

Statistika za jezikoslovno istraživanje

Damir Ćavar
Sveučilište u Zadru
21. travnja 2010.

Domaći

- Varijabla X

```
x <- c(3, 4, 8, 4, 2, 1, 0, 6)
```

- Varijabla Y

```
y <- c(1, 2, 4, 2, 2, 0, 1, 4)
```

- Izračunajte:

- Standardnu mjeru za sve rezultate varijabli
- Korelacijski koeficijent za X i Y

Dodatni domaći

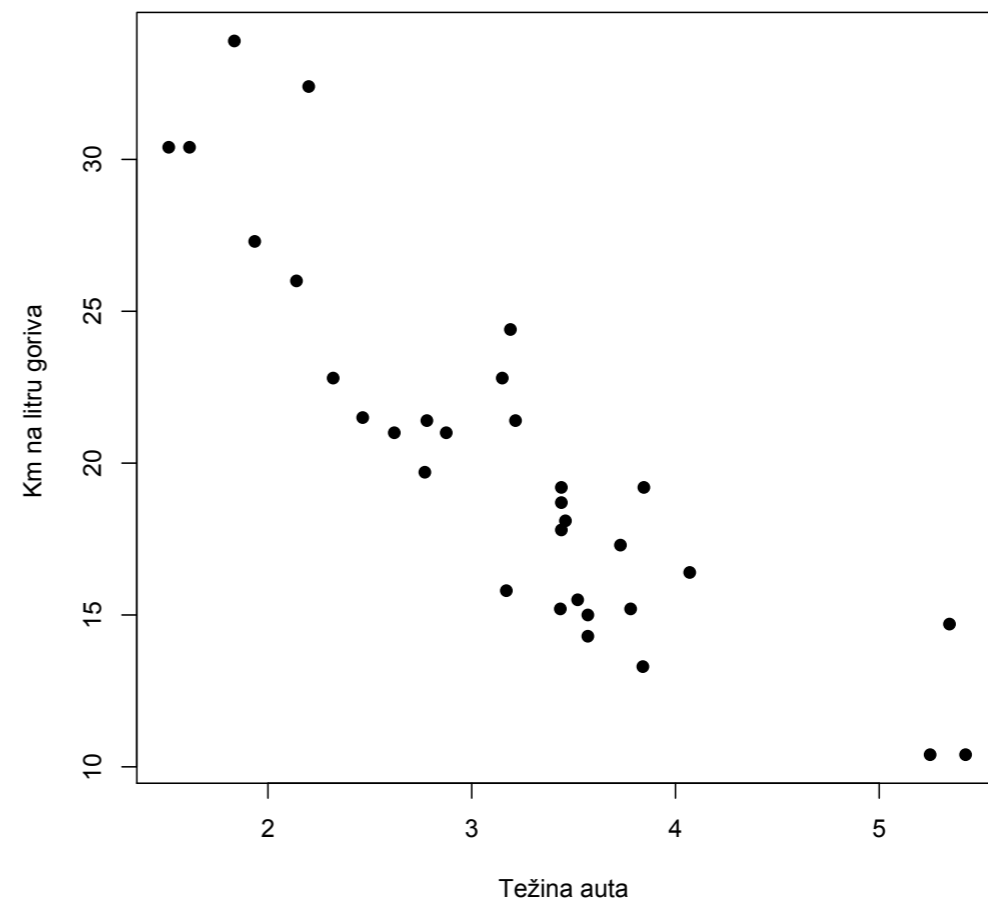
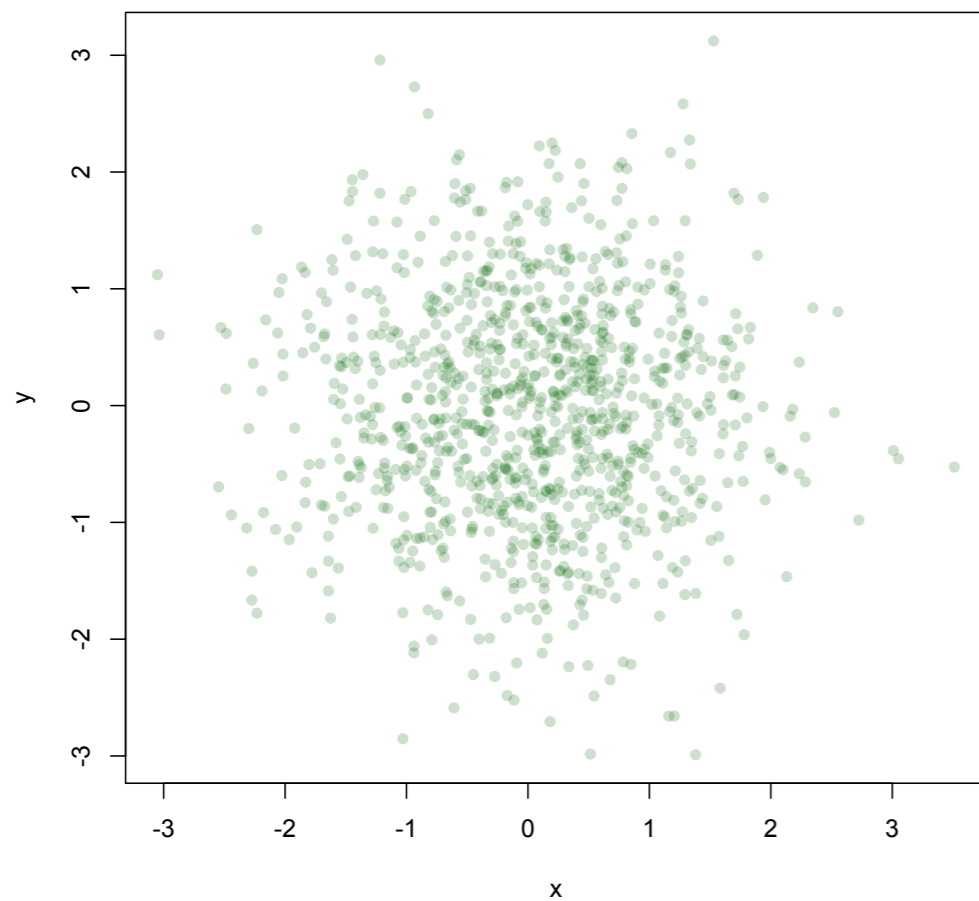
- Prevedite sljedeću jednadžbu u R:

$$- \sum_{i=1}^n p(x_i) \frac{\log_b p(x_i)}{\log_b(n)}$$

Raspršeni grafikon

- Što ako dodamo svakoj vrijednosti jedne varijable konstantni broj?

Primjer raspršenog grafikona



Korelacijski koeficijent

- Što ako dodamo svakoj vrijednosti jedne varijable konstantni broj?
 - Standardne mjere se ne mijenjaju,
 - r se ne mijenja.
- Što ako množimo svaku vrijednost jedne varijable s konstantnim brojem?

Korelacijski koeficijent

- Što ako množimo svaku vrijednost jedne varijable s konstantnim pozitivnim brojem?
- Standardne mjere se ne mijenjaju,
- r se ne mijenja.
- Što ako množimo svaku vrijednost jedne varijable s -1 ?

Korelacijski koeficijent

- Što ako množimo svaku vrijednost jedne varijable s -1 ?
- r postaje negativan,
- tako i linearna korelacija okreće smjer
- Što ako zamjenjujemo varijable, tj. relacija između visine i broja cipela, umjesto broja cipela i visine?

Korelacijski koeficijent

- Korelacija ne implicira kauzalne relacije
 - Asocijacija nije kauzalna

Raspršeni grafikon

- Oblici:
 - linearne korelacije: kao lopta, oval itd.
 - r dobro opisuje linearne korelacije
 - ne-linearne korelacije: krivulje itd.
 - r ne opisuje dobro krivulje
- ekstremne vrijednosti

Ekološke korelacije

- Korelacije o prosjecima grupa
 - Prosjek grupe (po kvartovima u gradu) u raspršenim grafikonom preko prosječne ocjene i zarade

Korelacijski koeficijent

- Pogrešne korelacije...
 - npr. konfundirani faktor

Korelacijski koeficijent

- Predviđanje Y , ako znamo X
 - ako je korelacija blizu 0
 - ako je korelacija slaba
 - ako je korelacija blizu 1

Regresijska analiza

- Koraci:
 - Pretvaramo X u standardne mjere
 - Umnožavanje s r
 - predviđeni Y u standardnim mjerama
 - Vratiti Y iz standardnih u realne mjere

Regresijska analiza

- Predviđanje vrijednosti, primjer u R-u...

Domaći

- Poglavlje 21 u “Statistics”
- Prvo poglavlje u: Baayen “Analyzing Linguistic Data”