

Domača naloga 10

16.11.2020

Domačo nalogo shranite v `.Rmd` datoteki z imenom **dn10_priimek.Rmd** (kjer namesto besede *priimek* uporabite vaš priimek).

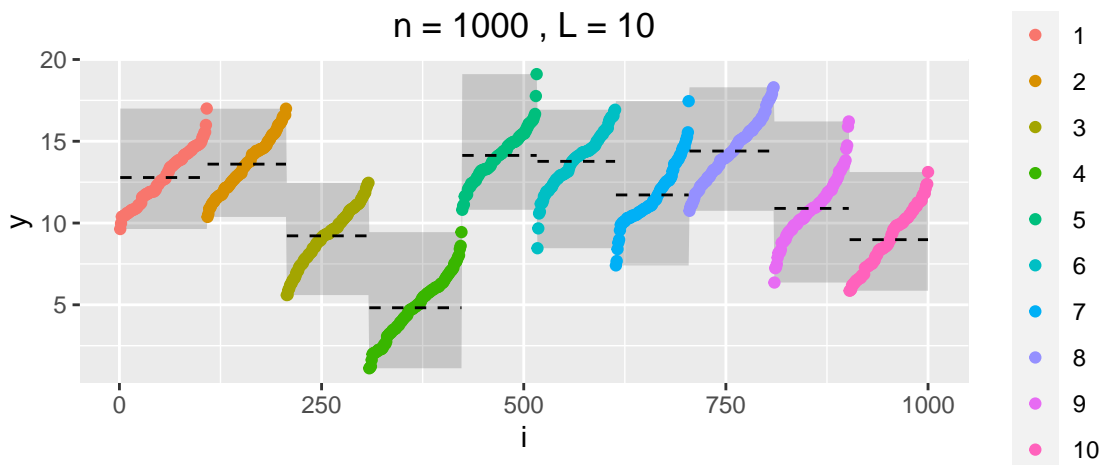
Naloga zahteva, da generirate podatke za model linearne regresije, ki vključuje n enot, ki so razdeljene v L skupin. Model lahko zapišemo kot:

$$y_{i,j} = 2 + 5 * x_{i,j} + 10 * b_i + \epsilon_{i,j},$$

kjer so $\epsilon_{i,j}$ ostanki, ki so neodvisni in definirani kot $\epsilon_{i,j} \sim N(0, 1)$.

- S knjižnicami znotraj paketa `tidyverse` generirajte `tibble` z n statističnimi enotami, za katere poznamo številski spremenljivki y in x ter skupino b . Velikost skupin naj bo generirana naključno.
- Spremenljivki x in b generirajte iz enakomerne zvezne spremenljivke z intervala $[0, 1]$.
- Statistično tabelo uredite najprej skupini b , nato pa še po y .
- Podatke nato narišite s pomočjo paketa `ggplot2`, in sicer tako, da bodo na sliki prikazane točke za dejanske vrednosti y (pobarvane glede na skupino), s črtkano črto bo označena povprečna vrednost skupine, pobarvano pa naj bo tudi območje med najmanjšo in največjo vrednostjo y v skupini.

Rezultat naj zgleда približno tako kot na sliki 1.



Slika 1: Pričakovan rezultat.

Namigi

- Dvojni indeks (npr. $y_{i,j}$) v modelu razumemo tako, da indeks i teče po skupinah od 1 do L , indeks j pa teče po enotah znotraj skupine i . Vrednost b_i je tako enaka za vse enote znotraj skupine i .
- Pri tej nalogi si lahko večkrat pomagata z funkcijo `group_by`. Po uporabi le te se lahko zgodi, da potrebujete funkcijo `ungroup`.
- Pred uporabo `ggplot2` je zelo dobro imeti vse potrebne podatke in vrednosti že v statistični tabeli.

Dodatno (za znanje, ne za točke)

- Funkcijo v nalogi poskušajte napisati brez definicije dodatnih lokalnih in globalnih spremenljivk - uporabite samo nekaj zaporednih klicev funkcij, ki so ločeni z `%>%`.
- V statistično tabelo na dodajte spremenljivko `y_pred` - napovedane vrednosti. Pri tem uporabite funkcijo `lm`.
- Spremenljivko iz prejšnje točke dodajte na sliko (npr. s tanko črno črto).