

Luku 4 - Sattuma ja satunnaisuus tilastotieteessä

Tiivistelmä

Luvun ydinviesti

Luku käsittelee satunnaisilmiöitä ja satunnaismuuttujia sekä niiden roolia erityisesti tilastotieteessä, mutta myös kaikessa tieteenteossa sekä elämässä muutenkin.

Tämän satunnaisuuden roolin sekä luonteen ymmärtäminen on ympäröivää maailmaa koskevan havainnoinnin kannalta tärkeää, sillä me ihmiset olemme alttiita kognitiivisille vinoumille.

- ▶ Miten erottaa systemaattinen ja oikeasti merkityksellinen vaihtelu satunnaisvaihtelusta? Ensin täytyy ymmärtää satunnaisuutta ja se on keskeinen osa tilastotiedettä.

Keskeiset termit: satunnaisilmiö

Satunnaisilmiö

- ▶ Reaalimaailman ilmiö on satunnaisilmiö, jos seuraavat ehdot pätevät:
 - ▶ Ilmiöllä on useita erilaisia tulosvaihtoehtoja.
 - ▶ Sattuma määrää mikä tulosvaihtoehtoista toteutuu, eli yksittäistä tulosta ei voida tietää etukäteen.
 - ▶ Vaikka tulos vaihtelee ilmiön toistuessa satunnaisesti, käyttäytyy tulosvaihtoehtojen suhteellisten osuuksien jakauma tilastollisesti stabiilisti ilmiön toistokertojen lukumäärän kasvaessa.
- ▶ **Tilastollisella stabiiliudella** tarkoitetaan sitä, että on mahdollista arvioida kuinka **todennäköisiä** erilaiset tapahtumat, eli satunnaisilmiön tulosvaihtoehdot ovat. Täten näihin tulosvaihtoehtoihin on liityttävä säännönmukaisuutta, jonka on tultava esille ilmiön toistuessa.

Keskeiset termit: satunnaismuuttuja

Satunnaismuuttuja

- ▶ Satunnaismuuttuja (sm.), Y , on todennäköisyyslaskennan peruskäsite, jolla tarkoitetaan satunnaisilmiön määräämää lukua.
 - ▶ Satunnaismuuttujan Y realisoituvaa arvoa y kutsutaan realisaatioksi tai toteumaksi.
 - ▶ Tilastollinen aineisto muodostuu useiden satunnaismuuttujien realisoituneista arvoista.
 - ▶ Realisoituneiden arvojen vaihtelua tilastoyksiköiden välillä kutsutaan satunnaisvaihteluksi.
- ▶ Satunnaismuuttuja kuvaa tilastollisen muuttujan vaihtelua tilastoyksiköiden joukossa ja sen todennäköisyysjakauma määrää erilaisten tulosvaihtoehtojen todennäköisyyden ja mahdollistaa tilastollisen analyysin ja päättelyn.

Keskeiset termit: jatkuvat ja diskreetit sm:t

Jatkuvat ja diskreetit satunnaismuuttujat

- ▶ Satunnaismuuttuja Y on
 - ▶ Jatkuva, jos se voi saada ylinumeroituvan määrän arvoja tai toisin sanoen minkä tahansa arvon joltain reaalitylukuväliltä.
 - ▶ Diskreetti, jos se voi saada vain joitain mahdollisia arvoja (yksittäisiä, äärellisen tai numeroituvasti äärettömän määrän, arvoja)

Satunnaisilmiöt ja satunnaismuuttujat tilastotieteessä

Tilastotieteessä tutkimuskohteena on aina tilastoyksikköjen tilastollisista muuttujista, tutkimusmuuttujista, koostuva havaintoaineisto, jonka pohjalta pyritään tekemään perusjoukkoa/populaatiota koskevia päätelmiä.

Nämä tilastolliset muuttujat tulkitaan satunnaisiksi, eli ne ovat **satunnaismuuttujia**, ja täten tilastollisen tutkimuksen tavoite on tutkia satunnaisilmiötä, joka on generoinut nämä havaitut eli toteutuneet arvot.

Satunnaisilmiöiden mallintaminen tilastotieteessä tapahtuu kyseisen satunnaisilmiön tilastollisella mallilla, jolla tarkoitetaan sitä todennäköisyysjakaumaa, jonka ajatellaan generoineen havainnot.

Haluamme oppia tästä todennäköisyysjakaumasta siis jotain!

Satunnaisuus ja todennäköisyydet

Tilastotieteessä havaintoaineistosta pyritään löytämään säännönmukaisuuksia tai ns. tilastollista stabiliteettia, eli pyritään arvioimaan sitä kuinka todennäköisiä erilaiset tilastollisen muuttujan arvot ovat.

Tilastotiede voidaan jakaa kahteen merkittävään paradigmaan sen mukaan, miten **tilastolliseen päättelyyn** suhtaudutaan.

- ▶ Frekventistinen, eli klassinen tilastotiede
 - ▶ Tilastollinen päättely perustuu yksinomaan havaittuun aineistoon ja siihen liitettävään tilastolliseen malliin.
- ▶ Bayesilainen tilastotiede
 - ▶ Tilastollisessa päättelyssä hyödynnetään havaitun aineiston lisäksi tutkimuskysymystä koskevia ennakkokäsityksiä, joita päivitetään kun aineisto havaitaan.

Tilastolliset mallit, jakaumat ja parametrit

Tilastolliset mallit approksimoivat “todellista” aineiston generoinutta ilmiötä. Keskeinen oletus onkin, että havaitun aineiston on generoinut jokin satunnaisilmiö, jota voidaan kuvata parametrisella todennäköisyysjakaumalla, mutta jonka parametrit ovat tuntemattomia.

Tilastotieteen tehtävä on pyrkiä jollain tavalla selvittää mitkä nämä parametrit ovat. Yleisin tapa tehdä tämä, on käyttää nk. uskottavuusfunktioita, joka riippuu niin tilastollisen mallin parametreista, kuin havaintoaineistosta.

Ns. “suurimman uskottavuuden estimoinnissa” pyritään löytämään sellaiset parametriarvot, jotka maksimoivat kyseisen uskottavuusfunktion. Nämä parametriarvot ovat siten tietyssä mielessä paras arvio jakauman todellisista, mutta tuntemattomista parametreista, annettuna malli- eli jakaumaoletus.

Odotusarvo ja varianssi

Odotusarvo

Satunnaismuuttujan Y odotusarvo $E(Y)$ kuvaa satunnaismuuttujan odotettavissa olevaa arvoa. Muodostetaan satunnaiskokeen tulosten painotettuna keskiarvona, jossa kunkin tuloksen painona on vastaavan tapauksen todennäköisyys.

Varianssi

Satunnaismuuttujan Y hajontaa voidaan mitata varianssilla, $Var(Y)$ joka kuvaa satunnaiskokeen tulosten vaihtelua odotusarvon ympärillä (ks. tarkempi määrittely monisteesta). Mitä lähempänä varianssi on nollaa, sitä todennäköisempää on että satunnaismuuttujan realisoituvat arvot ovat lähellä sen odotusarvoa.

Joitain jakaumia

Jos satunnaismuuttuja Y noudattaa normaalijakaumaa, odotusarvolla $E(Y) = \mu$ ja varianssilla $Var(Y) = \sigma^2$, niin tällöin merkitään $Y \sim N(\mu, \sigma^2)$.

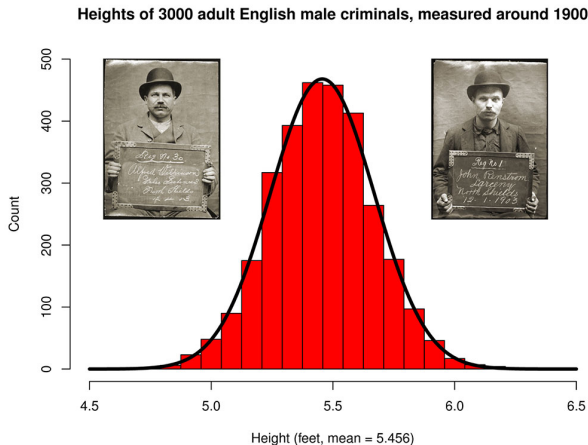


Figure 1: Kuva: A. H. Dekker.

Sattuman rooli tieteenteossa

Sattuma vaikuttaa tieteessä monin eri tavoin, eikä sen vaikutusta voida koskaan sulkea täysin pois.

Hyvin suunnitellussa ja luotettavia menetelmiä käyttävässä tutkimuksessakin voidaan sattuman vuoksi saada tuloksia, jotka ovatkin vääriä.

Täten sattuman ja satunnaisuuden rooli tiivistyy erityisesti tieteenteossa ja tieteellisten tutkimusten lukemisessa: yksi, vaikka merkittäväkin, tutkimustulos ei ole vielä tae luotettavuudesta.

Tieteessä muistamme onneksi luottaa vain pitkän ajan saatossa kumuloituvaan, eri tavoin koeteltuun ja tutkittuun tietoon!
Erityisen tärkeää tieteen popularisoinnissa!