

Kvalitetssikre og analysere egne data - Nødvendig kunnskap

Lena Ringstad Olsen

29. januar 2019

Oversikt

- ▶ Tallknuserhjelp fra Servicemiljøet; tilbud og krav
- ▶ Tallbehandling: verktøy, metoder

Minimumshjelp fra tallknuserne i Servicemiljøet

Region	Ant.stat/an.	ant.registre
Nasjonalt	1	-
Nord (Tromsø)	1	8 (9)
Midt (Trondheim)	1	6
Vest (Bergen)	2,2	18
Sør-Øst (Oslo)	2	11 (19)
Tot	7,2	43 (51)

- ▶ Planlegging av variable (format, mål, samarbeid m/IKT) ved oppstart av register.
- ▶ Bistand med design av valideringsstudie
- ▶ Beregning og visualisering av kvalitetsindikatorer, ifm. årsrapport og på Resultatportalen.

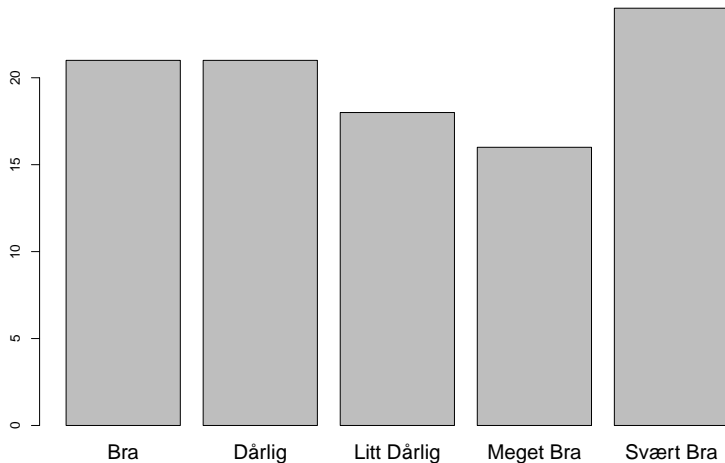
Krav til registre/data for å starte analyse

- ▶ Hva er et forløp/hendelse
- ▶ Tilgjengelige elektroniske data på hensiktsmessig format
- ▶ Beskrivelse av datastrukturen - sammenheng mellom tabeller
- ▶ Vise enkel relasjonsmodell
- ▶ Kunne levere data på csv
 - ▶ Gode navn? *RevaskulariseringFoerInfarkt, Kjønn, Blod2, Ms3*

Krav til registre/data for å starte analyse

- ▶ Hva er et forløp/hendelse
- ▶ Tilgjengelige elektroniske data på hensiktsmessig format
- ▶ Beskrivelse av datastrukturen - sammenheng mellom tabeller
- ▶ Vise enkel relasjonsmodell
- ▶ Kunne levere data på csv
 - ▶ Gode navn? *RevaskulariseringFoerInfarkt, Kjønn, Blod2, Ms3*
- ▶ Variabelnavn: Maskinlesbare, entydige, tolkbare, ikke for lange
- ▶ Kodede variable (heltall, naturlig ordnet)
 - ▶ Variabelinnhold: Svært Bra, Meget Bra, Bra, Litt Dårlig, Dårlig

Eksempel, uhensiktsmessig format



Ønsker: 1,2,3,4,5

Krav, forts. . .

- ▶ Beskrivelse av alle variable (kodebok)
- ▶ Beskrivelse av beskrivelsen (kodeboka)
- ▶ Klart definerte kvalitetsindikatorer (finnes alle nødvendige variabler?)
- ▶ Hva ønsker vi å vise (budskap)
- ▶ Unngå endringer i datasettet. . .
 - ▶ Pass på at det utarbeides **god dokumentasjon av ALLE endringer**
 - ▶ Endringer i “kjerneversjon” av innregistreringsløsninga?
 - ▶ Finnes sjekkliste

Fra handlingsplanen: *Det er ønskelig at registrene skal å holde ei fullstendig oversikt over hver enkelt variabel med begrunnelse for hva de har tenkt å bruke variabelen til og hvordan den blir brukt, samt kvaliteten av den (utfyllingsgrad, pålitelighet).*

Endringslogg, sjekkliste

- ▶ Endringer i variabelnavn? Gjelder også navneendring for id-variable/koblingsnøkler.
- ▶ Er det nye variabler og er disse fullstendig beskrevet
- ▶ Er variabler fjernet
- ▶ Har variabler endret innhold, f.eks. fått en ny kategori
- ▶ Har variabler endret format, f.eks. fra tekststrenger til kodede variabler
- ▶ Har kodeboka fått versjons-spesifisering av variabel som er lagt til/fjernet/endret
- ▶ Er det fjernet eller lagt til tabeller? I så fall:
 - ▶ Hvilke variable er koblingsnøkler
 - ▶ Hvordan kobles den nye tabellen til eksisterende tabeller (i hierarkiet)
- ▶ Er standardverdier endret? Eks. endring i representasjon av tomme felter
- ▶ Har man gått gjennom diff mellom gammel og ny kodebok?

Tallbehandling: Verktøy

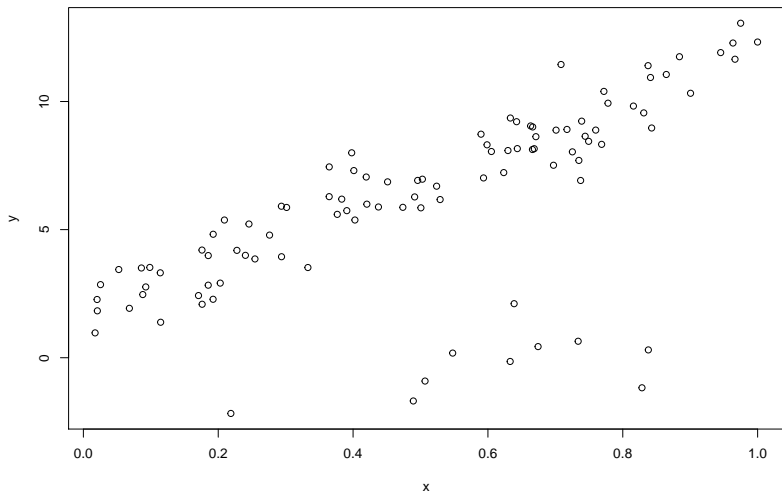
- ▶ Personlige favorittverktøy. . .
- ▶ Excel: sortere, gjøre utvalg, sammenlikne variable, . . .
- ▶ Muligheter for å se på data i resultattjenester
- ▶ Reproduserbar kode

Tallbehandling: Vurdere data

- ▶ Kontrollere innhold i variable.
 - ▶ Koding/tekststrenger
 - ▶ Inneholder variabelen andre enn de dokumenterte verdiene
 - ▶ Hvordan representeres manglende verdier
 - ▶ Max/min-verdier, er de ok
 - ▶ Format (eks. . og ,)

Vurdere data, forts...

- ▶ Plott variablene for å få overblikk

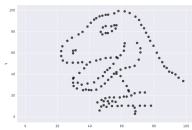


Vurdere data, forts...

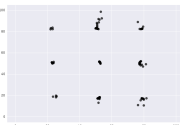
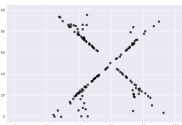
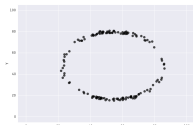
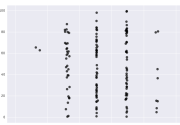
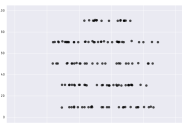
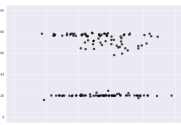
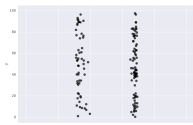
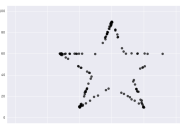
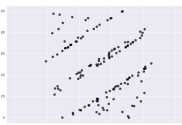
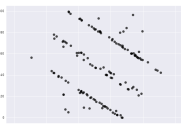
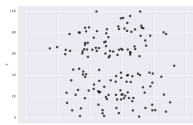
- ▶ Lag rutiner for datasjekk

```
##      isl_1          vekt          blod1
##  Min.    :298.0   Min.    :  -4.00   Min.    :107
##  1st Qu.:300.0   1st Qu.:  59.75   1st Qu.:109
##  Median :300.0   Median :   74.00   Median :110
##  Mean   :300.1   Mean   : 100.94   Mean   :110
##  3rd Qu.:301.0   3rd Qu.:  96.00   3rd Qu.:111
##  Max.   :303.0   Max.   :2444.00   Max.   :112
##  NA's   :13
```

Vurdere data, forts...



X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06



https:

[//www.autodeskresearch.com/publications/samestats](https://www.autodeskresearch.com/publications/samestats)

Presentasjon/fremstilling

- ▶ Servicemiljøet har anbefalte standarder:
 - ▶ Standard for figurer og tabeller
 - ▶ Resultatportalen for offentlige resultater (kvalitetsindikatorer)
- ▶ Tilbakemeldinger fra ekspertgruppe og publiseringsråd

Oppsummering

- ▶ Gode metadata, dvs. variabelbeskrivelser.
- ▶ Klare mål med registeret/dataene
- ▶ Gode datavaskrutiner og logikk i innregistreringa.
- ▶ Automatiserte/gjenbrukbare rutiner.
- ▶ Eierforhold til dataene, lek med tall
- ▶ Spør om hjelp, men ha kontroll på dataene