

Primjer izrade seminara

(PTF – Primijenjena i inženjerska matematika)

Za primjer promatramo bazu fat s web stranice

https://ww2.amstat.org/publications/jse/jse_data_archive.htm

Varijable su objašnjene u datoteci fat.txt:

<http://ww2.amstat.org/publications/jse/datasets/fat.txt>

Radi se o različitim tjelesnim mjerama zabilježenim za 252 muškarca. U seminaru treba najprije objasniti što predstavlja svaka od varijabli te objasniti problematiku seminara. Primjerice, objasniti što je postotak tjelesne masti (BF), kako se izračunava, kako se mjeri, zašto je važan itd.

Datoteku s podacima treba najprije učitati u Excel i srediti. Ovoj bazi nema nijedne kvalitativne varijable, ali možemo kategorizirati neku od numeričkih. Prema *American Council on Exercise*¹, osobe se na sljedeći način mogu kategorizirati u 5 kategorija po vrijednostima BF:

DESCRIPTION	WOMEN	MEN
ESSENTIAL FAT	10–13%	2–5%
ATHLETES	14–20%	6–13%
FITNESS	21–24%	14–17%
AVERAGE	25–31%	18–24%
OBESE	32%+	25%+

Na osnovu vrijednosti za muškarce, dodat ćemo u bazu varijablu kategorija s vrijednostima 1 (osnovna masnoća) do 5 (pretilost). To je najlakše napraviti u Excelu koristeći funkciju IF. U nastavku su navedene samo neke od analiza koje bi se mogle provesti na ovoj bazi.

Bazu učitati u Statisticu (*Open -> Import selected sheets from Excel*). Za početak analizirati bitne varijable i njihove vrijednosti, utvrditi kakvog su tipa, izračunati osnovne statistike i napraviti grafičke prikaze. Primjerice, distribuciju dobi ispitanih, prosječnu dob, najstariji i najmlađi. Isto tako za masu, visinu, BF, gustoću... Za kategorija promotriti tablicu frekvencija, stupčasti dijagram i ostale prikaze za kvalitativne varijable, komentirati primjerice koliko je pretilih.

Od statističkih testova na jednom uzorku, neka od pitanja koja bi imalo smisla promatrati su: je li proporcija pretilih statistički značajno veća od 20%, ili uspoređivati neku od ostalih varijabli s tipičnim vrijednostima. Analizirati normalnosti distribucija za BF, visinu, masu i dr.

Varijable se mogu analizirati po različitim vrijednostima varijable kategorija. Primjerice, prikazati kategorizirane histograme i kutijaste dijagrame za masa, godine, visina u ovisnosti o kategorija. Usporediti primjerice prosječnu dob po kategorijama.

Testirati razlikuje li se očekivana masa (visina, godine ili nešto drugo) za prosječne (kategorija 4) i pretile (kategorija 5) - t-test za nezavisne uzorke po grupama (u Statistici staviti oznake *Code for group 1: 4* i *Code for group 2: 5*). Dalje, npr. je li proporcija pretilih veća kod starijih od 50 godina nego kod mlađih? Radi se o testu usporedbe proporcija za dva nezavisna uzorka. Zatim usporediti očekivane vrijednosti BF1 i BF2, radi se sada o zavisnim uzorcima.

¹ <https://www.acefitness.org/education-and-resources/lifestyle/tools-calculators/percent-body-fat-calculator>

Za ovu bazu posebno je važna analiza veza među varijablama, posebno odnos s BF. Primjerice postoji li korelacija između mase (godina, visine ili nešto drugo) i BF, kakva je kvaliteta odgovarajućeg linearnog modela, koliko uspješno možemo predvidjeti BF na osnovu svake od tih varijabli. Koja od tjelesnih mjera je najjače korelirana s BF? Koja s masom?

Ovo su samo neki od primjera pitanja koja se mogu postaviti vezano za ove podatke.