



5. TILASTOAINEISTON ANALYSOINTIA

5.1 HAJONTALUKUJA

5.2 REGRESSIO JA KORRELAATIO

HAJONTALUVUT

VAIHTELUVÄLI

- Ilmaistaan antamalla havaintoarvojen pienin ja suurin arvo

VAIHTELUVÄLIN PITUUS

- Suurimman ja pienimmän havaintoarvon erotus

KESKIHAJONTA

- Kuvaa sitä, kuinka kaukana keskiarvosta havaintoarvot ovat

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

x = havaintoarvo

\bar{x} = keskiarvo

n = havaintoarvojen lukumäärä

Esim. Määritä lukujoukon 2, 7, 3, 10, 15 vaihteluväli, vaihteluvälin pituus ja keskihajonta laskukaavalla sekä sopivalla ohjelmalla.

LUOKITELLUN AINEISTON KESKIHAJONTA:

$$s = \sqrt{\frac{\sum f \cdot (x - \bar{x})^2}{n - 1}}$$

Esim: <https://vimeo.com/227019957/aaab673ae2>

5.2 REGRESSIO JA KORRELAATIO

KORRELAATIOKERROIN r

- Kuvaa kahden tilastomuuttujan välisen lineaarisen riippuvuuden voimakkuutta.
- Lasketaan sopivalla ohjelmalla.
- Korrelaatiokertoimen ohjeellinen voimakkuustaulukko:

Voimakkuus	r	r
merkityksetön	0,0...0,3	-0,3...0,0
kohtalainen	0,3...0,6	-0,6...-0,3
huomattava	0,6...0,8	-0,8...-0,6
voimakas	0,8...1,0	-1,0...-0,8

REGRESSIO

- Kahden tilastomuuttujan lineaarista riippuvuutta kutsutaan lineaariseksi regressioksi ja suoraa regressiosuoraksi.
- Sovitetaan sopivalla ohjelmalla (opetettu mab4 kurssissa).
- Aina aineisto ei ole lineaarinen. Silloin pitää sovittaa epälineaarisia malleja.

SELITYSASTE

- Korrelaatiokertoimen toinen potenssi r^2 .
- Ilmaisee, kuinka monta prosenttia toisistaan riippuvat tilastomuuttujat selittävät toistensa arvojen vaihtelusta.

Esim. Sini oli tammikuusta kesäkuuhun Roomassa. Samalla hän tutki sademäärän ja keskilämpötilan välistä riippuvuutta. Tulokset on taulukossa.

- a) Arvioi riippuvuuden voimakkuutta korrelaatiokertoimella ja muodosta regressiosuora.
- b) Arvioi mallin avulla, kuinka suuri sademäärä on, jos kuukauden keskilämpötila on 30 °C.

Sademäärä (mm)	71	62	57	51	45	37
Lämpötila (°C)	11	13	15	18	23	28