

08 - Korrelasjon 1

Forelesningsmål

Være bevisst på om sammenhenger er sanne og informative - og hvilke sammenhenger som er villedende og uriktige.

Relevante eksamensoppgaver

- *Vår 2022*

En gruppe forskere finner en interessant sammenheng der 25 % av den norske befolkningen opplever vårslapphet og økt tretthet når det går mot varmere og lysere dager.

a. Beskriv forholdet mellom A: symptomer på slapphet og tretthet og B: lyse og varme dager om korrelasjonen mellom disse to variablene var positiv, negativ og kurvlineær.

b. En observert sammenheng mellom variabel A og variabel B kan skyldes innvirkning fra en tredje variabel: Z. Forklar med eksempler på mulige tredjevariabler som kan være årsaken bak sammenhengen mellom økt tretthet/slapphet og varmere/lysere dager. Bruk gjerne ulike kausalmodeller når du forklarer.

c. Det viser seg at sammenhengen skyldes ren tilfeldighet og at disse variablene ikke har et kausalt forhold til hverandre. Hva kaller vi en slik sammenheng som enten skyldes en tredjevariabel eller oppstår ved ren tilfeldighet? Kom med tre eksempler fra din hverdag på slike sammenhenger.

- *Vår 2021*

Du ønsker å finne ut om det er en kausal sammenheng mellom hukommelsestap og Covid-19 sykdom blant et stort utvalg av personer.

c) Du finner i tillegg at det er en positiv korrelasjon på 0.35 mellom hukommelsessvikt og Covid-19 sykdom. Hva betyr det?

d) Hvorfor skal man være forsiktig med å trekke slutninger om årsak-virkning fra korrelasjonsdata?

- *Vår 2020*

Du finner at tid brukt på å spille voldelige dataspill og aggresjonsnivå hos barn ser ut til å ha en sammenheng med hverandre.

- a) Korrelasjonen mellom de to variablene er $r = 0,30$. Hva betyr det?
- b) En observert sammenheng mellom variabel x og variabel y kan skyldes innvirkning fra en tredje variabel: z. Forklar med tre ulike kausalmodeller hvordan z kan virke inn på voldelige dataspill og aggresjon. Tegn de tre modellene på et ark som du legger ved som en egen fil.
- c) Hva er en spuriøs sammenheng? Forklar hvordan en sammenheng mellom dataspill og aggresjon kan være spuriøs.
- d) Hva forstår du med delt varians (Coefficient of determination) i denne sammenhengen, og hvor stor er den i dette eksempelet?

- *Vår 2019*

Det ser ut til å være en sammenheng mellom røyking og høyt skolefravær blant ungdom.

- a) Korrelasjonen mellom de to variablene er $r = 0.40$. Hva betyr det i denne sammenhengen?
- b) Hva forstår du med delt varians (Coefficient of determination) i denne sammenhengen, og hvor stor er den?
- c) En observert sammenheng mellom variabel x og variabel y kan skyldes innvirkning fra en tredje variabel: z. Forklar med ulike kausalmodeller hvordan z kan virke inn på røyking og skolefravær.

- *Høst 2017*

Du finner at tid brukt på voldelige dataspill og aggresjonsnivå hos barn ser ut til å ha en sammenheng med hverandre.

- a) Korrelasjonen mellom de to variablene er $r = 0.45$. Hva betyr det?
- b) En observert sammenheng mellom variabel x og variabel y kan skyldes innvirkning fra en tredje variabel: z. Forklar med ulike kausalmodeller hvordan z kan virke inn.

Begreper o.l

- **korrelasjon** - når det er en sammenheng i utviklingen til to variabler

“Korrelasjon er et statistisk mål på hvor mye to målbare størrelser henger sammen med hverandre.” (Store Norske Leksikon)

“Definisjonsmessig er korrelasjonen egentlig et mål på manglende variasjon rundt en rett linje.” (Fra forelesningsslide)

- **samvariasjon** - synonym for korrelasjon
- **kausalitet** - når en ting forårsaker noe annet

“Kausalitet kommer fra latinsk «causa», som betyr årsak eller grunn, og brukes om en årsakssammenheng, det vil si når det foreligger et påvirkningsforhold mellom to fenomener. En vitenskapelig forklaring innebærer som regel å vise til en årsakssammenheng.

Vi skiller mellom en deterministisk og en stokastisk forståelse av kausalitetsbegrepet. Den deterministiske forståelsen, som er mest vanlig i naturvitenskapen, går ut på at et bestemt fenomen (B) alltid har samme årsak (A), og at en bestemt årsak (A) alltid har samme virkning (B).

I samfunnet finnes det få deterministiske årsaksforhold. I samfunnsvitenskapen er det derfor vanlig å legge vekt på en stokastisk kausalitetsforståelse, som innebærer at et fenomen (A) betraktes som årsak til et annet fenomen (B) dersom det er en viss sannsynlighet for at A fører til B, eller at A øker sannsynligheten for B. Denne forståelsen kalles også tendensiell, siden det er en tendens til at A fører til B.”
(Store Norske Leksikon)

- **variabel** - noe målbart som kan endre på seg/ha forskjellige verdier, feks hvor mange nordmenn som er millionærer eller hvor høy en person er

“En variabel er et kjennetegn eller en egenskap ved en analyseenhet. Begrepet variabel brukes særlig i kvantitativ forskning. Variasjoner mellom enhetene med hensyn til en bestemt variabel uttrykkes i form av ulike tallmessige verdier på denne variabelen. Variabelbegrepet kan også brukes i kvalitativ forskning, men da uten at variasjonene mellom enhetene blir tallfestet.

I en studie med individer (personer) som analyseenheter kan utdanningsnivå være et eksempel på en variabel. Den kan for eksempel ha verdiene (1) lav utdanning, (2) middels utdanning og (3) høy utdanning.

Variabler kan ha ulike målenivå, avhengig av hvor nyansert og detaljert informasjon variabelen og dens verdier gir, samt hvilke metoder som kan brukes i analyser av denne informasjonen.”

(Store Norske Leksikon)

- **metodetilnærminger** - et ord som brukes når man snakker om hvilken type vitenskapelig metode/forskningsmetode man bruker for å studere noe, feks deskriptive metoder, korrelasjonsmetoder eller eksperimentelle metoder

Fra Wikipedia-siden til Foucauldiansk diskursanalyse: “Foucauldiansk diskursanalyse (FDA) er en kvalitativ metodetilnærming som vektlegger språkets rolle i oppbygningen av sosialt og psykologisk liv.”

- **deskriptive metoder** - “I deskriptive studier forsøker man å beskrive virkeligheten uten å gi forklaringer, slik man gjør i analytiske studier som forsøker å avdekke årsak-virkning-forhold eller kausale forhold.

I en deskriptiv eller beskrivende studie er formålet å beskrive fordelingen av ulike faktorer, for eksempel utbredelsen av kreftsykdom etter kjønn, alder og bosted, uten å identifisere årsaker og/eller teste hypoteser. Slike studier kan blant annet være grunnlag for beslutninger om fordeling av ressurser, eller for å utvikle hypoteser.

Motsatsen til en deskriptiv studie er en analytisk studie.”

(Store Norske Leksikon)

Eksempler på deskriptive metoder: kasusstudier, observasjonsmetodikk, surveymetoder

- **korrelasjonsmetoder** - metodetilnærminger der du ser på korrelasjonen mellom fenomener/variabler
- **eksperimentelle metoder** - metodetilnærminger der du kan isolere en variabel og studere kun denne. Forskning i en lab, mao.

Eks: Du vil studere effekten av rusmiddel X på balanse, så du skaffer 100 deltagere som alle har like god balanse, gir halvparten en placebo og halvparten rusmiddel X, får alle til å ta en balansetest og sammenligner resultatene. Fordi den eneste forskjellen mellom gruppene var at den ene gruppen hadde tatt rusmiddel X, har du isolert rusmiddel X som variabel og kan med høy grad av sikkerhet si noe hvor mye rusmiddel X spiller inn på balanse.

- **statistisk kontroll** - når man bruker statistikk/data til å kontrollere for en variabel

“The use of statistical procedures to remove the influence of a particular factor that could not be eliminated or controlled by the experimental design in order to better analyze the relationship between two variables. For example, the relationship between age (x) and

income earned (y) could be influenced by a third variable, years of education (z). Thus, if a researcher did not first remove the effects of education he or she might derive erroneous conclusions about the influence of age on income from his or her analysis.”
(American Psychological Association)

- **eksperimentell kontroll** - den kontroll man har over ulike variabler/faktorerers påvirkning på utfallet når man utfører et eksperiment i en lab

Annen tolkning: kontrollgruppe, kontrollprøve, osv (den/det man ikke gjør noe med, som likevel må være med i forskningen for å kontrollere resultatet, feks placebogruppe)

“A controlled experiment is a scientific test done under controlled conditions, meaning that just one (or a few) factors are changed at a time, while all others are kept constant.”
(Khan Academy)

“An experimental control is used in scientific experiments to minimize the effect of variables which are not the interest of the study. The control can be an object, population, or any other variable which a scientist would like to “control.””
(ScienceTrends.com)

- **kasusstudier** - “Kasusstudier er kartlegging av psykologisk interessante forhold, typisk hos et enkeltindivid, eller på en arbeidsplass, skole, bedrift, med mer.

Kasusstudier gjennomføres ofte på personer som på en eller annen måte skiller seg ut, som for eksempel en fremragende vitenskapsmann eller komponist.

Den kanskje mest kjente kasusstudie ble gjennomført på HM (Henry Molaison), som i 1953 gjennomgikk en hjerneoperasjon for å kurere epilepsi. Operasjonen reduserte epilepsianfallene, men etter operasjonen viste det seg at den hadde gitt HM omfattende hukommelsestap. Oppfølgende studier av HM viste seg å være svært verdifulle for forståelsen av hukommelse, eksempelvis betydningen av hippocampus.

Kasusstudier må ikke forveksles med N=1-eksperimenter.”
(Store Norske Leksikon)

- **observasjonsmetodikk** - når man benytter en form for observasjon som forskningsmetode

“A type of study in which individuals are observed or certain outcomes are measured. No attempt is made to affect the outcome (for example, no treatment is given).”
(National Cancer Institute, USA)

“In fields such as epidemiology, social sciences, psychology and statistics, an observational study draws inferences from a sample to a population where the independent variable is not under the control of the researcher because of ethical

concerns or logistical constraints.”

Wikipedia

“Observasjon innebærer at vi ser på hva mennesker gjør i ulike situasjoner på en særlig oppmerksom måte. Det ligger i ordet hva observasjonsmetoden går ut på: Vi observerer atferden til dem vi ønsker å studere.”

(estudie.no)

- **surveymetoder** - spørreundersøkelser/skjemaer som forskningsmetode

“In research of human subjects, a survey is a list of questions aimed for extracting specific data from a particular group of people. Surveys may be conducted by phone, mail, via the internet, and also at street corners or in malls. Surveys are used to gather or gain knowledge in fields such as social research and demography.

Survey research is often used to assess thoughts, opinions and feelings. Surveys can be specific and limited, or they can have more global, widespread goals. Psychologists and sociologists often use surveys to analyze behavior, while it is also used to meet the more pragmatic needs of the media, such as, in evaluating political candidates, public health officials, professional organizations, and advertising and marketing directors.”

(Wikipedia)

- **retrospektiv (måling)** - når man måler ved å spørre deltakerne om noe som ligger i fortiden
- **tversnitt (måling)** - tversnitt er egentlig et matematisk ord (snitt gjennom et geometrisk legeme som er vinkelrett på lengderetningen til legemet), men brukes også som synom på et representativt utvalg. Ved en tversnittmåling måler man et representativt utvalg av en populasjon.

“Tverrsnittsstudie er en studie av fenomener, mønstre eller sammenhenger på ett bestemt tidspunkt. Dette kalles også synkron forskning, i motsetning til diakron forskning, som innebærer studier av fenomener, mønstre eller sammenhenger på flere tidspunkter. Tverrsnittsstudier kan gi grunnlag for omfattende analyser av de forholdene som finnes på undersøkelsestidspunktet, men viser ikke hvordan disse forholdene opprettholdes, utvikles eller endres over tid.”

(Store Norske Leksikon)

- **krysseksjonelt (måling)** - “Kryss-seksjonell metode er en tilnærming som studerer utvikling i egenskaper ved å sammenligne individer fra ulike alderstrinn. Eksempelvis kan en utviklingspsykolog studere språkutvikling ved å sammenligne grupper av barn på ulike alderstrinn (for eksempel toåringer, treåringer og fireåringer).”

(Store Norske Leksikon)

- **tidsserie/tidsrekke(analyse) (måling)** - “statistisk metode for å analysere serier av observasjoner på etterfølgende tidspunkter over en viss tid. Eksempler på tidsrekker er daglige temperaturmålinger, månedlige prisindekser, og arbeidsledighet fra måned til måned. Formålet med en tidsrekkeanalyse kan være å forstå den utvikling som har funnet sted eller lage prognoser for den videre utvikling.”

(Store Norske Leksikon)

- **longitudunelt (måling)** - “Longitudinell metode, også kalt langsgående metode, en tilnærming i forskningen hvor man følger et individ eller en gruppe av individer over en lengre periode, ofte over flere år.”

(Store Norske Leksikon)

- **predikere - å forutsi**

“Predikere betyr å forutsi. I vitenskapelige sammenhenger betyr predikere å anta hvordan fremtidige forhold vil bli ved hjelp av sannsynlighetsberegninger. Begrepet kan brukes som substantiv - prediksjon - eller som adjektiv - prediktiv.”

(Store Norske Leksikon)

- **årsaksretning** - hvilken retning årsakssammenhengen går. Er det X som fører til Y, eller Y som fører til X?

- **tredjevariabler** - variabler som ikke er en del av de to man i utgangspunktet studerte, men som påvirker en eller begge

“Tredjevariabel er en betegnelse på en mulig faktor som kan forklare at to observerte variabler samvarierer. Om det er tilfelle er det en såkalt spuriøs sammenheng mellom de to observerte variablene.

Hvis en forsker påviser at det er en samvariasjon eller korrelasjon mellom trimming og helse, kan dette fortolkes som at trimming fører til bedre helse, alternativt at det å ha god helse gjør at man er mer tilbøyelig til å trimme. Dette kalles retningsproblemet. Men i denne situasjonen er det også tenkelig at vi har å gjøre med en annen faktor som ikke er målt, for eksempel helsebevissthet, som kan bidra til å forklare både at man trimmer og har god helse. En slik faktor kalles en tredjevariabel.”

(Store Norske Leksikon)

- **Mills kriterium** - Mills 3 krav til hva som må være på plass for å kunne si at X forårsaker Y. X må komme før Y, X og Y må samvariere, tredjevariabler må være utelukket.

“According to John Stuart Mill’s classical formulation, establishing a causal relationship requires three criteria: (a) temporal precedence (i.e., the cause precedes the effect), (b) covariance (i.e., the cause and effect are related), and (c) disqualification of alternative explanations (i.e., no third variable accounts for the observed relationship).

Random-assignment, double-blind, placebo-controlled experimental designs meet all

three criteria when their assumptions are met (e.g., no attrition and perfect compliance)."
[*\(Duckworth, Tsukayama, May\)*](#)

- **metoden om overensstemmelse** - hvis X så Y. En av fem induktive metoder utviklet av John Stuart Mill og publisert i boka "A System of Logic" (1843).

"If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon."
(John Stuart Mill, "A System of Logic", 1843)

- **metoden om ulikhet** - en av fem induktive metoder utviklet av John Stuart Mill og publisert i boka "A System of Logic" (1843)

"The method of difference is a method of comparing cases where a social phenomenon is present with cases where this phenomenon is absent, keeping most circumstances similar. By comparing these cases, researchers are able to identify the single or the few conditions that explain the occurrence of the social phenomenon."
(Encyclopedia of Case Study Research)

"If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or an indispensable part of the cause, of the phenomenon."
(John Stuart Mill, "A System of Logic", 1843)

- **induksjonsproblemet** - uansett hvor mange observasjoner man har av noe, kan man aldri konkludere med at det nødvendigvis er slik, f.eks at svaner nødvendigvis er hvite fordi alle svaner noen har sett til nå er hvite. Observasjoner er ikke-generaliserbare.
- **spuriøs sammenheng** - når to variabler korrelerer uten at det er kausal sammenheng mellom dem, samvariasjonen skyldes en tredjevariabel

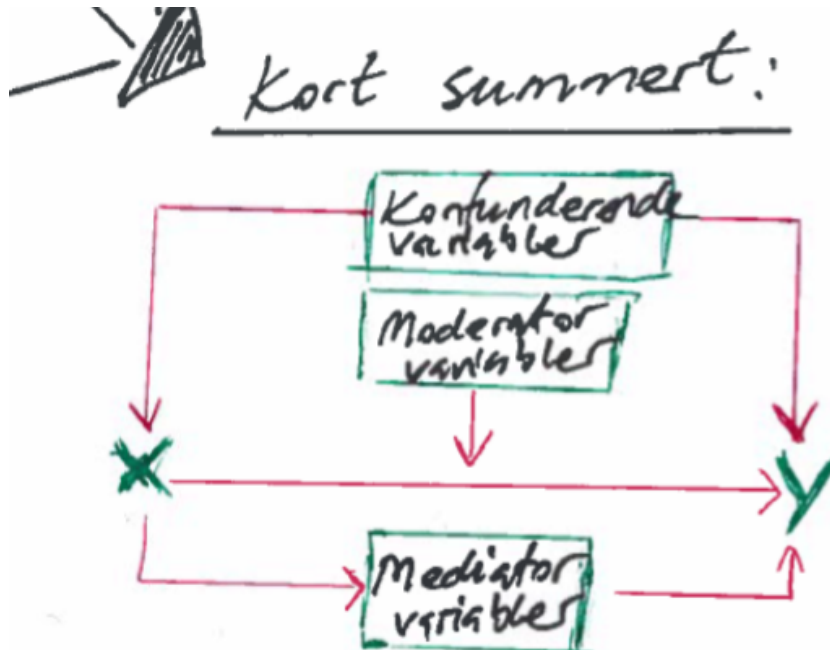
"Spuriøs er en sammenheng som framstår som et årsak–virkning-forhold, men ikke er det. Ved en spuriøs sammenheng er det en bakenforliggende variabel som er årsaken til sammenhengen."
(Store Norske Leksikon)

- **kausalitetsproblemet** - omtrent det samme som induksjonsproblemet. Mest kjent fra Hume, som bemerket at selv om det virker åpenbart for oss at det at biljardball A treffer biljardball B er årsaken til at biljardball B flytter på seg, så kan vi ikke vite det sikkert. En "ekstrem tolkning" av at korrelasjon ikke er lik kausalitet.

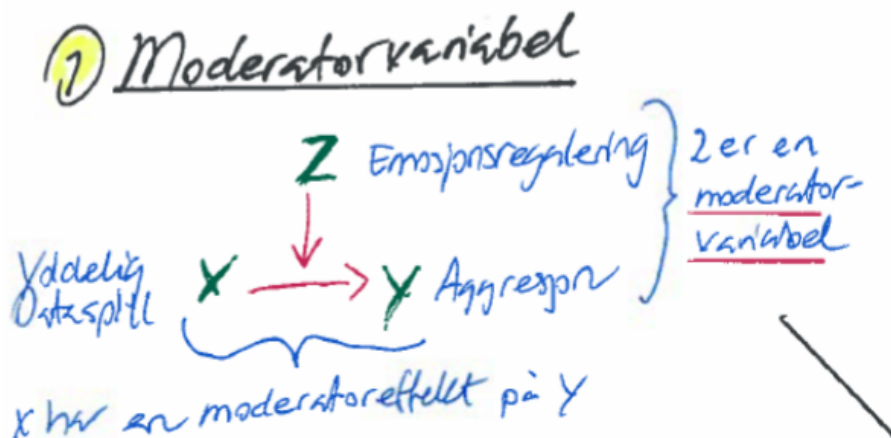
Lengere forklaring:

<https://www.quora.com/What-is-the-problem-of-causality-in-philosophy>

- **kausallmodeller** - ulike modeller for kausale relasjoner mellom X og Y, evt. med innvirkning fra Z



- **moderatorvariabel** - når en variabel Z regulerer X sin påvirkning av Y

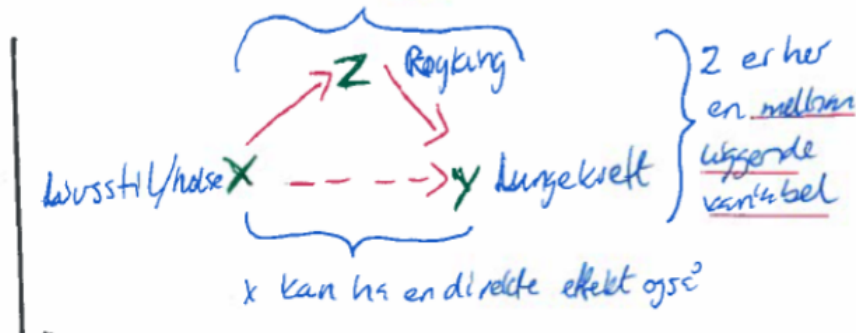


- **mediatorvariabel** - når X hovedsakelig påvirker Y gjennom Z

②

Mediator variabel

x har indirekte effekt på y



- **kontunderende variabel** - når X ikke påvirker Y i det hele tatt, men det er en tredjevariabel Z som fører til samvarians mellom X og Y

③ Kontunderende variabel

Sommer



- **varians** - statistikk som kvantifiserer variasjon i en variabel

$$s^2 = \frac{\sum (X - \bar{X})^2}{N-1}$$

- **kovarians** - statistikk som kvantifiserer samvariasjon mellom to variabler (x og y)

$$Cov_{xy} = \frac{\sum (Y - \bar{Y}) \cdot (X - \bar{X})}{N-1}$$

$$Cov_{xy} = \frac{\sum XY - \frac{\sum Y \sum X}{n}}{n-1}$$

<i>X</i> <i>Depression</i>	<i>Y</i> <i>Hours of sleep</i>	<i>XY</i>
35	7.5	262.5
38	7.5	285
40	8.0	320
42	7.0	294
42	6.5	273
42	5.5	231
44	6.0	264
49	5.0	245
$\bar{X} = 41.5$	$\bar{Y} = 6.625$	
$s_x = 4.14$	$s_y = 1.06$	
$\Sigma X = 332$	$\Sigma Y = 53$	$\Sigma XY = 2174.5$

$$Cov_{xy} = \frac{\sum XY - \frac{\sum Y \sum X}{n}}{n-1}$$

$$Cov_{xy} = \frac{2174.5 - \frac{53 \cdot 332}{8}}{8-1}$$

$$Cov_{xy} = -3.5$$

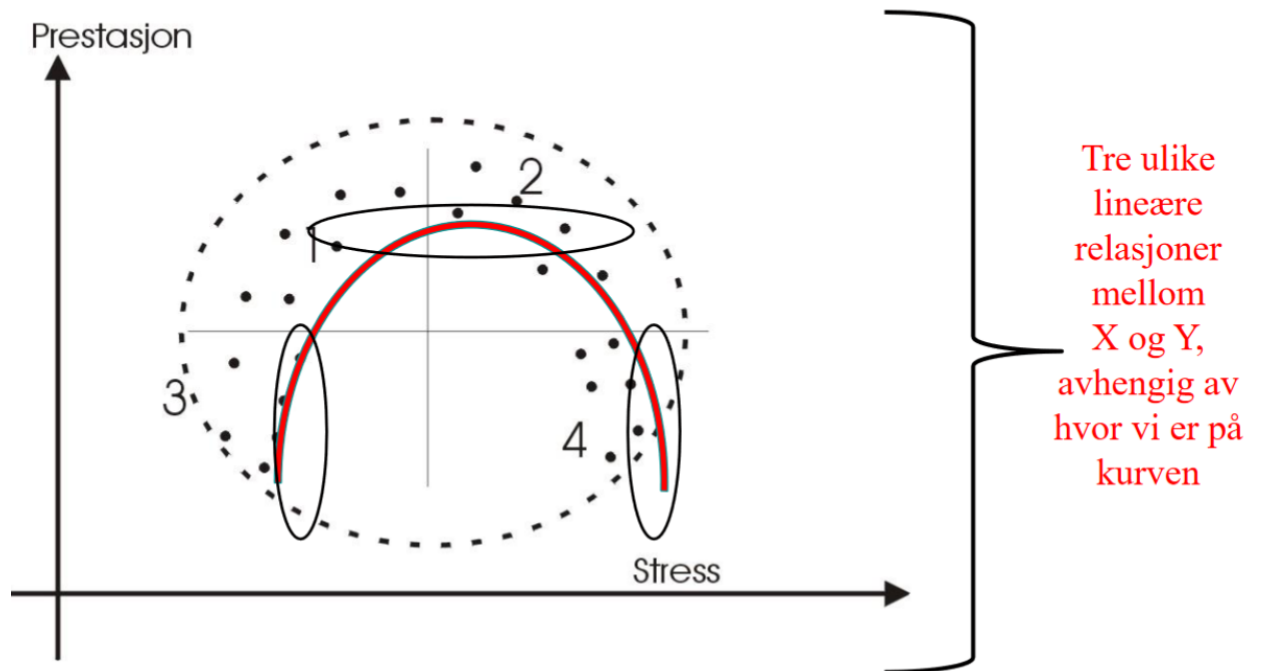
Vi forventer en **negativ lineær sammenheng** mellom søvn og depresjon

30

- **bivariat analyse** - analyse av to variabler og hvordan de forholder seg til hverandre.
- **spredningsdiagram** - "Et spredningsdiagram (også kalt scatter plot eller punktdiagram) viser forholdet mellom to sett av kvantitative verdier. Alle verdiene er plottet inn på diagrammet, plassert i henhold til deres respektive posisjonene langs både x- og y-aksen. De plottede punktene lar deg observere hva slags forhold som finnes mellom de to variablene. F.eks. kan et spredningsdiagram vise hvor mye en film har tjent inn på x-aksen og antall priser den har vunnet på y-aksen. Med dette kan du se hvorvidt filmer som tjener inn mye penger, også vinner mange priser." (Tallgrafikk.no)
- **kurvelineær samvariasjon** - det vi får når X samvarierer positivt med Y inntil et visst punkt, deretter negativt.

Eksempel: Stress inn mot eksamen samvarierer positivt med prestasjon på eksamen inntil et visst punkt, deretter samvarierer de to negativt (både de som stresser 0 og de som

stresser masse får i snitt dårligere karakterer enn de som stresser middels).



- **korrelasjonskoeffisienten (r)** - sier noe om hvor sterk sammenhengen er mellom to variabler. Befinner seg på ordinalnivå (ikke-kvantifiserte forskjeller mellom nivåer), sier derfor ikke noe direkte om forskjeller i styrke.
0 = Ingen relasjon
+1 = Max positiv relasjon
-1 = Max negativ relasjon
- **Pearsons produktmoment korrelasjonskoeffisient** - den mest vanlige formen for korrelasjonskoeffisient, ofte forkortet liten r .

Utgangspunkt:

$$r = \frac{COV_{xy}}{S_x S_y}$$

r = Pearsons produktmomentkorrelasjonskoeffisient

COV_{xy} = Kovariansen til variabel X og Y

S_x = Standardavviket til variabel X

S_y = Standardavviket til variabel Y

$$Korrelasjon_{xy} = \frac{\frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\cancel{n}}}{\sqrt{\frac{\Sigma(X - \bar{X})^2}{\cancel{n}}} \sqrt{\frac{\Sigma(Y - \bar{Y})^2}{\cancel{n}}}}$$

Kovariansen

Standardavviket til X Standardavviket til Y

Definisjonsformelen:

$$r_{xy} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2 \Sigma(Y - \bar{Y})^2}}$$

Husk: Ved å stryke bort N får vi definisjonsformelen

$$r_{xy} = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{n}}{\sqrt{[\Sigma X^2 - \frac{(\Sigma X)^2}{n}] [\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}]}}$$

r_{xy} = korrelasjonen mellom x og y
 n = Utvalgsstørrelse
 X = individuell skåre på variabel X
 Y = Individuell skåre på variabel Y

*For spesielt interesserte:
 Her kan vi ikke
 stryke bort n, i stedet
 ganges n i teller og
 kvadratroten av n² i nevner³³

$$r_{xy} = \frac{n \Sigma XY - \Sigma X \Sigma Y}{\sqrt{[n \Sigma X^2 - (\Sigma X)^2] [n \Sigma Y^2 - (\Sigma Y)^2]}}$$

r_{xy} = korrelasjonen mellom x og y
 n = utvalgssstørrelse
 X = individuell skåre på variabel X
 Y = Individuell skåre på variabel Y

<i>X</i> <i>Depression</i>	<i>Y</i> <i>Hours of sleep</i>	<i>XY</i>
35	7.5	262.5
38	7.5	285
40	8.0	320
42	7.0	294
42	6.5	273
42	5.5	231
44	6.0	264
49	5.0	245
$\bar{X} = 41.5$	$\bar{Y} = 6.625$	
$s_x = 4.14$	$s_y = 1.06$	
$\Sigma X = 332$	$\Sigma Y = 53$	$\Sigma XY = 2174.5$

Forts. forrige eksempel:

$$Cov_{xy} = -3.5$$

$$r_{xy} = \frac{Cov_{xy}}{s_x \cdot s_y}$$

$$r_{xy} = \frac{-3.5}{4.14 \cdot 1.06} = -0.80$$

Er dette en sterk eller svak sammenheng?

Eksempel 2 med Pearson r på egenhånd

$$r_{xy} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2] [n \sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{(10)(506) - (55)(103)}{\sqrt{[(10)(385) - (55)^2] [(10)(1401) - (103)^2]}}$$

$$= \frac{5060 - 5665}{\sqrt{3850 - 3025} \sqrt{14010 - 10609}} = \frac{-605}{\sqrt{825} \sqrt{3401}} = \frac{-605}{(28.723)(58.318)} = \frac{-605}{1675.0679} = -0.36$$

37

Korrelasjon mellom regning og lesing					
Student	Regning	Lesing	X ²	Y ²	XY
1	3	11	9	121	33
2	7	1	49	1	7
3	2	19	4	361	38
4	9	5	81	25	45
5	8	17	64	289	136
6	4	3	16	9	12
7	1	15	1	225	15
8	10	9	100	81	90
9	6	15	36	225	90
10	5	8	25	64	40
Sum	55	103	385	1401	506

Konvensjoner for styrke:

0,00 - 0,30: Liten eller ingen korrelasjon

0,30 - 0,50: Lav korrelasjon

0,50 - 0,70: Moderat korrelasjon

0,70 - 0,90: Høy korrelasjon

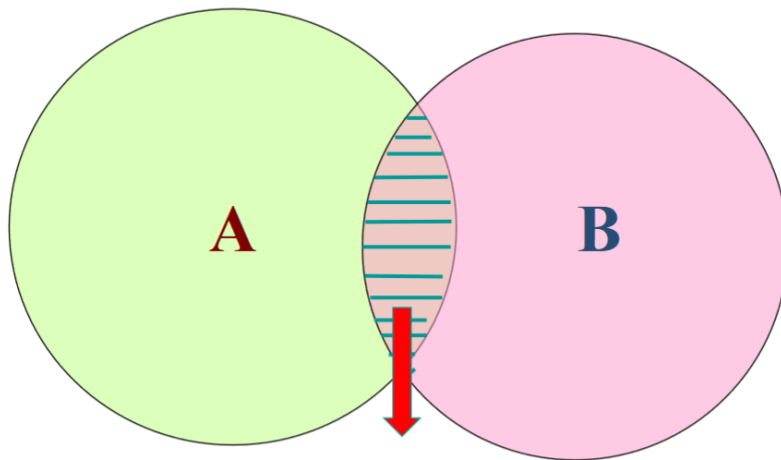
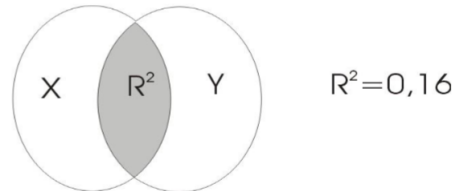
0,90 - 1,00: Svært høy korrelasjon

- **delt varians (r^2 - coefficient of determination)** - Andel varians i en variabel som kan forklares av den lineære relasjonen mellom x og y.

En r på 0,80 uttrykker ikke dobbelt så sterk sammenheng som en r på 0,40.



Opphøyer vi derimot r til r^2 , får vi et uttrykk for sammenhengens styrke.



A) Karakterer på videregående og B) karakterer på psykologistudiet:
 $r = 0,35 \rightarrow r^2 = 0,12 \rightarrow$ Det betyr 12% delt varians (nokså lite).

Altså: Det er andre variabler som påvirker hvordan dere vil prestere her på universitetet ☺