$\frac{1}{2}$	0,034	0,06	0,045
10	5	$\frac{1}{2}$	0,089	0,059	0,074*
15	5	$\frac{1}{2}$	0,096	0,077	0,081*
20	10	$\frac{1}{2}$	0,07	0,069	0,053
30	20	$\frac{1}{2}$	0,076	0,106	0,055
10	10	$\frac{1}{2}$	0,072	0,066	0,069
20	20	$\frac{1}{2}$	0,049	0,076	0,048
30	30	$\frac{1}{2}$	0,05	0,079	0,05
5	10	$\frac{1}{4}$	0,038	0,044	0,041
5	15	$\frac{1}{4}$	0,030	0,048	0,048
10	20	$\frac{1}{4}$	0,038	0,096	0,053
10	5	$\frac{1}{4}$	0,119	0,108	0,105*
15	5	$\frac{1}{4}$	0,175	0,128	0,121*
20	10	$\frac{1}{4}$	0,121	0,154	0,082
30	20	$\frac{1}{4}$	0,083	0,207	0,062
40	50	$\frac{1}{4}$	0,044	0,297	0,051
10	10	$\frac{1}{4}$	0,067	0,093	0,065
20	20	$\frac{1}{4}$	0,062	0,168	0,061
30	30	$\frac{1}{4}$	0,055	0,227	0,055

Tablica 5: Usporedba vjerojatnosti pogrešaka 1. vrste koristeći *Studentov t*-test, *MWW* test i *Welchov t*-test u ovisnosti o različitim dimenzijama uzoraka i različitim omjerima varijanci uzoraka iz Gamma distribucije

Welchov t-test više ne kontrolira dovoljno dobro pogrešku prve vrste. Naime, ako je veća dimenzija uzorka ona s manjom varijancom, čak i on postaje preliberalan (rezultati označeni zvjezdicom u tablici). Kontrola se poboljšava porastom dimenzije uzoraka. Veća liberalnost uočena je i kod jednakih, a manjih, dimenzija uzoraka. *MWW* test prikladan je za manje uzorke jednakih varijanci.

Vidljivo je da simetričnost distribucija uzoraka itekako utječe na jačinu testova. Navedene *Gamma* distribucije pozitivno su asimetrične, odnosno imaju duži desni rep. Na sljedećoj su slici prikazane simulirane distribucije.



Slika 3: Histogrami generiranih Gamma distribucija

Na osnovu prethodnih simulacija, zaključujemo sljedeće:

1. *Studentov* t-test zahtijeva pretpostavku o jednakosti varijanci i normalnosti distribucije uzorka. Ako je pretpostavka o normalnosti narušena, prikladan je za velike uzorke ($n > 30$). Ako su varijance različite, prikladan je u slučaju jednakih dimenzija uzoraka, [7].

2. *MWW* test alternativa je *Studentovom* t-testu u slučaju da pretpostavka normalnosti nije zadovoljena, za male uzorke. Izrazito je osjetljiv je na nejednakost varijanci te se u takvom slučaju ne preporuča, [7].

3. *Welchov* t-test rješava problem nejednakih varijanci. Također zahtijeva normalnost distribucije uzoraka ili dovoljnu veličinu uzoraka. Za različite dimenzije i populacijske varijance uzoraka, pokazuje se najučinkovitijim, [4]. Za male uzorke i različite varijance bitnu ulogu ima simetričnost distribucije. Ako je distribucija asimetrična, a uzorci su mali i varijance su različite, neće dati zadovoljavajuće rezultate, [7].

Asimetričnost distribucije uzorka igra veliku ulogu u odabiru najjačeg lokacijskog testa za dva uzorka. Pri odabiru testa, uvijek je potrebno provjeriti asimetričnost distribucije uzoraka te jednakost populacijskih varijanci, [7]. Problem asimetričnosti može se pokušati riješiti različitim transformacijama podataka objašnjenim u poglavlju 4.

Za testiranje hipoteza o lokacijskom parametru za jedan uzorak koristit ćemo, [10],:

1. *Studentov* t-test o populacijskom očekivanju, ako je zadovoljena pretpostavka o normalnosti distribucije uzorka ili je uzorak dovoljno velik, ($n > 30$);

2. *Wilcoxonov test ranga i predznaka* o populacijskom medijanu, ako je uzorak mali i populacija slijedi simetričnu distribuciju;

3. *Test predznaka* o populacijskom medijanu, ako populacijska distribucija nije simetrična, a uzorak je mali.

Literatura

[1] L.E.BAIN, M.ENGELHARDT, *Introduction to Probability and Mathematical statistics,* Brooks/Cole Cangage Learning, 1992.

[2] M.BENŠIĆ, N.ŠUVAK, *Primijenjena statistika,* Sveučilište J. J. Strossmayera, Odjel za matematiku, Osijek, 2013.

[3] R.C.BLAIR, J.HIGGINS, *A Comparison of the Power of Wilcoxon's Rank-Sum Statistic to That of Student's t Statistic under Various Nonormal Distributions,* Journal of Educational Statistics vol.5., No.4., 1980.

[4] M.W.FAGERLAND, L.SANDVIK, *Performance of five two-sample location tests for skewed distributions with unequal variances,* Contemporary Clinical Trials, 2009.

[5] M.P.FAY, M.A.PROSCHAN, *Wilcoxon-Mann-Whiteny or t-test? On assumptions for hypothesis test and multiple interpretations of decision rules,* Statistics Surveys, 2010.

[6] Ž.PAUŠE, *Uvod u matematičku statistiku,* Školska knjiga, Zagreb, 1993.

[7] D.A.PENFIELD, *Choosing a Two-Sample Location Test,* The Journal of Experimental Education vol.62, No.4, 1994.

[8] N.SARAPA, *Teorija vjerojatnosti,* Školska knjiga, Zagreb, 2002.

[9] J.SHAO, *Mathematical Statistics (2nd edn.),* Springer, New York, 1994.

[10] D.J. SHESKIN, *Handbook of Parametric and Nonparametric Statistical Procedures (3rd edn.),* Chapman & Hall/CRC, Boca Raton, 2003.

Sažetak

U ovome radu objašnjeni su statistički testovi za lokacijski parametar kao mjeru centralne tendencije distribucije. Parametarski testovi podrazumijevat će testiranje hipoteza o populacijskom očekivanju, a zahtijevat će određene pretpostavke o tipu distribucije, kao što su normalnost distribucije ili jednakost varijanci u slučaju usporedbe lokacija dvaju nezavisnih uzoraka. Nakon parametarskih, opisane su neparametarske procedure korištene ukoliko su pretpostavke parametarskih testova narušene. Odabir prikladnog testa za testiranje lokacijskog parametra ovisit će o raznim čimbenicima, kao što su veličina uzoraka, jednakost varijanci, asimetričnost distribucija. S obzirom na navedene čimbenike, nakon teorijskog dijela, provedene su simulacije za usporedbu u kojim je situacijama koji test bolji u smislu bolje kontrole pogreške prve vrste.

Ključne riječi: Lokacijski problem, Welchov t-test, simulacije, Behrens-Fisherov problem

Summary

This paper contains an explanation of statistical procedures to test location parameter as a measure of central tendency of distribution. Parametric methods will involve testing hypothesis for population mean, and also will require certain assumptions about the distribution type, such as distribution normality or equality of variances in the case of locations comparison of two independent samples. After parametric, nonparametric procedures are described which are used if listed assumptions were violated. The selection of appropriate test for testing location parameter will depend on various factors, such as sample sizes, equality of variances, skewness of distribution. Following the theoretical part, simulations were performed to compare which test is better in terms of type one error rate control in different situations.

Key words: Location problem, Welch's t-test, simulations, Behrens-Fisher problem

Životopis

Rođena sam 11. studenog 1995. godine u Rijeci. Nakon završene Osnovne škole Tar-Vabriga u Taru, upisujem jezičnu gimnaziju u Poreču. Srednjoškolsko obrazovanje završavam 2014. godine te iste godine upisujem preddiplomski studij matematike na Odjelu za matematiku u Osijeku. Preddiplomski studij završavam 2018. godine sa završnim radom na temu "Polinomijalne matrice" te time stječem akademski stupanj prvostupnice matematike. Iste godine upisujem diplomski studij na Odjelu za matematiku u Osijeku, smjer Financijska matematika i statistika. Tijekom završne godine studija, privremeno sam bila zaposlena kao predavač matematike u Osnovnoj školi Tar-Vabriga.