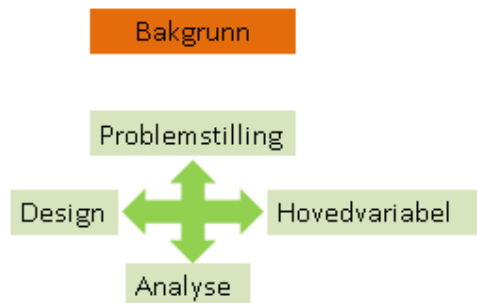
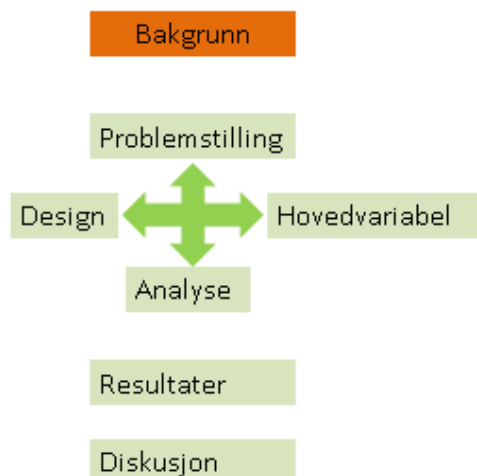


Planlegging av studier, design, og innsamling av data

Forskningsprosjektet



Forskningsprosjektet



Bakgrunn

Problemstilling

Ch 3.1-3.3+3.5

Design: **Plan** for å besvare
problemstillingen

Eksperimentelt
RCT

Observasjonelt
«Sampling»
Case-Control
Cohort

Ch 1.1-1.3

Main outcome: Tallfesting av
hovedinteresse

Kategorisk
Type sitrusfrukt
Kjønn
Keisersnitt
Kreft

Glidende
overgang

Kontinuerlig
Appelsinvekt
Skrelletid
BMI
Blodsukker
CRP

Statistisk analyse

Eksperimentelle design

I eksperimentelle design er det en **intervensjon**, eller behandling

Vi undersøker **responsen** på behandlingen, enten ved å sammenligne en gruppe som har fått en type behandling med en som ikke har fått det (en **kontrollgruppe**), ved å sammenligne flere grupper og flere typer behandling, eller ulike grader av behandling, såkalt dose – respons.

Enhetene vi studerer kan være bakteriekulturer, celler, stein, kjemiske løsninger, dyr, mennesker,...

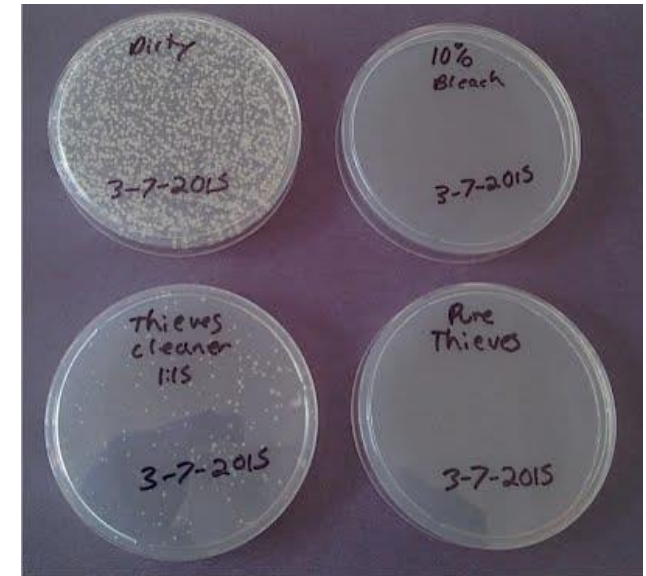
I eksperimentelle design jakter vi på **årsakssammenhenger**.

Vi kaller variabelen som tallfester responsen for «**Responsvariabel**» (eller endepunkt, avhengig variabel, y, utfall, outcome, response variable, dependent variable, exogen variable),

Variabelen som tallfester intervensjonen kalles «**Forklaringsvariabel**» (eller kovariat, uavhengig variabel, explanatory variable, independent variable).

En kategorisk forklaringsvariabel kalles ofte en *faktor*.

I eksperimentelle studier vil vi sørge for at gruppene er så like som mulig, slik at den eneste forskjellen er den behandlingen vi utsetter dem for.



Snu deg mot sidemannen:

Tran i svangerskapet. Trankapsler eller kapsler med soyaolje. Ingrid Helland, Rikshospitalet

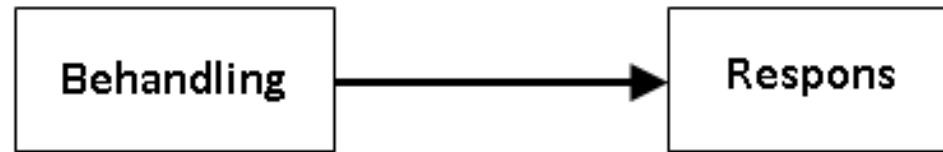


**Forklaringsvariabel:
Respons:**

Vaksine mot smittsom hjernehinnebetennelse



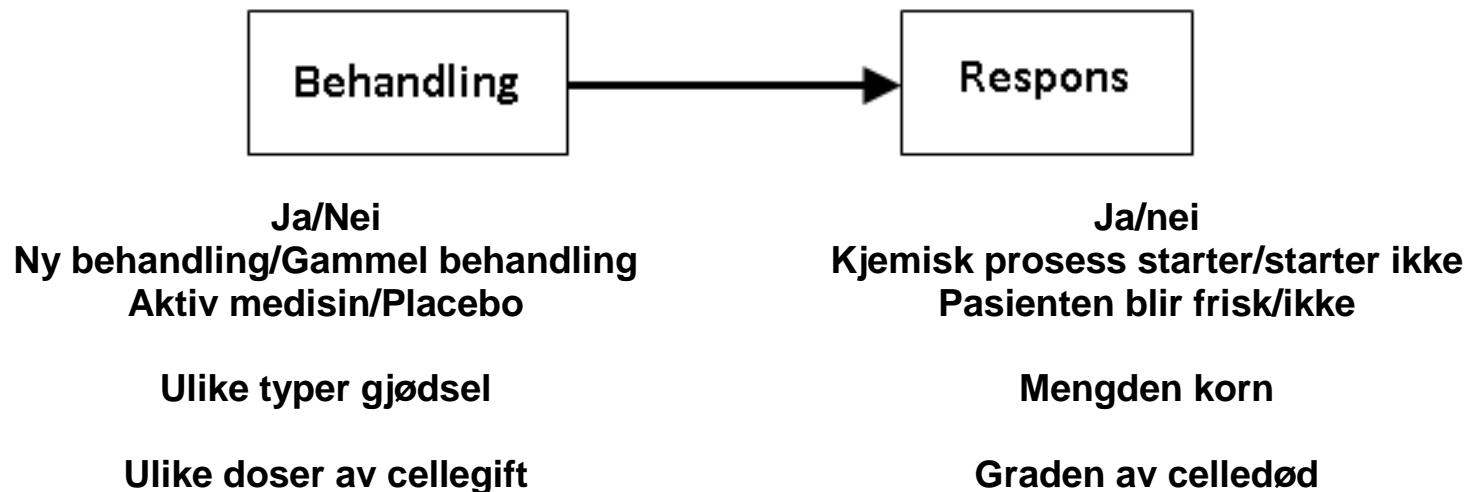
**Forklaringsvariabel:
Respons:**



Hvis gruppene er sammenlignbare, og vi analyserer sammenhengen mellom variabelen som beskriver behandlingen, og variabelen som beskriver responsen, kan vi tallfeste effekten av behandlingen.

Analyser der vi finner sammenhenger mellom to variabler, kalles ofte bivariate analyser.

Både forklaringsvariabelen og responsvariabelen kan være kategorisk eller kontinuerlig (eller en annen type data):



Ulike typer data krever ulike analyser, akkurat som for deskriptiv statistikk. Se analyseoversikt bakerst i disse notatene.

Hvordan oppnår vi sammenlignbare grupper?

Vi kan **randomisere**, altså trekke lodd om hvilke enheter som skal være i de ulike gruppene.

Hvorfor virker randomisering?

Virker randomisering alltid?

Randomisering i praksis: Gjør ukesoppgavene.

Randomisering

- ✓ på individnivå «completely randomized»
- ✓ eller på gruppenivå: Blokk-randomisering

Stratifisering

Viktigheten av kontrollgruppe

Placeboeffekt

Hawthorne-effekt

Det randomiserte, kontrollerte forsøket: Randomized, Controlled Trial, RCT

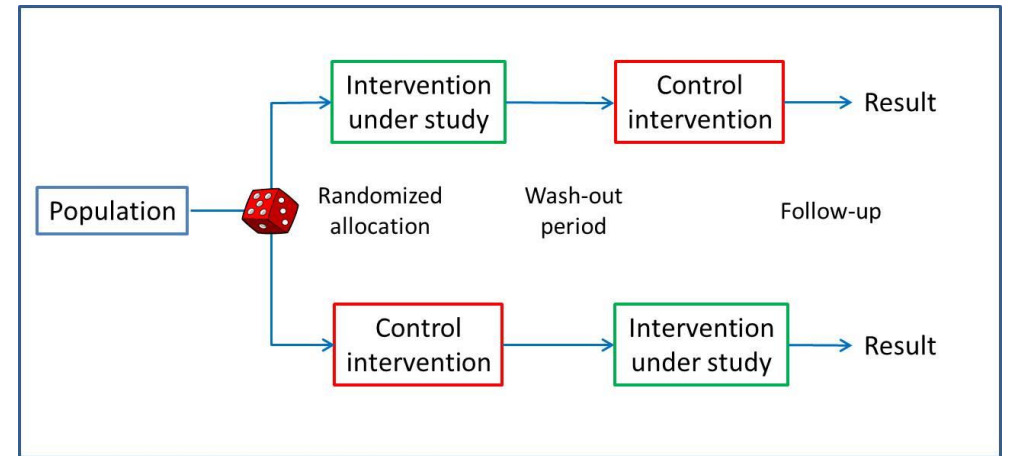
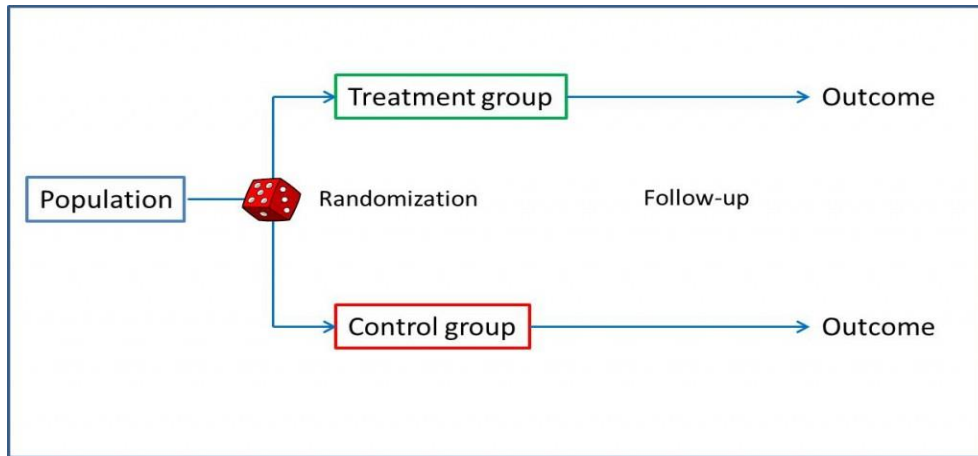
Blinding

Matchede par

Ulemper: Må ha god kunnskap om det man studerer for å vite hvilke kriterier man skal matche på.

Har vi mange nok med i studien? Utvalgsberegning er viktig. Detaljer kommer senere i kurset.

Behandlingsperioder: Parallellgruppe-design eller cross-over-design



Hentet fra <http://www.cienciasinseso.com/en/>

Bias. På norsk: Skjevhet

Når studiedesignet gir en systematisk feil i resultatene.

Eksempler: Eksperimentelt design uten kontrollgruppe: Placeboeffekt?
Ekskluderer dem som får bivirkninger og som ikke fortsetter med medikamentet: Intention-to-treat.

CONSORT statement/sjekkliste er retningslinjer for den som skal rapportere resultater fra RCT.

Snu deg mot sidemannen:

Tredemølleterapi etter hjerneslag, sammenlignet med tradisjonell fysioterapi. Ellen Høyer, Sunnaas



Forklaringsvariabel:

Responsvariabel:

Hvordan randomisere? Hvor mange må vi ha med?

Behandlingsreiser til Beitostølen eller Syden, for pasienter med nevromuskulære sykdommer, noen i rullestol.



Forklaringsvariabel:

Respons:

Randomisering?

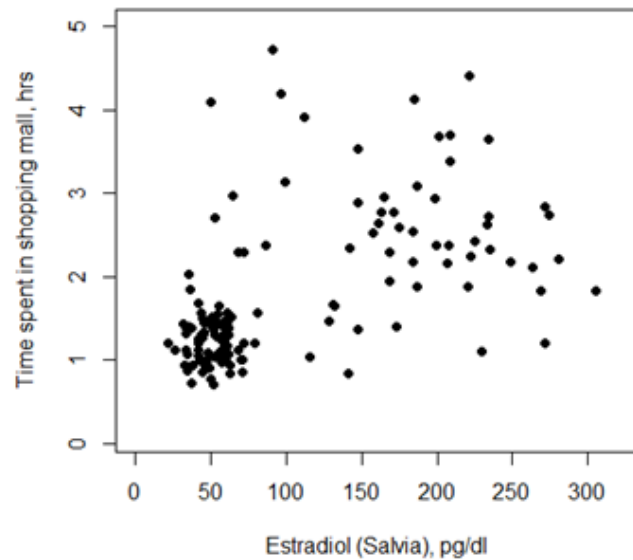
Stratifisering?

Hvor mange må vi ha med?

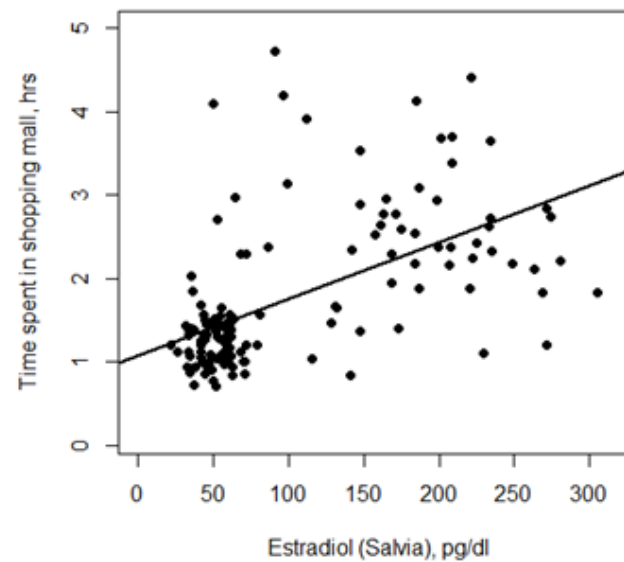
Observasjonelle design



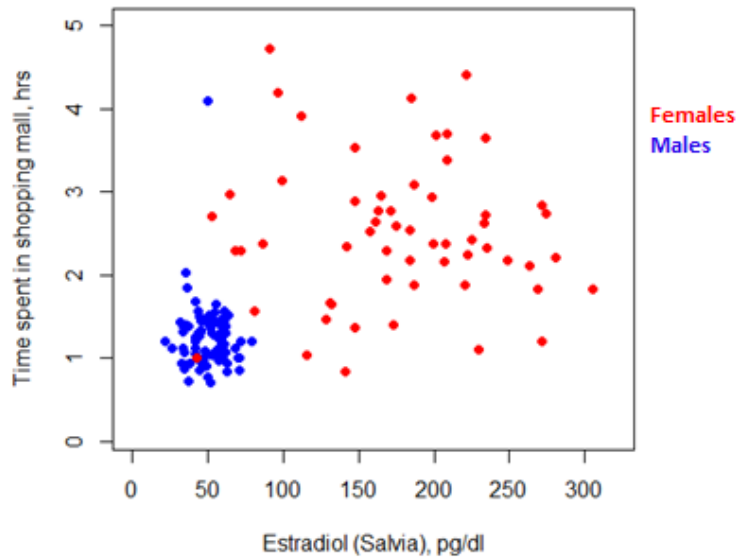
Eksempel: Tid brukt på shopping vs østrogen-nivåer. Simulerte data.



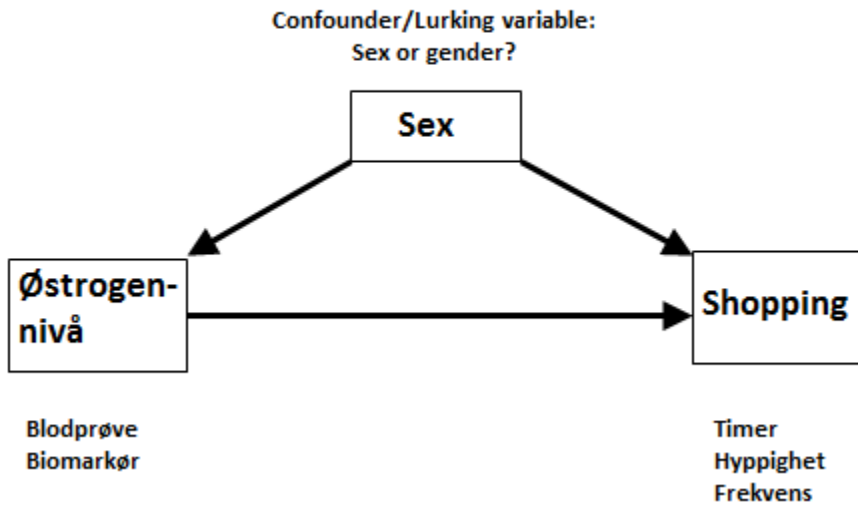
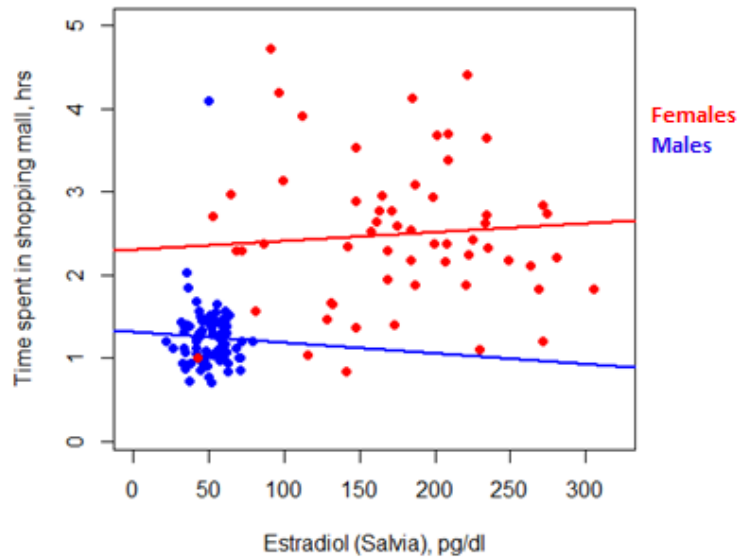
Eksempel: Tid brukt på shopping vs østrogen-nivåer. Simulerte data.



Eksempel: Tid brukt på shopping vs østrogen-nivåer. Simulerte data.



Eksempel: Tid brukt på shopping vs østrogen-nivåer. Simulerte data.



ResponsvARIABLE:

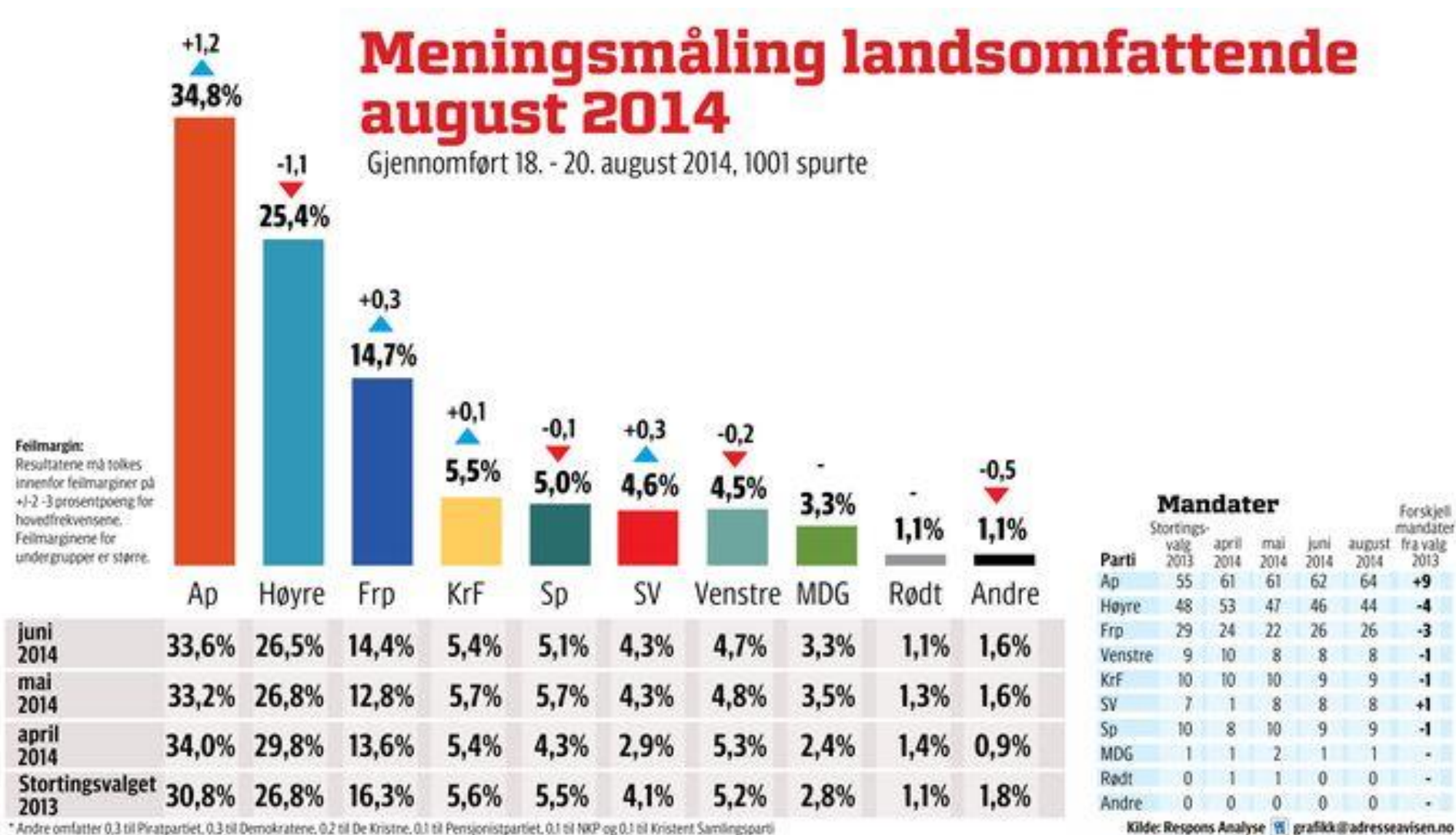
Forklaringsvariable?

Eksponeeringsvariable

Årsakssammenheng?

I observasjonelle design er det ingen inngripen fra forskeren (ingen intervensjon, ingen behandling), kun observasjon

Observasjonelle design spenner fra meningsmålinger, observasjoner av stjernehimlen og telling av hval langs norskekysten, til store oppfølgingsstudier av kreftpasienter.



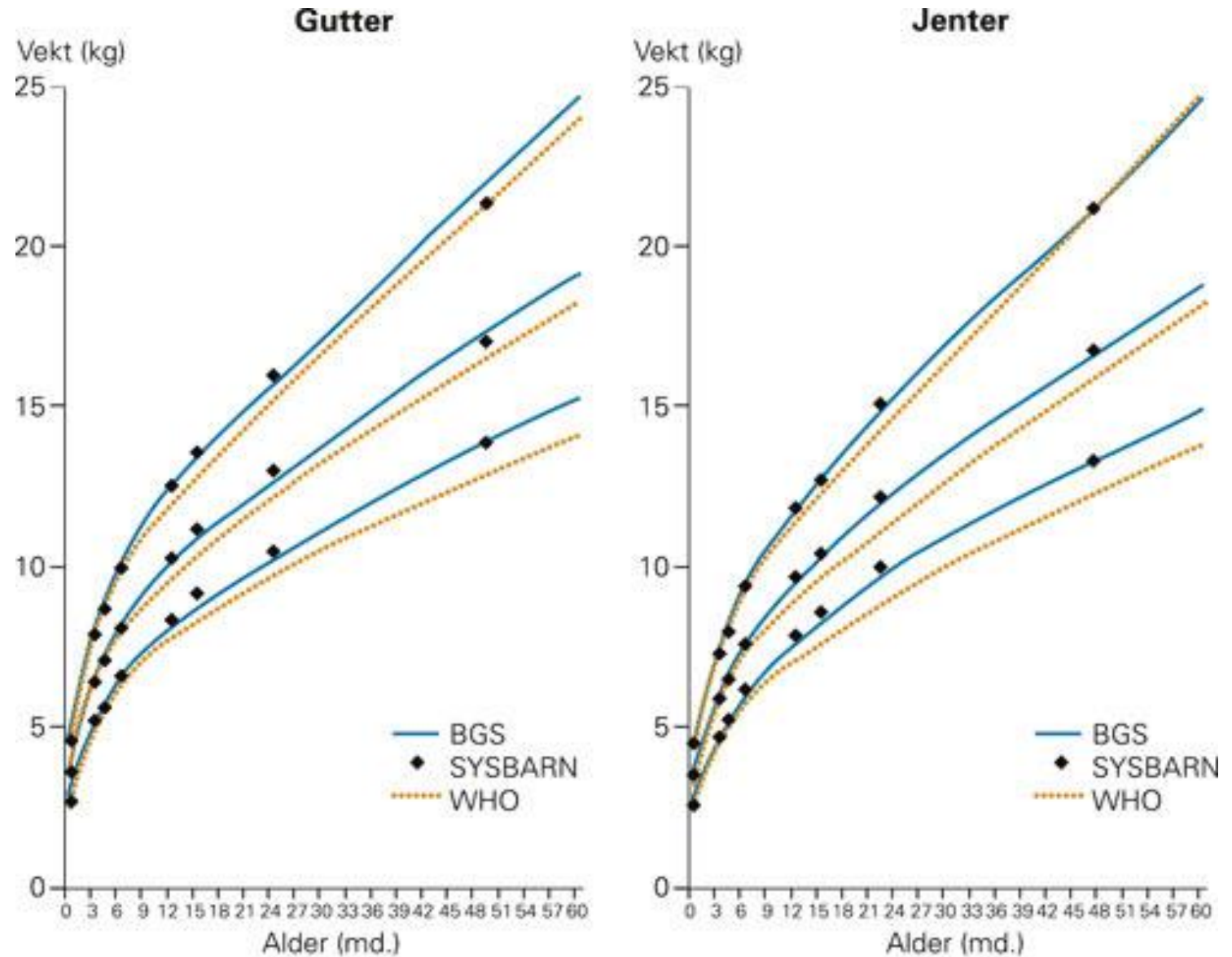
* Andre omfatter 0,3 til Piratpartiet, 0,3 til Demokratene, 0,2 til De Kristne, 0,1 til Pensjonistpartiet, 0,1 til NKP og 0,1 til Kristent Samlingsparti

Eksempel: Vekstkurver.

Sammenlikning av vekt mot alder mellom nye norske vekstkurver fra Vekststudien i Bergen (BGS), SYSBARN-undersøkelsen og WHO's internasjonale vekstkurver for gutter og jenter i alderen 0 - 5 år.

Linjene viser 2,5-, 50- og 97,5-prosentilene.

Figuren er hentet fra Tidsskr Nor Legeforen 2009; 129: 281-6



Bruk kurven til å finne ut hvor tunge 95% av 4-årige gutter forventes å være, og hvor tunge 95% av 4-årige jenter forventes å være. Bruk dette sammen med figuren til å finne gjennomsnitt og sd for vekten til 4-årige gutter og jenter.

Utvalgsundersøkelse: Sample survey

Det er kostbart/vanskelig å undersøke hele populasjonen. I stedet trekkes et representativt **utvalg** fra **populasjonen**, og undersøker dette.

Idé:

Sånn som et er i utvalget, er det omtrent i populasjonen også.

Utvalg = «Sample»



Undersøker og beskriver disse



...for å si noe om alle nordmenn

Undersøker og beskriver disse



...for å si noe om alle spedbarn

Undersøker og beskriver disse



...for å si noe om alle eldre

Vi bruker gjennomsnittet både til å
beskrive dem vi har undersøkt



...og til å si noe generelt om dem vi
ikke har undersøkt; populasjonen

Utvalgsdesign

Frivillig deltakelse: Voluntary response

Simple random sample (SRS): Jfr Completely randomized design

Stratifisert utvalg

Cluster design

Hva er populasjonen, og hvilken type utvalg er dette:

- ✓ **Innringere forteller sin mening om temaet som debatteres på et tv-program**
- ✓ **Meningsmålingsbyrå trekker 1000 tilfeldige personer fra telefonkatalogen og gjennomfører en spørreundersøkelse per telefon**
- ✓ **Meningsmålingsbyrå trekker 1000 tilfeldige personer fra folkeregisteret og sender ut spørreskjemaer per post**
- ✓ **Forskere på Rikshospitalet inviterer gravide kvinner til å delta i studie om fødselskomplikasjoner**
- ✓ **Forskere som er interessert i spiseforstyrrelser trekker tilfeldig 100 jenter og 100 gutter fra et idrettsmiljø, og fra et musikkmiljø**
- ✓ **Forskere inviterer tenåringer med en spesiell sykdom til å delta i en studie om fritidsaktiviteter, sammen med sine fem nærmeste venner.**

Responsrate og representativitet

Bias. På norsk: Skjevhet

Når studiedesignet gir en systematisk feil i resultatene.

Er utvalget representativt?

STROBE statement/sjekkliste er retningslinjer for den som skal rapportere resultater fra observasjonelle studier.

Reliabilitet

Repetierbarhet av målinger/kategoriseringer

Validitet

Metodisk kvalitet: Er resultatene preget av bias?

Er tallene gode beskrivelser av det vi ønsker å fange opp?

Er utvalgene representative?

Er de statistiske metodene brukt riktig?

Kan resultatene generaliseres?

Etikk

Spesielt viktig når det samles inn data om mennesker

Etiske komiteer og instanser:

- ✓ Regionale forskningsetiske komiteer (REK)
- ✓ Norsk samfunnsvitenskapelig datatjeneste (NSD)
- ✓ Datatilsynet
- ✓ Forskningsutvalg
- ✓ Styringsgrupper

Skal beskytte enkeltpersoners rettigheter og velferd.

Skal vurdere den metodiske kvaliteten på studier

Gjør deg kjent med de organene som gjelder i ditt eget fagfelt og på din egen institusjon.

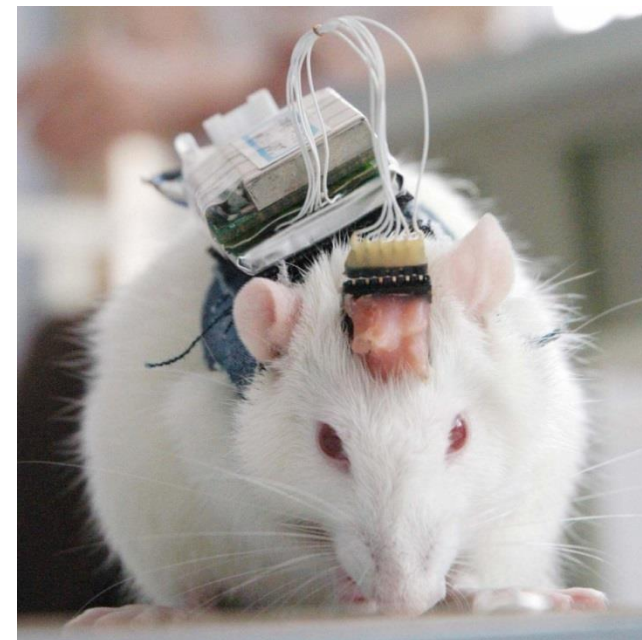
Informert samtykke

Informere om potensielle skadevirkninger, tidsbruk, stress, inngripen i privatliv etc

Konfidensielle data vs anonymiserte data

Etikk i kliniske studier

- ✓ Kan man sammenligne med placebo?
- ✓ Sammenligne med eksisterende praksis?
- ✓ Non-inferiority studies
- ✓ Dyrevelferd



Diskusjonsoppgave: IQ

En forsker er interessert i universitetsstudenters intelligens. Hun ber studenter i et innføringskurs i statistikk om å ta en IQ-test på nett. Responsrate: 58 av de 137 påmeldte i kurset svarer: 42% respondenter.

Eksperimentelt eller observasjonelt design?

Populasjon?

Hvilken type utvalg?

Hvor representativt er dette?

Hvor reliabel er en slik test?

Hvor valid er en slik test?

Hvilke etiske betraktninger burde forskeren gjøre seg?

Er denne datainnsamlingen i det hele tatt lovlig?

Diskusjonsoppgave: Botox mot migrene. Ukesoppgave denne uken.

Anectotal data:

Middelaldrende kvinne på Manhattan går til plastikk-kirurgen sin for å glatte ut sinnarynken. Opplever å bli kvitt migrene. Forteller det til venninne, som har samme erfaring. Intervjues i kvinneblad.

Available data:

En spesifikk lege har diverse data tilgjengelig i sine pasientregistre

Observasjonell studie:

Spørreskjema om «korrigerende kirurgi» og overgangsplager/migrene sendes ut.

Snu deg til sidemannen:

- 1) Hvordan velger man hvem som bør inviteres til å delta i studien?**
- 2) Diskutér styrker og svakheter med dette studiedesignet.**



RCT-studie:

Hva er en RCT-studie?

Diskutér med sidemannen hvordan en slik studie bør gjøres.

Diskutér styrker og svakheter med det studiedesignet du har valgt.

Analyseoversikt

Versjon 1

Hva skal vi gjøre?

Oppsummere/presentere/
beskrive data

Inferens

Beregne
konfidens-intervall

Hypotesetesting:

Sammenligne to eller flere grupper
(Undersøke om det er sammenheng
mellom to variabler, der den ene er en
en kategorisk variabel)

H_0 : Gruppene er like
Konfidensintervall kan også brukes

Undersøke om det er sammenheng
mellom to variabler, der den ene er en
kontinuerlig variabel
 H_0 : Ingen sammenheng

Analysere forskjell på par av data
 H_0 : Ingen forskjell

Kvantifisere samsvar.
Utgangspunkt: Det er samsvar
Reliabilitet

Analysere sammenheng mellom en
responsvariabel og en eller flere
kovariater (forklaringsvariabler eller
prediktorer)
 H_0 : Ingen effekt/ingen prediktiv verdi

H v i l k e n t y p e d a t a h a r v i ?

Kategoriske data

Tabeller, andeler
Søyle(stolpe)/kakediagram

95% KI for p , som er
populasjonens andel

$$\hat{p} \pm 1.96 \cdot SE(\hat{p})$$

«Tabellanalyse»
Krysstabeller
Andeler

Differanse av andeler m/ 95% KI
Relativ risiko (RR) m/ 95% KI
Odds-ratio (OR) m/ 95% KI
Pearsons χ^2 -test (kji-kvadrat-test)

To kategorier: t -test /Wsr-test
Flere kategorier: ANOVA/KW

Krysstabeller
Andelen samsvar & McNemars test

Krysstabeller
Cohens kappa

Binær responsvariabel:
Logistisk regresjon
Effekt mål: OR m/ 95% KI
 H_0 : OR = 1

Kontinuerlige data

Skjeve



Median, kvartiler

95% KI for
populasjonens median

Tabeller? Bootstrapping?

To grupper:
Median, kv i hver gruppe
Wilcoxon rank sum test

Flere grupper:
Median, kv i hver gruppe
Kruskal-Wallis test
Posthoc-tester=Wrs-tester

Spearman's
korrelasjonskoeffisient

Skjevfordelte differanser:
Wilcoxon signed-rank test

Symmetriske



\bar{x} , SD

95% KI for μ , som er
populasjonens gjennomsnitt

$$\bar{x} \pm 1.96 \cdot SE(\bar{x})$$

To grupper:
 \bar{x} , SD i hver gruppe
Diff. av gj.sn. m/ 95% KI
To-utvalgs t -test

Flere grupper:
 \bar{x} , SD i hver gruppe
Enveis ANOVA. Hvis $p < 0.05$
Posthoc-tester = t -tester

Pearsons
korrelasjonskoeffisient r

Normalfordelte differanser:
Paret t -test

Scatterplott med $y=x$,
Bland-Altman-plott
ICC

Kontinuerlig responsvariabel:
Lineær regresjon
Effekt mål: B
 H_0 : B = 0
(Normalfordelte residualer)

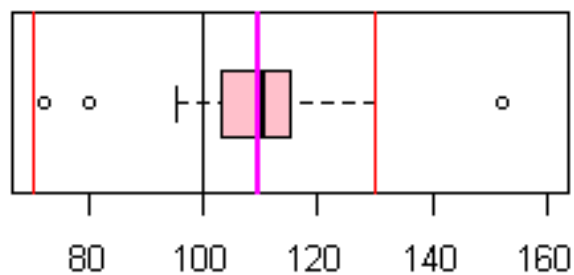
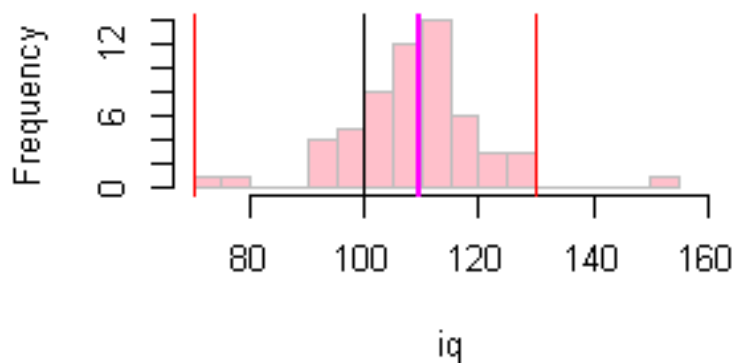
Beskrive enkle (bivariate) sammenhenger mellom variabler

Demonstrasjon i Rstudio:

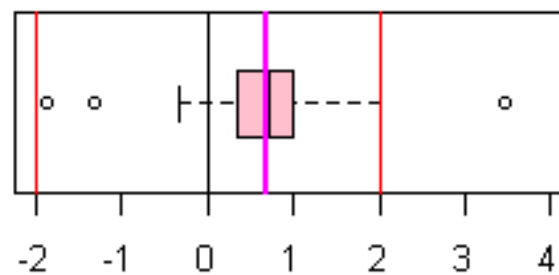
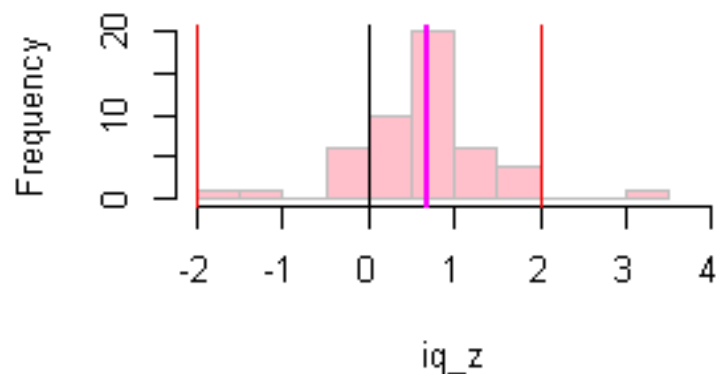
IQdata, z-scorer, kjønn, vgs-snitt, studietid, foreldres utdanningsnivå.

Vi starter med deskriptiv statistikk

Histogram of iq



Histogram of iq_z



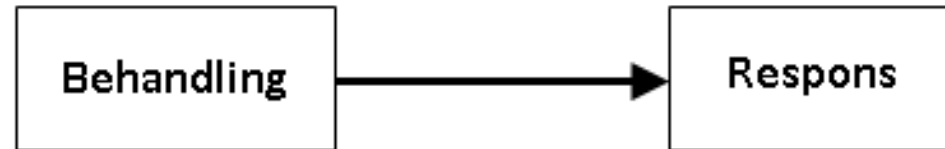
Sammenhenger mellom **to variabler** kalles **bivariate sammenhenger**, og analyser av sammenhenger mellom to variabler, kalles ofte **bivariate analyser**.

Fortunately, statistical analysis of several-variable data builds on the tools used for examining individual variables. The principles that guide our work also remain the same:

- Start with a graphical display of the data.
- Look for overall patterns and deviations from those patterns.
- Based on what you see, use numerical summaries to describe specific aspects of the data.

- Beskrive sammenheng med **figur**
- Beskrive sammenheng med **ord**
- Beskrive sammenheng med **oppsummeringstall**

Bivariate analyser og eksperimentelle design



Begge variablene kan være kategorisk eller kontinuerlig (eller en annen type data):

Ja/Nei
Ny behandling/Gammel
behandling
Aktiv medisin/Placebo
Ulike typer gjødsel
Ulike doser av cellegift

Ja/nei
Kjemisk prosess
starter/starter ikke
Pasienten blir frisk/ikke
Mengden korn
Graden av celledød

Ulike typer data medfører at det er ulike måter å beskrive sammenhengen på, og ulike tall for å oppsummere den, akkurat som for deskriptiv statistikk.

Analyseoversikt

Versjon 1

Hva skal vi gjøre?

Oppsummere/presentere/
beskrive data

Hvilken type data har vi?

Kategoriske data

Tabeller, andeler
Søyle(stolpe)/kakediagram

Skjeve



Median, kvartiler

Kontinuerlige data



Symmetriske

\bar{x} , SD

Inferens

Beregne
konfidens-intervall

95% KI for p , som er
populasjonens andel

$$\hat{p} \pm 1.96 \cdot SE(\hat{p})$$

95% KI for
populasjonens median

Tabeller? Bootstrapping?

95% KI for μ , som er
populasjonens gjennomsnitt

$$\bar{x} \pm 1.96 \cdot SE(\bar{x})$$

Hypotesetesting:

Sammenligne to eller flere grupper
(Undersøke om det er sammenheng
mellom to variabler, der den ene er en
en kategorisk variabel)

H_0 : Gruppene er like

Konfidensintervall kan også brukes

«Tabellanalyse»
Krysstabeller
Andeler

Differanse av andeler m/ 95% KI

Relativ risiko (RR) m/ 95% KI

Odds-ratio (OR) m/ 95% KI

Pearsons χ^2 -test (kji-kvadrat-test)

To grupper:
Median, kv i hver gruppe
Wilcoxon rank sum test



Flere grupper:
Median, kv i hver gruppe
Kruskal-Wallis test
Posthoc-tester=Wrs-tester



To grupper:
 \bar{x} , SD i hver gruppe
Diff. av gj.sn. m/ 95%KI
To-utvalgs t -test



Flere grupper:
 \bar{x} , SD i hver gruppe
Enveis ANOVA. Hvis $p < 0.05$
Posthoc-tester = t -tester



Undersøke om det er sammenheng
mellom to variabler, der den ene er en
kontinuerlig variabel

H_0 : Ingen sammenheng

To kategorier: t -test /Wsr-test
Flere kategorier: ANOVA/KW

Spearman's
korrelasjonskoeffisient

Pearson's
korrelasjonskoeffisient r

Analysere samsvar mellom par av data
 H_0 : Ingen forskjell

Krysstabeller
Andelen samsvar & McNemars test

Skjevfordelte differanser:
Wilcoxon signed-rank test

Normalfordelte differanser:
Paret t -test

Kvantifisere samsvar.
Utgangspunkt: Det er samsvar
Reliabilitet

Krysstabeller
Cohens kappa

Scatterplott med $y=x$,
Bland-Altman-plott
ICC

Analysere sammenheng mellom en
responsvariabel og en eller flere
kovariater (forklaringsvariabler eller
prediktorer)
 H_0 : Ingen effekt/ingen prediktiv verdi

Binær responsvariabel:
Logistisk regresjon
Effekt mål: OR m/ 95% KI
 H_0 : OR = 1

Kontinuerlig responsvariabel:
Lineær regresjon
Effekt mål: B
 H_0 : B = 0
(Normalfordelte residualer)

Kategorisk variabel mot kategorisk variabel (Ch 2.6)

(Husk at både ordinale variabler (ordnede kategorier) og diskrete variabler (telldata) av og til må behandles som kategoriske variabler, og av og til som kontinuerlige variabler.)

Beskrive med **figur**: Krysstabell (altså frekvenstabell for hver gruppe)
 Kalles også Toveis-tabell, kontingenstabell, krysstabell

Beskrive med **ord**:

Beskrive med **oppsummeringstall**:

Eksempel:

Blodtype og yrke registrert for 160 RH-ansatte ved ansatte-inngangen mellom 8 og 9 en kald høstdag:

		Yrke				Tot
		Lege	Spl	Adm	Stat	
Blod- type	A	9	20	55	0	84
	AB	1	3	5	1	10
	B	2	6	17	0	25
	0	5	11	33	2	51
Tot		17	40	110	3	160

Snu deg til sidemannen og kommenter totaltallene og enkelttallene og hva de betyr

Deskriptiv statistikk: Fordelingen til verdiene til enkeltvariabler

Bivariate sammenhenger: Felles fordeling (joint distribution)

Marginale fordelinger (det er to av dem)

Betinget fordeling

Er vi interessert i de observerte antallene eller i (prosent)andelene?

Eksempel: Resultater fra to forskjellige meningsmålinger

		Meningsmåling		Total
		1 VG	2 Aftenposten	
stemme_til	1 FrP	272	161	433
	2 Ikke FrP	734	569	1303
Total		1006	730	1736

Snu deg til sidemannen og kommenter eventuelle endringer og hva du tenker om dette.

Eksempel: Risiko for kolektomi i løpet av de 5 første årene etter diagnose for pasienter med Ulcerøs kolitt. Prospektiv studie.

		Røykestatus		Total
		1 Røyker	2 Ikke-røyker	
Kolektomi	1 Kolektomert	2	40	42
	2 Ikke kol	59	251	310
Total		61	291	352

Snu deg til sidemannen og kommenter eventuelle sammenhenger og hva du tenker om dette.

Eksempel: Risiko for uforklart fosterdød for førstegangsfødende sammenlignet med de som har født før. Retrospektiv studie.

		Røyking		Total
		1 10 eller flere sigaretter daglig	2 under 10 sigaretter daglig	
Fødselsgruppe	1 Dødfødsel	16	7	23
	2 Kontroll, levendefødt	41	81	122
Total		57	88	145

Snu deg til sidemannen og kommenter eventuelle sammenhenger og hva du tenker om dette.

Oppsummeringstall for sammenheng mellom to kategoriske variabler:

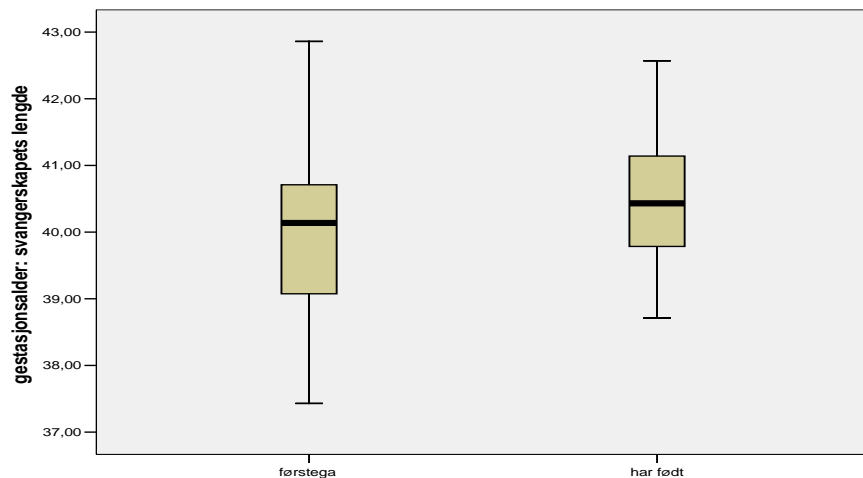
Tallene i kysstabellene, evt uttrykt i (prosent)andeler

Odds ratio (OR), Relativ risiko (RR)

Kontinuerlig variabel mot kategorisk variabel (eller kategorisk mot kontinuerlig)

Beskrive med **figur**: Boksplott for hver gruppe
(Side 97 + Ch1.3: «Side-by-side boxplot»)

Eksempel: Svangerskapslengde for førstegangsfødende og de som har født før

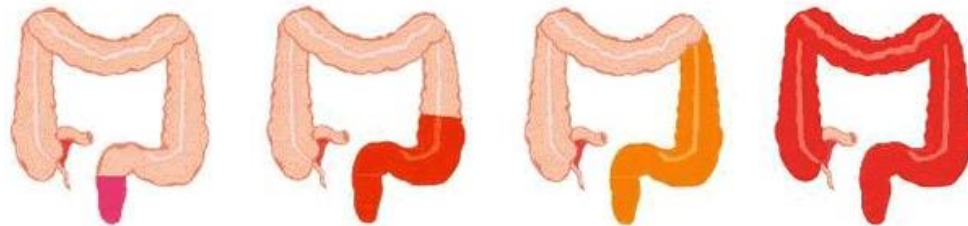
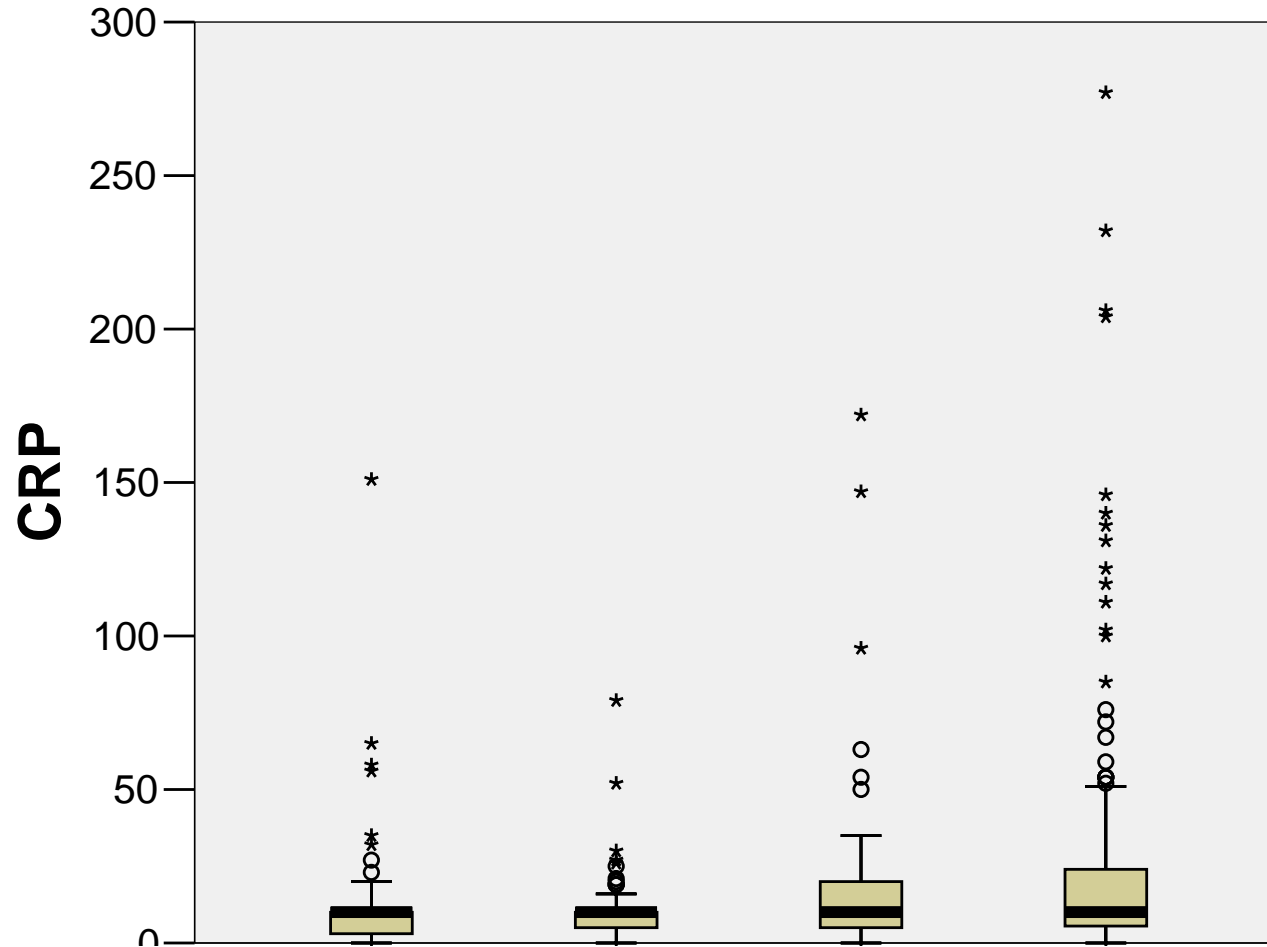


Beskrive med **ord**:

Beskrive med **oppsummeringstall**:

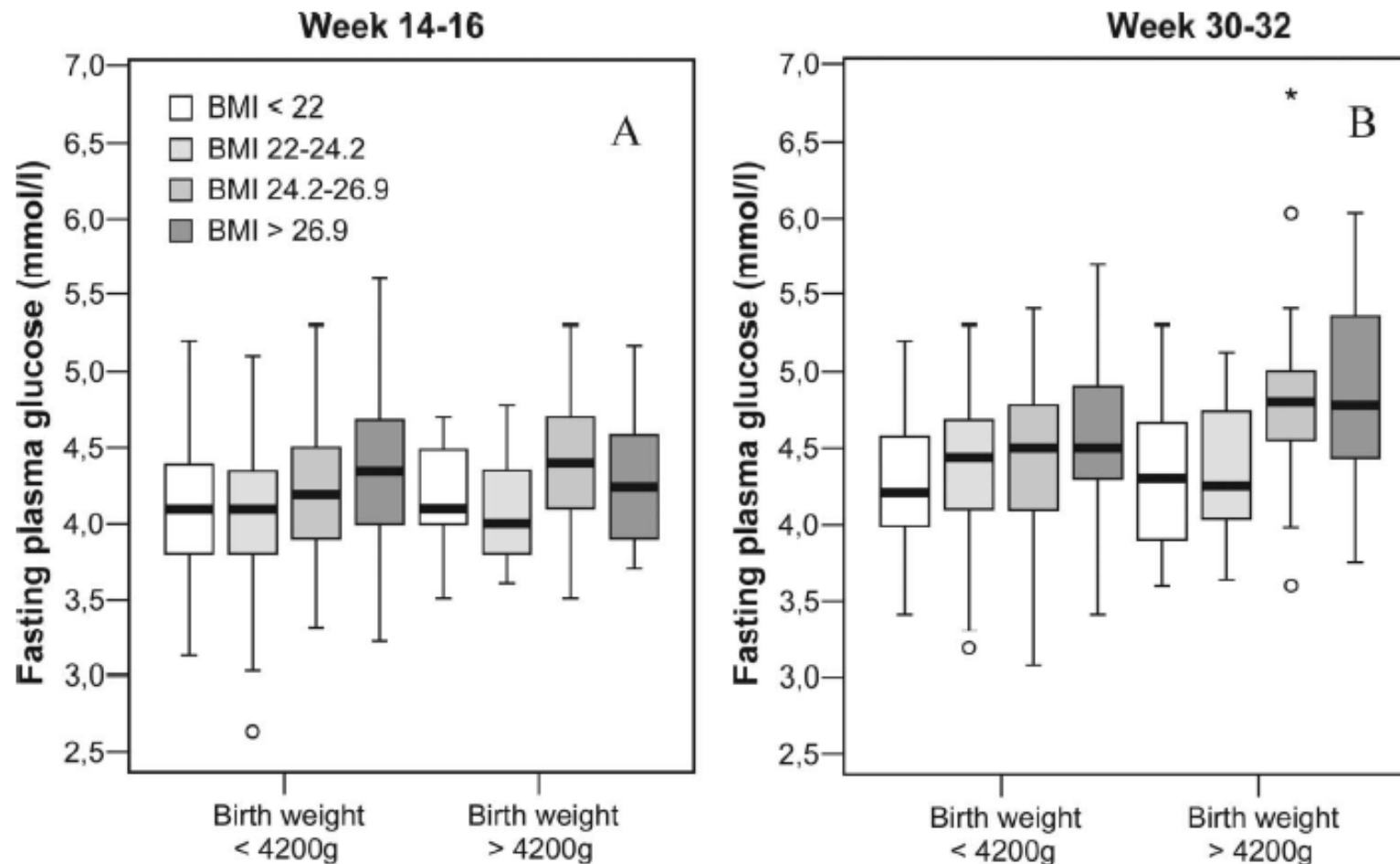
Deskriptiv statistikk for hver gruppe, differanser mellom gjennomsnitt eller medianer.

Eksempel: CRP-målinger for ulike grupper pasienter med kronisk betent tarm



Eksempel: Blodsukkerverdier for gravide kvinner på to tidspunkt i svangerskapet, gruppert etter deres BMI-kvartil, og om de fødte et stort barn eller ikke. Stort barn ble definert som over 4200g.

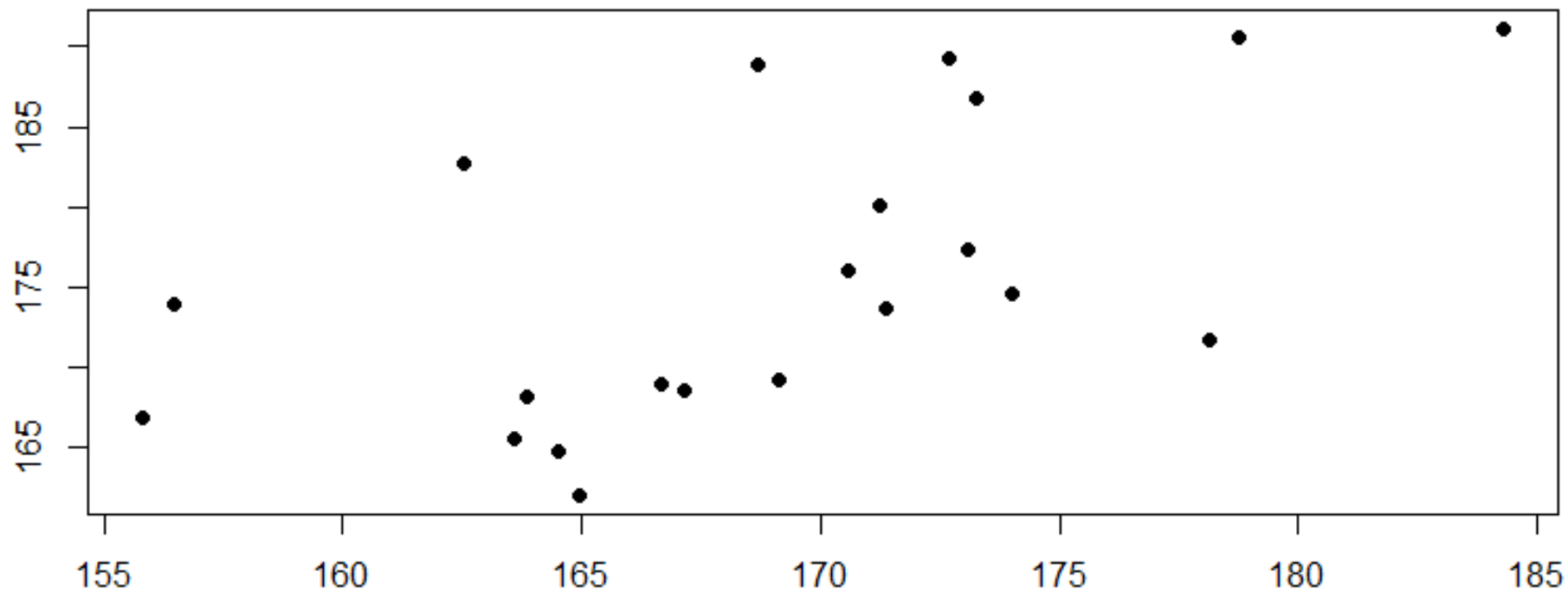
N, Qvigstad E, Frøslie KF, Godang K, Henriksen T, Bollerslev J: Increased risk of macrosomia among overweight women with high gestational rise in fasting glucose. Journal of Maternal-Fetal and Neonatal Medicine 2010;23:74-81



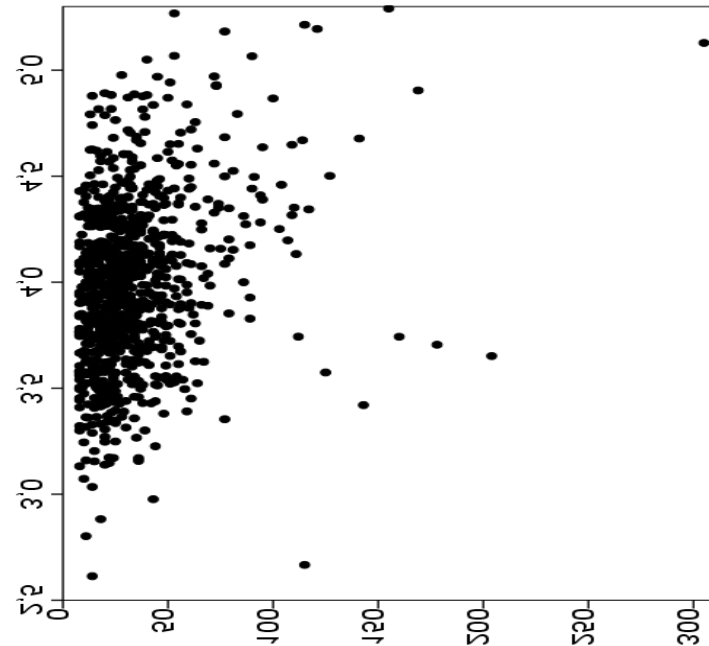
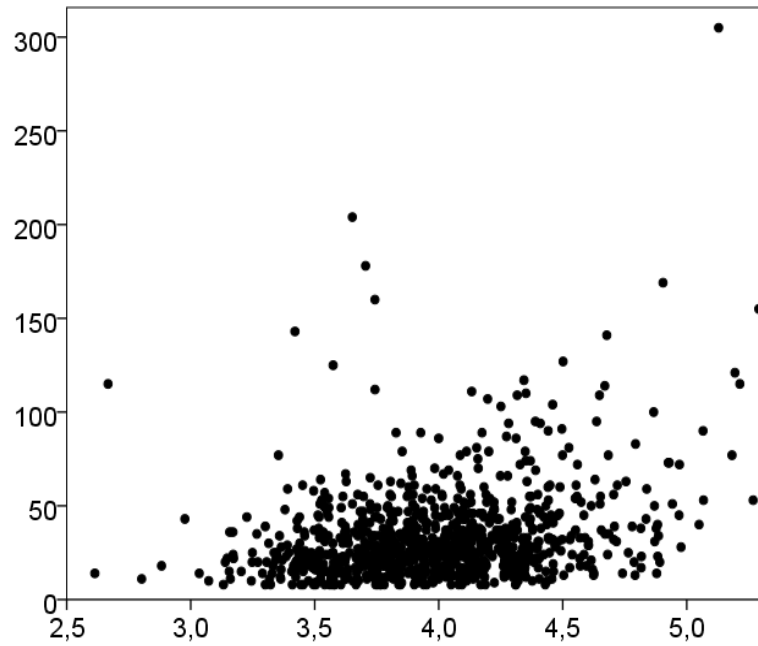
Kontinuerlig variabel mot kontinuerlig variabel (Ch 2.2)

Beskrive med **figur**: Scatterplot («Spredningsdiagram»)

Velg selv hva som skal være på x-aksen og y-aksen. Men: Hvis (du mistenker at) den ene variabelen er en forklaringsvariabel og den andre en responsvariabel, bør forklaringsvariabelen være på x-aksen.



Eksempel: Fastende blodsukker og insulin for gravide kvinner (fra STORK-studien ved Rikshospitalet)

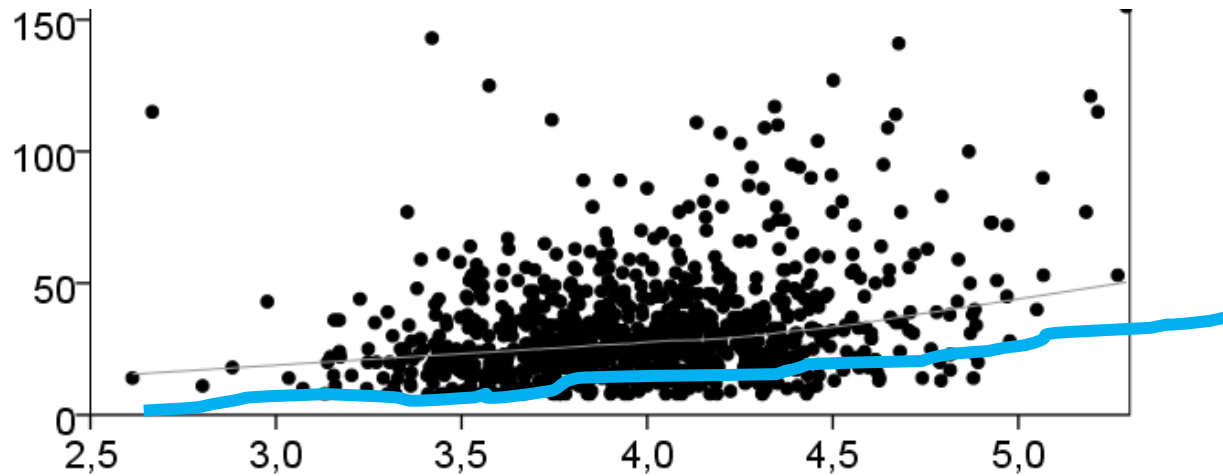


Beskrive med **ord**:

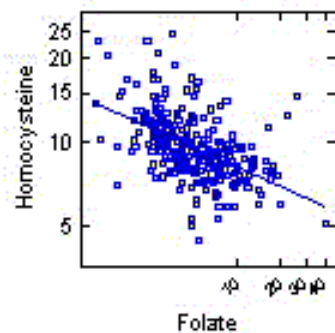
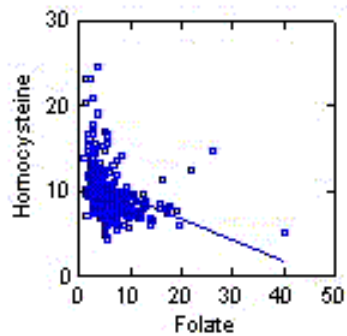
Beskrive med **oppsummeringstall**:

Hva ser vi etter?

- ✓ Lineærsammenheng? (en glattet linje (boka: Smoother) kan være nyttig)



- ✓ Positiv/negativ sammenheng?
- ✓ Outliere?
- ✓ Potensiale for at transformasjoner vil gjøre det mer lineært?



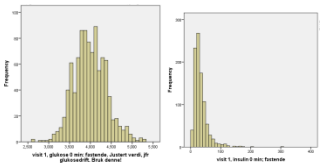
Beskrive med **oppsummeringstall:**

Pearsons korrelasjonskoeffisient r

Tallfester graden av lineær sammenheng mellom to kontinuerlige variabler.

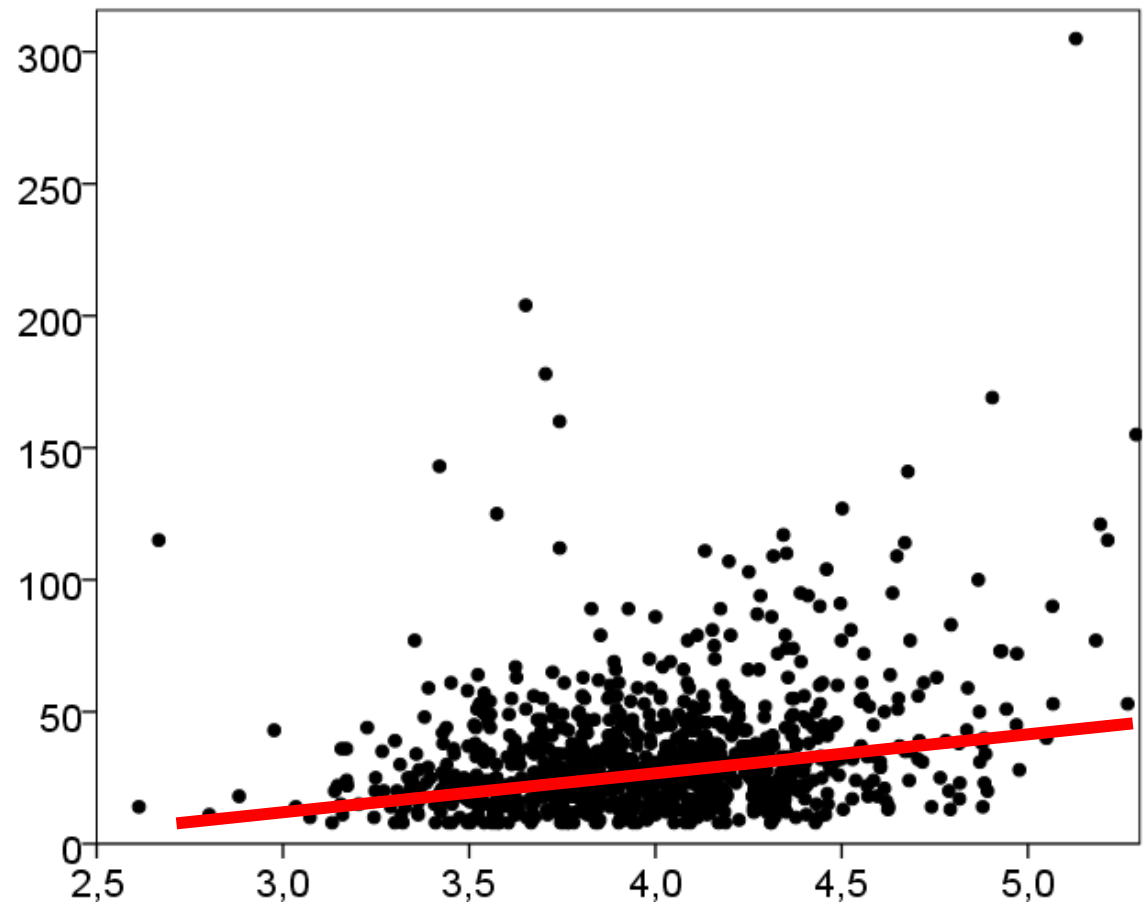
Tolkbarhet forutsetter at minst en av variablene er normalfordelte.

Sjekk v/ scatter-plott (ellipseform?),
histogram, QQ-plott



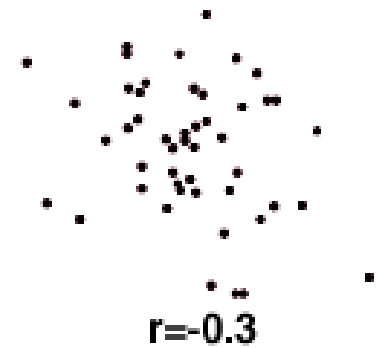
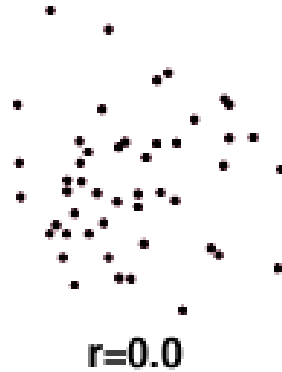
Eksempel:

Blodsukker og insulin: $r = 0.28$



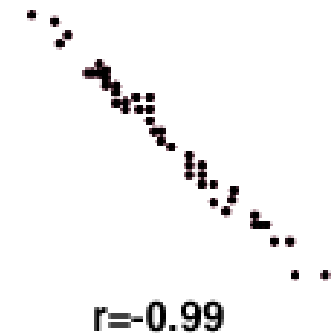
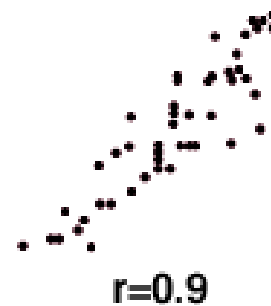
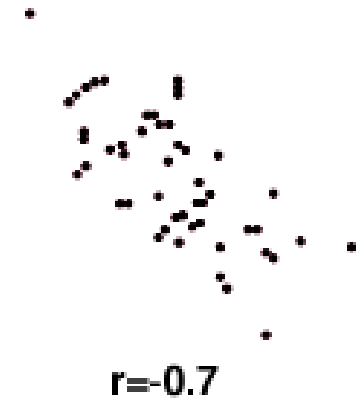
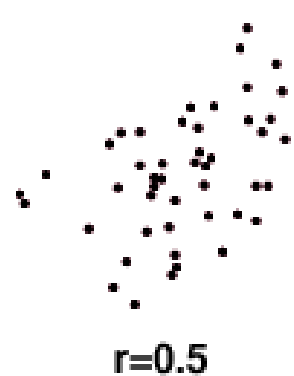
Egenskaper for korrelasjonskoeffisienter

- ✓ Korrelasjonskoeffisienten er alltid et tall mellom -1 og 1
- ✓ Hvis uavhengighet: Korrelasjon = 0
- ✓ Korrelasjon -1.0 eller 1.0 : Eksakt lineær sammenheng



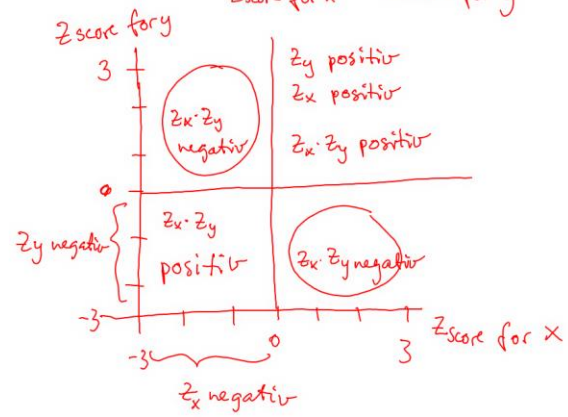
Negativ og positiv korrelasjon

www.guessthecorrelation.com



$$\text{Korrelasjon } r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_{dx}} \right) \cdot \left(\frac{y_i - \bar{y}}{s_{dy}} \right)$$

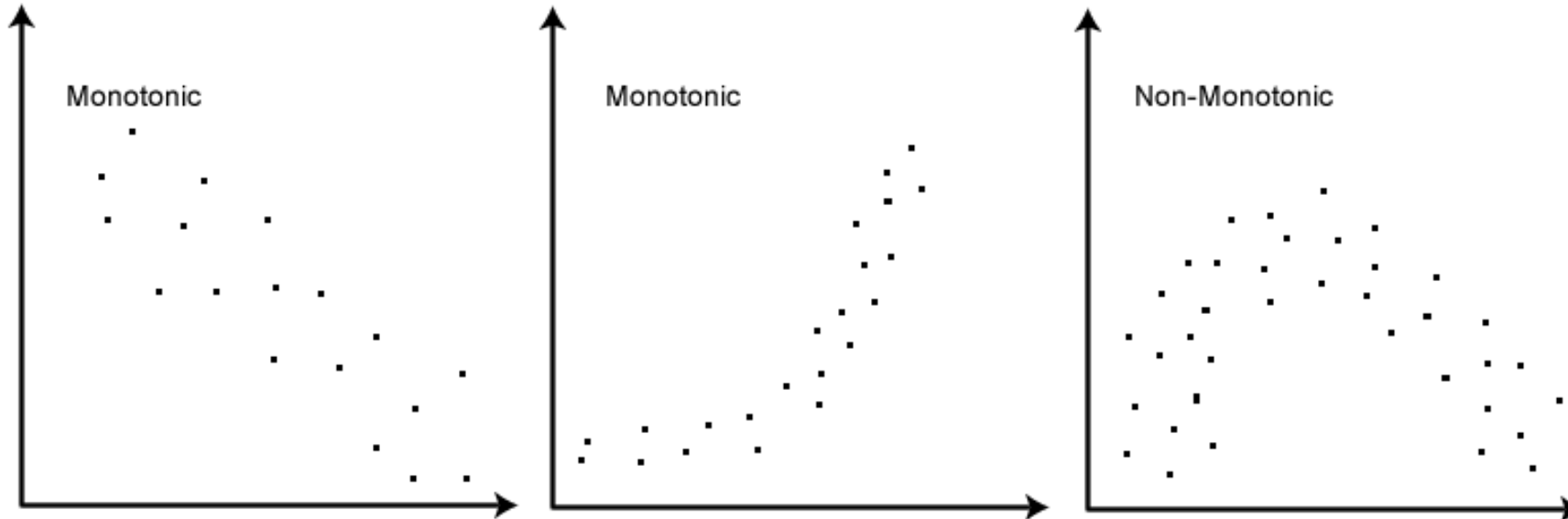
Zscore for x Z-score for y

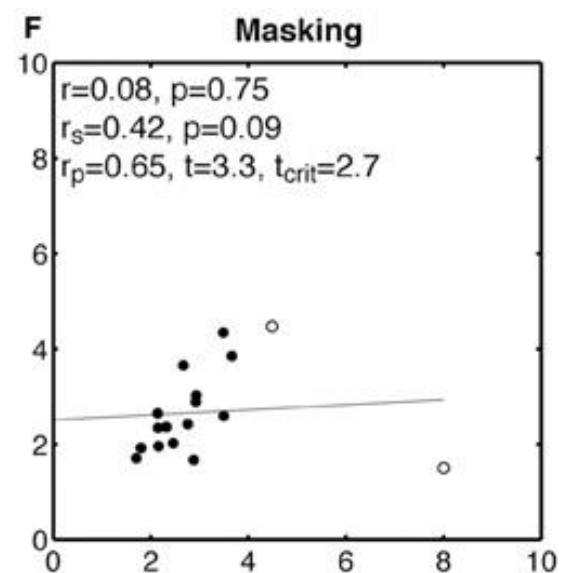
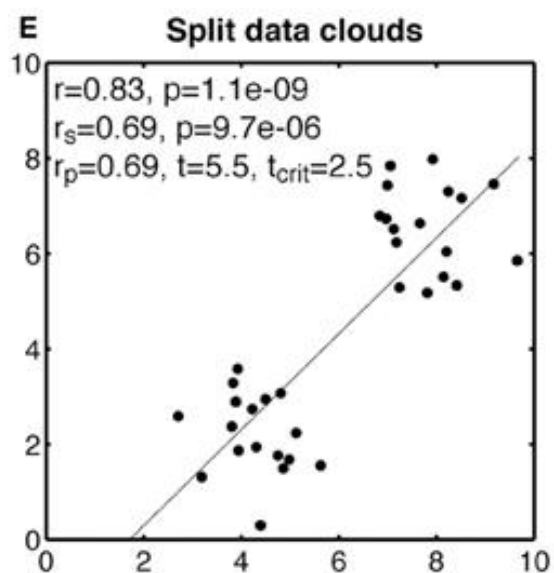
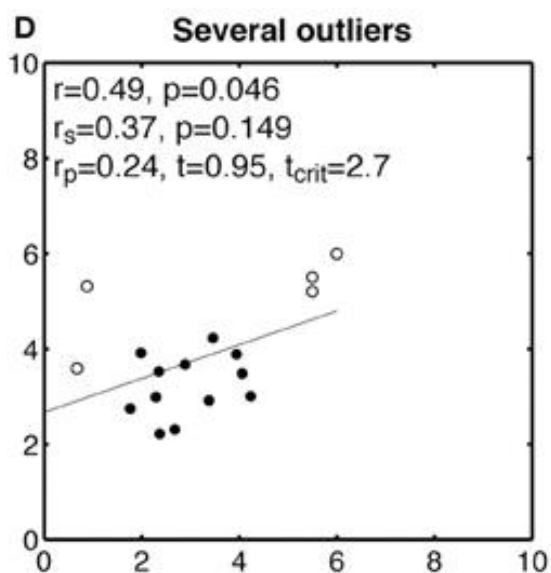
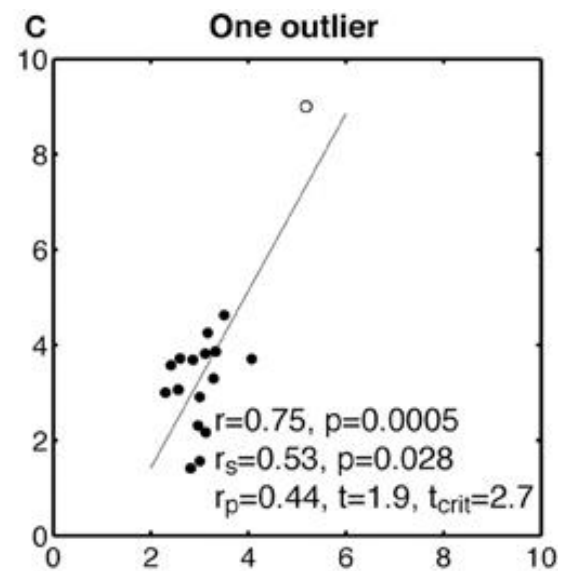
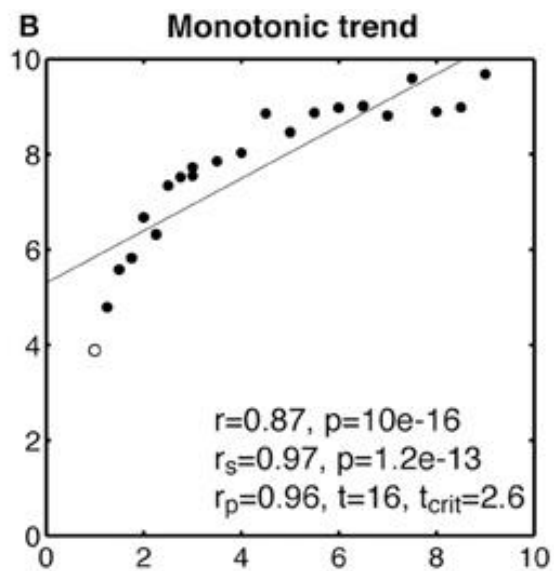
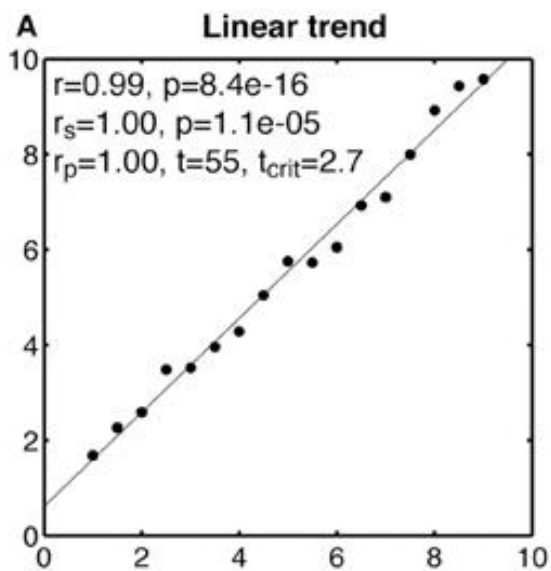


Noen vanlige korrelasjonsfeller:

- ✓ Tidsutvikling
- ✓ Målinger fra uavhengige individer?
- ✓ Manipulering av utvalg? (Ikke tilfeldig?)
- ✓ Ikke-lineære sammenhenger?

Hvordan unngå dem? Scatter, scatter og atter scatter!





En spesialutgave av scatter pots: «Time plots», der x-aksen viser tid. (Boka, side 23-24)

Eksempel: Månedlig overvåking av sykehusinfeksjoner forårsaket av bakterien «*Pseudomonas Aeruginosa*». Statistisk prosesskontroll med plotting av antall infeksjoner per måned.

«Dent-O-Sept-saken»: <http://www.aftenposten.no/norge/50-dode-etter-bakteriesmitte-fra-Dent-O-Sept-524351b.html>

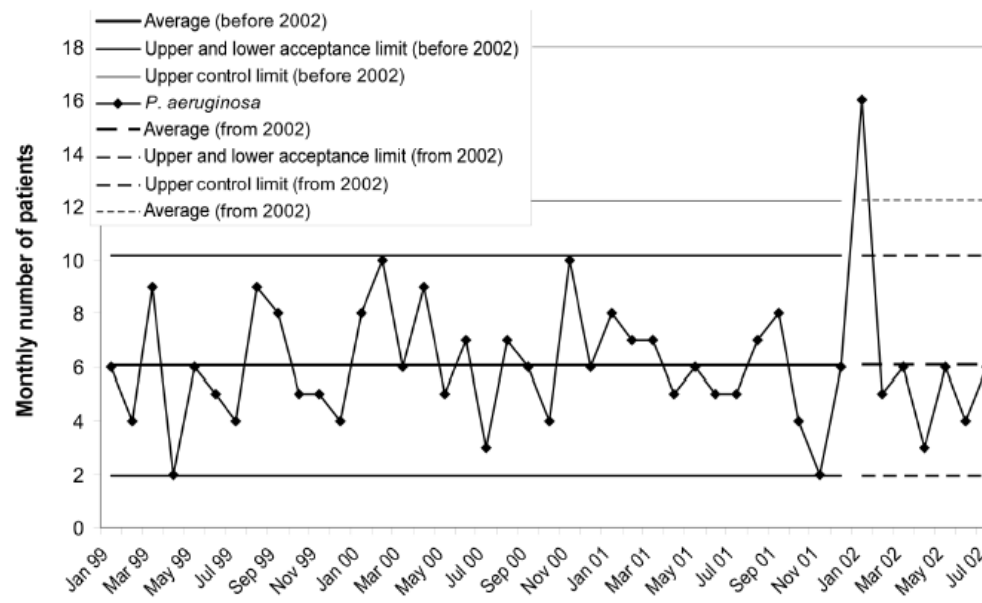


FIGURE 1. A C chart of the monthly incidence rate of *Pseudomonas aeruginosa* infection during the period from November 1997 to July 2002. Included in the figure are the average and the upper and lower control and acceptance limits (3σ and 2σ , respectively). The calculations of the mean and the limits were based on 30 observations from 1999 to 2001.

Eksempel: Månedlig overvåking av sykehusinfeksjoner forårsaket av bakterien «Pseudomonas Aeruginosa». Statistisk prosesskontroll med plotting av antall dager mellom nye infeksjoner.

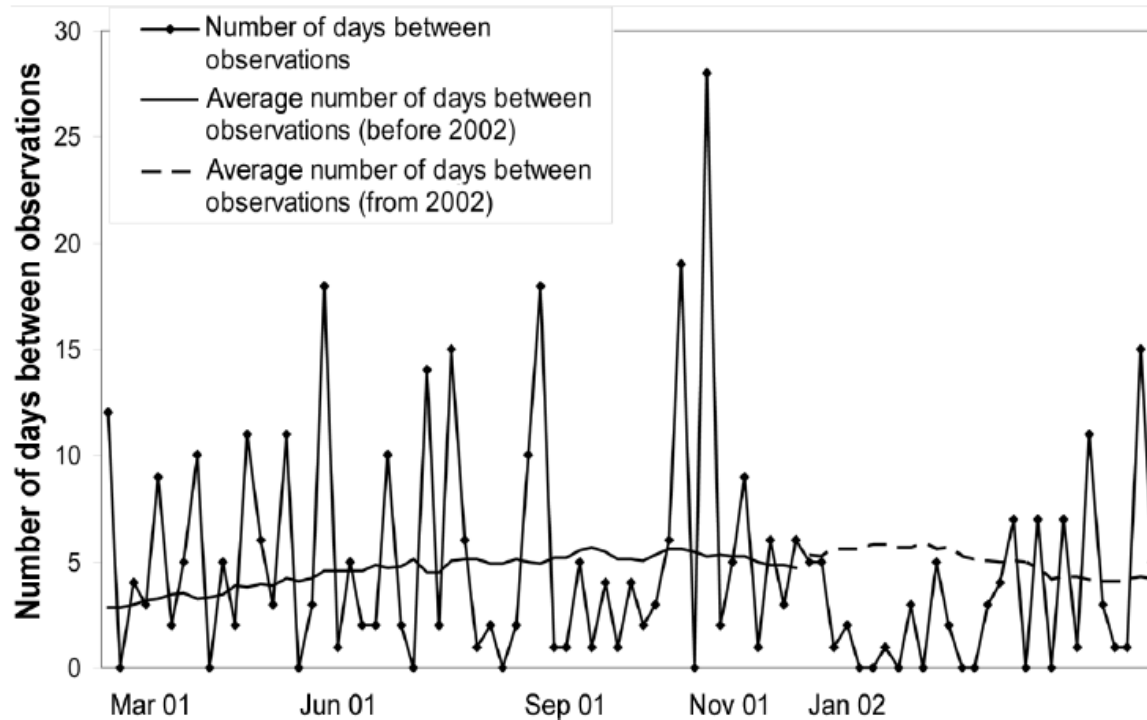


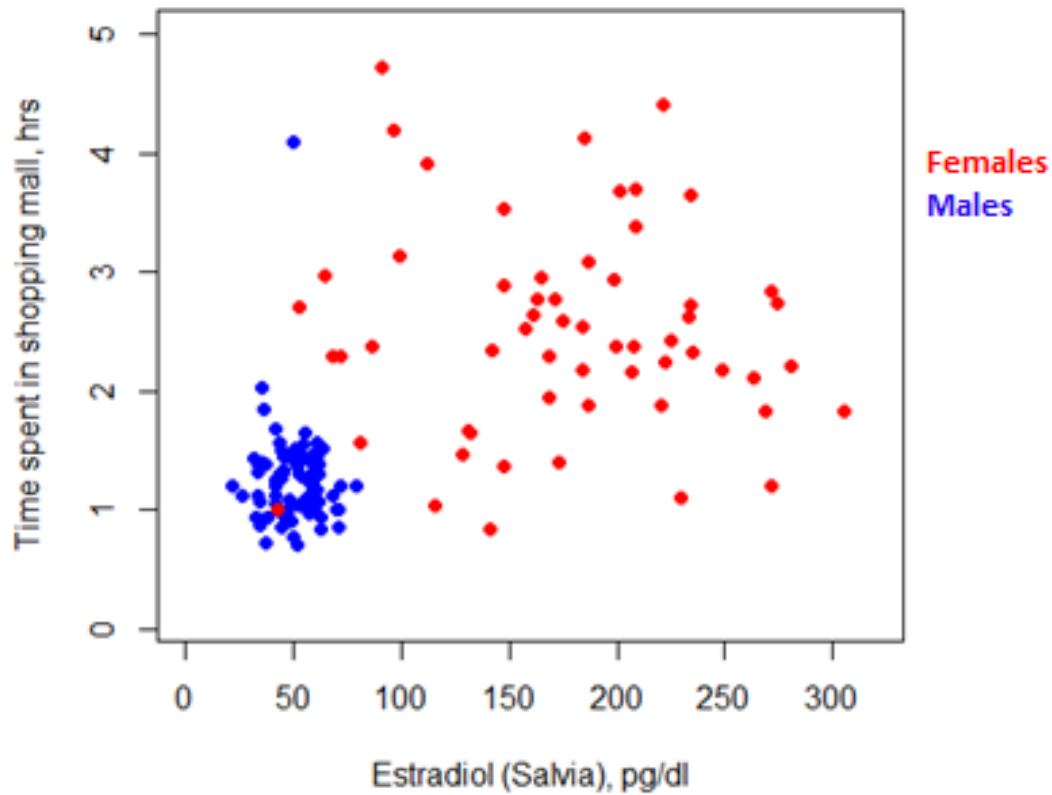
FIGURE 2. A g chart of 83 consecutive observations of the number of days between *Pseudomonas aeruginosa* infections during the period from March 2001 to March 2002. The calculations of the mean and the limits were based on 30 observations, with a delay of 10 observations. Observation 62 occurred on January 11, 2002.

Boka, side 95

Markering av undergrupper i dataene (inkludere en kategorisk variabel i scatterplottet)

På denne måten inkluderes en ekstra variabel, og det er ikke lenger en bivariat sammenheng.

Eksempel: Tid brukt på shopping vs østrogen-nivåer. Simulerte data.

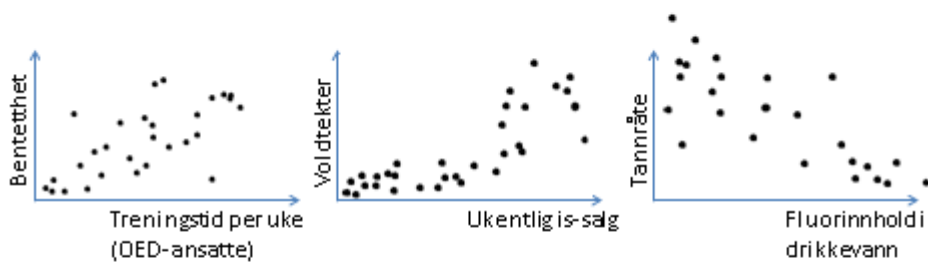


Vær varsom-plakaten for bivariate analyser.

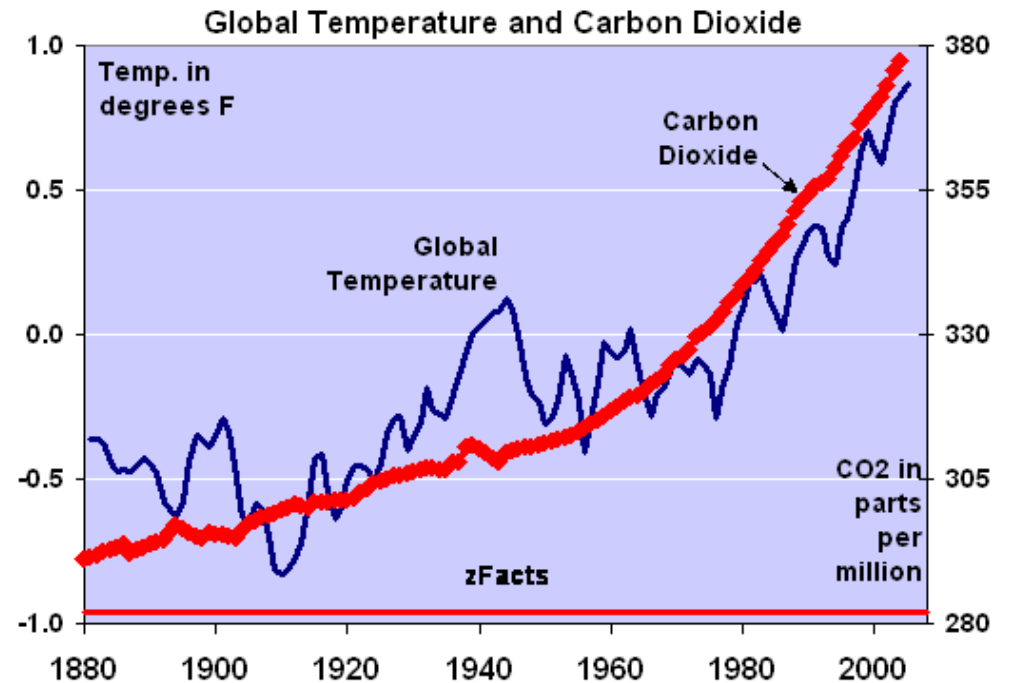
Assosiasjon, hva betyr det?

Assosiasjon er ikke det samme som kausalitet. Gjelder alle bivariate analyser.

Årsakssammenhenger?



- Bør man anbefale trening mot benskjørhet?
- Bør man forby is?
- Bør man anbefale fluor-tilskudd?



Denne mannen har en teori: Når renten stiger, øker antall dødsulykker i trafikken.



TILFELDIG? Denne tabellen viser to årg. rentenivået på side 15 årene og antallet drepte i trafikken. Synes du kurvene ser like ut?

Av LINDA NILSEN METH og TOVE BERNTSEN (foto)

Trafikkentusiasten Arild Christiansen har sett på forhold rundt trafikkuulykker som fører til døden på norske veier. Han mener rentenivået har en betydning...

Det er ikke tilfeldig, mener Christiansen, tilhittsavlagt i NAF Oslo.

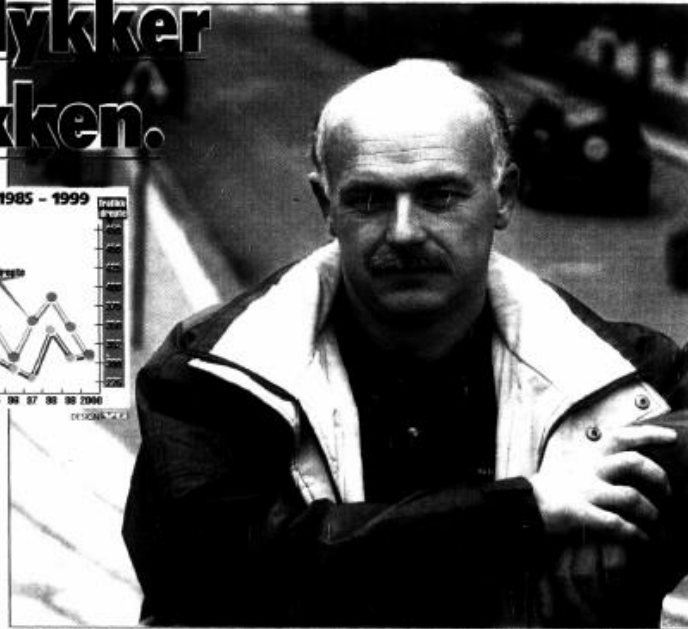
Hans funn, som du hun se i tabellen på denne siden, kan sies å være oppsiktsvekkende, om ikke akkurat statistisk utvilsomt. Kurven for rentenivået er påfallende lik kurven for trafikkdrepte de siste 15 årene.

- Kurven er litt for like til at det skal være tilfeldig, sier Christiansen, som har lagt tallene frem for generalforsamlingen i NAF.

Bekymret og ukonsentrert

Christiansens teori er enkel: - Stiger renten, stiger folks bekymringer. Dess mer bekymret folk er, dess mer ukonsentrert blir de i det de holder på med. Ukonsentrerte trafikkantør gjør at faren for å bli innblandet i ulykker øker.

- Er ikke dette en litt enkel forklaring? Rentnivået forklarer ikke ulykkesfrekvensen i sin helhet, men renten er faktisk noe av det alvorligste folk har å bekymre seg over. De tenker ikke på det, og da mister de konsentrasjonen om det de ellers gjør - for eksempel kjøre bil, sier Christiansen.



RENTEN DREPER: - Antallet trafikkdrepte følger rentenivået på en bemerkelsesverdig, sier Arild Christiansen, tilhittsavlagt for Norges Automobil-Forbund (NAF) i Oslo og Ormskog.

Han tror også at bekymringer om rentenivå, ekte tapfrelle og for lav lønn holder folk våkne om natten.

Folk som sover dårlig, blir trøste, og risikoen for å sove bak rattet blir større, sier Christiansen, og viser til at veldig mange ulykker i Norge skjer forth gjuloren sover bak rattet.

Han mener det er på høy tid at statlige myndigheter ser mere på omstendighetene rundt dødsfall på norske veier.

Da vil de finne ut det ikke bare er fart, promille, handilskille, mobiltelefoner og manglende bruk av seletelte som drøper. Det finnes andre faktorer som ulykker. Rentnivået er en av dem, sier Christiansen.

- Hvordan fikk du ideen om å sjekke sammen rentenivået og trafikkdrepte? Det er ikke lett å

forklare, men det var et resultat av interesse for trafikksikkerhet. Jeg er ikke sikker på at myndighetene gjør det riktige for å begrense ulykker. Et er det jeg har funnet, kan det kanskje være mer gunstig å senke renten enn å senke farten, smiler Christiansen.

Og hva gjør han selv for å unngå å bli rammet av rentestigning og trafikkuulykker? - Jeg har fastrenter og kjøreturpikop. Det er min beste investering

- Bilene er tryggere

Av ELVE DALUM

En nedgang i antall trafikkdrepte på 90-tallet henger sammen med at bilene er blitt tryggere. Rentnivået betyr mindre, mer informasjonssjef Stein Haakonson i Finansringens Hovedorganisasjon.

Uendret rente

Engstelige boligselgere fikk ingen hjelp av Norges Bank. Det ble ingen rentenedgang etter gjensidig hovedtilløp. Og samtidig fortsetter prisnedgangen på brukte boliger - i alle fall i Oslo.

Fra september til oktober sank den gjennomsnittlige kvadratmeterprisen for OBOS-omsatte boliger med 3,6 prosent - til 15 722 kroner, går det frem av en oversikt fra OBOS. Det har også vært en dramatisk nedgang i tallet på solgte boliger i den samme perioden - fra 801 i september til 450 femte måned.

ring, gløse Arild Christiansen. Tre nasjonalistiske partier støtter ikke Christiansens teori.

- Vi har nemlig sett at folk blir over forsiktede og at ulykkestallene går ned i dårlige tider - og at det er under høy konjunktur ulykkestallene øker, sier Harald Aas i TD.

© 2000. Alle rettigheter reservert.

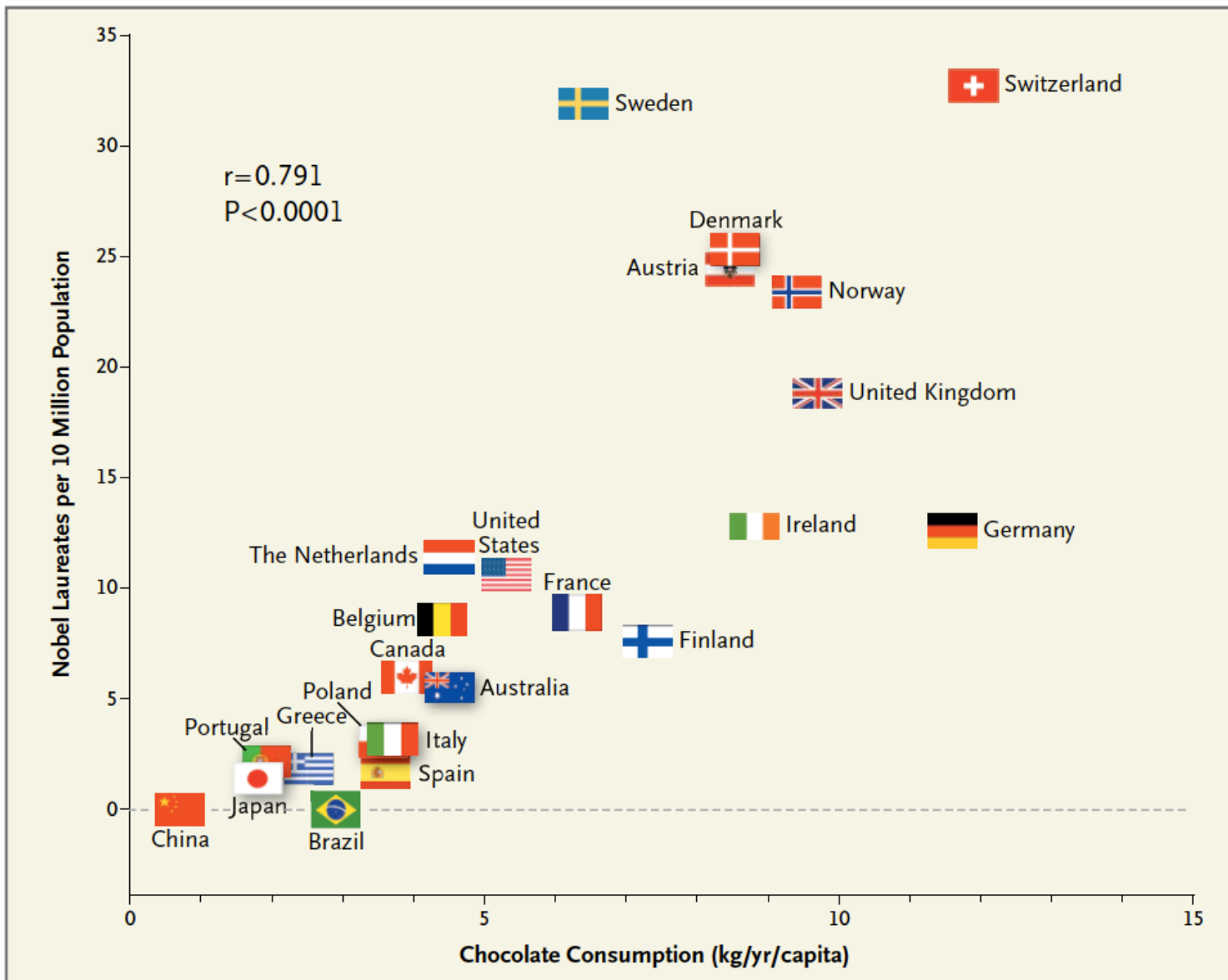


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

En annen måte å oppsummere lineære sammenhenger på: Beregn linja $y = ax + b$ som passer best til punktene (Ch 2.4 i boka) Dette gjennomgår vi i forbindelse med regresjon, Ch 10+11

Etter dette bør dere **repetere**, for dette (Ch 1, 3 og deler av Ch 2) er den overordnede introduksjonen i kurset. Vi må forstå hvordan forskningsprosessen fungerer og hvilke valg vi har når det gjelder å samle inn data (Ch 3), hvordan vi presenterer og oppsummerer informasjon vi har samlet inn (Ch 1), og hvordan vi beskriver enkle sammenhenger mellom variabler (Ch 2). Dette er i stor grad diskusjons- eller lesestoff.

Hvis vi derimot skal forstå essensen i statistiske analyser, altså hva et konfidensintervall, en p-verdi eller en statistisk test er og tolkes, må vi grave oss lenger ned i tekniske begreper og mer matematikk. Det starter med Ch 3.4, og fortsetter i Ch 4, 5 og 6.

STK1000 neste uke: Les Ch 3.4 og Ch 4 i læreboka (MMC).