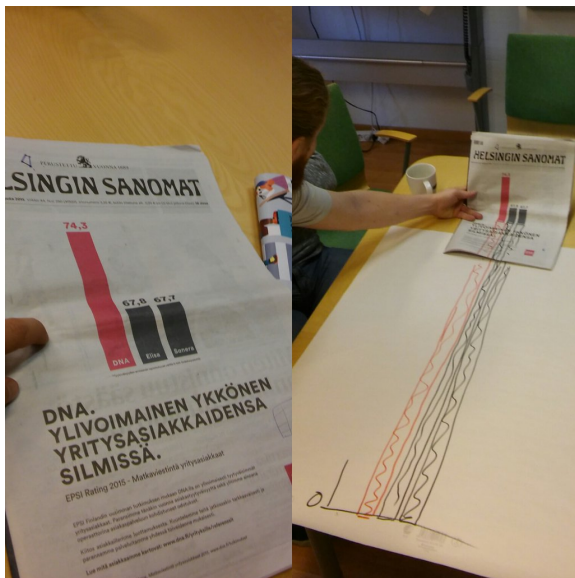


Tilastotieteestä

Tapio Hansson

Osa tilastoista on huonoja



Siksi niiden tekemiseen ja tulkitsemiseen on syytä perehtyä!

Tilastollinen muuttuja

- ▶ Lähes mistä tahansa voi tehdä tilaston. Muuttujan valinta on kuitenkin äärimmäisen tärkeä osa sitä, miten tilasto kuvaa taustalla olevaa ilmiötä.
- ▶ Havaintoaineisto koostuu tilastollisen muuttujan tutkimuksessa saamista arvoista.
- ▶ Näitä arvoja kutsutaan jakaumaksi.
- ▶ Mikäli muuttuja voi saada ainoastaan tiettyjä arvoja (esim. kokonaislukuja) puhutaan diskreetistä muuttujasta. Jatkuvalle välillä reaalilukuja arvokseen saava muuttuja on jatkuva.

Frekvenssi

- ▶ Muuttujan tietyn arvon havaintojen lukumäärää aineistossa kutsutaan kyseisen arvon frekvenssiksi.
- ▶ Arvon esiintymislukumäärän prosentuaalista osuutta kaikista arvoista kutsutaan suhteelliseksi frekvenssiksi.
- ▶ Esim. Erään FY1-kurssin arvosanat

Arvosana	lukumäärä
10	3
9	7
8	8
7	1
6	3
5	0
4	1

Yllä olevaa esitystä kutsutaan frekvenssitauluksi.

Tunnuslukuja

- ▶ Aineiston *tyyppi-arvo*, eli *moodi* on aineistossa useimmin esiintyvä arvo. Edellisessä esimerkissä arvosana 8 oli useimmalla opiskelijalla.
- ▶ *Keskiarvo* on havaintoaineiston arvojen summa jaettuna havaintojen lukumäärällä. Edellisessä esimerkissä:

$$\frac{3 \cdot 10 + 7 \cdot 9 + 8 \cdot 8 + 1 \cdot 7 + 3 \cdot 6 + 0 \cdot 5 + 1 \cdot 4}{23} \approx 8,1$$

- ▶ *Mediaani* on suuruusjärjestykseen asetettujen arvojen keskimäinen arvo. Edellisessä esimerkissä suuruusjärjestys on:
10, 10, 10, 9, 9, 9, 9, 9, 9, 9, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 7, 6, 6, 6, 4
Jossa keskimäinen luku on 8.

Hajonta

- ▶ Usein keskimääräisten arvojen lisäksi oleellista on, kuinka paljon arvot keskimääräisestä poikkeavat. Tällöin tarvitaan hajontalukuja.
- ▶ Tyypillisin on *keskihajonta*, s , joka on keskiarvosta laskettujen poikkeamien neliöiden keskiarvon neliöjuuri.

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

- ▶ Keskihajonnan neliötä, eli keskipoikkeamien neliöiden keskiarvoa kutsutaan *varianssiksi* s^2 .

Esimerkki

- ▶ Aiemman arvosana-aineiston keskiarvo $\bar{x} = 8,087$.

Arvosana	lukumäärä	$(x - \bar{x})^2$
10	3	3,660
9	7	0,834
8	8	0,008
7	1	1,182
6	3	4,356
5	0	9,529
4	1	16,70

Näin ollen keskihajonnaksi saadaan

$$s = \sqrt{\frac{3 \cdot 3,660 + 7 \cdot 0,834 + 8 \cdot 0,008 + 1 \cdot 1,182 + 3 \cdot 4,356 + 1 \cdot 16,70}{23}}$$
$$\approx 1,44$$