

Težina

Statističke metode za informatičare

Filip Antunović

fantunovi@foi.hr

Petar Matišić

pmatistic@foi.hr

Fakultet organizacije i informatike
Sveučilište u Zagrebu

14. lipnja 2023.

Sadržaj

1 Uvod

2 Zadatci

- Zadatak A
- Zadatak B
- Zadatak C
- Zadatak D
- Zadatak E
- Zadatak F
- Zadatak G

3 Zaključak

Uvod

Skup podataka

- Važnost potkožnog masnog tkiva je ključno za procjenu zdravlja.
- Razmatranje alternativnih metoda procjene s ciljem smanjenja financijskog i vremenskog opterećenja.
- Primjena antropometrijskih instrumenata kao što su kaliperi, uz kvantifikaciju obujma i kožnih nabora.
- Skup podataka¹ uključuje 252 ispitanika s varijablama uključujući dob, težinu, visinu, i obujmove različitih segmenata tijela.

¹<http://lib.stat.cmu.edu/datasets/bodyfat>

Varijable

Varijabla (HR)	Varijabla (EN)	Tip varijable	Modaliteti
Gustoća (podvodno mjerenje)	Density (underwater)	Kontinuirana	-
Postotak tjelesne masti (Siri)	Percent body fat (Siri)	Kontinuirana	-
Dob (godine)	Age (years)	Kontinuirana	-
Težina (lbs)	Weight (lbs)	Kontinuirana	-
Visina (inči)	Height (inches)	Kontinuirana	-
Opseg vrata (cm)	Neck circ. (cm)	Kontinuirana	-
Opseg prsa (cm)	Chest circ. (cm)	Kontinuirana	-
Opseg trbuha (cm)	Abdomen 2 circ. (cm)	Kontinuirana	-
Opseg bokova (cm)	Hip circ. (cm)	Kontinuirana	-
Opseg bedra (cm)	Thigh circ. (cm)	Kontinuirana	-
Opseg koljena (cm)	Knee circ. (cm)	Kontinuirana	-
Opseg gležnja (cm)	Ankle circ. (cm)	Kontinuirana	-
Opseg bicepsa (istegnut, cm)	Biceps (extended) circ. (cm)	Kontinuirana	-
Opseg podlaktice (cm)	Forearm circ. (cm)	Kontinuirana	-
Opseg zapešća (cm)	Wrist circ. (cm)	Kontinuirana	-

Tablica: Popis i opis varijabli

Proces čišćenja podataka



Zadatci

Zadatak A

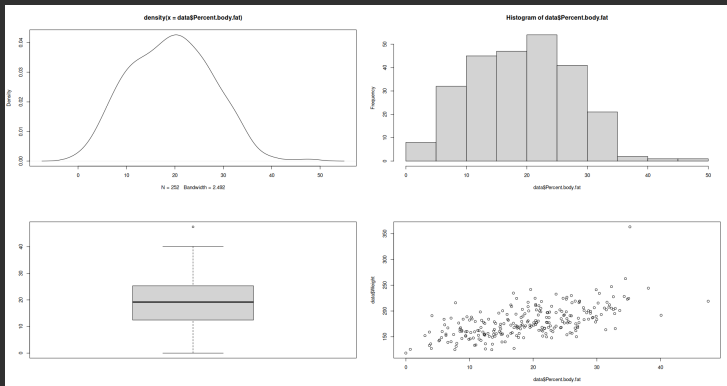
BMI (*Body Mass Index*) se izračunava kao omjer mase osobe u kilogramima i kvadrata visine osobe u metrima. Matematički, to može biti izraženo kao:

$$\text{BMI} = \frac{\text{Masa (kg)}}{\text{Visina (m)}^2}.$$

Međutim, kada je masa izražena u *pounds* i visina u inčima (*inches*), tada dolazimo do formule koja se koristi s jedinicama koje su uobičajene u Sjedinjenim Američkim Državama:

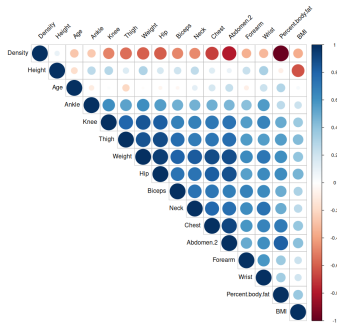
$$\text{BMI} = \frac{\text{Weight (lbs)}}{\text{Height (in)}^2} \times 703.$$

Zadatak B



Slika: Grafički prikaz varijable `Percent.body.fat`

Zadatak C



Slika: Grafički prikaz matrice korelacije

Zadatak D

Rezultati Shapiro-Wilk testa normalnosti:

- Density: $p = 0.66$, moguća normalna distribucija.
- Percent body fat: $p = 0.16$, moguća normalna distribucija.
- Age: $p = 0.001$, vjerojatno nije normalno distribuirano.
- Većina ostalih varijabli (uključujući Weight, Height, Neck, Chest i druge) imaju p -vrijednosti < 0.05 , ukazujući da vjerojatno nisu normalno distribuirani.

Zadatak E

Listing 1: Formiranje nove varijable

```
# e) Formiranje nove varijable BMI_faktor
data$BMI_faktor <- cut(data$BMI,
                        breaks = c(-Inf, 20, 25, 30, Inf),
                        labels = c("Pothranjenost",
                                   "Idealna tezina",
                                   "Prekomjerna tjelesna
                                   masa",
                                   "Pretilost"))
```

Zadatak F

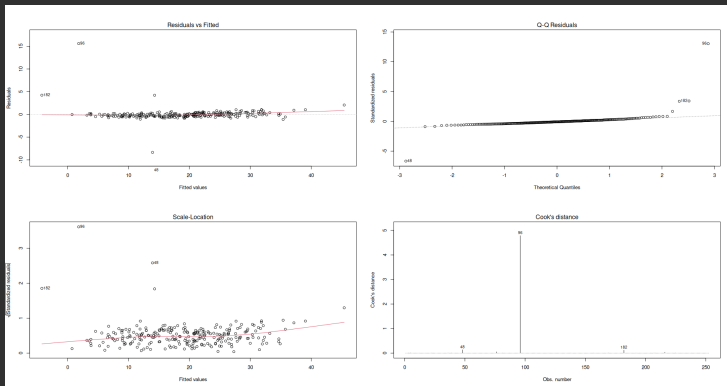
- `Percent.body.fat`:
 - Pretpostavke zadovoljene ($p > 0.05$).
 - ANOVA: značajne razlike ($p < 2 \times 10^{-16}$).
 - Tukey HSD za određivanje razlika među kategorijama.
- Ostale varijable (Neck, Chest, ...):
 - Pretpostavke nisu zadovoljene ($p < 0.05$).
 - Kruskal-Wallisov test korišten.
 - Značajne razlike u kategorijama BMI faktora.

Zadatak G

Varijabla	Koeficijent	Standardna greška	t-vrijednost	p-vrijednost
(Intercept)	450.4	10.73	41.964	$< 2e - 16$
Gustoća	-411.5	8.313	-49.500	$< 2e - 16$
Dob	0.0144	0.009688	1.486	0.139
Težina	0.00935	0.01604	0.583	0.561
Visina	-0.0117	0.03062	-0.383	0.702
⋮	⋮	⋮	⋮	⋮

Tablica: Rezultati linearne regresije

Zadatak G



Slika: Prikaz svih dijagnostičkih grafova

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥ 0.1	

Slika: xkcd

Q & A

Hvala na pozornosti!

Pitanja?

Literatura

1. Johnson, R. W. *Lists estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men.* Department of Mathematics & Computer Science, South Dakota School of Mines & Technology.
<http://lib.stat.cmu.edu/datasets/bodyfat> (2023.).