

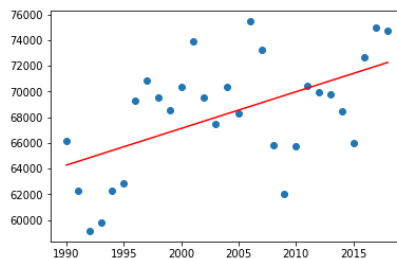
Opplegg 28 - Statistikk og korrelasjon

Korrelasjon tyder ikkje nødvendigvis at vi har kausalitet!

Dersom vi har to datasett (for to variabler), kan vi finne ut kor godt dei korrelerer med kvarandre. Altså om dei varierer på same måten. Det vil seie at dersom den eine variabelen aukar, så aukar den andre variabelen tilsvarande, eller kan hende minkar han tilsvarande. Vi kan seie at korrelasjon er det same som samvariasjon.

For å sjekke kor godt to variablar korrelerer utfører vi ein lineær regresjon. Da kan vi få berekna korrelasjonen, det er nemlig eit mål på kor store avstandene er mellom regresjonsgrafan og kvart av punkta i grafan. Vi reknar ikkje dette for hand, men geogebra eller Python kan gjere det for oss.

Litt forenkla vil geogebra eller Python i ein lineær regresjon lage ei rett linje som gjer at dei samla avstandene mellom alle punkta og sjølve regresjonslinja blir minst mogleg. Altså at summen av avstandane mellom dei blå punkta og den røde lineære grafan blir minst mogleg.

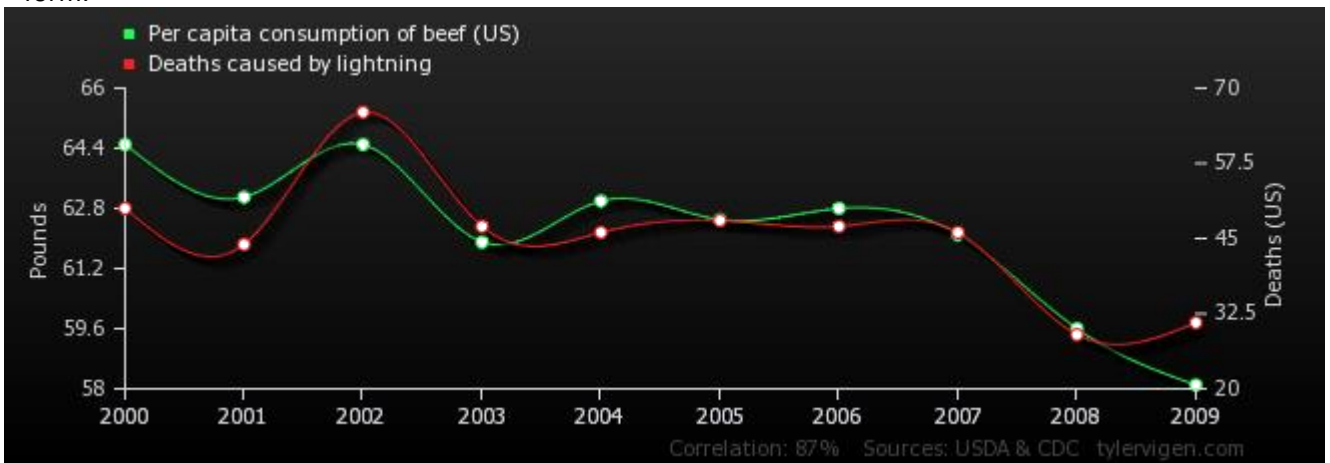


Kva er kausalitet?

Det tyder at det er ein årsak-verknad- samanheng mellom variablar. Dersom ein endrar på éin variabel, vil den andre variabelen og endre seg. Eit døme på en kausal samanheng kan vere at talet på timar du søv om natta påverkar kor trøtt du er neste dag. Eit anna ord for kausalitet er årsakssamanheng.

Ein kan samanlikne to variablar som varierer over tid, og sjekke om dei varierer på same måten. Til dømes kan ein velje to variablar som det er gjort målinger av årleg i perioden 1990 til 2020. For å sjekke kor godt dei to variablane korrelerer, vel ein den éine av variablane som x-verdi og den andre som y-verdi. Det er viktig å passe på at x-verdien og y-verdien er for same årstalet. Da kan ein utføre ein lineær regresjon på x- og y-verdiane og finne korrelasjonen.

Det er ikkje dette som er gjort med dei to grafane under. Her er begge variablane teikna inn som funksjon av årstal, slik at vi har to separate funksjonar. Men vi kan sjå at begge grafane har liknande form.



Korrelasjonskoeffisient: 0,87

Kilde: www.tylervigen.com

Korrelasjonskoeffisienten – kan variere mellom -1 og 1.

Dersom han er 0, eller nær 0, tyder det at vi har ingen eller liten korrelasjon.

Dersom han er nær -1, minkar den eine variabelen når den andre aukar.

Dersom han er nær 1, vil begge variablane auke/minke på same måten.

Det er tilsvarande for R^2 , men han kan ikkje vere negativ.

Nokre moglege årsaker til korrelasjon:

- Ein variabel kan direkte forårsake eller avhenge av verdien til ein annan variabel.
- Ein variabel kan vere laust knytt til ein annan variabel.
- To variablar kan begge avhenge av ein ukjent tredje variabel.
- Det kan vere tilfeldig – størst sjanse for at dette skjer ved få målingar.

Lag ei villeiande aktualitetssak

- med bruk av korrelasjon

Oppgåve

Finn to datasett hos statistisk sentralbyrå (<https://www.ssb.no/statbank/>), og finn korrelasjonen med bruk av Geogebra eller Python. Lag eit aktualitetsoppdrag som spekulerer i kausalitet (at den eine variabelen er årsak til at den andre blir endra).

Fase 1: Undersøk kva for nokre data som kan vere fornuftig å bruke. Kva for nokre data kan tenkest å variere på same måten over tid? Hugs at de skal lage ei falsk aktualitetssak til slutt, så det kan vere lurt å tenke på kva vinkling de ønsker å ha i saka.

Fase 2: Ha ei idémyldring for deg sjølv. Kva for nokre data ønsker du å bruke? Kvifor akkurat desse? Teikne gjerne ei skisse over saka før du diskuterer med dei andre. Deretter må gruppa samla avgjere kva for nokre data de skal bruke og korleis aktualitetssaka dykkar skal sjå ut.

Fase 3: Lag første versjon av grafen for den matematiske modellen. Kva blir korrelasjonen mellom variablane? Lag aktualitetssaka basert på korrelasjonen.

Fase 4: Får de en korrelasjon som verkar fornuftig? Spør andre i klassen om aktualitetssaka dykkar er tydeleg og god.

Nokre døme på villeiande aktualitetssaker og konspirasjonsteoriar - basert på korrelasjonar

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/rpsas5Wfm2z> / Twitter

Fase 5: Samanlikne gjerne med dei andre i klassen, er det nokon som har ei betre aktualitetssak? Kvifor meiner de at deira er betre? Kan de bruke noko av dei same grepa i dykkar sak?

Fase 6: Hopp gjerne attende til tidlegare punkt, og gjer endringar for å få ei best mogleg aktualitetssak. Gjer gjerne endringar om det trengst.

Fase 7: Pass på å ta vare på bilete og notat de har gjort undervegs, slik at de kan vise kva de har tenkt. I denne oppgåva går dokumenteringa ut på å lage sjølv aktualitetssaka.