

Helsingin yliopisto
Matematiikan ja tilastotieteen laitos
Tilastotiede tutuksi
Timo Keränen

Essee: **Statistics - A very short introduction. Hand, D.**

Tämä on ensimmäinen esseeni Tilastotiede tutuksi kurssilla ja samalla ensimmäinen Matematiikan ja tilastotieteen laitoksella kirjoittamani. Käyn läpi kirjan seitsemän lukua.

Tilastojen keskellä

Moderni tilastotiede auttaa meitä näkemään mitä ei paljain silmin näe, samoinkuin teleskoopit, mikroskoopit, lääketieteen röntgen ja tutka. Kirjan tekijä ei mainitse että mittalaitteilta tuleva viesti voi olla jänniteviesti, virtaviesti tai analoginen viesti. Jos rikosmäärää kuvataan kaupoista näpistelyllä, niin poliisin voimien kohdistaminen näpistelyyn tietenkin vähentää näpistelyä, mutta kokonaistilanne saattaa huonontua (Goldhartin laki). Psykoanalyysi ja astrologia eivät ole tieteellisiä teorioita, niitä ei voida verifioida eikä falsifioida. Kiintoisena esimerkkinä tekijä käyttää roskapostin (spam) suodatusta. Esimerkkinä on myös opiskelijoiden tenttitulokset, joissa joku opiskelija on voinut jäädä saapumatta tenttiin.

Yksinkertaiset kuvaukset

Aritmeettinen keskiarvo ei aina ole hyvä mittari, koska jakauma voi olla epätasainen (esimerkkinä vaikkapa työntekijöiden palkat ja johtajien palkat). Moodi on tällaisissa tapauksessa parempi, ja tietenkin keskihajonta. Keskihajontaa laskiessa arvot koroitetaan toiseen potenssiin, jotta negatiiviset ja positiiviset arvot eivät kumoaisi toinen toisiaan. Tärkeää on myös vaihteluväli ja kvantiilit.

Hyvän datan kerääminen

Data voi olla hyvää tai enemmän tai vähemmän huonoa. Huonoutta aiheuttaa mm. puuttuvat tiedot tai näytteen harha (bias). Mielestäni bias on parempi sana kuin harha. Esimerkkinä mainitaan vuoden 1936 USAn presidentinvaalien kysely, joka lähetettiin vain niille joilla oli tuona aikana harvinaiset puhelimet ja auto. Toinen esimerkki on vuonna 1986 Challenger avaruussukkulan räjähtäminen, jossa koko miehistön sai surmansa, koska testeissä ei oltu otettu huomioon huonon sään vaikutusta. Jos datassa on selvä virhe, esimerkiksi oppilaiden ikä on: 20, 18, 25, 230, 22 vuotta, on helppo havaita että 230 vuotias on vika ja se voidaan poistaa datasta. Lääketieteessä käytetään usein kaksoissokkotutkimusta. Hyvänä esimerkkinä on myös maanviljelijä, joka

tilastomatematiikan avulla saa simuloitua kastelun ja lannoitteen määrän vaikutukset hyvällä koejärjestelyllä helposti. Keskeinen raja-arvolause ja suurten lukujen laki kertovat että suurempi näyte on parempi. Usein näytekoon kasvattaminen on kallista.

Todennäköisyys

Suurten lukujen lain perusteella mitattu suure lähestyy oikeaa arvoa kun näytekoko kasvaa. Toinen tärkeä laki on riippumattomuus, kolikon heitto on täydellisen riippumaton koska tuloksena on kruuna tai klaava. Ehdollinen todennäköisyys vastaa siihen miten käy jos esimerkiksi sataa lunta, ja ihminen liukastuu. Bernoullin jakauma on yksinkertainen, jos on tapahtumat p ja q , joiden yhteenlaskettu todennäköisyys on 1, niin $p=1-q$. Diskreetti muuttuja voi saada vain tiettyjä arvoja, jatkuvat muuttujat mitä tahansa arvoja tietyltä väliltä. Tunnetuin jakauma on Gaussin kellokäyrä eli normaalijakauma.

Estimointi ja interferenssi

Suurimman uskottavuuden menetelmä (maximum likelihood function) kuvaa havantodatan todennäköisyyttä eri lopputuksiin. Pienimmän neliösumman menetelmässä käytetään saatujen arvojen ja lasketun mallin neliöiden summan pienintä eli parasta arvoa.

Bayesin teoreemaa voi soveltaa tässäkin. Mikä estimaatti on paras: Jos tiedämme mikä menetelmä antaa tavallisesti tarkimman arvon tosiarvon suhteen, niin menetelmä on hyvä. Estimaatti itsekin voi olla jakauma! Ja usein emme tiedä mikä on tosiarvo. Pienin ja suurin mittatulos voidaan jättää pois ennen keskiarvon laskemista. Kirjassa käsitellään myös tilastollisen merkittävyyden ja hypoteesien testaamista. Jos saadaan muodostettua utiliteettifunktio, voidaan siitä johtaa päätössääntöjä.

Tilastolliset mallit ja menetelmät

Tilastolliset mallit voidaan jakaa empiirisiin ja mekaanisiin. Mekaanisten mallien takana on jokin teoria esimerkiksi siitä kuinka putoavan kappaleen nopeus lisääntyy tai kuinka lääkeaine jakautuu ihmisen kehossa. Empiiriset mallit soveltuvat esim mittausdatan mallittamiseen, kuten esim. talouden kehityksen seuraamiseen. Toinen tapa on jakaa mallit kuvaaviin (description) ja ennustaviin (exploration). Usein käytetään grafiikkaa, piirtoa kaksiulotteiseen koordinaatistoon, josta ihmisen on helppo nähdä riippuvuus.. Usein käytetään lineaarista korrelaatioanalyysiä, mutta on syytä selvittää onko korrelaatio todellista vai onko vain kaksi ryhmää, joiden välillä ei ole

yhtetyttä. Varianssianalyysi ja faktorianalyysi sekä aikasarjat voivat olla hyödyllisiä työkaluja oikein käytettyinä. Muita keinoja ovat puumallit (tree classifiers), naapurimallit (nearest neighbour) ja klusterianalyysi (cluster analysis).

Tilastot ja tietokoneet

Jos on käytössä vain yksi näyte, se voidaan kahteen alinäytteeseen: Toinen näyte voi olla harjoitus- tai suunnittelunäyte (training or design sample), ja toinen sitten suorituskyky- ja mallin valintanäyte (validation sample). Bioinformatiikkaa voidaan käyttää ihmisen kehon tai lääkkeiden tutkimiseen.

Helsingin yliopisto
Tilastotiede tutuksi

Timo Keränen

Referaatti artikkelista
The virtues of being blind by Sandy Zapell
Chance vol 29:1 2016

Artikkelissa kerrotaan kuinka tärkeitä sokko- ja kaksoissokkotutkimukset ovat:

Vuonna 1954 tehdyssä poliorokotustutkimuksessa annettiin osalle lapsia suola-liuosta rokotteen sijaan. Lapset, vanhemmat ja lääkärit eivät tieneet kuka sai rokotteen tai kuka plasebon.

Vuonna 1976 testattiin ranskalaista ja yhdysvaltalaisista viiniä sokkotutkimuksessa, ja ranskalaisten kauhuksi yhdysvaltalainen viini oli parempaa.

Vuonna 1774 Frank Mesmer tutki ja hoiti potilaita mesmerismin keinoin. Vuonna 1784 tehtiin sokkotutkimus jossa todettiin että mesmerismiä oli vain Frank Mesmerin mielessä. Vuonna 1903 Blondlot luuli löytäneensä uuden ilmiön, N-säteet, prisman avulla, mutta kun prisma poistettiin, Blondlot näki säteitä.

Statiineja käytetään kolestrolin hoitoon, mutta jotkut potilaat valittivat lihassärkyä ja muistihäiriöitä. Kun tehtiin sokkotutkimus 20000 potilaalla, havaittiin että oireita ei ollut statiinilääkkeitä saaneilla enempää kuin kontrollissa. Potilaiden oireet olivat kuitenkin tosia.

Vuonna 1994 multipeliskleroosia tutkittiin mutta sokkotutkimus paljasti että eroja ei ollut verrattuna placeboon.

Vuonna 1915 gyromagneettista suhdetta (kulmamomentin suhdetta hiukkasen massaun) tutkineet Albert Einstein ja de Haas erehtyivät luulemaan että ovat he olivat saaneet järkeviä tuloksia, mutta hiukkastasolla ei klassinen fysiikka ei toiminut. Oikea arvo oli kaksinkertainen.

Robert Millikan sai tutkimuksessaan liian pieniä arvoja elektronin massaun, mutta toisissa kokeissa saatuja suurempia arvoja pidettiin väärinä.

Rikostutkimuksessa sormenjälkien tutkiminen ei ole eksaktia. Vuonna 2004 Madridin terrorismiskussa juna-asemalla kuoli 191 henkeä ja haavoittui noin 2000. Espanjalaiset poliisit löysivät läheltä tapahtumapaikkaa nalleja, joista saatiin sormenjälki mutta sitä ei löytynyt Espanjan poliisien tietokannasta. FBI löysi tietokannastaan samankaltaisen sormenjäljen ”100 % todennäköisyydellä”

ja yksi henkilö pidätettiin. Myöhemmin FBI ja Espanjan poliisi vertasivat sormenjälkiä, FBI löysi 15 osakohtaa ja espanjalaiset vain 7. Lisätutkimuksessa sormenjälki todettiin kuuluvan algerialaiselle nimeltä Ouhnane Daod, ja Yhdysvalloissa pidätetty mies vapautettiin. FBI:n tutkijat eivät olleet käyttäneet sokkotestiä, vaan tutkijat olivat tienneet aikaisempien testin tulokset. Myöhemmin FBI alkoi toimia sitten että sormenjälkitutkijat eivät tienneet toistensa tuloksista

Helsingin yliopisto
Tilastotiede tutuksi

Timo Keränen

Referaatti artikkelista Actual risk prediction and the criminal justice system. Star, S., Chance 2016:1

Tämä on kolmas referaattini kurssilla Tilastotiede tutuksi.

Kirjoittaja pohtii seuraavaa: Kaksi miestä, Robert ja William, ovat varastaneet 10000 USD arvosta timantteja. Robert on myymäläpäällikkö, 35 vuotta vanha, koulutettu. Hän kasvoi keskiluokkaisessa vakaassa perheessä. Hän on kuukausien ajan vaihtanut timantteja väärennöksiin hankkiakseen uuden hienon auton. William on työtön, 18 vuotta vanha, hänen äitinsä on huumeiden käyttäjä, isä ja veli ovat vankilassa. Hän lopetti koulun kesken ja etsii töitä saadakseen rahaa uuteen kotiin. Hän ei löydä töitä. Kulkiessaan korukaupan ohi hän huomasi että yksi kaappi on auki ja nappasi itselleen muutaman kaulanauhan. Ostoskeskuksen vartija sai hänet heti kiinni.

Pitääkö Robert ja Williams tuomita samanlaiseen rangaistukseen? William on epätoivoinen nuori aikuinen, jolla on ollut vaikea lapsuus. Moraalisesti hän ansaitsee siis lyhyemmän tuomion. Toisaalta, Williamin elämässä on monta riskitekijää rikoksen uusintaan: köyhyys, kodittomuus ja työttömyys. Jos halutaan suojata yhteiskuntaa tulevilta rikoksista, pitäisikö hänet tuomita pitkäksi aikaa vankilaan.

Riskienhallinnassa otetaan huomioon sukupuoli, ikä, siviilisäätö, asunotilanne, suhde vanhempiin ja kesken jäänyt koulutus. Koitetaan laskea todennäköisyys sille että rikollinen uusii tekonsa..

Riskienhallinnassa käytetään AOC (Area under the Curve). Tämä menetelmä ennustaa tuloksen 75 % todennäköisyydellä, kun taas pelkkä satunnaisluku ennustaa 50 %.

Rotua ei saa käyttää riskienhallinnassa, vaikka Yhdysvalloissa lähes 10 % tummista alle 35 vuotiaista on vankilassa. Vangeille pitäisi järjestää työelämään sopeutumista ja mahdollisuus työpaikkaan.

Helsingin yliopisto
Tilastotiede tutuksi
Timo Keränen

Referaatti **Mandatory drug testing in Canadian Workplace.** Hurley, W.
Chance 2015:1

Tämä on neljäs kirjoittamani essee Tilastotiede tutuksi kurssilla.

Vuonna 2013 Kanadan korkein oikeus päätti että työntekijä voidaan määrätä huumeetesteihin vasta kun syytä epäillä. Urheilijoille on jo aikaisemmin ehdotettu huumeetesteihin.

Kirjoittaja keskittyy väriin positiivisiin (Non drug User).

Huumeetstit tehdään kahdessa vaiheessa: Ensin immunologinen testi virtsata. Testistä voidaan usein havaita myös huumeen metaboliitit. Positiiviset näytteet tutkitaan kaasukromatografilla ja massaspektrometrillä. Menetelmä on tarkka mutta kallis.

Ehkä työntekijät voitaisiin määrätä huumeetestein vain jos on syytä epäillä: outoa käytöstä tai löytyy huumeen käyttövälineitä. Oma mielipiteeni on että kuitenkin epäilystä voi herättää esimerkiksi vedenpainetauti (pyöräyttää jo nousee pystyyn nopeasti) tai sokeritauti (jos verensokeri heittelee liikaa).

Kirjoittaja on laskenut että jos 10000 testiä, saataisiin 950 positiivista tulosta, joista puolet olisi huumeita käyttämättömiä.

Kirjoittaja ei kerro mikä huume on haitallisimmin. Mielestäni voida tehdä lista

1. Vaarallisimmin PCP Fensyklidiini eli enkelipöly
2. Toiseksi vaarallinen. opiaatit, amfetamiini ja sen johdokset
3. Vähiten vaarallinen. Cannabis

Alkoholia on vaikea sijoittaa tähän luokitukseen, kohtuukäyttö ei ole haitallinen, liikakäyttö vaarallisempi kuin cannabis.

Psilosybiini (taikasienet) ja LSD voitaisiin sijoittaa luokkaan 3 tai 2, mutta aika vaarallisia ne ovat jos työntekijä alkaa nähdä hallusinaatioita työpaikalla.

Helsingin Yliopisto
Tilastotiede tutuksi
Timo Keränen

Referaatti artikkelista ”Flawed forensics: Statistical failing of microscopic hair analysis
Significance 2016:2

Tämä on viides esseeni Tilastotiede tutuksi -kurssilla. Ensimmäisen kerran rikostutkimuksessa käytettiin hiusnäytettä vuonna 1855 ja vuonna 1934 käytettiin ensimmäisen kerran mikroskooppia vertailussa rikospaikalta löydettyyn hiuksiin. FBI on käyttänyt hiustutkimusta osana työkalupakkia.

Mikroskoopilla nähdään hiusnäytteessä yli 20 eri piirrettä, kuten muoto, paksuus ja pigmentti. Jotkut näistä on subjektiivisia: onko hius kellertävä, vaalean ruskea vai ruskea, ohut, tavallinen tai paksu, tai pigmentin jakautuma. Uusi tutkimus voi tuottaa eri tuloksia, jopa saman tutkijan tekemänä. Ei myöskään ole koko väestön kattavaa tietokantaa.

Vuonna 1987 Jimmy Ray Bromgard tuomittiin raiskauksesta jota hän ei ollut tehnyt. Tytön kuvauksen perusteella piirrettiin kuva, joka muistutti Bromgardia. Kun Bromgard oli tunnistusjonossa, uhri arveli tunnistavansa 60-65 % todennäköisyydellä. Kun uhrilta pyydettiin sanomaan asia ilman numeroita hän sanoi että ei ole täysin varma. Syyttäjän rikostutkija todisti että rikospaikalta löydettyä hiusnäyte voi olla jonkin muu vain todennäköisyydellä 1: 10000. Bromgard tuomittiin vankilaan ja vasta 14 vuoden päästä DNA-testi osoitti että hän ei ollutkaan tekijä. Bromgard vapautettiin vuonna 2012. Perustettiin työryhmä tutkimaan uudestaan noin 3000 FBI:n juttua. Hiustutkimuksissa todettiin niin paljon puutteita että vuonna 1996 FBI alkoi käyttää mitokondria-DNA -analyysiä hiustutkimuksessa.

Liittovaltion ja osavaltioiden oikeusistuimet käyttävät menetelmää valitessaan Daubertin faktoreita: Onko menetelmä tai teoria testattu, onko sitä testattu *peer review* -menetelmällä tai tieteellisissä julkaisuissa, mikä on menetelmän virheen todennäköisyys ja onko sillä mahdollinen tiedeyhteisön hyväksyntä ja onko menetelmän standardi- ja kontrollinäyte kunnossa.

Nykyään käytetään DNA-tutkimusta, koska se on toistettavissa, laboratorio voidaan verifioida ja tietokanta on suuri.

Vuoden 1989 jälkeen on 74 vankia, joiden tuomioon liittyi hiustutkimus, vapautettu uuden DNA-tutkimuksen jälkeen. Näistä oli rikoksena murha tai tappo (62 %), raiskaus (37 %) tai ryöstö (1 %). Väärään tuomioihin oli syynä silminnäkijöiden väärä tunnistus (64 %) tai väärä tunnustus (36 %).

Miehiä oli 73 ja naisia vain 1.

Kirjoittaja ei kerro voidaanko afroamerikkalaisen hiusnäyte heti erottaa valkoihoisen hiusnäytteestä.

Helsingin yliopisto
Tilastotiede tutuksi
24.4.2016

Referaatti artikkelista

Deterrence and the death penalty- Why the statistics should be ignored.
by Daniel Nagin. *Significance*. Vol 11(2014)2

Tämä on kuudes esseeni Tilastotiede tutuksi kurssilla. Pitäisikö rikoksista, erityisesti henkirikoksista, tuomita kuolemanrangaistukseen. Toisaalta kuolemanrangaistus voi olla henkirikoksia estävä pelote, ja ainakaan kyseinen rikollinen enää rikoksia tee. Toisaalta kuolemanrangaistus on ihmisoikeuksia loukkaava, julma ja peruuttamaton rangaistus. Väärää positiivista tuomiota ei voida perua. Voisiko tilastotieteestä olla apua päätöksentekoon?

Jotkut Yhdysvalloissa tehdyt tutkimukset osoittavat että kuolemantuomiot vähentävät henkirikoksia, ja siten säästävät ihmishenkiä mutta jotkut että henkirikosten määrä kasvaa. Kuolemantuomio ei ole käytössä kaikissa osavaltioissa. Joissain tutkimuksissa on todettu että kuolemanrangaistuksella ei ole vaikutusta murhiin, mutta nollahypoteesin vääräksi todistamisen epäonnistuminen ei tarkoita että nollahypoteesi on oikein.

Paneelidatutkimuksessa tarkastellaan useamman maakunnan tai osavaltion tilannetta ajan funktiona, yleensä vuosina 1970-2000, käyttäen erilaisia regressioanalyysimenetelmiä. Toinen tapa on aikasarja-analyysi: siinä tarkastellaan vain yhtä maantieteellistä aluetta. Alue voi olla kaupunki tai koko valtio: onko kuolemantuomiolla lyhytaikaista pelotusvaikutusta. Molemmissa ongelma on, onko otettu huomioon muut tuomiot kuin kuolemanrangaistus ja murhaajien mielen toiminta.

Mahdolliset murhaajat ei tiedä millä todennäköisyydellä jäävät kiinni ja tulevat kuolemantuomituiksi. Heillä ei ole riittävästi dataa eikä kykyä analysoida dataa.

Lopputuloksena on, että tilastotiede ei tarjoa ratkaisuja kaikkiin kuolemantuomion ongelmiin

Kirjoittaja ei kerro kuolemantuomioista esimerkiksi Pohjois-Koreassa tai Kiinassa.

Suomessahan ei ole käytössä kuolemantuomiota. Mutta suuri osa Suomessa tapahtuneista henkirikoksista on sellaisia että Arska ja Pena ryypiskelevät, ja Pena ottaa liian pitkän huikan pullosta, ja Arska puukottaa Penan. Vasta seuraavana aamuna Arskalle selviää että tuli tapettua

ainoa ja paras ystävä. Tällöin tuomion ankaruus ei vaikuta.

Helsingin yliopisto
Tilastotiede tutuksi
Timo Keränen

Referaatti artikkelista The Erebus, the Terror and the North-West Passage. By Millar, K, Bowman, A., Battersby
Significance vol 11 (2014)2

Tämän on seitsemäs esseeni Tilastotiede tutuksi kurssilla.

Vuonna 1845 kaksi laivaa ja 129 miestä lähti tutkimaan magneettikentän vaihtelua pohjoisnavan läheisyydellä ja samalla etsimään merireittiä Atlantin ja Tyynen valtameren kautta. Laivat katosivat ja kaikki miehet kuolivat. Kummankaan laivan hylkyä ei ole löydetty. Oliko syynä huonojen säilykepurkkien aiheuttama lyijymyrkytys? Artikkelissa asiaa pyritään selvittämään tilastotieteen keinoin.

Laivat olivat kolmimastoiset Erebus ja Terror. Purjeiden lisäksi niissä oli apua höyrykoneet ja uudenlainen voimansiirto. Muonaa oli varattu kolmen vuoden ajaksi, säilykepurkkeja, tavanomaisten suolalihan sijasta. Molempien laivojen kapteeneilla oli kokemusta arktisista oloista.

Kahden vuoden kuluttua alkoivat etsinnät. Löydettiin tyhjiä säilykepurkkeja, nuotiopaikkoja ja rekien jälkiä. Jotkut seikat viittasivat kannibalismiin. Esiin tuli luurankoja, mutta ei tiedetty kenen, ja kolme jään säilömaa ruumista. Löydettiin lokikirja, jossa oli tieto että laivat olivat olleet jäissä kiinni puolitoista vuotta ja miehistö lähti jalkaisin etelään. Osa miehistä palasi laivaan, osa menehtyi. Osa miehistä kuoli uupumukseen, nälkään, keripukkiin ja tuberkuloosiin. C-vitamiinin puute aiheuttaa keripukkia.

Sata vuotta tämän jälkeen analyttinen kemia oli kehittynyt niin että nyt pystyttiin mittaamaan luiden lyijypitoisuus. Se oli korkea. Myös ruumiiden hiusten lyijypitoisuus mitattiin. Lyijy saattoi olla peräisin säilykepurkeista, mutta myös lyijyputkista joilla miehet olivat sulattaneet jäätä juomavedeksi. Laivan lääkevarastossa oli myös lyijyasetattia, jolla oli mahdollisesti käytetty ripuliin ja kylmänkyhmyihin.

Lyijy keräytyy elimistöön hitaasti, yleensä vuosien ajan. Kerääntyminen ja poistuminen ja myrkytyksen oireet riippuvat myös yksilöstä. Lyijy haittaa munuaisia, päätöksentekoa, muistia ja näppäryyttä ja isoilla annoksilla aiheuttaa päänsärkyä, väsymystä ja ummetusta. Stressi, C-vitamiinin ja muiden ravinteiden puutos lisäävät lyijymyrkytyksen oireita. Kolmessa ruumiissa ei ollut selviä lyijymyrkytyksen merkkejä (mm. ikenien värjäntymiä), vaikka lyijyä oli runsaasti.

Onkin mahdollista että nämä kolme kuolivat keuhkokuumeeseen tai tuberkuloosiin.

Vaikka lyijyä oli paljon, vaihteluväli ja hajonta olivat suuria. Pelkkä lyijymyrkytys ei selitä näiden urheiden miesten kuolemia, vaikka saattoi olla osatekijä. Paikan ilmasto ja maasto vaikuttivat, 1850-luvulla meri oli auki vain yhtenä vuotena viidestä.

Helsingin Yliopisto
Tilastotiede tutuksi
Timo Keränen
26.4.2016

Referaatti artikkelista ”**The Statistics of the rare earth industry**”
by Tyrer, M and Sykes, J. Significance vol 10 (2013)2

Tämä on kahdeksas esseeni Tilastotiede tutuksi -kurssilla.

Harvinaiset maametallit ovat: lanthanum, cerium, praseodymium, neodymium, promethium, samarium, europium, gadolinium, terbium, dysprosium, holmium, erbium, thulium, ytterbium ja lutetium sekä scandium ja yttrium. Prometium-metallia ei esiinny luonnossa, sitä on muodostettu ydinreaktiossa. Nimestään huolimatta ne eivät ole kovin harvinaisia. Niitä käytetään elektroniikassa ja voimakkaissa magneeteissa. Voimakkaita magneetteja käytetään tuulienergian tuotossa. Auton katalysaattorissa on platinan lisäksi cerumia ja rodiumia. Niitä käytetään myös hybridi-autoissa. Lasiteollisuus käyttää niitä väriaineina. Jotkut niistä fluoresoivat eli fotonin saatuaan lähettävät uuden fotonin ja siten näkyvät väreinä eli eri aallonpitoisuuksilla esimerkiksi LCD- ja plasmatelevisioissa. Lanthanumia käytetään nikkelimetallihybridiakuiissa. Yhdysvaltojen asevoimat käyttävät niitä älypommeissa, lennokeissa (drone) ja ohjuksissa (Tomahawk). Melkein 97 % harvinaisista maametalleista tuotetaan Kiinassa. Osa näistä menee salakuljetukseen tai varastoidaan, lopusta 35 % (hinnan mukaan) menee Japaniin, 25 % Yhdysvaltoihin ja noin 30 % Eurooppaan. Eräiden harvinaisten maametallien hinta on noussut vuodessa 200-kertaisesti.

Tilastotieteilijät ovat yrittäneet ennustaa harvinaisten maametallien hintaa: Kaivostoiminnassa nykyarvo kymmenen vuoden sisällä on viidestä neljäänkymmeneen prosenttia. Maassa olevalla malmilla ei ole paljoakaan arvoa, rikastus ja puhdistus maksavat. Yhdysvallat, Euroopan unioni ja Japani väittävät Kiinan rajoittavan harvinaisten maametallien vientiä ulkomaille, mutta Kiina puolustautuu sillä väitteellä että sillä on oikeus rajoittaa vientiä rajallisten resurssien ja ympäristösyiden takia. Maailman kauppajärjestö (WTO) ei ole vielä antanut ratkaisuaan kiistassa. Jos Kiina voittaa oikeusjutun, ehkä harvinaisia maametalleja olisi liiketaloudellisesti kannattavaa louhia muuallakin: satoja kaivosyrityksiä on perustettu mm. Yhdysvaltoihin, Kanadaan ja Australiaan.

Harvinaisia maametalleja voitaisiin ottaa talteen teollisuuden sivuvirroista ja jätteistä.

Suomessa on löydetty harvinaisia maametalleja sisältäviä mineraaleja, mutta kaivostoimintaa ei

vielä ole.

Helsingin Yliopisto
Tilastotiede tutuksi
Timo Keränen
27.4.2016

Referaatti artikkelista ”**To boldly and safely go: Biostatistics in space**
Significance vol 9(2112)1 by Robert Ploutz-Snyder

Tämä on yhdeksäs esseeni Tilastotiede tutuksi -kurssilla.

Pidän suuresti tieteiskirjallisuudesta, paras tv-sarja on tietenkin Star Trek.

Mutta mitä muita avaruusmatkailussa tarvitaan? Tietenkin tilastotieteilijöitä!

Pöly Maapallolla on suhteellisen harmitonta. Avaruusmatkoilla se kuitenkin haittaa terveyttä ja turvallisuutta. Avaruuspölyssä on usein rautaa ja ultraviolettisäteilyn aiheuttamia muutoksia. Pölyhiukkaset ovat teräviä. Pölyllä on usein sähköinen lataus, mikä aiheuttaa sen kiinnittymisen avaruuspukuihin ja laitteisiin. Maasta avaruuteen tullessa miehistössä on mikro-organismeja, bakteereja ja viruksia, joiden reaktiota suljetussa tilassa alhaisessa painovoimakentässä ei tunneta. Maan magneettikenttä suojaa meitä säteilyltä, mutta säteily avaruudessa voi aiheuttaa säteily sairautta, syöpää ja muutoksia soluissa ja geneeissä. Astronauttien luut voivat heiketä, ja voi tulla myös muita luihin liittyviä sairauksia, kuten munuaiskiviä. Lihasvoima heikkenee. Sydän ja verenkierto voi heiketä: kokonaisveren määrä, nesteet ja alhainen verenpaine voivat heikentää miehistön toimintakykyä hätätilanteissa. Aistitoiminta voi muuttua, kuten liikkeiden kontrolli, näköhäiriöt ja häiriöt kävelyssä. Immunologian muutoksia ei tunneta tarpeeksi, avaruusmiesten vastustuskyky bakteereille ja viruksille voi heiketä.

Kansainvälisellä avaruusasemalla (ISS) on yleensä 3-6 henkilöä, kuuden kuukauden rotaatiolla ja aikataulu on hyvin tiukka, mikä vaikeuttaa biotilastollista tutkimusta. Ja kun datan keruu on valmis, datassa on puuttuvia arvoja, laajoja kliinisiä tutkimuksia on hankala tehdä maassa olevan kontrollijoukon kanssa. Kaikkea dataa ei tarvitse kerätä avaruudessa, maassa voidaan tehdä simulointeja esimerkiksi pitämällä koehenkilöitä 90 päivää sairaalasängyssä mahdollisimman liikkumatta. Sängyn jalkopäätä voidaan nostaa nesteiden kerääntyminen mallittamiseen. Myös psykologia vaivoja voidaan tutkia esimerkiksi tarkkailemalla Etelämantereen talven ajan eristyksissä olevaa tutkijoita. Veden alla kelluminen on todettu hyväksi harjoitukseksi. NASA on rakentanut 20 metriä merenpinnan alle kopion avaruusasemasta, jossa astronautit voivat harjoitella avaruuskävelyä. Siellä testataan myös avaruuspukuja, kuten myös Arizonan autiomaassa Desert RATS (Research and Technology Studies) -ohjelmassa.

Vaikka NASAlla on käynyt onnettomuuksia (Challenger, Apollo 1, Apollo 13), turvallisuuteen kiinnitetään huomiota. Ja tähänkin tarvitaan tilastotieteilijöitä.

Helsingin yliopisto

Tilastotiede tutuksi

28.4.2016

Tämä on kymmenes esseeni Tilastotiede tutuksi -kurssilla

Referaatti artikkelista:

Chance at the Bar - A Statistical Portrait of the Handgun Industry

By Finkelstein, M. and Levin, B.

Chance vol. 14(2003)4

Aseet tappavat ihmisiä? Ei, aseet eivät tapa, ihmiset tappavat ihmisiä. Aseenkäsitteilyn kolme perussääntöä ovat: Kohtele kaikkia aseita kuin ne olisivat ladattuja. Älä osoita aseella mitään mitä et ole valmis tuhoamaan. Älä laita sormeasi liipaisimelle ennen kuin olet tähdännyt. Älä jätä aseita lasten saataville, aseet eivät ole leikkikaluja.

Artikkelissa selvitetään kuinka aseet päätyvät rikollisille. New Yorkin tuomioistuin on todennut että asia on monimutkainen: Syytettyjen (plaintiffs), määrä on suuri, uhrien määrä on suuri, aseitehtaiden ja rikosten välinen yhteys on kaukainen, aseet päätyvät liittovaltion hyväksymille tukkukauppiaille ennen kuin päätyvät kauppiaille. Markkinointi ei estä aseiden väärinkäyttöä.

Tuomari Einstein määräsi BAFT:n (Firearm Bureau of Alcohol, Tobacco and Firearms) tutkimaan miten aseet ovat päätyneet rikollisille. Kun saadaan selville, mitä asetta on käytetty, BATF pyytää valmistajaa selvittämään sarjanumeron perusteella mille tukkurille ase on myyty, ja edelleen vähittäiskauppiaille. Kuitenkin Kongressi katkaisi projektin rahoituksen. BAFTin tutkimusta on kritisoitu sillä että se ei ole edustava satunnaisnäyte

Ekonomisti Lucy Allen arveli vuonna 1990 että rikolliset ovat kolmanneksi suurin ryhmä, jolla on aseita, kaksi suurinta ovat kansalaiset, joilla on henkilökohtaisia suojeleuseita ja tai ampuma-urheiluaseita. On arveltu että laillisista käsiaseista joutuu rikollisten käsiin 13 % yhden vuoden aikana ja 30 % kolmen vuoden aikana. Joissain osavaltioissa käsiaseita saa ostettua helpommin.

Afroamerikkalaisia on noin 12 % Yhdysvaltojen väestöstä, mutta he tekivät puolet Yhdysvaltojen murhista. Mustat ovat köyhiä ja elävät huonoissa oloissa, slummeissa, olisi kiintoisaa selvittää millainen tilanne olisi sosioekonomisesti vakioituna.

On ehdotettu että, asemessuilla ei saisi myydä aseita, asekauppiaille tulee olla liikehuoneisto, aseita saisi myydä samalle asiakkaalle vain yhden ase kuukaudessa, kauppiaan täytyy pitää kirjaa kaupoista ja osallistua turvallisuuskoulutukseen. Kukaan ei tiedä miten nämä vaikuttaisivat, tarvitaan tilastotieteilijöitä ja tilastollista dataa. Yksi osavaltio, Virginia, määräsi vuonna 1993 että yksi asiakas saa ostaa vain yhden ase kuukaudessa, mikä laski aserikollisuutta.

Suomessa tavallisilla kansalaisella on hyvin pienet mahdollisuudet saada luvallisia suojeleuseita. Suomessa suurin osa aseista on metsästyskiväärejä ja rata- tai metsästyshaulikkoja.

Helsingin yliopisto
Tilastotiede tutuksi
Timo Keränen

Tämän on yhdestoista esseeni Tilastotiede tutuksi -kurssilla.

Referaatti artikkelista:
Visual Revelations. Howard Wainer & Paul Velleman
CHANCE Vol 21 (2008)4

Diabetes on lisääntynyt eksponentiaalisesti Yhdysvalloissa. Syynä saattaa olla lihavuus, liikunnan väheneminen, prosessoidun ruuan lisääntyminen ja väestön ikääntyminen. Diabetes voi aiheuttaa munuaisen ja maksan vaurioita, ja verenkierron häiriöitä jotka voivat johtaa sokeuteen ja raajojen menettämiseen.

Lääkäri voi määrätä lääkkeitä ja neuvoa potilasta laihduttamaan ja liikkumaan enemmän, mutta vain potilas voi toteuttaa nämä muunnokset.

Kirjoittaja esittää että verensokerin keskiarvo voi vaihdella, ja mediaani olisi parempi mittari. Keskiarvohan kertoo ”Tämä on tyypillinen arvo, muut arvot ovat lähellä tätä” jos kyseessä on normaalijakautuma. Verensokeri kuitenkin kannattaa esittää graafisesti ajan funktiona. Vielä parempi olisi liukuva keskiarvo. Datasta voidaan laskea residuaalit. Graafisesta esityksestä potilas näkee selvästi poikkeavat arvot ja voi pohtia niiden syitä, tuliko syötyä liikaa jäätelöä tai juotua liikaa olutta.

John Tukey on kehittänyt 53h twice -säännön: ensin otetaan datasta liukuva mediaani joka viidennestä ja sitten joka kolmas ja ”hoonataan”:

$$x(i) = x(i-1) + 2x(i) + x(i+1)/4$$

Kirjoittaja ei erota ykkös ja kakkos-tyypin diabetesta. Hän käyttää Yhdysvaltalaisia yksikköjä.

Oma kuva:



Helsingin yliopisto

Tilastotiede tutuksi

Timo Keränen

2.5.2016

Referaatti julkaisusta: Visual Revelations: Defeating Deception: Escaping the Shackles of Truthiness by learning to think like a Data Scientist. Wainer, H. Chance 29(2016)1

Tämä on kahdestoista esseeni Tilastotiede tutuksi -kurssilla.

Kirjoittaja käsittelee tilastotieteen tarpeellisuutta esimerkkien voimalla. Vuonna 2012 Yhdysvaltain presidenttivaaleissa toiset ennustivat tulosta ”asiantuntijoin lausunnoin” ja huonolaatuisen tv-kanavan perusteella. Uskoivat he itse omaan ennusteensa vai teeskentelivätkö saadakseen palkkansa.

Renee Haynes on esittänyt hyvän käsitteen, *boggle threshold*: Mieli epäröi kun tulee uusia ideoita. Stanfordin yliopiston antropologi Tanya Luhrmann on esittänyt sen näin: Jumala, jolla oli oma poika ja jonka Jumala antoi tapettavaksi, on mielestämme aivan uskottava, kun taas jumala jolla on kahdeksan kättä ja valtava seksinälkä on outo juttu.

Tiede voi käsitellä vain asioita joita voidaan havainnoida ja ilmiöitä jotka ovat toistettavissa.

Sikiölle puhuminen voi aluksi tuntua hyvältä idealta, kyllähän ”maha-asukki” on osa perhettä, mutta tarkemmin ajatellen, tieteellisen tutkimuksen nimessä, olisi mietittävä vaikuttaako se millä kielellä, sanavarastolla ja kieliopilla puhutaan. Ja miten mitataan tulos? Onko lapsi pitkä aikuisena, oppiiko hän nopeammin uuden kielen tai onko hän onnellisempi. Viiden, kymmenen tai kahdenkymmenen vuoden päästä.

Entä onko esikoulu (kindergarden) hyödyllinen. Millä hyödyllisyys mitataan, oppiiko lapsi lukemaan tai laskemaan myöhemmin.

Miesten ympärileikkaus saattaa vähentää naisten kohdunkaulan syöpäriskiä – vai vähentääkö. Tähän yleiseen väitteeseen olisi tarvittu puoli miljoonaa naista sekä testi- ja kontrolliryhmään.

Entä kuinka laskea lammessa olevat kalat? Pitäisikö ne kaikki nostaa verkolla ylös ja laskea. Vai

voitaisiinko pyydystää 100 kalaa, merkitä ne, ja viikon päästä pyydystää 100 kalaa. Jos näissä sadassa kalassa on kymmenessä merkki, merkitseekö tämä että lammessa on $10 \cdot 100 = 1000$ kalaa.