



UNIVERSITY
OF TAMPERE

STATISTICS

DATA ANALYTICS

BIG DATA

RESEARCH

TEACHING

CONSULTANCY



www.uta.fi/cast/

CAST:n rooli tutkimuksessa

Ansa Lilja

CAST - Centre for Applied Statistics and Data
Analytics –Tilastotieteen keskus

Arvokas tutkimusaineisto -seminaari 3.5.2016



Tavoitteita

- TaYsta vahva kansallinen ja kansainvälinen toimija kvantitatiivisten menetelmien osaajana
 - Koulutus, tutkimus, konsultointi
- Jo olemassa olevan osaamisen parempi hyödyntäminen TaYssa
 - Tutkimuksen ja artikkelien laadun kehittäminen kvantitatiivisten menetelmien laajemmalla käyttöönotolla
- Tampere3?



Keinoja...

- Tilastollisen yhteistyön tukeminen UTA:ssa, kansallisesti ja kansainvälisesti
 - Työpajat, säännölliset seminaarit
 - Vierailevat asiantuntijat ja tutkijat
 - Yhteistyö: tutkimus, julkaisut, tutkimussuunnitelmat
- Kvantitatiivisten menetelmien koulutus
 - Tarjonnan näkyvyyden parantaminen
 - Koordinointi ja päällekkäisyyksien minimointi



...Keinoja

- R:n laajemman käytön tukeminen
- Tutkijoiden/tutkijaryhmien konsultointi
 - Tutkimussuunnitelma ->...-> tulosten raportointi
 - Konsultoinnin toimintaohjeiden ja hinnoitteluperiaatteiden luominen
 - Yksi artikkelin kirjoittajista
 - Osaprojektirahoitus mahdollistaa CAST-resurssoinnin



Resurssit

- Kukin yksikkö on vastuussa oman alueensa riittävästä kvantimenetelmien resurssoinnista
 - koulutus, tutkimus, konsultointi
- Tilastolliseen tutkimukseen varattava riittävästi rahoitusta per tutkimusprojekti
- Tohtoriopiskelijoiden menetelmäohjauksesta vastaa pääosin oma(t) ohjaaja(t)
 - Tukena CAST/Tohtorikoulun/yksikön järjestämät kvanti-koulutukset



UNIVERSITY
OF TAMPERE

STATISTICS

DATA ANALYTICS

BIG DATA

RESEARCH

TEACHING

CONSULTANCY



CAST
Centre for Applied Statistics
and Data Analytics - Tampere

www.uta.fi/cast/

Raakadatan esikäsittely



Raakadata - primääridata

- Kerättyä käsittelemätöntä/kevyesti käsiteltyä aineistoa, mikä vaatii muokkausta ennen analysointia
 - Erilaiset aineiston muunnokset
 - Tiedon tiivistys
 - Järjestäminen, yhdistely
 - Tiedon puutteellisuus, virheellisyys, ristiriitaisuus
- Raakadatan käyttökelpoiseksi muunto usein merkittävä osuus kvantitatiivisesta tutkimuksesta



Raakadatan haasteita

- Arkistointimuoto -> analyysimuoto
 - Rakenne- ja sisältömuunnokset
- Raakadatan suuri koko
 - Talletustila, käsittelyvaatimukset, valinta
 - Analyyseissä käytetään vain osaa raakadatatista
 - Volyymi, kompleksisuus ja dimensio laskee
- Yksiköt voivat vaihdella
 - Esim. kun raakadataa kerätty eri lähteistä



Raakadatan haasteita

- Mittaus/arkistointivirheet/poikkeavat arvot
 - virheellisen datan osuuden arviointi
 - poikkeavat arvot: outlier-havaintojen merkityksen testaaminen (mallin sovitus ilman/kanssa -> residuaalit&kertoimet)
 - puuttuva tieto: analysointi ja mahdolliset korjaustoimenpiteet
- *Piilossa oleva systematiikka datassa (keruuvaiheen prosessi?) voi vääristää tuloksia, jos sitä ei huomioida*



Kehittäminen

- Aineiston keräämis- ja arkistointivaiheiden paremmat käytännöt (Tuuli etc.)
 - Tunne aineistosi ja sen muuttujat (semantiikka)
 - Kvantitatiivisten menetelmien ja raakadatan arkistoinnin osaja mukaan tutkimussuunnitelman tekoon tai viimeistään ennen aineistonkeruuta
- Jos raakadatan keruu-aika on pitkä, aineiston muuttujien tunnuslukujen seuranta mahdollisten muutosten varalta
- Puuttuvan tiedon analyysiosaaminen