

Domača naloga 1

Neža Kržan, Tom Rupnik Medjedovič

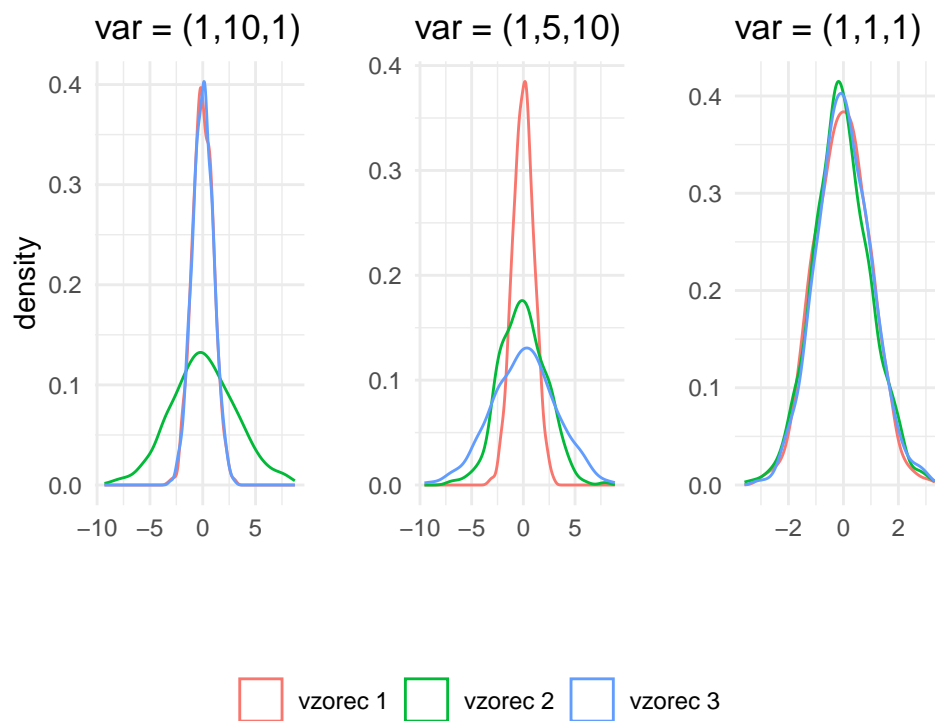
Navodilo

Estimate the validity of a test for the difference of arithmetic means from three independent samples. Compare method where you assume:

- equal variance (classic ANOVA)
- unequal variances (Welch's ANOVA)
- an option where you assume equal variances if you cannot reject the assumption of equality of variances (at 5% risk) with some test for equality of variances (e.g. Levene's), and unequal variances if you can.

Generate the data so that the variances are equal in some places and different in others.

1 Generiranje podatkov



Slika 1: Primer gostote porazdelitev vzorcev za različne kombinacije varianc.

Podatke sva generirala tako, da je porazdelitev znotraj posamezne porazdelitve normalna s povprečjem 0 ($\mu_i = 0$). Variabilnost posamezne skupine sva nekoliko 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)

Pričakujeva, da bo klasični ANOVA test dobro deloval le v primeru, ko je predpostavka enakih varianc izpolnjena. Preostala dva testa pa pričakujeva, da dobro delujeta tako v primeru izpolnjene kot tudi kršene predpostavke.

2 Simulacije (p-vrednost)

V simulacijah, kjer preverjava enakomerno porazdelitev p -vrednosti, sva spreminjala varianco v posamezni skupini in velikost posamezne skupine. Varianca, ki sva si izbrala, zavzame vrednosti:

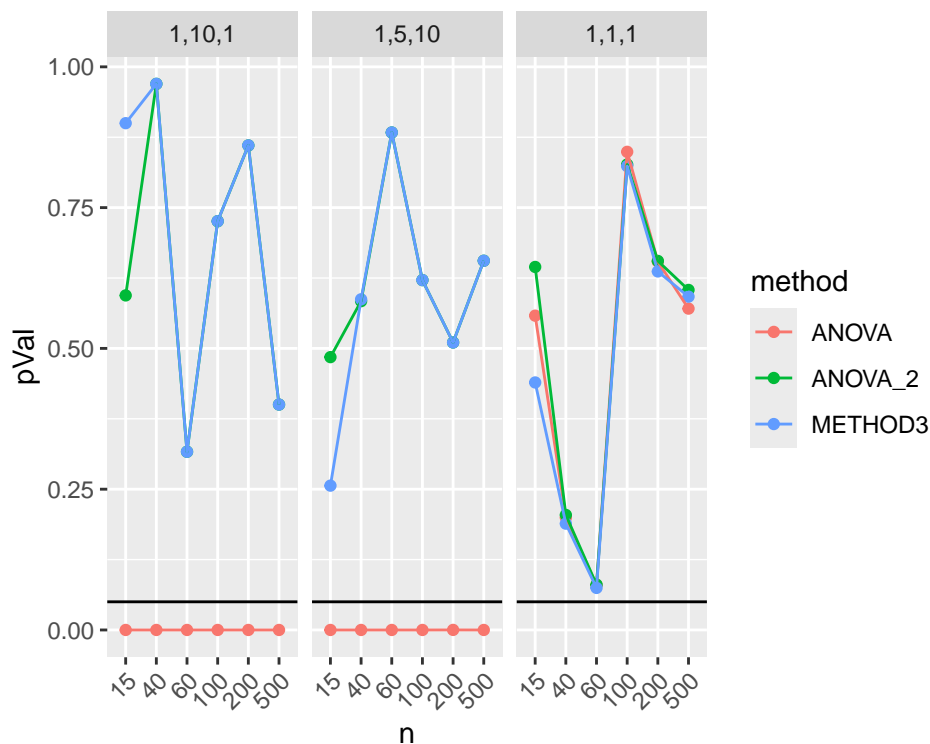
- (1, 10, 1)
- (1, 5, 10)
- (1,1,1)

Za velikost posamezne skupine (vse tri skupine imajo enako vrednost) pa sva si izbrala $n \in \{15, 40, 100, 200, 500\}$.

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 $[0,1]$.

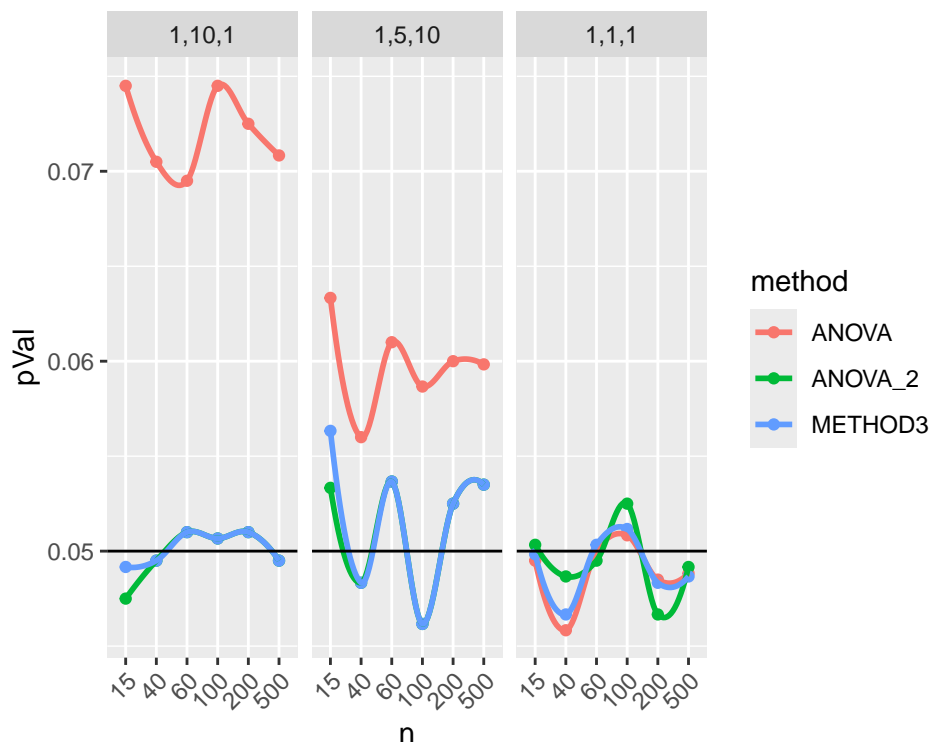


Kot sva že na začetku predvidela, klasična ANOVA v primeru različnih varianc ni veljavna, saj ima p -vrednost enako oz. blizu 0. To pomeni, da zavrnemo ničelno hipotezo.

V primeru, ko so variance v vseh treh skupinah enake, so vse p -vrednosti nad mejo 0.05, torej ne moremo zavrniti ničelne hipoteze.

3 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.



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.