

Domača naloga 1

Neža Kržan, Tom Rupnik Medjedovič

1 Cilj naloge

Ocenjujemo veljavnost testa za razliko aritmetičnih sredin iz treh neodvisnih vzorcev, kjer primerjava naslednje tri metode:

- predpostavka enakih varianc(klasični ANOVA test),
- predpostavka različnih varianc(Welchow ANOVA test) in
- kombiniran pristop z Levenovim testom.

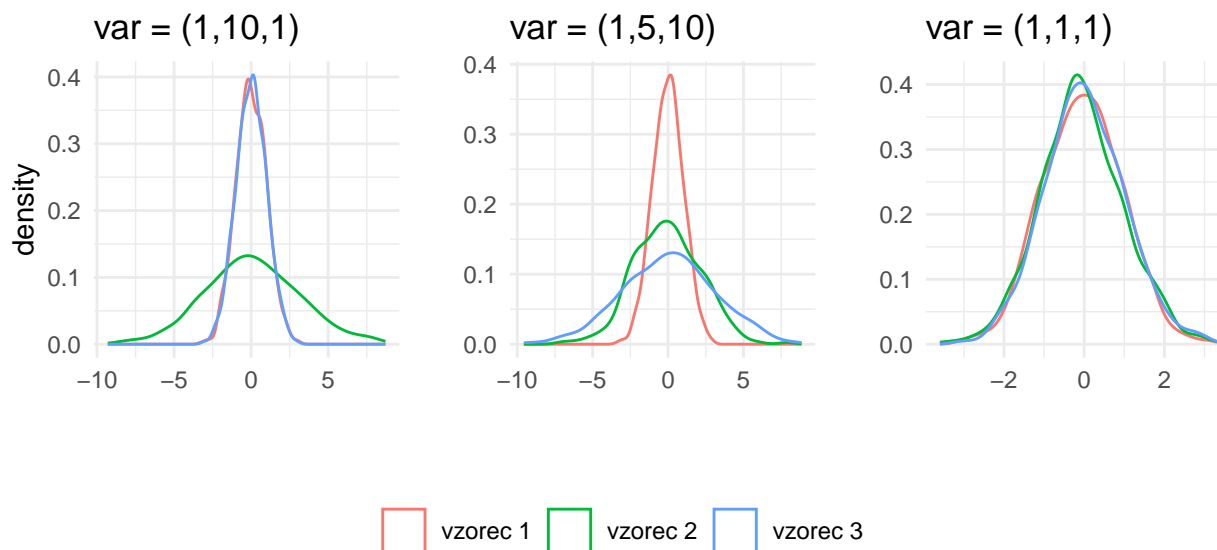
Pri kombiniranem pristopu z Levenovim testom gre za možnost, kjer predpostaviva enake variance, če ne moreva zavrniti predpostavke o enakosti varianc(pri 5 % tveganju) s testom za enakost varianc in neenake variance, če lahko. Tukaj sva za testiranje predpostavke o enakosti varianc torej uporabljala Levenejev test.

2 Generiranje podatkov

Podatke sva generirala tako, da je porazdelitev znotraj posamezne skupine normalna s povprečjem 0 ($\mu_i = 0$). Variabilnost posamezne skupine sva spreminjala, tako da sva lahko preverila veljavnost testa za razliko aritmetičnih sredin v primeru, da je predpostavka enakih varianc izpolnjena oz. kršena. V ta namen sva si izbrala primere treh skupin z naslednjo varianco:

- (1, 10, 1),
- (1, 5, 10),
- (1, 1, 1).

Poleg tega sva spreminjala tudi velikost skupin, izbrala sva si velikosti $n \in \{15, 40, 100, 200, 1000\}$, pri čemer imajo vse skupine vedno enako velikost, torej npr. [15, 15, 15].



Slika 1: Primer gostote porazdelitev skupin(velikost skupin 1000) za različne kombinacije varianc.

Pričakujeva, da bo klasični ANOVA test dobro deloval le v primeru, ko je predpostavka enakih varianc izpolnjena in pričakujeva, da velikost skupin ne bo veliko vplivala na rezultate tega testa. Za druga dva testa(Welch's ANOVA test in kombinirani pristop z Levenovim testom) pa pričakujeva, da dobro delujeta tako v primeru izpolnjene kot tudi kršene predpostavke. Moč testov bo pri enakih variancah verjetno približno enaka za vse 3, s tem ko pri različnih pa bo večja pri drugih dveh testih(Welch's ANOVA in p-vrednost na podlagi Levene's testa). Pričakujeva, da se bo moč večala z večanjem vzorca.

3 Simulacije

Izvedla sva simulacije s 5000 ponovitvami(da so bile dokaj hitro končane).

V prvi simulaciji sva računala p - vrednost in preverjala njeno porazdelitev; zanimalo naju je ali je p -vrednost porazdeljena enakomerno. Na podlagi p - vrednosti pa sva potem določila tudi velikost testa.

Izvedla sva tudi simulacijo za ugotavljanje kakšna je moč testa pri enakih in različnih variancah. V simulaciji sva uporabila tri metode, klasični ANOVA test, Welchov ANOVA test in možnost, kjer predpostavimo enake variance, če ne moremo zavrniti predpostavke o enakosti varianc (pri 5 % tveganju) z Levenejevim testom za enakost varianc, in neenake variance, če to lahko.

Na koncu sva izvedla še simulacijo, kjer imajo vse skupine enako varianco, vendar so različnih velikosti. V simulaciji naju je zanimala razlika med klasičnim ANOVA testom in Welchovim ANOVA testom.

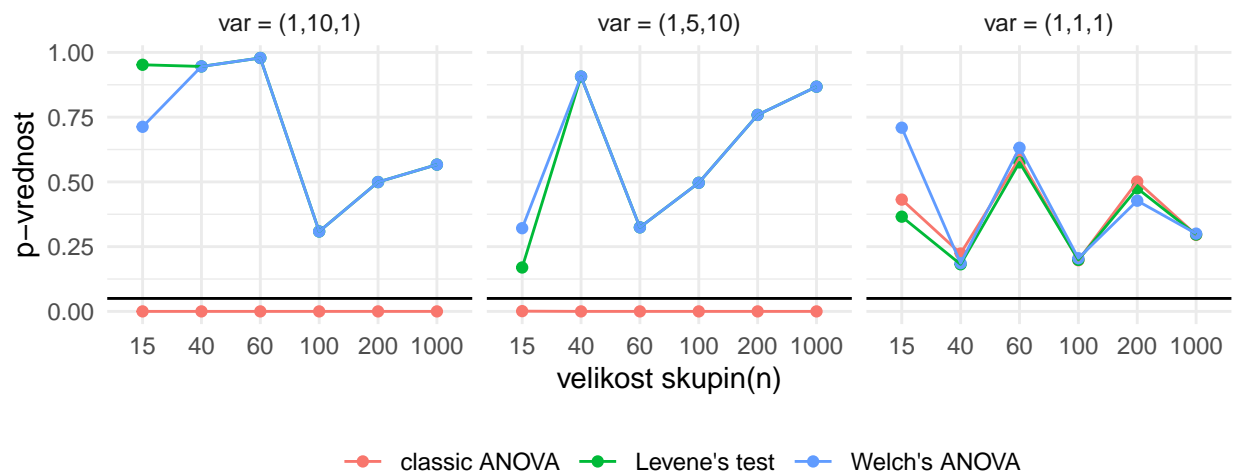
3.1 p - vrednost

V simulacijah, kjer preverjava enakomerno porazdelitev p -vrednosti, sva spreminjala varianco v posamezni skupini in velikost posamezne skupine.

Na grafih je prikazana p -vrednost Anderson-Darling testa z ničelno hipotezo

$$H_0 : \text{vzorec sledi zvezni enakomerni porazdelitvi.}$$

Želeli bi, da je porazdelitev p -vrednosti podobna zvezni enakomerni porazdelitvi, torej da enakomerno zavzame vse vrednosti na intervalu $[0,1]$.

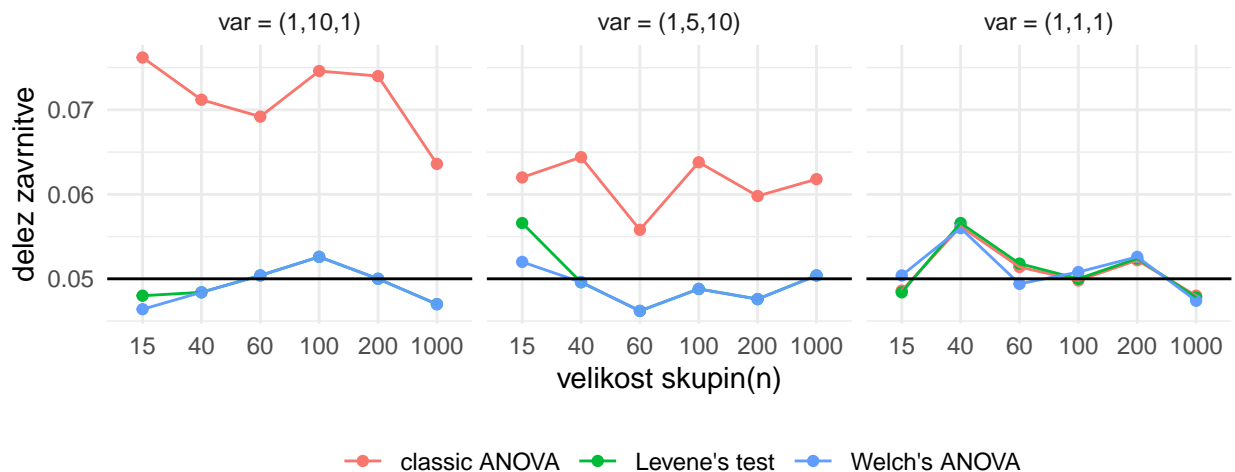


Slika 2: Prikaz p -vrednosti Anderson-Darling testa.

Kot sva že na začetku predvidela, klasična ANOVA (classic ANOVA) v primeru različnih varianc ni veljavna, saj ima p -vrednost enako oz. blizu 0 za vse velikosti vzorca. To pomeni, da zavrnilo ničelno hipotezo. V primeru, ko je varianca v vseh treh skupinah enaka, so vse p -vrednosti nad mejo 0.05, torej ne moremo zavrniti ničelne hipoteze.

3.2 Velikost testa

Na podlagi zgornjega grafa pričakujemo, da velikost testa v primeru klasičnega ANOVA testa in različnih varianc ne bo enaka oz. blizu vrednosti 0.05. Zaradi kršenja predpostavk bomo prevečkrat oz. premalokrat zavračali ničelno hipotezo, ko ta drži.



Slika 3: Prikaz velikosti testa pri različnih parametrih.

Res smo potrdili zgornje predvidevanje. V primeru različnih varianc klasični ANOVA test prevečkrat zavrne ničelno hipotezo. Ko pa je varianca v vseh treh skupinah enaka 1, so vrednosti dokaj podobne pri vseh treh testih.

Opaziva tudi pri klasični ANOVI, da se pri večjih razlikah v variancah (npr. $var = (1, 10, 1)$) delež zavrnitev

ničelne hipoteze zmanjša pri zelo veliki skupini ($n = 1000$), ampak pri manjših razlikah v varianc (npr. $var = (1, 5, 10)$) pa te razlike v deležu zavrnitve ničelne hipoteze med najmanjšo ($n = 15$) in največjo skupino ($n = 1000$) ni.

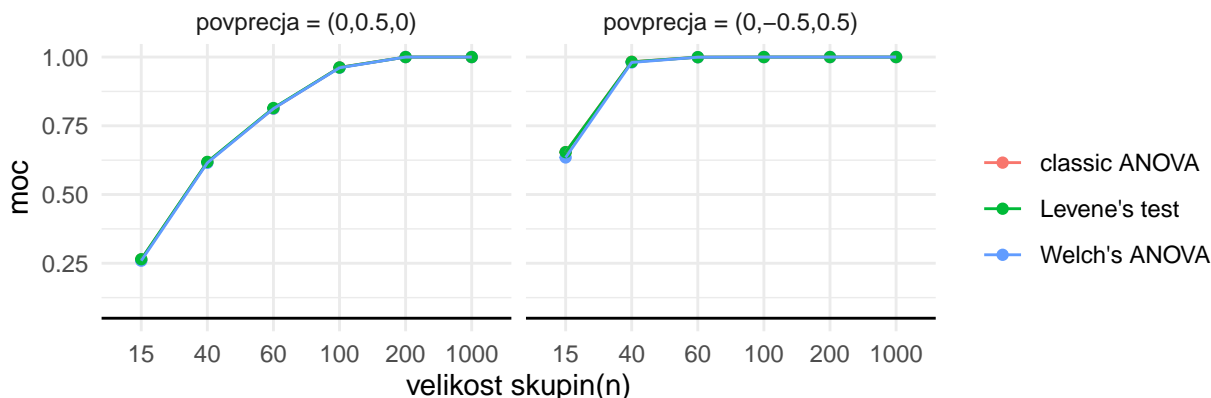
Glede na velikost skupin delež zavrnitve ničelne hipoteze niha, ampak med najmanjšo skupino ($n = 15$) in največjo ($n = 1000$) pri vseh treh tipih izbranih varianc, ni večjih razlik v deležu zavrnitve ničelne hipoteze.

3.3 Moč testa

Glede na moč testa lahko primerjamo vse tri teste le v primeru enakih varianc. V preostalih primerih klasični ANOVA test ni veljaven, zato bomo primerjali le preostala dva. Najprej sva primerjala med seboj teste v primeru enakih varianc, nato pa še v primeru različnih, kjer sva si pod alternativno domnevo (*povprečja med skupinami so različna*) izbrala povprečja

- $(0, 0.5, 0)$ in
- $(0, -0.5, 0.5)$

Enake variance



Slika 4: Prikaz moči testa v primeru enakih varianc.

Tabela 1: Moč testa v primeru enakih varianc in povprečij $(0, 0.5, 0)$.

	n = 15	n = 40	n = 60	n = 100	n = 200	n = 1000
classic ANOVA	0.2638	0.6176	0.8126	0.9624	1	1
Welch's ANOVA	0.2592	0.6152	0.8124	0.9610	1	1
Levene's test	0.2646	0.6186	0.8146	0.9622	1	1

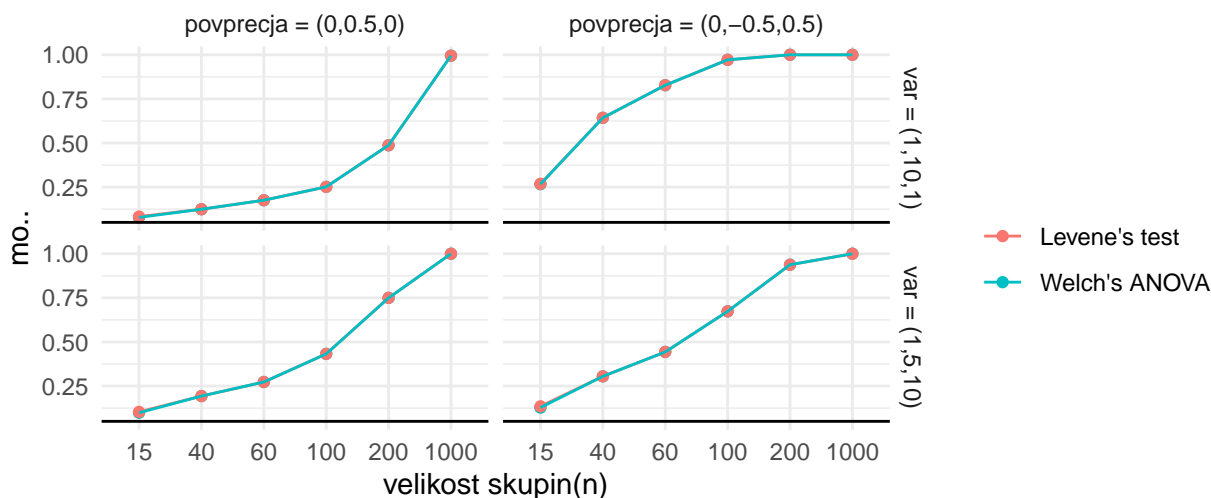
Tabela 2: Moč testa v primeru enakih varianc in povprečij $(0, -0.5, 0.5)$.

	n = 15	n = 40	n = 60	n = 100	n = 200	n = 1000
classic ANOVA	0.6542	0.9828	0.9996	1	1	1
Welch's ANOVA	0.6348	0.9804	0.9994	1	1	1
Levene's test	0.6542	0.9826	0.9996	1	1	1

Na podlagi zgornjega grafa in tabel ne moremo govoriti o velikih razlikah med testi. Vsi testi imajo zelo podobno moč testa. Pričakovano z naraščanjem velikosti skupin narašča tudi moč testa. Opazimo lahko tudi,

da je v primeru različnega povprečja le v eni skupini manjša moč kot v primeru različnega povprečja v vseh skupinah - v drugem primeru je tudi razlika med povprečjema 1, pri prvi pa 0.5, kar je verjetno razlog za take rezultate.

Različne variance



Slika 5: Prikaz moči testa v primeru različnih varianc(vrstice predstavljajo varianco skupin, stolpci pa povprečje posamezne skupine).

Tabela 3: Moč testa v primeru različnih varianc (1, 10, 1) in povprečij (0, 0.5, 0).

	n = 15	n = 40	n = 60	n = 100	n = 200	n = 1000
Welch's ANOVA	0.0784	0.1246	0.1756	0.2514	0.4872	0.995
Levene's test	0.0830	0.1246	0.1756	0.2514	0.4872	0.995

Ponovno imata metodi zelo podobno moč testa(v tabeli vidimo, da celo za skupine velike $n = 1000$ enako). Opazimo, da se moč testa spreminja z velikostjo skupine in kombinacijo variabilnosti skupine in razlike povprečja. V primeru, da je varianca skupine velika in povprečje malo odstopa od preostalih dveh, je moč testa dokaj majhna tudi pri večjih skupinah(npr. povprečja = (0, 0.5, 0), var = (1, 10, 1) in $n = 200$). Torej manj, kot se porazdelitve skupin med seboj prekrivajo in večje kot so, večja je moč testa.

3.4 Razlika med klasičnim ANOVA testom in Welchovim ANOVA testom

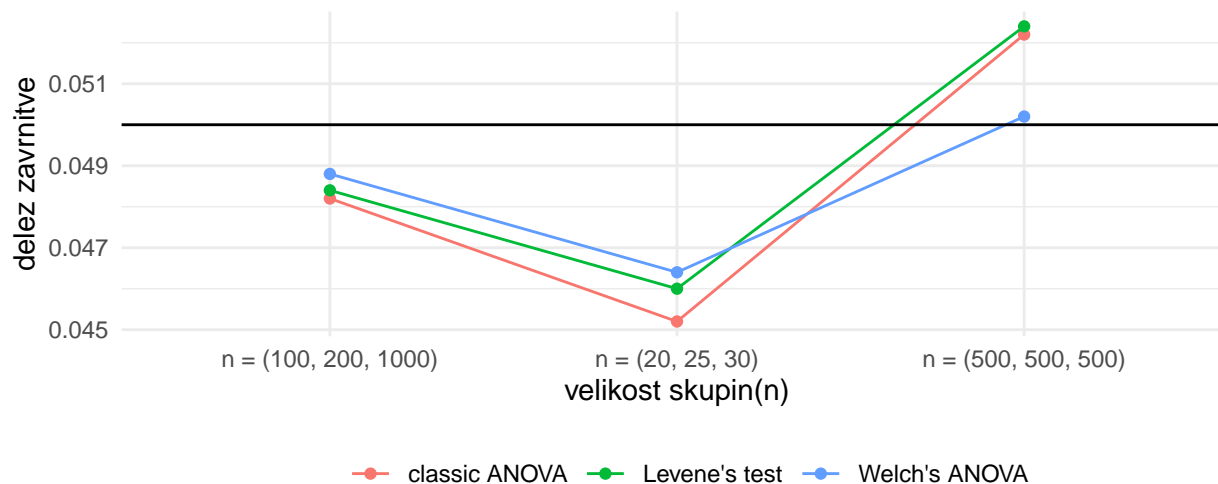
Zanima naju kakšna je razlika med klasičnim in Welchovim ANOVA testom glede na to, da so moči vseh treh testov pri enakih variancah($var = (1, 1, 1)$) dokaj podobne. Za simulacijo sva si izbrala tri različne kombinacije velikosti skupin, in sicer

- $n = (100, 200, 1000)$, torej primer, ko so velikosti skupin med seboj precej različne,
- $n = (20, 25, 30)$, torej primer, ko so velikosti skupin med seboj zelo malo različne in
- $n = (500, 500, 500)$, primer, ko so velikosti skupin med seboj enake.

Pričakujeva, da bosta testa dokaj podobna, ampak verjetno bo Welchov ANOVA test v primeru različnih velikosti skupin boljši, ker je glede tega bolj fleksibilen.

3.4.1 Velikost testa

Na podlagi zgornjih rezultatov pričakujemo, da bo velikost testa v primeru Welchovega ANOVA testa bližje vrednosti 0.05 kot klasična ANOVA, saj naj bi test boljši za različne velikosti skupin.



Slika 6: Prikaz velikosti testa pri različnih velikostih skupin.

Takoj opazimo, da v primeru enakih varianc v vseh treh skupinah, ko so velikosti le teh različne, največkrat zavrača Welchov ANOVA test, v primeru enakih velikosti skupin pa najmanjkrat in je velikost testa najbolj blizu vrednosti 0.05.

Večja razlika v deležu zavrnitev pri klasičnem in Welchovim ANOVA testom je, če so razlike v velikosti skupin manjše ($n = (20, 25, 30)$).

4 Zaključek

Kot je bilo pričakovano, pri različnih variancah klasični ANOVA test ni ustrezen, v primerih enakih varianc med skupinami pa rezultati kažejo ustreznost klasičnega ANOVA testa. Velikost vzorca nima nekega pomembnega vpliva na veljavnost klasičnega ANOVA testa, saj delež zavračanja ničelne hipoteze ostaja visok, kadar variabilnost v skupinah ni enaka.

Welchov ANOVA test in kombinirani pristop (kjer sva uporabljala Levenov test za preverjanje enakosti varianc) se izkažeta za bolj ustrezna tako v primerih enakih kot tudi različnih varianc. Pri enakih variancah je delež zavrnitev ničelne hipoteze pri vseh treh testih podoben, vendar Welchov ANOVA in kombinirani pristop pa imata manjšo velikost testa pri neenakih variancah.

Klasični ANOVA test ni primeren test za razliko aritmetičnih sredin iz treh neodvisnih vzorcev z neenakimi variancami, saj, zaradi kršenja predpostavk, prepogosto nepravilno zavrne ničelno hipotezo. Welchov ANOVA test in kombinirani pristop z Levenovim testom sta v takšnih primerih bolj ustrezna in zagotavljata pravilnejše rezultate, ne glede na velikost vzorca. Sta tudi močnejša na večjih vzorcih oz. skupinah in skupinah, kjer se porazdelitve med seboj manj prekrivajo.