

*”There are three kinds of lies:
lies, damned lies, and statistics.”*

–Mark Twain (?)

Tilastotiede – mitä se on?

Oppiaine ja välineistö, jolla...

- ...selvitetään tekemiemme *havaintojen* merkitys ja sisältö.
- ...hallitaan *satunnaisuus* ja epävarmuus.
- ...tehdään satunnaisuutta sisältävien havaintojen perusteella *johtopäätöksiä* (reaalimaailman) olosuhteista ja seikoista, jotka ovat tuntemattomia ja suoran havainnoinnin tavoittamattomissa.
- ”tieteellinen menetelmä käytännössä”

Mitä se ei ole?

- ”temppukokoelma” tai ”työkalupakki” yksittäisten tutkimusongelmien ratkaisemiseksi
- osa matematiikkaa, vaikkakin se käyttää matematiikkaa, esim. todennäköisyyslaskentaa satunnaisuuden mallintamiseen.

Havainnot eli aineisto (data)

Numeerisia (tai sellaisiksi muunnettavissa olevia)

- mittaustuloksia
- lukumääriä
- järjestyslukuja
- luokittelevan muuttujan arvoja
- jne.

Tilastotieteen osa-alueita

1. Otantateoria ja koesuunnittelu

2. Mittaaminen

- miten kerätä tutkittavasta ilmiöstä mahdollisimman laadukas aineisto?
- otanta-asetelmat: miten turvata, että otos edustaa hyvin perusjoukkoa?
- mittaamisen ongelma (esim. kyselylomakkeet)

3. Deskriptiivinen tilastotiede

- tilastollisen aineiston tiivistäminen, tulkinta ja kuvailu
- ...*tunnuslukujen* (keskiarvot, hajonnat, korrelaatiokertoimet, jne.) avulla
- ...*graafisten menetelmien* (histogrammit, hajontakuviot, jne.) avulla
- usein välttämätöntä, jotta löydetään tutkittavaan aiheeseen liittyvät mielenkiintoiset piirteet ja hypoteesit

4. Tilastollinen päättely (inferenssi) – KURSSIMME AIHE

- tilastollisen tutkimuksen ”kova ydin” (?)
- miten ja millaisia päätelmiä aineiston perusteella voi tehdä tutkittavasta ilmiöstä?
- luonteeltaan *induktiivista*: yksittäisistä havainnoista yleisiä päätelmiä (vrt. deduktiivinen päättely esim. matematiikassa)
- yleensä etenee *seurauksista syihin* (”inversio-ongelma”)
- perustuu todennäköisyysmallien käyttöön
- (lähes) aina *epävarmaa* ja sisältää virheellisen päätelmän riskin
- kuitenkin tyypillisesti tuottaa *arvion epävarmuuden suuruudesta*

5. Tilastollinen tietojenkäsittely