
Johdatus tilastotieteeseen
Tilastotiede tieteenalana

Tilastotiede tieteenalana

Mitä tilastotiede on?

Tilastotieteen sovellukset

Tilastotiede tieteenalana: Mitä opimme?

- Pyrimme tässä luvussa vastaamaan seuraaviin kysymyksiin:
 - **Mitä tilastotiede on ja mitä se ei ole?**
 - **Mihin tilastotiedettä käytetään?**
- Saamme tietää, että tilastotiede on **yleinen menetelmätiede**, jota voidaan soveltaa aina, kun *reaalimaailman ilmiöistä halutaan tehdä johtopäätöksiä ilmiöitä kuvaavien kvantitatiivisten tietojen perusteella* sellaisissa tilanteissa, joissa *tietoihin liittyy epävarmuutta*.
- Näemme myös, että tilastotiede pyrkii muodostamaan matemaattisia, **todennäköisyyslaskentaan** perustuvia **malleja** niille *prosesseille*, jotka *generoivat tiedot*.

Tilastotiede tieteenalana

>> **Mitä tilastotiede on?**

Tilastotieteen sovellukset

Mitä tilastotiede on?

Avainsanat

Arvonta

Epävarmuus

Johtopäätösten tekeminen

Kuvaileva tilastotiede

Kvantitatiivinen tieto

Matemaattinen malli

Matematiikka

Menetelmätiede

Numeerinen tieto

Reaalimaailman ilmiö

Satunnaisilmiö

Satunnaisuus

Soveltava tilastotiede

Teoreettinen tilastotiede

Tieto

Tilasto

Tilastotiede

Tilastollinen päättely

Tilastollinen stabiliteetti

Tilastollinen tutkimusasetelma

Tilastollinen malli

Tilastollinen menetelmä

Todennäköisyyslaskenta

Tulosvaihtoehto

Tunnusluku

Mitä tilastotiede on?

Tilastotiede ei ole oppi tilastoista!

- Tilastotiede **ei ole**
 - nimestään huolimatta –**oppi tilastoista tai tilastojen tuotannosta!**
- Mikä sen sijaan on totta, on se, että *tilastojen tuotannon, jalostuksen ja analysoinnin menetelmien kehittäminen muodostaa keskeisen osan tilastotiedettä.*

Mitä tilastotiede on?

Tilastotiede ei ole matematiikkaa!

- Tilastotiede **ei ole matematiikan osa-alue!**
- Mikä sen sijaan on totta, on se, että *tilastotieteen menetelmät ja mallit ovat matemaattisia ja perustuvat todennäköisyyslaskentaan:*
 - Matematiikalla on tilastotieteessä *välineellinen rooli.*
 - Tilastotiede käyttää *matematiikan kieltä.*

Mitä tilastotiede on?

Tilastotiede on yleinen menetelmätiede

- Tilastotiede on **yleinen menetelmätiede**.
- Tilastotiede kehittää ja soveltaa menetelmiä ja malleja, joiden avulla *reaalimaailman ilmiöistä* voidaan tehdä **johtopäätöksiä** ilmiöitä kuvaavien **numeeristen** tai **kvantitatiivisten tietojen** perusteella tilanteissa, joissa tietoihin liittyy **epävarmuutta** ja **satunnaisuutta**.

Tilastolliset menetelmät ja mallit

- **Tilastollisten menetelmien** avulla reaali maailman ilmiöitä kuvaavat numeeriset tai kvantitatiiviset *tiedot jalostetaan* sellaiseen muotoon, että ilmiöitä koskevat *johtopäätökset tulevat mahdollisiksi*.
- Tietojen jalostaminen merkitsee *tietojen tiivistämistä graafiseksi esityksiksi ja tunnusluvuiksi* sekä **tilastollisten mallien** rakentamista *tiedot generoineille prosesseille tai mekanismeille*.

Tilastolliset tutkimusasetelmat

- **Tilastollisissa tutkimusasetelmissa** reaali maailman ilmiötä kuvaaviin numeerisiin tai kvantitatiivisiin tietoihin liittyy aina *epävarmuutta* ja *satunnaisuutta*.
- Reaali maailman ilmiötä kuvaavien tietojen **tilastollinen analyysi** perustuu siihen, että tietoihin liittyvän epävarmuuden ja satunnaisuuden ajatellaan johtuvan *tiedot generoineesta prosessista* tai *mekanismista*.
- Epävarmuuden ja satunnaisuuden generoijana voi olla *ilmiö itse* tai ne voivat olla seurausta *menetelmästä, jolla tutkimuksen kohteet valitaan*.

Satunnaisilmiöt

- Reaalimaailman ilmiö on **satunnaisilmiö**, jos seuraavat ehdot pätevät:
 - (i) Ilmiöllä on useita erilaisia **tulosvaihtoehtoja**.
 - (ii) **Sattuma** määrää mikä tulosvaihtoehtoista toteutuu.
 - (iii) Vaikka *ilmiön tulos vaihtelee* ilmiön toistuessa *satunnaisesti*, ilmiön tulosvaihtoehtojen *suhteellisten osuuksien jakauma käyttäytyy tilastollisesti stabiilisti*, kun ilmiön toistokertojen lukumäärä kasvaa.
- **Todennäköisyyslaskennan** tehtävänä on tuottaa *matemaattisia malleja* satunnaisilmiöissä havaittavalle *tilastolliselle stabiliteetille*.

Mitä tilastotiede on?

Satunnaisilmiöt:

Kommentteja

- Satunnaisilmiöihin liittyy aina *ennustamattomuutta*:
Satunnaisilmiön yksittäistä tulosta ei voida tietää etukäteen.
- Satunnaisilmiöihin on kuitenkin liityttävä *säännönmukaisuutta*, jonka on tultava esille ilmiön toistuessa:
Vaikka satunnaisilmiön tulos vaihtelee satunnaisesti ilmiön toistokerrasta toiseen, ilmiön tulosvaihtoehtojen suhteellisten osuuksien jakauman on käyttäydyttävä stabiilisti, kun toistokertojen lukumäärä kasvaa.

Mitä tilastotiede on?

Satunnaisilmiöt: Esimerkkejä

- Esimerkkejä satunnaisilmiöistä:
 - Kvanttimekaniikan ilmiöt
 - Hiukkasfysiikan ilmiöt
 - Luonnontieteellisiin mittauksiin liittyvien mittausvirheiden syntymekanismit
 - Uhkapelit: *arpajaiset*, lotto, ruletti, kortti- ja noppapelit
 - Perinnöllisyys
 - Eliöiden ja eliöpopulaatioiden käyttäytyminen
 - Ihmisten, ihmisryhmien ja ihmisten muodostamien organisaatioiden sosiaalinen ja taloudellinen käyttäytyminen
 - Teknisten prosessien tuloksien ominaisuudet

Tietojen kerääminen satunnaisilmiönä 1/2

- Voimme ajatella, että tilastollisissa tutkimusasetelmissä tutkimuksen kohteet valitaan **arpomalla**.
- Arvonta on *satunnaisilmiö*:
 - (i) Arvontaan liittyy aina *ennustamattomuutta*, koska *yksittäisen arvonnän tulosta ei voida tietää etukäteen*.
 - (ii) Arvonta noudattaa kuitenkin *todennäköisyyden lakeja*.

Tietojen kerääminen satunnaisilmiönä 2/2

- Koska arvonnän tulos vaihtelee satunnaisesti arvontakerrasta toiseen, myös *tutkimuksen kohteita kuvaavat tiedot vaihtelevat satunnaisesti arvontakerrasta toiseen*.
- Tutkimuksen kohteita kuvaavien tietojen käyttäytymisessä havaitaan kuitenkin arvontaa toistettaessa sitä *säännönmukaisuutta*, jota kutsutaan **tilastolliseksi stabiliteetiksi**.
- Juuri tämä säännönmukaisuus on tilastollisen tutkimuksen kohde.

Mitä tilastotiede on?

Tietojen kerääminen satunnaisilmiönä: Esimerkkejä

- Esimerkkejä tietojen keräämisen menetelmistä, jotka perustuvat arvontaan:
 - **Satunnaistetut kokeet**
 - **Satunnaisotanta**
- Huomautus:

Koesuunnittelu ja otantateoria ovat keskeisiä tilastotieteen menetelmiä.

Teoreettinen ja soveltava tilastotiede 1/2

- **Teoreettinen tilastotiede** kehittää *matemaattisia malleja* prosesseille, jotka *generoivat* reaalimaailman ilmiöitä kuvaavia numeerisia tai kvantitatiivisia *tietoja*, joihin liittyy *epävarmuutta ja satunnaisuutta*.
- Teoreettisen tilastotieteen kehittämät mallit perustuvat *todennäköisyyslaskentaan* ja niitä kutsutaan **tilastollisiksi malleiksi, stokastisiksi malleiksi** tai **todennäköisyysmalleiksi**.
- Tilastollisten mallien avulla reaalimaailman ilmiöitä kuvaaviin tietoihin liittyvät *systemaattiset ja satunnaiset piirteet* voidaan *erottaa ja kuvata*.

Teoreettinen ja soveltava tilastotiede 2/2

- **Soveltava tilastotiede** soveltaa teoreettisen tilastotieteen kehittämiä matemaattisia malleja reaalimaailman ilmiöitä kuvaavien numeeristen tai kvantitatiivisten *tietojen analysointiin*.
- Teoreettinen ja soveltava tilastotiede kulkevat tilastollisessa tutkimuksessa käsi kädessä:
 - Teoreettinen tilastotiede *kehittää tilastomatemattisia malleja* soveltavan tilastotieteen empiiristen ongelmien ratkaisemiseksi.
 - Soveltava tilastotiede *käyttää hyväkseen teoreettisen tilastotieteen kehittämiä malleja*.

Kuvaileva tilastotiede ja tilastollinen päättely 1/2

- **Deskriptiivinen eli kuvaileva tilastotiede** kehittää ja soveltaa menetelmiä, joiden avulla tutkimuksen kohteena olevasta ilmiöstä kerättyjä numeerisia tai kvantitatiivisia tietoja voidaan *kuvailla* ja *esitellä*.
- Kuvailevan tilastotieteen työkaluja:
 - **Tilastografiikka**
 - **Tilastolliset tunnusluvut**
 - **Tilastolliset mallit**

Kuvaileva tilastotiede ja tilastollinen päättely 2/2

- **Tilastollinen inferenssi** eli **päättely** kehittää ja soveltaa menetelmiä, joiden avulla tutkimuksen kohteena olevasta ilmiöstä voidaan *tehdä johtopäätöksiä* ilmiöstä kerättyjen numeeristen tai kvantitatiivisten tietojen perusteella.
- Tilastollisen päättelyn työkaluja:
 - **Tilastolliset mallit**
 - **Tilastollinen testaus**
- Kuvaileva tilastotiede ja tilastollinen päättely kulkevat tilastollisessa tutkimuksessa käsi kädessä.

Tilastotiede tieteenalana

Mitä tilastotiede on?

>> Tilastotieteen sovellukset

Tilastotieteen sovellukset

Avainsanat

Empiirinen tutkimus

Havaintoaineisto

Laadunvalvonta

Lääketieteellinen koe

Koe

Kyselytutkimus

Menetelmätiede

Otanta

Päätöksenteko

Tilasto

Tilastoala

Tilastollinen aineisto

Tilastollinen tutkimus

Tilastollinen tutkimusasetelma

Tilastotiede

Tilastotieteen osa-alueet

Tilastotieteen reuna-alueet

Tilastotoimi

Tutkimus

Tutkimusaineisto

Yhteiskunta

Missä tilastotiedettä voidaan soveltaa?

- Tilastotiedettä voidaan **yleisenä menetelmätieteenä** soveltaa – ja myös pitäisi soveltaa – kaikkialla, missä *tuotetaan* reaali maailmaa ja sen ilmiöitä kuvaavaa *numeerista* tai *kvantitatiivista tietoa*.
- Tilastollisia menetelmiä voidaan soveltaa tietojen *keruun*, *jalostuksen* ja *analysoinnin* jokaisessa vaiheessa.
- Tilastollisia menetelmiä sovellettaessa päämääränä on jalostaa tiedot muotoon, joka *mahdollistaa* reaali maailmaa ja sen ilmiöitä koskevien *johtopäätösten tekemisen*.

Tilastotiede ja tieteellinen tutkimus

- Tilastotiedettä voidaan *yleisenä menetelmätieteenä* soveltaa kaikissa tieteissä, joiden **tutkimusaineistot** voidaan esittää *numeerisessa* tai *kvantitatiivisessa* muodossa.
- Jokainen tiede, jonka tutkimusaineistot voidaan esittää numeerisessa tai kvantitatiivisessa muodossa *voi soveltaa / voisi soveltaa / pitäisi soveltaa* tilastollisia menetelmiä sekä tutkimusaineistoja *kerättyinä* että niitä *analysoitaessa*.
- Jokainen **empiirisen tutkimuksen havaintoaineisto** on tilastollisen tutkimuksen mahdollinen kohde.

Tilastotieteen käyttöalueita

- **Biotieteet**

- biokemia
- biologia
- ekologia
- eläinlääketiede
- eläintiede
- kasvitiede
- lääketiede
- perinnöllisyystiede

- **Ihmistieteet**

- arkeologia
- kielitiede
- psykologia

- **Luonnontieteet**

- fysiikka
- kemia
- tähtitiede

- **Maatalous- ja metsätieteet**

- kasvinviljelytiede
- kotieläinten jalostustiede
- metsänarviointitiede
- metsänviljelytiede

- **Yhteiskuntatieteet**

- sosiaalitieteet
- taloustiede

Tilastotieteellä on monta nimeä

- **Biometria** tai **Biostatistiikka**
 - = Bio- ja lääketieteiden tilastotiede
- **Demometria**
 - = Väestötiede
- **Ekonometria**
 - = Taloustieteen tilastotiede
- **Epidemiologia**
 - = Tautien leviämismekanismia koskeva lääketieteen osa-alue
- **Kemometria**
 - = Kemian tilastotiede

Tilastotieteen osa-alueita

- **Aikasarja-analyysi**
- **Bayeslaiset menetelmät**
- **Biometria**
- **Demometria**
- **Ei-parametriset menetelmät**
- **Ekonometria**
- **Estimointiteoria**
- **Kemometria**
- **Koesuunnittelu**
- **Laadunvalvonta**
- **Lineaaristen mallien teoria**
- **Matemaattinen tilastotiede**
- **Monimuuttujamenetelmät**
- **Otantateoria**
- **Regressioanalyysi**
- **Robustit menetelmät**
- **Spatiaaliset menetelmät**
- **Testiteoria**
- **Tilastollinen päättely**
- **Tilastollinen tietojenkäsittely**
- **Varianssianalyysi**

Tilastotieteen reuna-alueita

- Finanssimatematiikka
 - Hahmontunnistus
 - Hermoverkot
 - Kaaosteoria
 - Katastrofiteoria
 - Kuvankäsittely
 - Kybernetiikka
 - Operaatioanalyysi
- Peliteoria
 - Päätösteoria
 - Riskiteoria
 - Signaalinkäsittely
 - Stokastiset prosessit
 - Todennäköisyyslaskenta
 - Tulevaisuudentutkimus
 - Vakuutusmatematiikka

Tilastotieteen sovelluksia teknisissä tieteissä

- Hahmontunnistus
 - Kalibrointi
 - Koesuunnittelu
 - Kuvankäsittely
 - Laadunvalvonta
 - Laskennallinen tekniikka
 - Lääketieteellinen tekniikka
- Neuroverkot
 - Päätöksentekomenetelmät
 - Prosessinvalvonta
 - Signaalinkäsittely
 - Spektroskopia
 - Tietoliikennetekniikka

Tilastotieteen eksoottisia sovelluksia 1: Dendrokronologia

Dendrokronologia

- *Arkeologiassa* puuesineiden ajoituksessa käytetään apuna mm. puiden vuosilustojen muodostamia (aika-) sarjoja.
- *Historiallisessa meteorologiassa* ilmastonmuutoksien tutkimuksessa käytetään apuna mm. puiden vuosilustojen muodostamia (aika-) sarjoja.
- Puiden vuosilustosarjojen analysoinnissa sovelletaan mm. tilastollista *aikasarja-analyysia*.

Tilastotieteen eksoottisia sovelluksia 2: Tietokonetomografia

Tietokonetomografia

- *Lääketieteellisissä tutkimuksissa* käytetään (esim. syöpäkasvaimia etsittäessä) apuna *tietokonetomografiaa*.
- *Tietokonetomografia* on menetelmä, jonka avulla ihmisen kudoksista tai elimistä tuotetaan *tomografi*-nimisellä laitteella ns. viipale- tai tasokuvia.
- Kuvat perustuvat sähkömagneettisen tai hiukkassäteilyn *mittaamiseen* säteilyn kulkiessa kudosten tai elinten läpi.
- Kuvaa muodostettaessa tomografiin ohjelmoitu algoritmi ratkaisee *inversio-ongelmaksi* kutsutun matemaattisen ongelman, joka voidaan luontevimmin tulkita *bayeslaisten tilastollisten menetelmien* muodostamassa kehikossa.

Tilastot ja tilastolliset aineistot 1/2

- Sana **tilasto** tuo useimmille ensimmäisenä mieleen *yhteiskuntaa ja sen toimintaa* kuvaavat *numeeristen tietojen järjestelmälliset kokoelmat*.
- Yhteiskuntaa ja sen toimintaa kuvaavien *tilastojen tuotannossa ja analysoinnissa tarvittavien menetelmien kehittäminen* on keskeinen osa tilastotiedettä, mutta tilastotieteen sovellusalue on paljon tätä laajempi.

Tilastot ja tilastolliset aineistot 2/2

- Tilastotieteen kannalta *mikä tahansa* reaali maailman ilmiötä kuvaava *numeeristen tai kvantitatiivisten tietojen järjestelmällinen kokoelma* muodostaa **tilastollisen aineiston** ja **tilastollisen tutkimuksen** mahdollisen kohteen.

Esimerkiksi *kaikki empiirisen tai kvantitatiivisen tutkimuksen tutkimus-* tai **havaintoaineistot** ovat tilastotieteen kannalta *tilastollisia aineistoja*.

Tilastoala, tilastotiede, tilastotoimi

- Terminologiaa:

Tilastoala = Tilastotiede +
Tilastotoimi

Tilastotiede = Teoreettinen tilastotiede +
Soveltava tilastotiede

Tilastotoimi = Tilastojen tuotanto +
Tilastojen hyödyntäminen

Tilastotiede, tilastot ja yhteiskunta 1/3

- Ihminen ei voi toimia nykymaailmassa järkevästi, ellei hän pysty muodostamaan *oikeata kuvaa maailmasta* ja sen *tilasta*.
- Rakennusaineeksi oikeata kuvaa varten tarvitaan mm. maailmaa ja sen tilaa *merkityksellisesti ja oikein kuvaavia, ajantasaisia (tilasto-) tietoja*.
- Merkityksellisesti ja oikein todellisuutta kuvaavat, ajantasaiset *(tilasto-) tiedot* ovat *välttämättömiä* modernin yhteiskunnan toiminnalle ja niiden saatavuutta voidaan pitää toimivan *demokratian edellytyksenä*.

Tilastotiede, tilastot ja yhteiskunta 2/3

- *Yhteiskunnan kaikilla sektoreilla toiminnan seuranta, päätöksenteko ja ennakointi perustuvat sekä yhteiskunnan eri sektoreita kuvaaviin (tilasto-) tietoihin että tilastollisiin menetelmiin.*
- *Päätöksenteko perustuu sekä julkisella että yksityisellä sektorilla (elinkeinoelämässä) yhteiskuntaa ja elinkeinoelämää kuvaaviin (tilasto-) tietoihin ja tilastollisiin menetelmiin.*

Esimerkiksi *tuotantoprosessien ohjaus ja laadunvalvonta* teollisuudessa sekä *markkinatutkimus* kaupan alalla perustuvat tilastollisiin menetelmiin.

Tilastotiede, tilastot ja yhteiskunta 3/3

- Koska todellisuutta kuvaaviin (*tilasto-*) *tietoihin* sisältyy (lähes) aina *epävarmuutta* ja *satunnaisuutta*, tilastotiede ja tilastolliset menetelmät luovat perustan *tilastojen tuotannolle, jalostukselle ja analysoinnille*.
- Tilastojen tuotannon, jalostuksen ja analysoinnin *menetelmien kehittäminen* on keskeinen osa tilastotieteen tehtäväkenttää.

Esimerkki 1:

Kyselytutkimukset – 1/4

- Päätöksentekijät ja tiedotusvälineet kartoittavat säännöllisten välein suomalaisten mielipiteet erilaisista yhteiskuntaa koskevista kysymyksistä.
- Esimerkkejä:
 - Miten suomalaiset suhtautuvat mahdolliseen NATO-jäsenyyteen?
 - Miten suomalaiset suhtautuvat ydinvoiman lisärakentamiseen?
 - Mitkä ovat poliittisten puolueiden kannatusosuudet?
- Mielipiteet selvitetään *kyselytutkimuksilla*, joiden kohteeksi poimitaan tyypillisesti 1000 – 2000 suomalaista.
- Kyselytutkimuksen **tavoitteena on tehdä kyselyn tulosten perusteella johtopäätöksiä mielipiteiden jakautumisesta *kaikkien* suomalaisten joukossa.**

Esimerkki 1:

Kyselytutkimukset – 2/4

- **Miten 1000 – 2000 suomalaiseseen kohdistetun kyselyn tulokset voidaan *yleistää* koskemaan kaikkia suomalaisia?**
 - Kyselyn tulokset voidaan *yleistää*, jos kyselyn kohteiksi poimittujen suomalaisten joukko muodostaa *edustavan pienoiskuvan* Suomen kansasta.
 - Pienoiskuva on *edustava*, jos mielipiteet jakautuvat kyselyn kohteiksi poimittujen joukossa *samalla tavalla* kuin kaikkien suomalaisten muodostamassa *perusjoukossa*.
 - Kyselyn kohteiden *poiminta arpomalla* on ainoa menetelmä, joka mahdollistaa edustavan pienoiskuvan saamisen.
 - Kyselyn kohteiden *poimintaa* kaikkien suomalaisten muodostamasta *perusjoukosta arpomalla* kutsutaan tilastotieteessä (**satunnais-**) **otannaksi** ja tutkimuksen kohteeksi poimittua perusjoukon osaa kutsutaan (**satunnais-**) **otokseksi**.

Esimerkki 1:

Kyselytutkimukset – 3/4

- *Arvonnan käyttö* kyselyn kohteiden poiminnassa merkitsee sitä, että *kyselyn tulokset ovat satunnaisia* seuraavassa mielessä:
Jos arvontaa toistettaisiin, kysely tuottaisi (suurella todennäköisyydellä) joka kerran (ainakin jonkin verran) erilaiset tulokset, koska eri arvonnoissa kyselyyn poimittaisiin (suurella todennäköisyydellä) eri henkilöt.
- Kysymyksiä:
 - Miten yhdestä otoksesta saadut ja satunnaiset kyselytulokset voidaan yleistää koskemaan koko sitä *perusjoukkoa*, josta otos poimitaan?
 - Miten *luotettava* tällainen yleistys on?

Esimerkki 1:

Kyselytutkimukset – 4/4

- Vastauksia:
 - Jos kyselyn kohteiden poiminnassa on käytetty *satunnaisotantaa*, kyselyn tuloksiin sisältyvälle epävarmuudelle ja satunnaisuudelle voidaan muodostaa *tilastollinen malli*, joka mahdollistaa sekä *kyselyn tulosten yleistämisen* että *yleistyksen luotettavuuden arvioinnin*.
 - Yleistyksen luotettavuutta ei pystytä arvioimaan, *ellei otoksen poiminnassa ole käytetty satunnaisotantaa*.
 - Kyselytutkimusten *suunnittelussa, toteutuksessa ja tulosten analysoinnissa* sovelletaan mm. seuraavia tilastollisia menetelmiä:
 - **otanta**
 - **estimointi**
 - **testaus**

Esimerkki 2:

Lääketieteelliset kokeet – 1/4

- Erään tappavan taudin hoitoon on kehitetty *uusi lääke*, jonka toivotaan parantavan enemmän potilaita kuin kauan käytössä ollut *vanha lääke*.
- Miten saadaan *varmuus* siitä, että uusi lääke on parempi kuin vanha lääke?
- Paranemistulosten *vertailemiseksi* järjestetään *tilastollinen koe*:
 - (i) Jaetaan joukko potilaita *arpomalla* kahteen ryhmään:
 - Ryhmälle 1 annetaan uutta lääkettä.
 - Ryhmälle 2 annetaan vanhaa lääkettä.
 - (ii) *Verrataan* parantuneiden *suhteellisia osuuksia* ryhmässä 1 ja 2.
- Kokeen **tavoitteena on tehdä kokeen tulosten perusteella yleisiä johtopäätöksiä uuden lääkkeen tehokkuudesta.**

Esimerkki 2:

Lääketieteelliset kokeet – 2/4

- **Miten yhdestä kokeesta saadut tulokset voidaan yleistää koskemaan kaikkia tautia sairastavia potilaita?**
 - Kokeen tulokset voidaan yleistää, jos kokeessa uutta ja vanhaa lääkettä saavien potilaiden ryhmät ovat *samankaltaisia* kaikissa muissa suhteissa paitsi siinä, että niihin kohdistetaan kokeessa *erilainen käsittely*.
 - Tällöin mahdolliset *erot* parantuneiden suhteellisissa osuuksissa *ovat seurausta erilaisista käsittelyistä*.
 - Kokeen kohteiden *jakaminen ryhmiiin arpomalla* on ainoa menetelmä, joka mahdollistaa samankaltaisten ryhmien saamisen.
 - Kokeen kohteiden *jakamista* erilaisen käsittelyn kohteiksi joutuviin *ryhmiin arpomalla* kutsutaan tilastotieteessä **satunnaistamiseksi**.

Esimerkki 2:

Lääketieteelliset kokeet – 3/4

- *Arvonnan käyttö* ryhmiin jaossa merkitsee sitä, että *koetulokset ovat satunnaisia* seuraavassa mielessä:

Jos arvontaa toistettaisiin, kokeesta saataisiin (suurella todennäköisyydellä) joka kerran (ainakin jonkin verran) erilaiset tulokset, koska eri arvonnoissa saataisiin (suurella todennäköisyydellä) erilaiset ryhmäjaot.

- Kysymyksiä:
 - Miten yhdestä kokeesta saadut ja satunnaiset koetulokset voidaan yleistää koskemaan kaikkia ko. tautia sairastavia potilaita?
 - Miten luotettava tällainen yleistys on?

Esimerkki 2:

Lääketieteelliset kokeet – 4/4

- Vastauksia:
 - Jos potilaiden jaossa ryhmiin on käytetty *satunnaistamista*, kokeen tuloksiin sisältyvälle epävarmuudelle ja satunnaisuudelle voidaan muodostaa *tilastollinen malli*, joka mahdollistaa sekä *koetulosten yleistämisen* että *yleistyksen luotettavuuden arvioinnin*.
 - Yleistyksen luotettavuutta ei pystytä arvioimaan, *ellei ryhmiin jaossa ole käytetty satunnaistamista*.
 - Tilastollisen kokeen *suunnittelussa, toteutuksessa ja tulosten analysoinnissa* sovelletaan mm. seuraavia tilastollisia menetelmiä:
 - **koesuunnittelu**
 - **estimointi**
 - **testaus**

Esimerkki 3:

Laadunvalvonta – 1/4

- Tehdas valmistaa korkealuokkaisia sulkimia kameroihin.
- Tehdas pyrkii siihen, että yli 90 % sulkimista kestää vähintään 100 000 laukaisua.
- Sulkimien *laadun valvonta* on toteutettu seuraavalla tavalla:
 - (i) Tuotantolinjalta *poimitaan arpomalla* joukko sulkimia rasituskokeeseen.
 - (ii) Rasituskokeessa *määrätään* vähintään 100 000 laukaisua kestävien sulkimien *suhteellinen osuus*.
- Kokeen **tavoitteena on tehdä kokeen tulosten perusteella yleisiä johtopäätöksiä sulkimien kestävydestä.**

Esimerkki 3:

Laadunvalvonta – 2/4

- **Miten vain osaan sulkimista kohdistetun rasituskokeen tulokset voidaan yleistää koskemaan kaikkia sulkimia?**
 - Kokeen tulokset voidaan yleistää, jos rasituskokeen kohteiksi poimittujen sulkimien joukko muodostaa *edustavan pienoiskuvan* kaikista valmistetuista sulkimista.
 - Pienoiskuva on *edustava*, jos sulkimien kesto jakautuu rasituskokeeseen poimittujen sulkimien joukossa *samalla tavalla* kuin kaikkien valmistettujen sulkimien muodostamassa *perusjoukossa*.
 - Rasituskokeen kohteiden *poiminta arpomalla* on ainoa menetelmä, joka mahdollistaa edustavan pienoiskuvan saamisen.
 - Rasituskokeen kohteiden *poimintaa* kaikkien valmistettujen sulkimien muodostamasta *perusjoukosta arpomalla* kutsutaan tilastotieteessä (**satunnais-**) **otannaksi** ja tutkimuksen kohteeksi poimittua perusjoukon osaa kutsutaan (**satunnais-**) **otokseksi**.

Esimerkki 3:

Laadunvalvonta – 3/4

- *Arvonnan käyttö* rasituskokeen kohteiden poiminnassa merkitsee sitä, että *koetulokset ovat satunnaisia* seuraavassa mielessä:
Jos arvontaa toistettaisiin, kokeesta saataisiin (suurella todennäköisyydellä) joka kerran (ainakin jonkin verran) erilaiset tulokset, koska eri arvonnoissa kokeeseen poimittaisiin (suurella todennäköisyydellä) eri sulkimet.
- Kysymyksiä:
 - Miten yhdestä kokeesta saadut ja satunnaiset koetulokset voidaan yleistää koskemaan kaikkia sulkimia?
 - Miten luotettava tällainen yleistys on?

Esimerkki 3:

Laadunvalvonta – 4/4

- Vastauksia:
 - Jos rasisuskokeen kohteiden poiminnassa on käytetty *satunnaisotantaa*, kokeen tuloksiin sisältyvälle epävarmuudelle ja satunnaisuudelle voidaan muodostaa *tilastollinen malli*, joka mahdollistaa sekä *koetulosten yleistämisen* että *yleistyksen luotettavuuden arvioinnin*.
 - Yleistyksen luotettavuutta ei pystytä arvioimaan, *ellei kokeen kohteiden poiminnassa ole käytetty satunnaisotantaa*.
 - Kokeen *suunnittelussa, toteutuksessa ja tulosten analysoinnissa* sovelletaan mm. seuraavia tilastollisia menetelmiä:
 - **koesuunnittelu ja otanta**
 - **estimointi**
 - **testaus**