

LOGISTIČNA REGRESIJA ZA KIFOZO

Kot podatke imamo vrednosti spremenljivk za 83 primerov otrok, ki so imeli operacijo hrbtenice. Podatki so

<i>Kyphosis</i>	Binarna spremenljivka, prisotnost kifoze (ukrivljenost hrbtenice)
<i>Age</i>	Starost pacienta v mesecih
<i>Number</i>	Število vretenc vključenih v operacijo
<i>Start</i>	Število začetnega vretenca pri operaciji

Podatke lahko najdete na domači strani predmeta.

Predpostavljamo, da velja naslednji posplošen regresijski model, ki je znan pod imenov *logistični regresijski model*:

$$P(Y = 1|\mathbf{X}) = \frac{\exp(\beta_0 + \sum_{k=1}^n \beta_k X_k)}{1 + \exp(\beta_0 + \sum_{k=1}^n \beta_k X_k)},$$

kjer so β_0, \dots, β_m parametri modela, Y odzivna binarna spremenljivka in X_1, \dots, X_m neodvisne spremenljivke.

- Zapišite splošno obliko funkcije verjetja pri logističnem modelu.
- Ocenite parametre $\beta_0, \beta_1, \beta_2, \beta_3$ po metodi največjega verjetja neodvisno od programov, ki vam izvedejo logistično regresijo. Primerjajte rezultate. Komentar?
- Izračunajte “na roko” standardne napake ocen parametrov. Razložite, zakaj standardne napake še vedno lahko dobimo z uporabo Fisherjeve matrike informacije. “Na roko” pomeni, da izpeljete obliko Fisherjeve matrike informacije, izračunate pa jo s pomočjo računalnika.
- Preverite domnevo $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ proti alternativni, da je vsaj kakšen parameter različen od 0.
- Preverite domnevo $H_0: \beta_1 = 0$ proti $H_1: \beta_1 \neq 0$. Pri tem parameter β_1 pripada spremenljivki *Age*. Primerjajte p -vrednost s p -vrednostjo, ki vam jo vrnejo t -testi iz programov za izvajanje logistične regresije. Komentar?
- Kakšni so vaši sklepi glede vpliva spremenljivk?
- S kakšnimi grafičnimi predstavitevami bi testirali pravilnost vašega modela?