

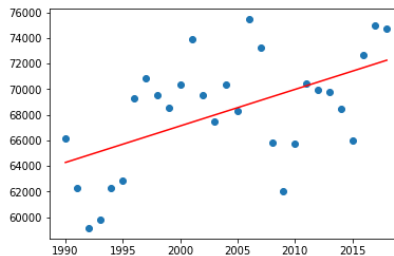
Opplegg 28 - Statistikk og korrelasjon

Korrelasjon betyr ikke nødvendigvis at vi har kausalitet!

Dersom vi har to datasett (for to variabler), kan vi finne ut hvor godt de korrelerer med hverandre. Altså om de varierer på samme måte. Det vil si at dersom den ene variabelen øker, så øker den andre variabelen tilsvarende, eller kanskje den minker tilsvarende. Vi kan si at korrelasjon er det samme som samvariasjon.

For å sjekke hvor godt to variabler korrelerer, utfører vi en lineær regresjon. Da kan vi få beregnet korrelasjonen, det er nemlig et mål på hvor store avstandene er mellom regresjonsgrafene og hvert av punktene i grafen. Vi regner ikke dette for hånd, men geogebra eller Python kan gjøre det for oss.

Litt forenklet vil geogebra eller Python i en lineær regresjon lage en rett linje som gjør at de samla avstandene mellom alle punktene og selve regresjonslinja blir minst mulig. Altså at summen av avstandene mellom de blå punktene og den røde lineære grafen blir minst mulig.

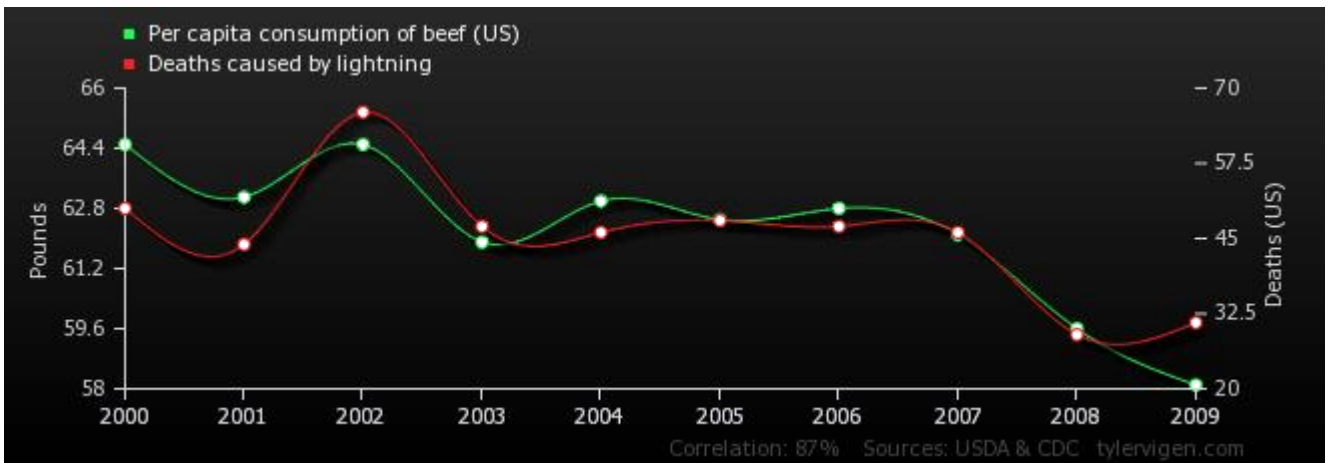


Hva er kausalitet?

Det betyr at det er en årsak-virkning sammenheng mellom variabler. Dersom man endrer på én variabel, så vil den andre variabelen også endre seg. Et eksempel på en kausal sammenheng kan være at antall timer du sover om natta påvirker hvor trøtt du er neste dag. Et annet ord for kausalitet er årsakssammenheng.

Man kan sammenligne to variabler som varierer over tid, og sjekke om de varierer på samme måte. For eksempel kan man velge to variabler som det er gjort målinger av årlig i perioden 1990 til 2020. For å sjekke hvor godt de to variablene korrelerer, velger man én av variablene som x-verdi og den andre som y-verdi. Det er viktig å passe på at x-verdien og y-verdien er for samme årstall. Da kan man utføre en lineær regresjon på x- og y-verdiene og finne korrelasjonen.

Det er ikke dette som er gjort med de to grafene under. Her er begge variablene tegnet inn som funksjon av årstall, slik at vi har to separate funksjoner. Men vi kan se at begge grafene har lignende form.



Korrelasjonskoeffisient: 0,87

Kilde: www.tylervigen.com

Korrelasjonskoeffisienten – kan variere mellom -1 og 1.

Dersom den er 0, eller nær 0, betyr det at vi har ingen eller liten korrelasjon.

Dersom den er nær -1, minker den ene variabelen når den andre øker.

Dersom den er nær 1, vil begge variablene øke/minke på samme måte.

Det er tilsvarende for R^2 , men den kan ikke være negativ.

Noen mulige årsaker til korrelasjon:

- En variabel kan direkte forårsake eller avhenge av verdien til en annen variabel.
- En variabel kan være løst knytta til en annen variabel.
- To variabler kan begge avhenge av en ukjent tredje variabel.
- Det kan være tilfeldig – størst sjanse for at dette skjer ved få målinger.

Lag en villedende nyhetssak

- ved bruk av korrelasjon

Oppgave

Finn to datasett hos statistisk sentralbyrå (<https://www.ssb.no/statbank/>), og finn korrelasjonen ved bruk av Geogebra eller Python. Lag et nyhetsoppslag som spekulerer i kausalitet (at den ene variabelen er årsak til at den andre endres).

Fase 1: Undersøk hvilke data som kan være fornuftig å bruke. Hvilke data kan tenkes å variere på samme måte over tid? Husk at dere skal lage en falsk nyhetssak til slutt, så det kan være lurt å tenke på hvilken vinkling dere ønsker å ha i saken.

Fase 2: Ha en idémyldring for deg selv. Hvilke data ønsker du å bruke? Hvorfor akkurat dette? Tegn gjerne en skisse over nyhetssaken før du diskuterer med de andre. Deretter må gruppa samlet bestemme hvilke data dere skal bruke og hvordan deres nyhetssak skal se ut.

Fase 3: Lag første versjon av grafen for den matematiske modellen. Hva blir korrelasjonen mellom variablene? Lag nyhetssaken basert på korrelasjonen.

Fase 4: Får dere en korrelasjon som virker fornuftig? Spør andre i klassen om nyhetssaken deres er tydelig og god.

Noen eksempler på villedende nyhetssaker og konspirasjonsteorier

- basert på korrelasjoner

Ny studie:

Mye egg koblet til hjertesykdom og død

Kanskje er ikke egg så ufarlig likevel?

[Ny studie: Mye egg koblet til hjertesykdom og død \(forskning.no\)](#)

Folk som drikker brus – både sukkerfri og vanlig – lever kortere

Brus hører ikke til i hverdagen, sier forsker.

[Folk som drikker brus – både sukkerfri og vanlig – lever kortere \(forskning.no\)](#)

Håndskrift hjelper hukommelsen

Barn som lærer seg å skrive for hånd, har lettere for å huske og utvikle egne ideer enn barn som lærer seg å skrive via et tastatur, viser ny forskning.

[Håndskrift hjelper hukommelsen \(aftenposten.no\)](#)

One is a 5G Map, One is the CoronaVirus Map in the USA
Always a bigger picture 📍



Theodore R. Smith på Twitter: «Correlation != Causation. Tho this probably will correlate with population density which will correlate with mass infections, and 5G could be ruining people's immune systems. <https://t.co/rpas5Wfm2z> / Twitter

Fase 5: Sammenlign gjerne med de andre i klassen. Er det noen som har en bedre nyhetssak? Hvorfor mener dere at den er bedre? Kan dere bruke noe av de samme grepene i deres sak?

Fase 6: Hopp gjerne tilbake til tidligere punkt, og gjør forandringer for å få en best mulig nyhetssak. Gjør gjerne endringer om det trengs.

Fase 7: Pass på å ta vare på bilder og notater dere har gjort underveis, slik at dere kan vise hva dere har tenkt. I denne oppgaven går dokumenteringen ut på å lage selve nyhetssaken.