

# Tilastotiede tutuksi

## Tiltu kertoo tilastotieteestä

esimerkein siitä miten tilastotiede liittyy jokapäiväiseen elämäämme, tauteihin, terveyteen, ravintoon, talouteen, ympäristöön jne.

laskuharjoituksilla, tutustuttamalla peruskäsitteisiin

Tervetuloa, 220 osallistujaa weboodissa!

Pieni keskustelutuokio naapurin kanssa:

*Kerro mikä on pääaineesi*

Kuullaan tämän jälkeen muutama esimerkki....

# Teitä on kymmenestä tiedekunnasta ja Avoimesta yliopistosta

## Matemattis-luonnontieteellinen

tilastotiede, matematiikka, tietojenkäsittelytiede, fysiikka, kemia, maantiede, geologia

## Bio- ja ympäristötieteellinen

biologia, akvaattiset tieteet, molekyylibiotieteet, ympäristönmuutos

## Maatalous-metsätieteellinen

mtsäekonomia, kotieläintiede, biotekniikka

## Farmasian

## Eläinlääketieteellinen

## Lääketieteellinen

## Teologinen

## Humanistinen

filosofia, kielitiede, kieliteknologia, musiikkitiede, Afrikan tutkimus, latinan kieli

## Käyttäytymistieteellinen

psykologia, kasvatustiede, kognitiotiede, käsityötiede

## Valtiotieteellinen

tilastotiede, taloustiede, sosiologia, käytännöllinen filosofia

## Svenska social- och kommunalhögskolan

## Miten löydät tietoja tilastotieteen opetuksesta

Matematiikan ja tilastotieteen laitoksen opetuksen ja opiskelun etusivu

<http://wiki.helsinki.fi/display/mathstatOpiskelu/Etusivu>

Mitä tarkoittaa tilastotieteen opiskelu pääaineena

<http://wiki.helsinki.fi/display/mathstatOpiskelu/Tilastotieteen+opiskeluohjeita>

Mitä muita kursseja on tilastotieteestä 2015-16, myös sivuaineopiskelijalle

syksy: <http://wiki.helsinki.fi/display/mathstatKurssit/Syksy+2015>

kevät: <http://wiki.helsinki.fi/pages/viewpage.action?pageId=161052172>

Käytä aina tiedonlähteenä laitoksen opetuksen ja oppimisen etusivua linkkeineen, *ei weboodia*.

Kurssien sisältö, toiminta, tehtävät jne esitetään wiki-kotisivuilla, ja sivuja päivitetään jatkuvasti.

Monilla kursseilla on wiki-sivujen "lisukkeena" Moodle-oppimisympäristön sivu, mutta Moodle *ei ole* millään kurssilla ensisijainen "kurssisivu".

Tiltu on nyt pilottivaiheessa.  
Osallistujia tulikin paljon, hienoa mutta haasteellista!

Miksi Tiltu?

Tilastotiede, jota lähes kaikki tarvitsevat, on huonosti tunnettu ja ymmärretty koulujen opetusohjelmien kautta:

Tarvitaan tieteenalan esittely opintojen alussa pääaineopiskelijoille, joille alan luonne alkaa yleensä avautua vasta ensimmäisen opiskeluvuoden lopulla.

Ja kaikki muut....mitä se tilastotiede, johon jossain vaiheessa yleensä "törmää", varsinaisesti on: ei se ole mikään temppukokoelma. Kannattaisiko sitä opiskella enemmänkin.....

Yliopiston Iso pyörä, laajoihin kandidutkintoihin tähtäävä uudistus.

Mielikuvia tilastotieteestä... *se on sitä millä saadaan p-arvoja.....*

# *P* values are just the tip of the iceberg

Ridding science of shoddy statistics will require scrutiny of every step, not merely the last one, say **Jeffrey T. Leek** and **Roger D. Peng**.

**T**here is no statistic more maligned than the *P* value. Hundreds of papers and blogposts have been written about what some statisticians deride as 'null hypothesis significance testing' (NHST; see, for example, [go.nature.com/pfvgqe](http://go.nature.com/pfvgqe)). NHST deems whether the results of a data analysis are important on the basis of whether a summary statistic (such as a *P* value) has

## DATA PIPELINE

The design and analysis of a successful study has many stages, all of which need policing.

Extreme scrutiny

*P* value

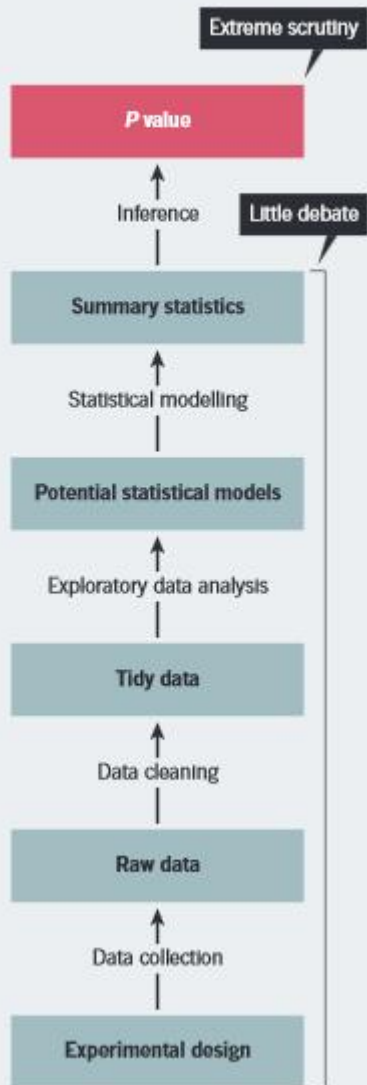
analysis is taught through an apprenticeship model, and different disciplines develop their own analysis subcultures. Decisions are based on cultural conventions in specific communities rather than on empirical evidence. For example, economists call data measured over time 'panel data', to which they frequently apply mixed-effects models. Biomedical scientists refer to the same type

Tieteen ykköslehtiin kuuluvasta Naturesta:

612 | NATURE | VOL 520 | 30 APRIL 2015

## DATA PIPELINE

The design and analysis of a successful study has many stages, all of which need policing.



Big data ?

Dataa on paljon....

Tietokantoja  $n+1$  kpl

(Esimerkki)

## Mitä on tilastotiede?

Tilastollinen päättely , *statistical inference*

“Kokoelma käsitteitä ja menetelmiä joiden on tarkoitus auttaa soveltajaa tekemään päätelmiä reaali maailman olosuhteista kun näitä olosuhteita ei havaita suoraan, vaan päätelmät pitää tehdä epävarmuutta sisältävien numeeristen havaintojen perusteella”

Matemaattinen päättely on *deduktiivista*.

Yleisistä säännöistä (aksiomeista) päätellään niiden seurauksia.

Tilastollinen päättely on *induktiivista*.

Yksittäisistä havainnoista pyritään kohti yleisiä sääntöjä.

Tilastollinen päättely on luonteeltaan *epävarmaa* ja sisältää aina virheellisen päättelyn mahdollisuuden. Epävarmuuden suuruutta on kuitenkin mahdollista kontrolloida ja arvioida.



Tilastollisessa päättelyssä käytetään matematiikkaa, erityisesti *todennäköisyyslaskentaa*.

Tilastollinen päättely on todennäköisyyslaskennalle *käänteinen ongelma*.

Todennäköisyyslaskenta tarjoaa työkaluja joilla voidaan laskea havaintojen jakauma ja niistä laskettujen tilastollisten tunnuslukujen jakauma – kun havaintoja generoiva todennäköisyysmalli on kiinnitetty.

Tilastollisessa päättelyssä pitää numeerisen aineiston perusteella yrittää arvioida, minkälainen todennäköisyysmalli olisi ne voinut generoida.

Havainnot vastaava todennäköisyysmalli

Tilastollisen päättelyn perusajatus on se, että havaittujen arvojen ajatellaan olevan *satunnaismuuttujien toteutuneita arvoja*.

Tilastollisen päättelyn tavoitteena on tehdä aineiston  $y$  perusteella johtopäätöksiä siitä todennäköisyysjakaumasta, jota satunnaisvektori  $Y$  noudattaa.

”Pallot kulhossa, nasta purkissa, binomikoe”

Parameterisessa mallissa havainnot vastaavan satunnaisvektorin  $Y$  jakauma tunnetaan täysin, jos mallin  $f(y; \theta)$  parametrin  $\theta$  arvo tunnetaan.

Tilastollisessa päättelyssä  $\theta$  on tuntematon luku:

Päättelyssä ensimmäinen vaihe on arvioida eli *estimoida* parametrin arvoa havaitun aineiston  $y$  perusteella sekä lisäksi yrittää kuvailla tähän arvoon liittyvää epävarmuutta.

Nyt siirrymme Tiltun kotisivulle,  
jonka kautta käymme perusteellisesti läpi  
kurssin suunnitellun toiminnan:

<http://wiki.helsinki.fi/display/mathstatKurssit/Tilastotiede+tutuksi%2C+syksy+2015>

Kyselkää paljon!

Kurssi voi vaikuttaa monimutkaiselta koska se sisältää vaihtoehtoisia toimintatapoja.

Luennolla esitetyistä kysymyksistä seuranneita muutoksia, esim. täsmennyksiä, löytyy kurssin kotisivulta merkittynä keltaisella tähdellä.