

Domača naloga 3

Neža Kržan, Tom Rupnik Medjedovič

1 Cilji naloge

Želiva preučiti uporabo metode ponovnega vzorčenja, v primeru ko klasičnimi testom ne moremo popolno zaupati zaradi kršenja predpostavk. Generirala bova podatke, na katerih bova izračunala intervale zaupanja koeficientov linearne regresije. Zaradi kršenja predpostavk bova s pomočjo testa `boxCox` izvedla primerno transformacijo odzivne spremenljivke (ta bo v vseh primerih `log` transformacija). Na koncu bova med seboj primerjala intervale zaupanja dobljene z linearno regresijo (`lm`) in metodo ponovnega vzorčenja. Primerjavo bova naredila tako na rezultatih pred in po transformaciji, vendar pa moramo paziti, saj rezultati pred in po transformaciji med seboj niso primerljivi.

Pričakujeva, da bomo z metodo ponovnega vzorčenja dobili boljše rezultate.

2 Generiranje podatkov

Podatke sva generirala tako, da je v linearnem regresiji kršena predpostavka linearne odvisnosti odzivne spremenljivke od napovednih in kršena homoskedastičnost (konstantna varianca napak). Enačba, ki sva jo uporabila za generiranje odzivne spremenljivke se glasi:

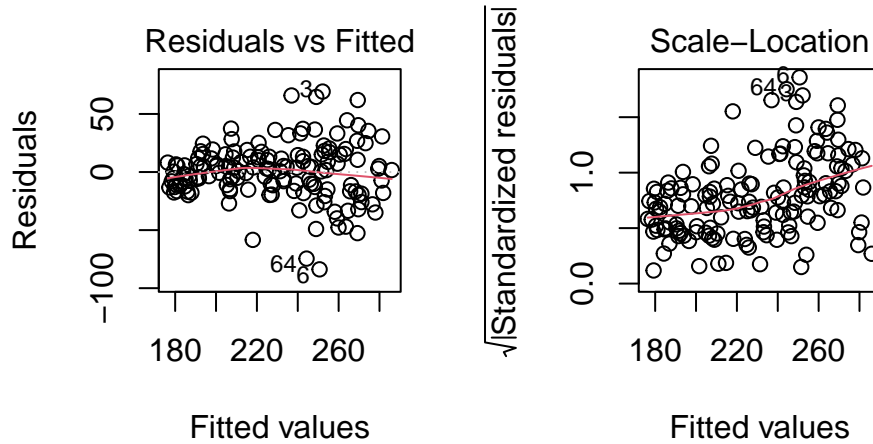
$$y_i = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i}^\gamma + \epsilon_i$$

pri čemer:

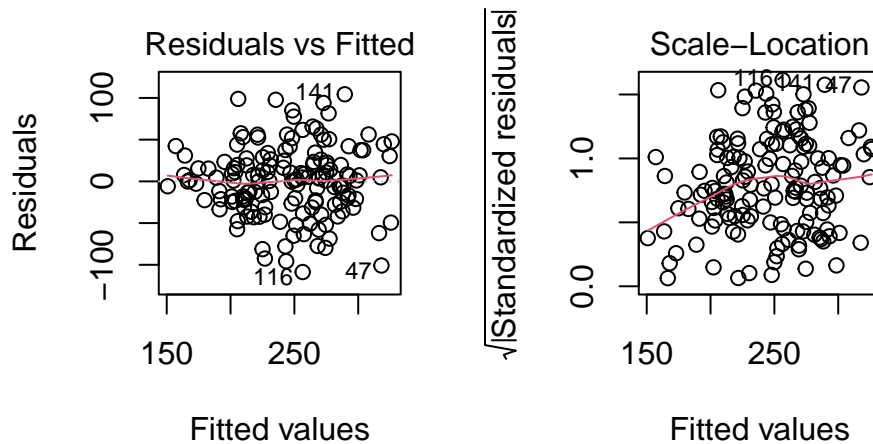
- β_0 -konstanta enaka 100
- β_1 -koeficient spremenljivke x_1 enak 3
- β_2 -koeficient spremenljivke x_2 enak 2
- γ -eksponent, ki ga bova spreminjala (določa nelinearno zvezo)
- ϵ -napaka, ki generirana iz porazdelitve $N(0, x_1 \cdot \alpha)$ (α določa povezanost s spremenljivko x_1)

Kot sva že omenila bova spreminjala faktorja γ in α . S faktorjem γ bomo kršili prespostavko o linearni zvezi, saj bo ta zavzel vrednosti 0.8 in 1.4. S faktorjem α pa bomo kršili predpostavko konstantne variance napak, saj se ta z večanjem vrednosti x_1 povečuje. Ta zavzame vrednosti $\alpha \in \{0.6, 1, 1.2\}$

Narišimo grafe ostankov za nekaj kombinacij faktorjev, da se prepričamo o kršenju predpostavk.



Slika 1: Grafi ostankov pri parametrih $\alpha=0.6$ in $\gamma=0.8$



Slika 2: Grafi ostankov pri parametrih $\alpha=1.2$ in $\gamma=1.4$

V obeh primerih lahko na desnem grafu opazimo, da se z večanjem vrednosti povečuje tudi variabilnost napak (naraščajoč trend). Na levem grafu lahko sorazmerno s faktorjem γ pričakujemo odstopanje vrednosti na robovih. V primeru $\gamma = 0.8$ opazimo, da so na robovih nekoliko nižje vrednosti, v primeru $\gamma = 1.4$ pa nekoliko višje vrednosti.

Za vsako kombinacijo parametrov preverimo ali se vrednost $\lambda = 0$ (log transformacija) nahaja znotraj 95% intervala optimalnega parametra λ , ki ga vrne funkcija `powerTransform`.

alpha	lambda	spodnji	zgornji
0.6	0.8	-0.831	0.534
0.6	1.4	-0.603	0.558
1.0	0.8	-0.454	0.541
1.0	1.4	-0.359	0.711
1.2	0.8	-0.663	0.326
1.2	1.4	-0.114	1.020

Res je $\lambda = 0$ vsebovana v vseh 95% intervalih zaupanja.

3 Primerjava vrednosti (brez transformacije)