

V R Commander-ju odprite datoteko `spornice.txt`, ki vsebuje podatke za slovenske športnice. Podatki so bili objavljeni v članku R. Vauhnik et al., *Rate and Risk of Anterior Cruciate Ligament Injury Among Sportswomen in Slovenia* (2011), ki ga najdete na spletni učilnici.

Sledite spodnjim navodilom. Vaše odgovore zapišite v predlogo poročila `AnalizaPodatkov_vzorec.docx`, ki jo najdete na spletni učilnici.

### 1. Opis populacije, vrste študije in vzorca.

- S pomočjo članka definirajte preučevano populacijo.
- Kako so bili pridobljeni podatki (za katero vrsto študije gre)?
- Opišite vzorec iz R datoteke, tako da navedete število enot, število spremenljivk in za vsako spremenljivko zapišete, katere vrste je.
- Izberite si eno opisno in eno številsko spremenljivko. Za vsako izračunajte ustrezno mero sredine in razpršenosti, ju smiselno grafično prikažite ter kratko komentirajte.

### 2. Raziskovalno vprašanje 1: Ali športnice, ki jemljejo kontracepcijo, trenirajo tedensko enako časa kot športnice, ki kontracepcije ne jemljejo?

- Katere spremenljivke potrebujemo, da bi lahko preverili raziskovalno vprašanje?
- Z ustreznim grafičnim prikazom ilustrirajte, ali na vzorcu med skupinama obstaja razlika in komentirajte.
- Za vsako skupino (tiste, ki jemljejo kontracepcijo, in tiste, ki je ne jemljejo) izračunajte povprečje, standardni odklon in 95% interval zaupanja za povprečno količino treninga. Rezultate komentirajte.
- Z ustreznim statističnim testom preverite domnevo, da v preučevani populaciji v povprečju ni razlik v obsegu treninga med športnicami, ki jemljejo kontracepcijo, in tistimi, ki je ne. Komentirajte dobljeno vrednost  $p$  in izračunani interval zaupanja.
- Kaj so predpostavke uporabljenega testa? Na podlagi slike, ki ste jo narisali zgoraj, komentirajte izpolnjenost predpostavk v vašem primeru.

### 3. Raziskovalno vprašanje 2: Ali obstaja povezanost med športom in jemanjem kontracepcije?

- Katere spremenljivke potrebujemo, da bi lahko preverili raziskovalno vprašanje?
- Uporabite ustrezen grafični prikaz, ki vam pomaga pri odgovoru na raziskovalno vprašanje. Komentirajte.
- Izračunajte delež športnic, ki jemljejo kontracepcijo, glede na šport, ki ga trenirajo.
- Z ustreznim testom preverite domnevo, da v preučevani populaciji med spremenljivkami ni povezanosti.
- Kaj so predpostavke uporabljenega testa? Ali so v našem primeru izpolnjene?

4. **Raziskovalno vprašanje 3:** Ali obstaja razlika med hiperekstenzijo levega in desnega kolena (spremenljivki `kextl` in `kextr`)?
- Katere spremenljivke potrebujemo, da bi lahko preverili raziskovalno vprašanje?
  - Uporabite ustrezen grafični prikaz, ki vam pomaga pri odgovoru na raziskovalno vprašanje. Komentirajte.
  - Z ustreznim statističnim testom preverite domnevo, da je povprečna razlika v hiperekstenziji med levim in desnim kolenom v preučevani populaciji enaka nič. Komentirajte dobljeno vrednost  $p$  in izračunani interval zaupanja.
5. **Raziskovalno vprašanje 4:** Ali obstaja povezanost med hiperekstenzijo levega in desnega kolena?
- Katere spremenljivke potrebujemo, da bi lahko preverili raziskovalno vprašanje?
  - Uporabite ustrezen grafični prikaz, ki vam pomaga pri odgovoru na raziskovalno vprašanje. Komentirajte.
  - Z ustreznim statističnim testom preverite to raziskovalno vprašanje. Komentirajte dobljeno vrednost  $p$  in korelacijski koeficient.
  - Napišite kratek povzetek rezultatov iz te točke in komentirajte, zakaj smo dobili drugačne rezultate kot pri raziskovalnem vprašanju 3.
6. **Regresijska analiza.** Glede na rezultate pri raziskovalnem vprašanju 3 lahko rezultate o hiperekstenziji levega in desnega kolena združimo. Izračunajte novo spremenljivko, ki bo povprečna hiperekstenzija levega in desnega kolena. Preverite, kako je ta spremenljivka povezana s količino treninga, starostjo, kontracepcijo in indeksom telesne mase (`bmi`).
- Univariatno povezanost hiperekstenzije kolena in vsake izmed naštetih spremenljivk prikažite v eni tabeli, v kateri poročajte regresijski koeficient, 95% interval zaupanja za regresijski koeficient in vrednost  $p$ . Vsakemu modelu dodajte vsebinsko interpretacijo regresijskega koeficienta in vrednosti  $p$ .
  - Ocenite model multiple linearne regresije, v katerega vključite vse naštete neodvisne spremenljivke. Rezultate prikažite v tabeli, v kateri poročajte regresijski koeficient, 95% interval zaupanja za regresijski koeficient in vrednost  $p$  ter interpretirajte regresijske koeficiente in dobljene vrednosti  $p$ . Komentirajte razlike in podobnosti med multiplo in univariatno analizo.