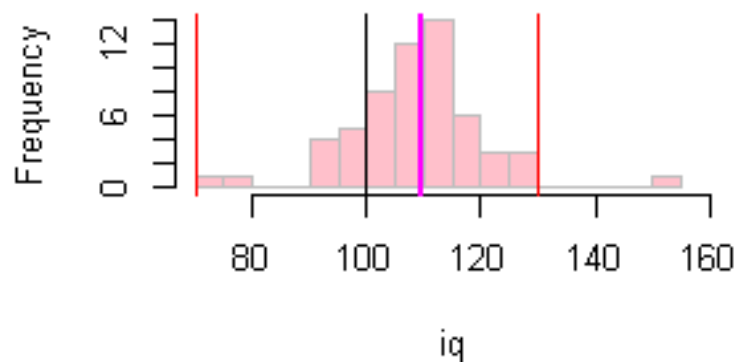


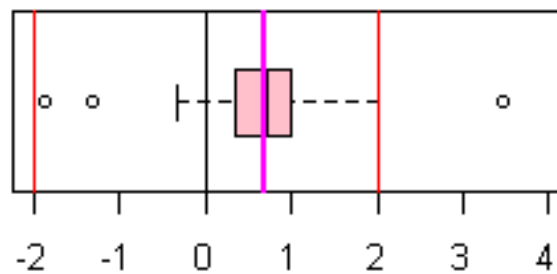
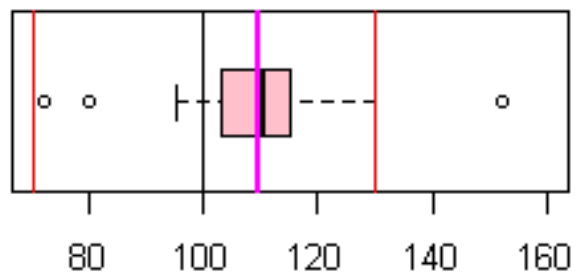
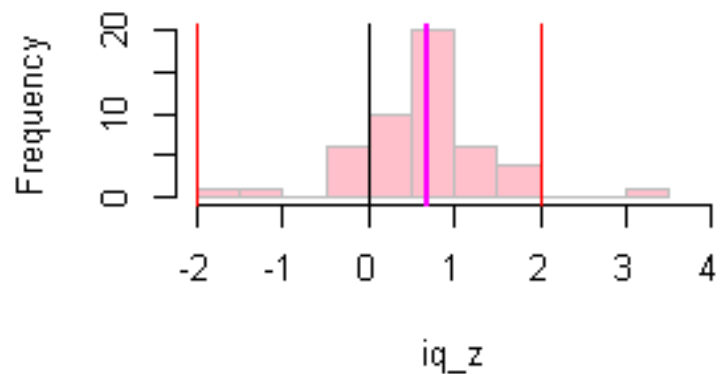
Beskrive enkle (bivariate) sammenhenger mellom variabler

Demonstrasjon i Rstudio: IQdata, z-scorer, kjønn, vgs-snitt, studietid, foreldres utdanningsnivå

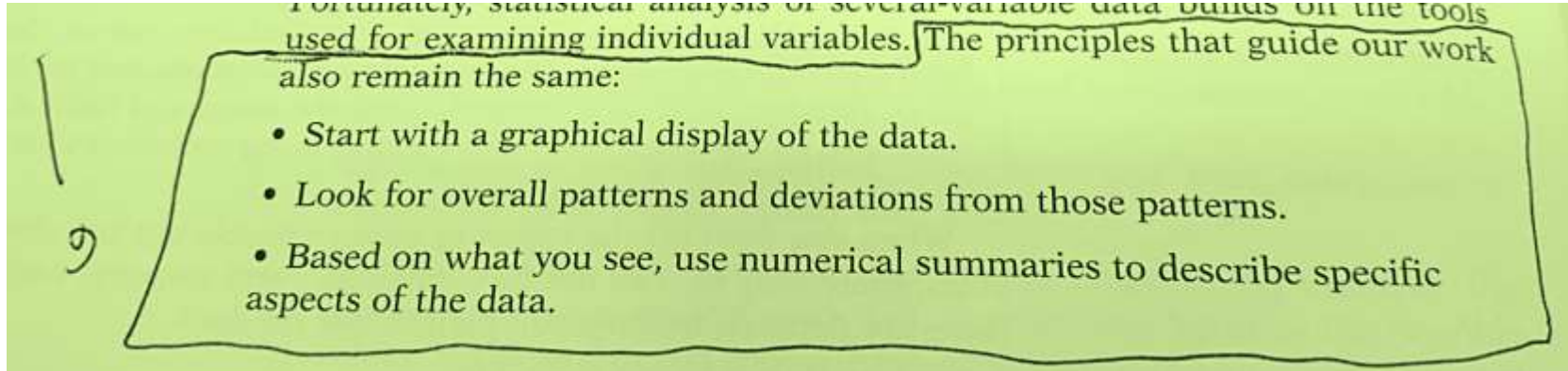
Histogram of iq



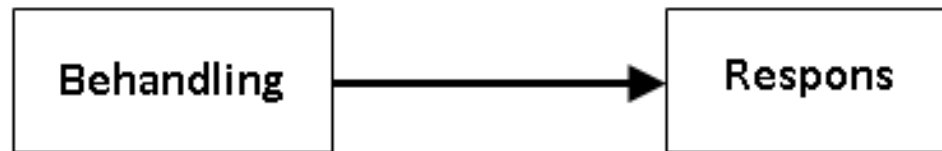
Histogram of iq_z



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



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, Uke 35

Versjon 1

Hva skal vi gjøre?

Oppsummere/presentere/
beskrive data

Hvilken type data har vi?

Kategoriske data

Tabeller, andeler
Søyle(stolpe)/kakediagram

Kontinuerlige data

Skjeve



Median, kvartiler

Symmetriske



\bar{x} , SD

Inferens

Beregne
konfidens-intervall

95% KI for p , som er
populasjonens andel

$$\hat{p} \pm 1.96 \cdot \sqrt{\hat{p}(1-\hat{p})}$$

95% KI for
populasjonens median

Tabeller? Bootstrapping!

95% KI for μ , som er
populasjonens gjennomsnitt

$$\bar{x} \pm 1.96 \cdot \frac{SD}{\sqrt{n}}$$

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

Undersøke forskjell på 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 sammenheng: Krysstabell (altså frekvenstabell for hver gruppe)

Kalles også Toveis-tabell, kontingenstabell, krysstabell

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

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

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

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

Oppsummeringstall for sammenheng:

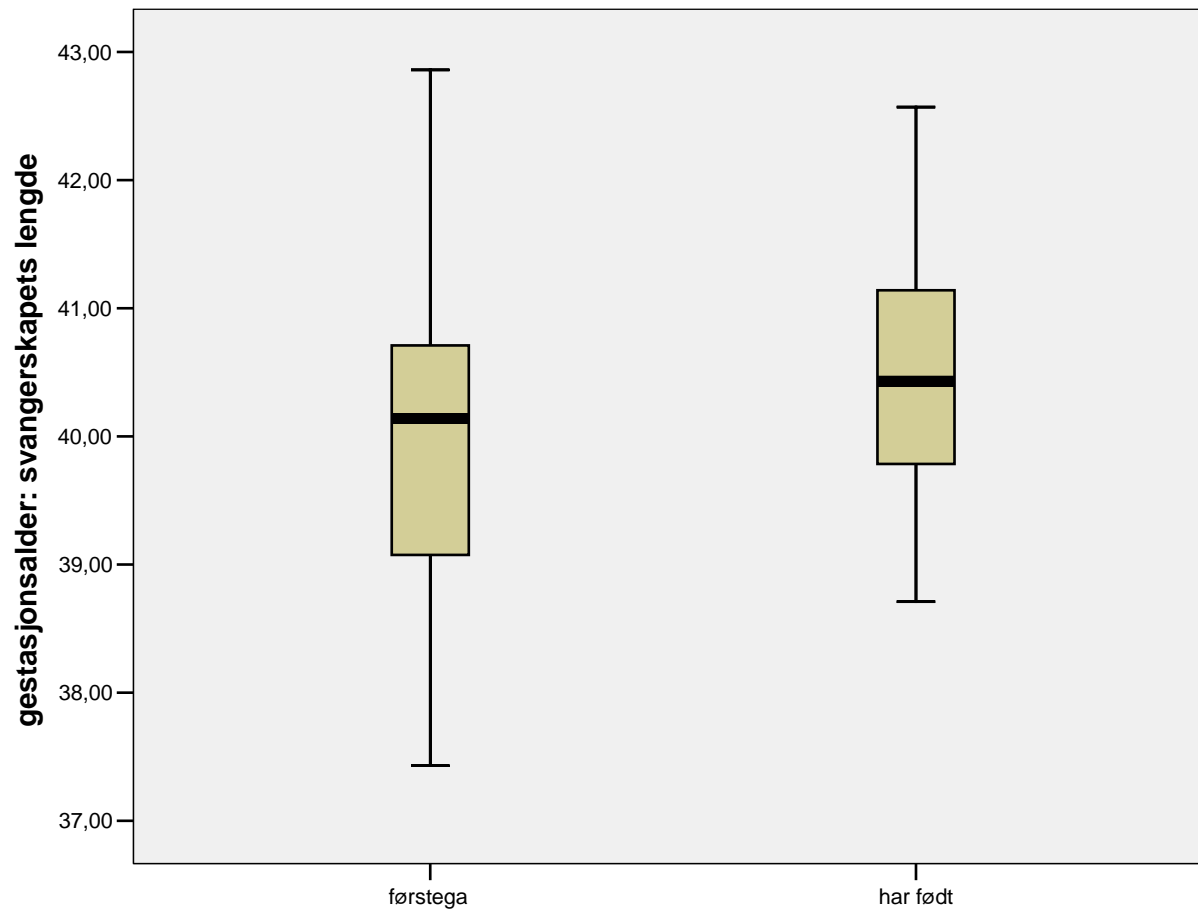
Tallene i kysstabellene, evt uttrykt i (prosent)andeler

Odds ratio (OR), Relativ risiko (RR)

Kontinuerlig variabel mot kategorisk variabel (eller kategorisk mot kontinuerlig)

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

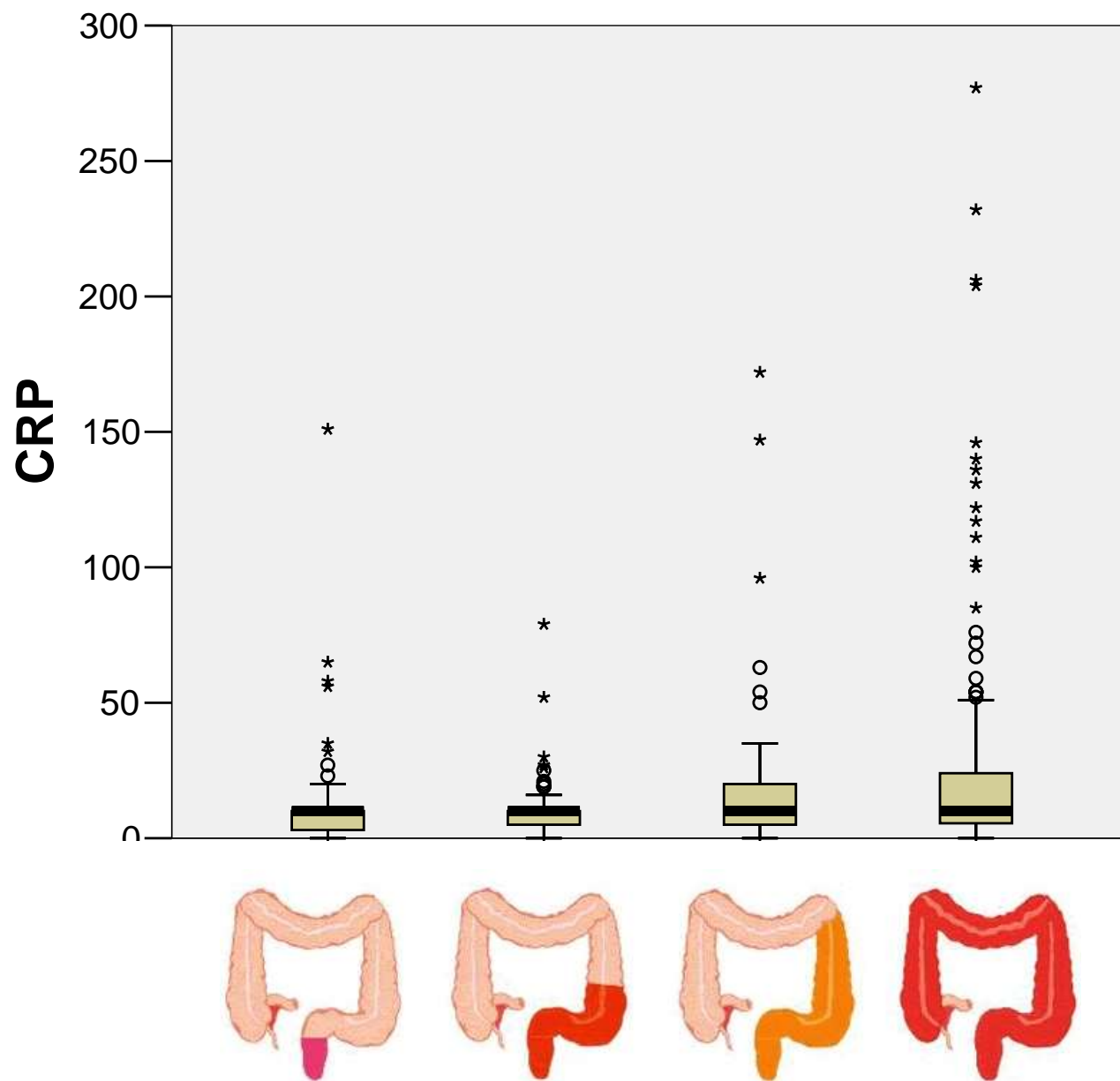
Eksempel: Svangerskapslengde for førstegangsfødende og de som har født før



Oppsummeringstall for sammenheng:

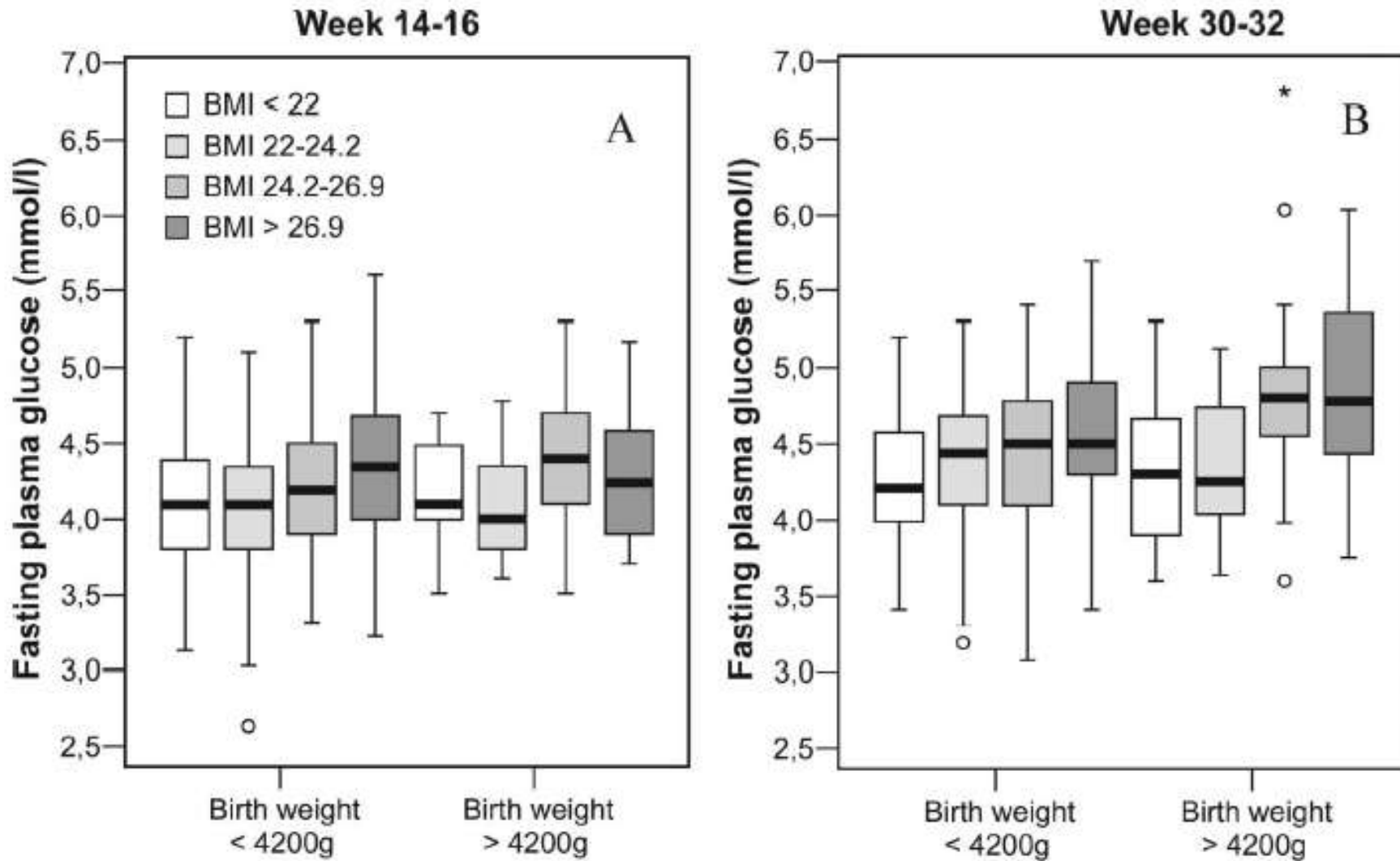
Deskriptiv statistikk for hver gruppe, differanser mellom gjennomsnitt eller medianer.

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.

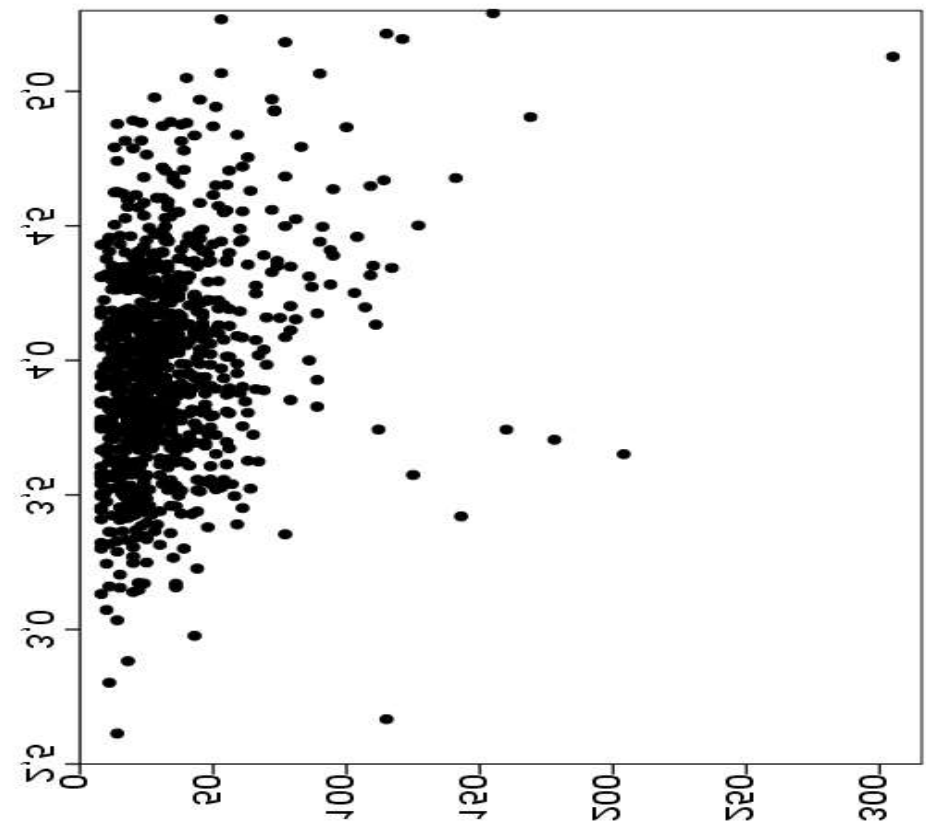
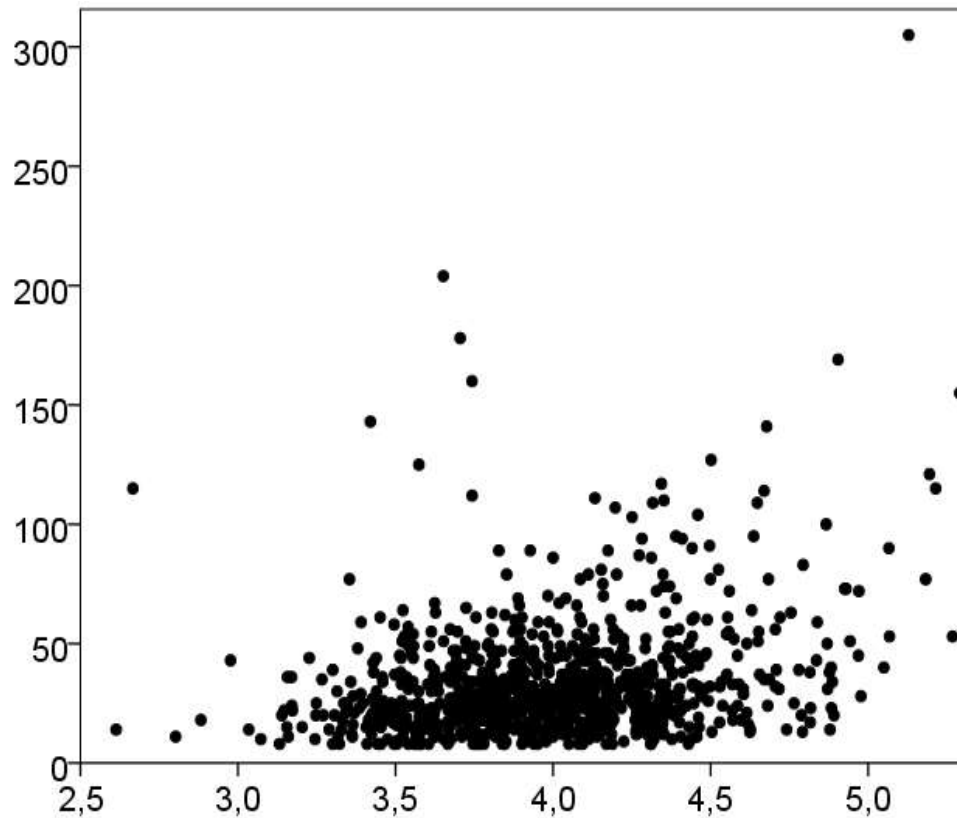
N, Qvigstad E, Frøslie KF, GodangK, Henriksen T, Bollerslev J: Increased risk of macrosomia among overweight women with high gestational rise in fasting glucose. Journal of Maternal-Fetal and Neonatal Medicine 2010;23:74-81



Kontinuerlig variabel mot kontinuerlig variabel (Ch 2.2)

Beskrive sammenheng: Scatterplot («Spredningsdiagram»)

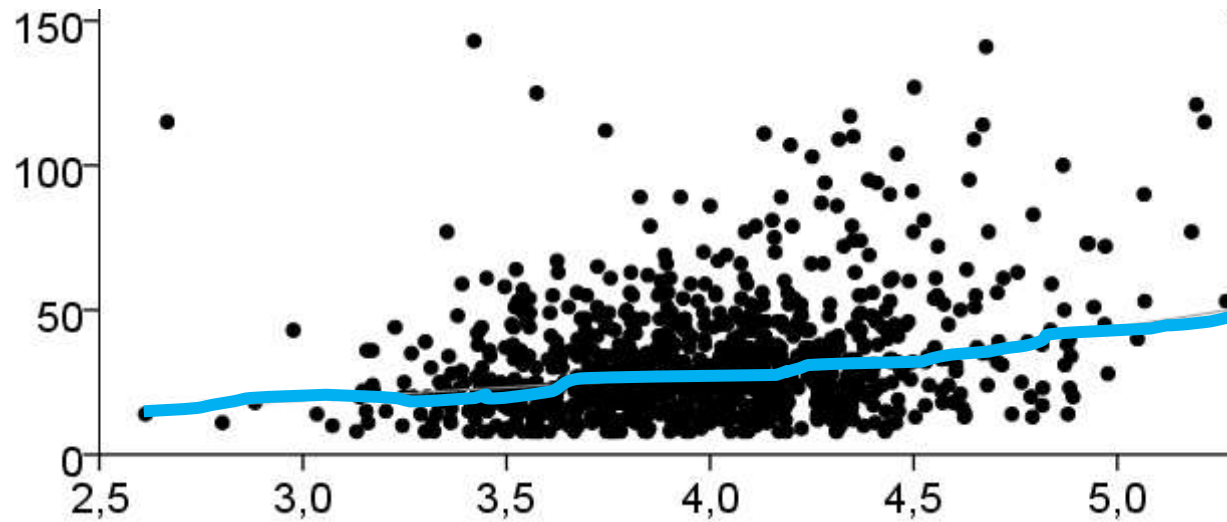
Eksempel: Fastende blodsukker og insulin for gravide kvinner (fra STORK-studien ved Rikshospitalet)



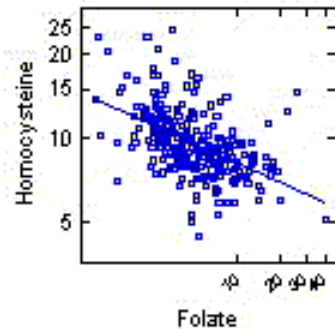
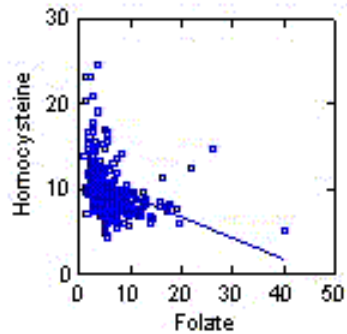
Velg selv hva som skal være på x-aksen og y-aksen. Men: Hvis (du mistenker at) den ene variabelen er en forklaringsvariabel og den andre en responsvariabel, bør forklaringsvariabelen være på x-aksen.

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?

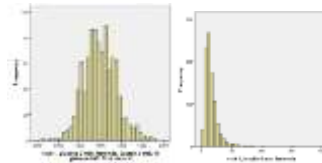


Oppsummeringstall for sammenheng:

Pearsons korrelasjonskoeffisient r

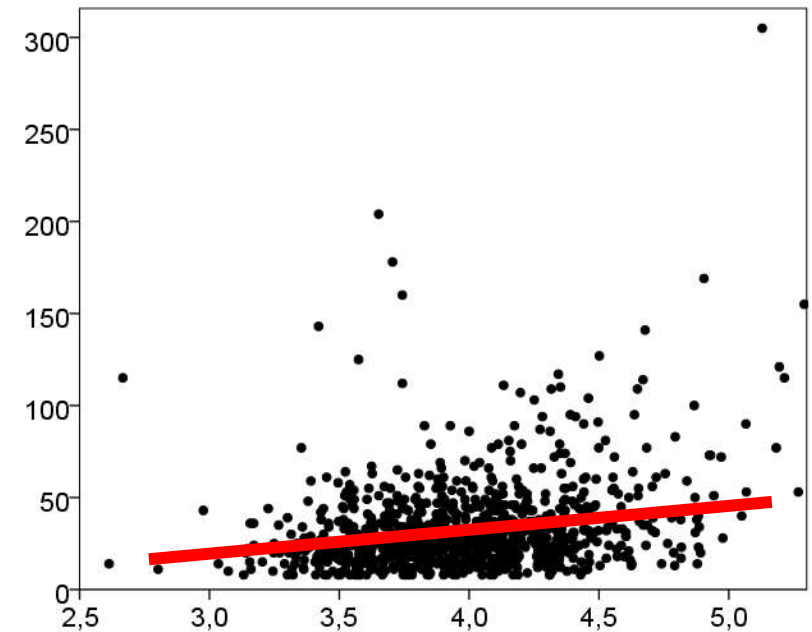
Tallfester graden av lineær sammenheng mellom to kontinuerlige variabler.

Tolkbarhet forutsetter at minst en av variablene er normalfordelte. Sjekk v/ scatter-plott (ellipseform?), histogram, QQ-plott



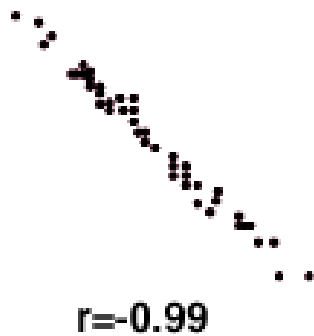
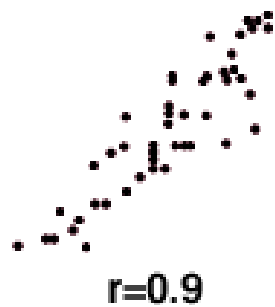
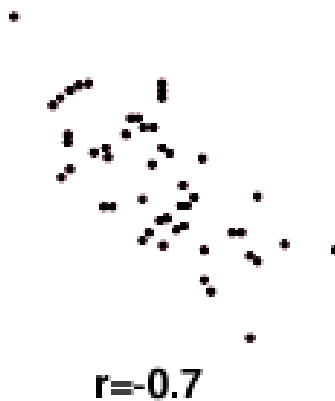
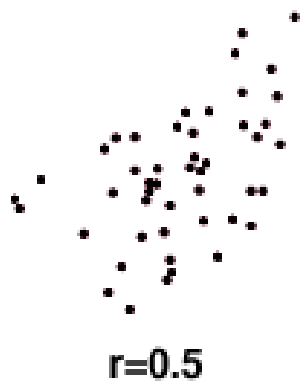
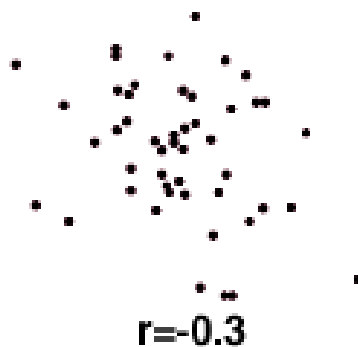
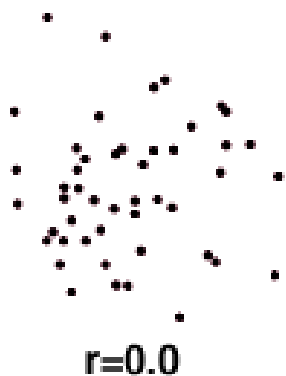
Eksempel:

Blodsukker og insulin: $r = 0.28$



Egenskaper for korrelasjonskoeffisienter

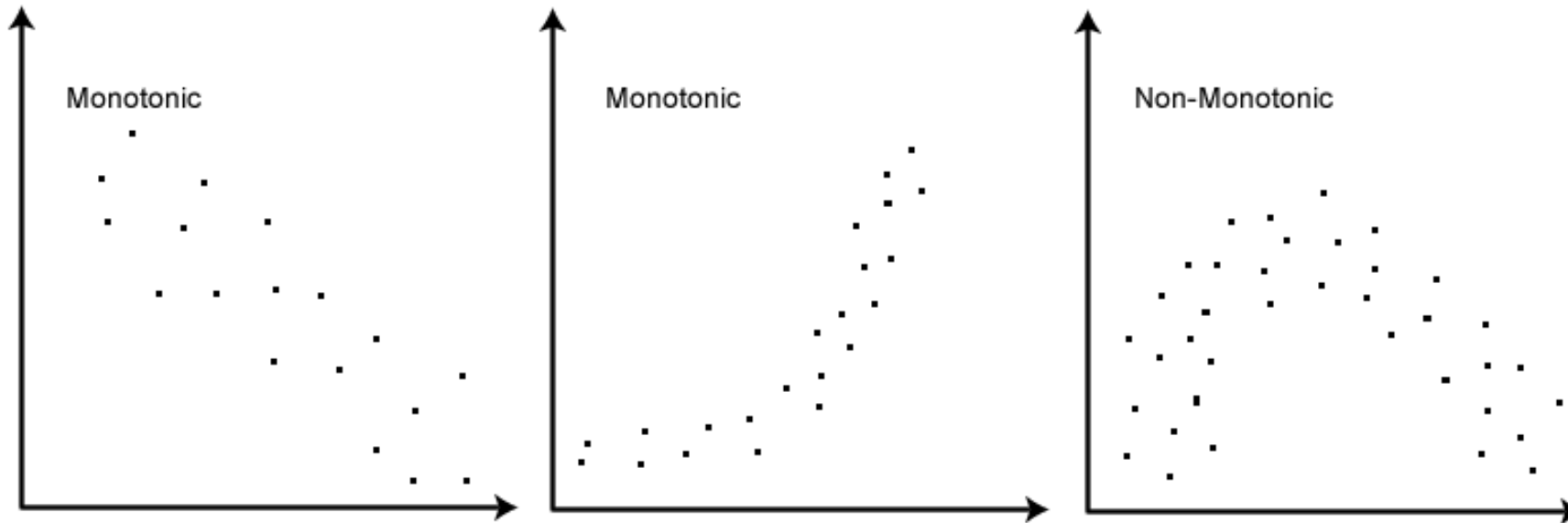
- ✓ Korrelasjonskoeffisienten er alltid et tall mellom -1 og 1
- ✓ Hvis uavhengighet: Korrelasjon = 0
- ✓ Korrelasjon -1.0 eller 1.0: Eksakt lineær sammenheng

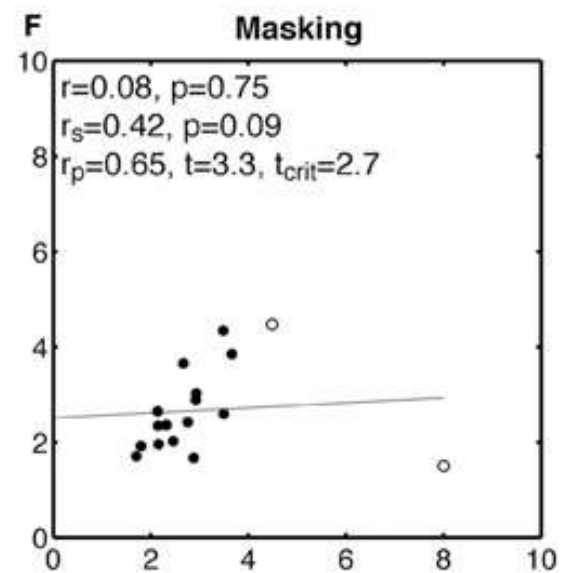
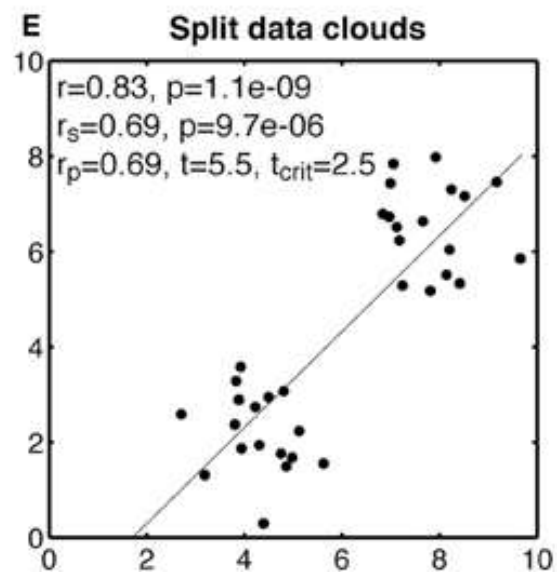
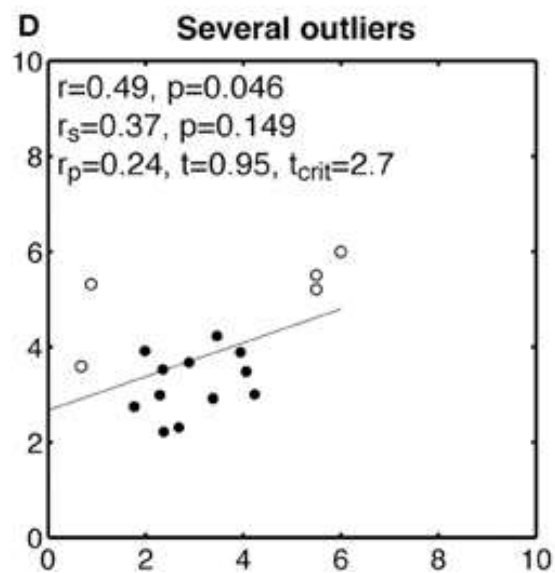
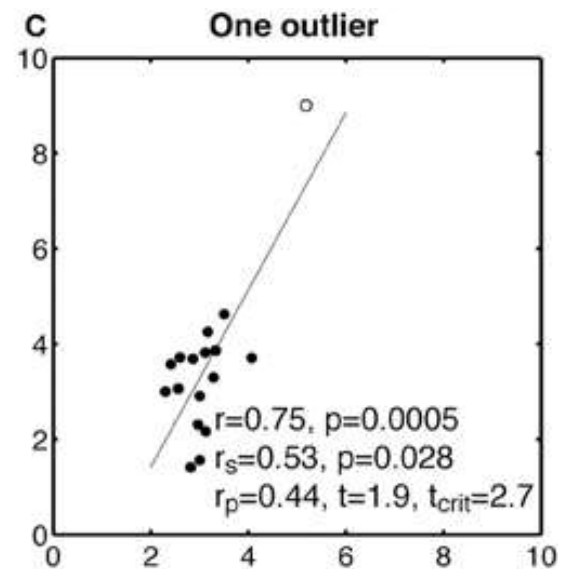
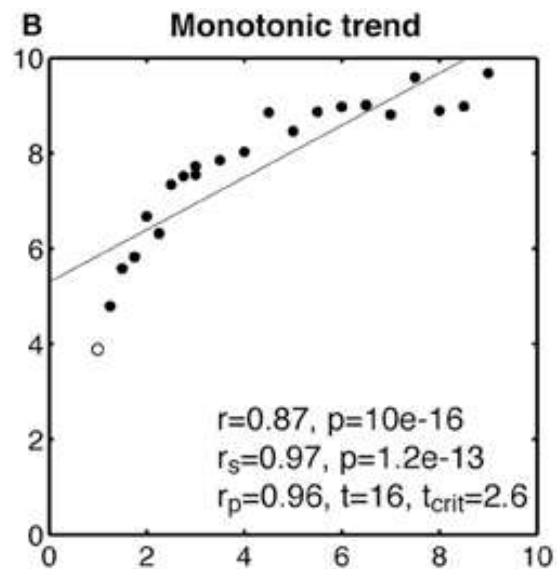
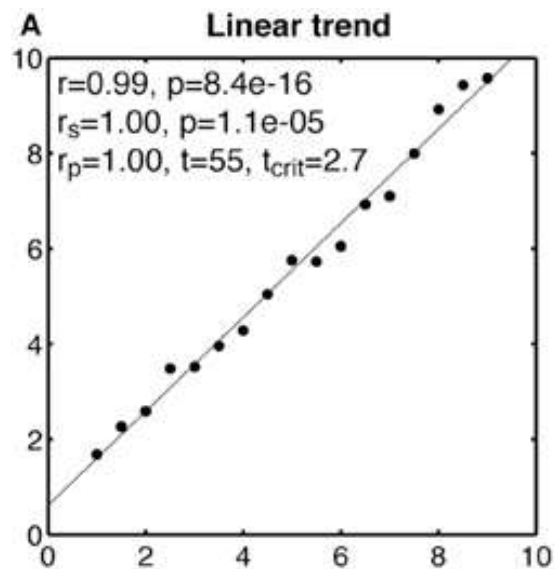


Noen vanlige korrelasjonsfeller:

- ✓ 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>

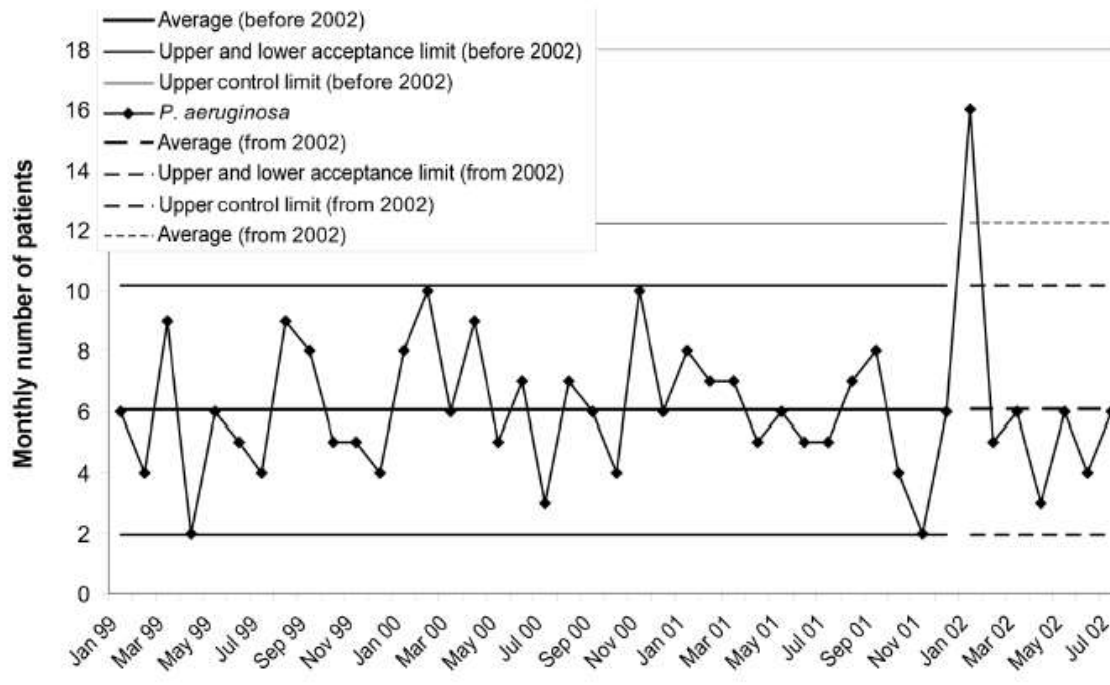


FIGURE 1. A c chart of the monthly incidence rate of *Pseudomonas aeruginosa* infection during the period from November 1997 to July 2002. Included in the figure are the average and the upper and lower control and acceptance limits (3σ and 2σ , respectively). The calculations of the mean and the limits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

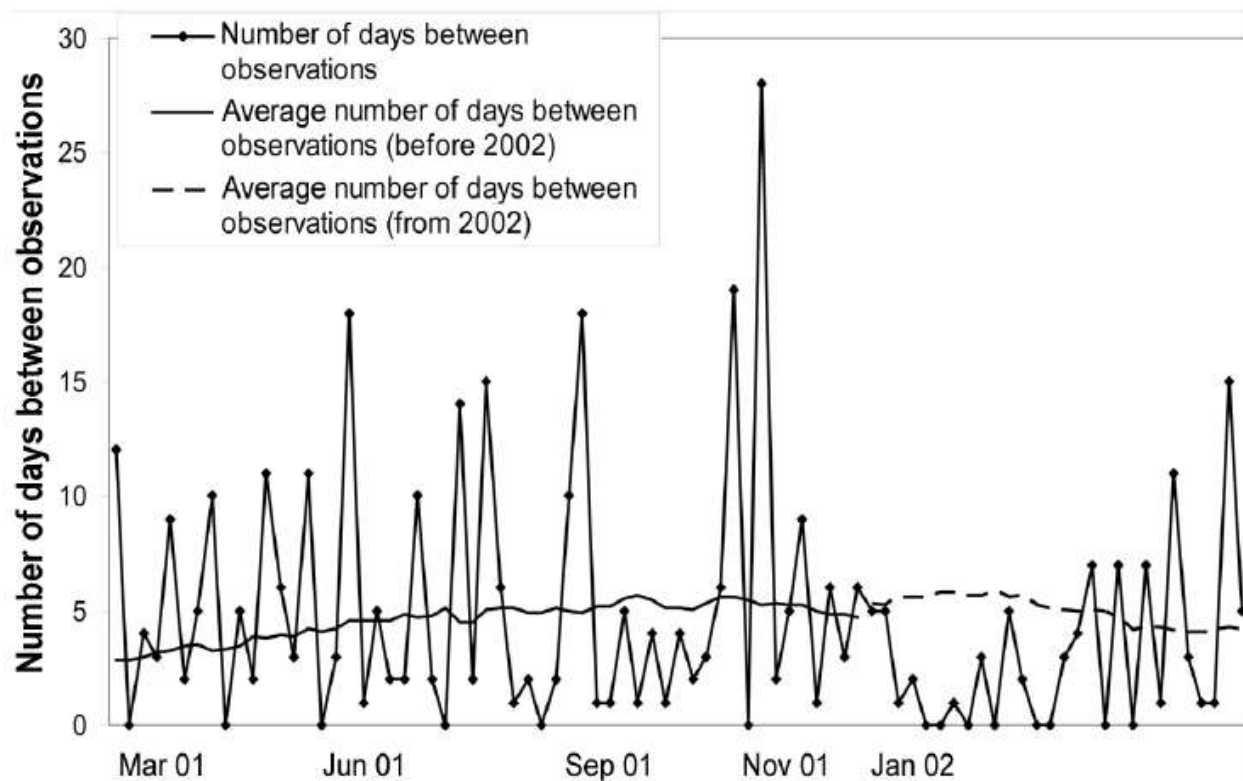


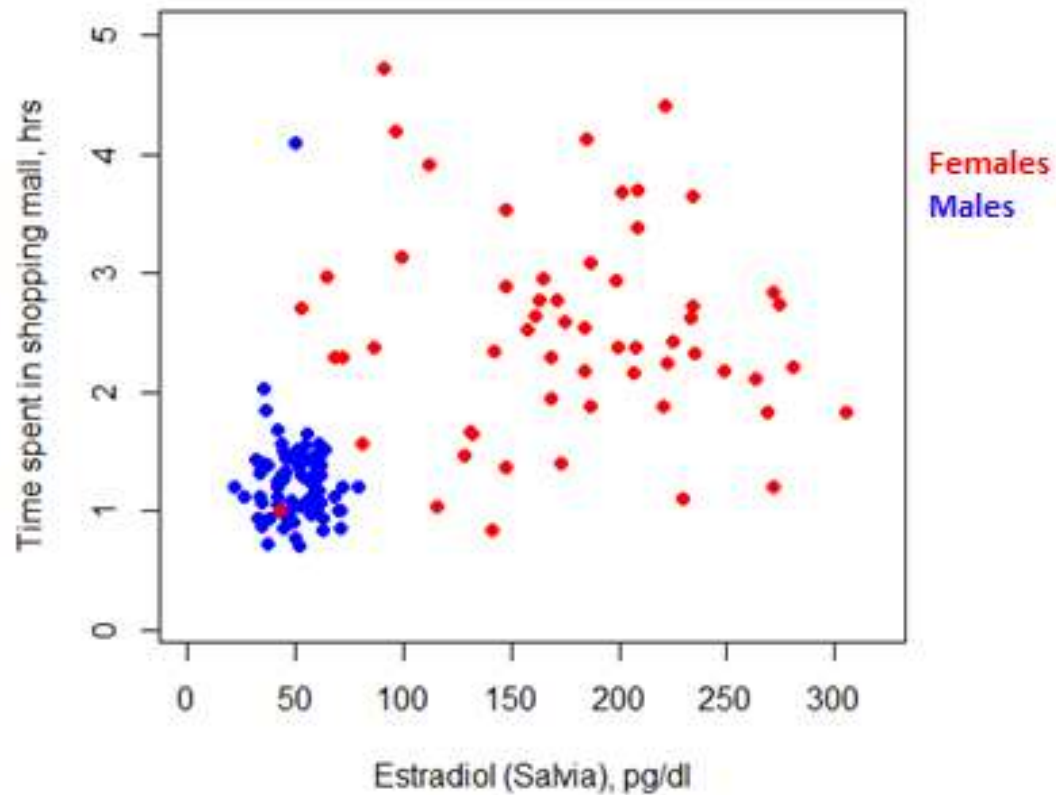
FIGURE 2. A g chart of 83 consecutive observations of the number of days between *Pseudomonas aeruginosa* infections during the period from March 2001 to March 2002. The calculations of the mean and the limits were based on 30 observations, with a delay of 10 observations. Observation 62 occurred on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

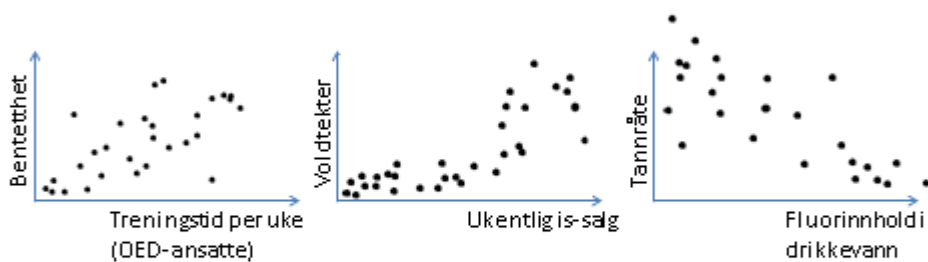


Vær varsom-plakaten for bivariate analyser.

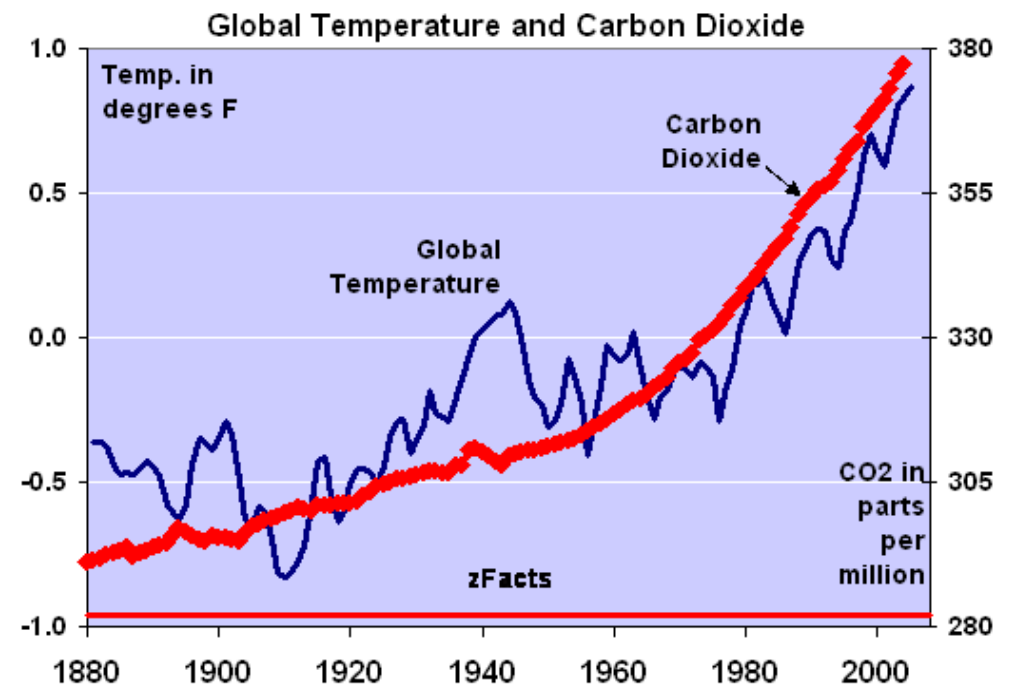
Assosiasjon, hva betyr det?

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

Årsakssammenhenger?



- B r man anbefale trening mot benskj rhet?
- B r man forby is?
- B r man anbefale fluor-tilskudd?



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



TILFELDIG? Denne tabellen viser de årlige dødsdrapene på side 15 denne og antallet drepte i trafikken. Synes du korrelasjonen er tilfeldig?

Av LINDA NILSEN METH og TORE 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, tilhøvsvalgt i NAF Oslo.

Hans funn, som du kan se i tabellen på denne siden, kan sies å være oppsiktsvekkende, om ikke akkurat statistisk utvetyklige: Korrelasjon mellom rentenivå og dødsdrapene i trafikken er påfallende lik kurven for trafikkdrepte de siste 15 årene.

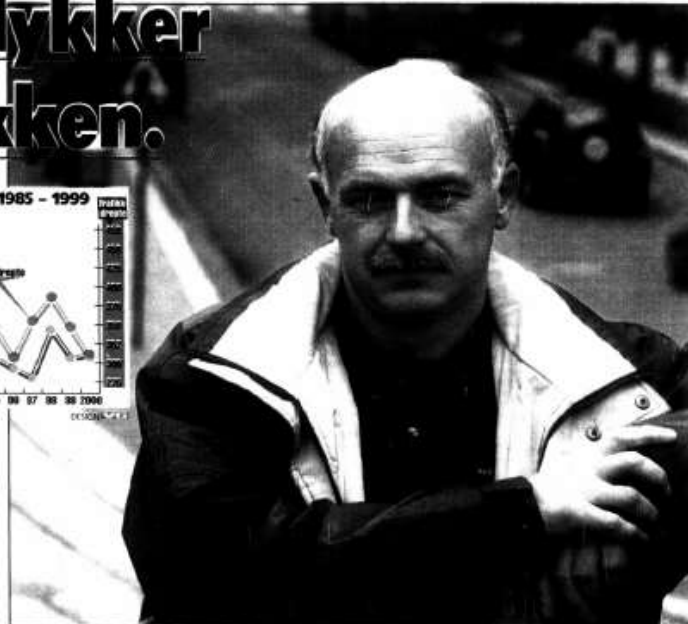
— Korrelasjon er like høy som i det såkalte søre tilfellet, sier Christiansen, som har lagt tallene frem for generalforsamlingen i NAF.

Bekymret og ukonsentrert
Christiansen (57) er enkelt.

— Stiger renten, stiger folks bekymringer. Disse mer bekymrede folk er dess mer ukonsentrerte blir de i det de holder på med. Ukonsentrerte trafikkantene gjør at faren for å bli involvert i ulykker øker.

— Er ikke dette en litt enkel forklaring?

Christiansen forklarer ikke ulykkesdriftsvesen i sin berber, men renten er faktisk noe av det alvorligste folk har å bekymre seg over. De tenker mye på det, og de mister de konsentrasjonen om det de ellers gjør. For eksempel kjøper bil, sier Christiansen.



RENTEN DREPER: — Antallet trafikkdrepte følger rentensvingningene bemerkelsesverdig, sier Arild Christiansen. Tilsvarende for Norges Automobil-Forbund (NAF) i Oslo og Ormskog.

Han tror også at bekymringer om rentenivå, ikke spiller og for lav lønn bedrer folk valgte sin bil.

Pått som sover dårlig, blir trette, og risikoen for å senere bak rattet blir større, sier Christiansen, og viser til at veldig mange ulykker i Norge skjer forth gjalden senere bak rattet.

Han mener det er på høy tid at statlige myndigheter ser mer på rentens betydning rundt dødsfall på norske veier.

— De vil de finne ut det ikke bare er fart, promille, handlingsuførelse og manglende bruk av sikkerhetsbelte som droper. Det finnes andre faktorer på ulykker. Rentenivået er en av dem, sier Christiansen.

— Hvordan ikke av årlige om å spore sammenheng mellom rente og trafikkuulykker?

— Det er ikke lett å

forklare, men det var et resultat av interesse for trafikkuulykker. Jeg er ikke sikker på at myndighetene gjør det riktige for å begrense ulykker. Er det jeg har funnet, kan det kanskje være noe grunnlag å søke reaksjon om å senke fart, sier Christiansen.

Og hva gjør han selv for å unngå å bli rammet av rentestigning og trafikkuulykker? — Jeg har fastrente og kjøretid. Det er min beste forsikring.

ring, gløer Arild Christiansen. Tre måneders arbeidsmarkedsstatistikk viser ikke Christiansen teori.

— Vi har nemlig sett at folk blir mer forsiktede og at slykkestallene går ned i daglig tilfelle — og at det er snarlig høy korrelasjon slykkestallene øker, sier Harald Aas i TDE.

Christiansen mener at folk blir mer forsiktede og at slykkestallene går ned i daglig tilfelle — og at det er snarlig høy korrelasjon slykkestallene øker, sier Harald Aas i TDE.

Christiansen mener at folk blir mer forsiktede og at slykkestallene går ned i daglig tilfelle — og at det er snarlig høy korrelasjon slykkestallene øker, sier Harald Aas i TDE.

Christiansen mener at folk blir mer forsiktede og at slykkestallene går ned i daglig tilfelle — og at det er snarlig høy korrelasjon slykkestallene øker, sier Harald Aas i TDE.

Christiansen mener at folk blir mer forsiktede og at slykkestallene går ned i daglig tilfelle — og at det er snarlig høy korrelasjon slykkestallene øker, sier Harald Aas i TDE.

Uendret rente

Engstelige boligkjøpere fikk ingen hjelp av Norges Bank. Det ble ingen rentenedgang eller gjensidige hovedlysmøte. 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 fram 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 forrige måned.

- Bilene er tryggere

Av ELNE DALLAM
En nedgang i antall trafikkuulykker på 50-tallet bergrer sammen med at bilene er blitt tryggere. Rentenivået betyr mindre, mer informasjonssøkt Stein Haakonson i Finansringens Hovedorganisasjon.

Etter årevis i trafikken kan bygge sammen med at bilene bygges sikrere enn tidligere, med mer sikkerhetsstyrer. Det er ingen tvil om at nye bilene er sikrere. Førre dat og flere lever videre, enkle friske eller sterkt innviduelt, sier Haakonson.
Tall fra Finansringens

© 2000. Alle rettigheter reservert.

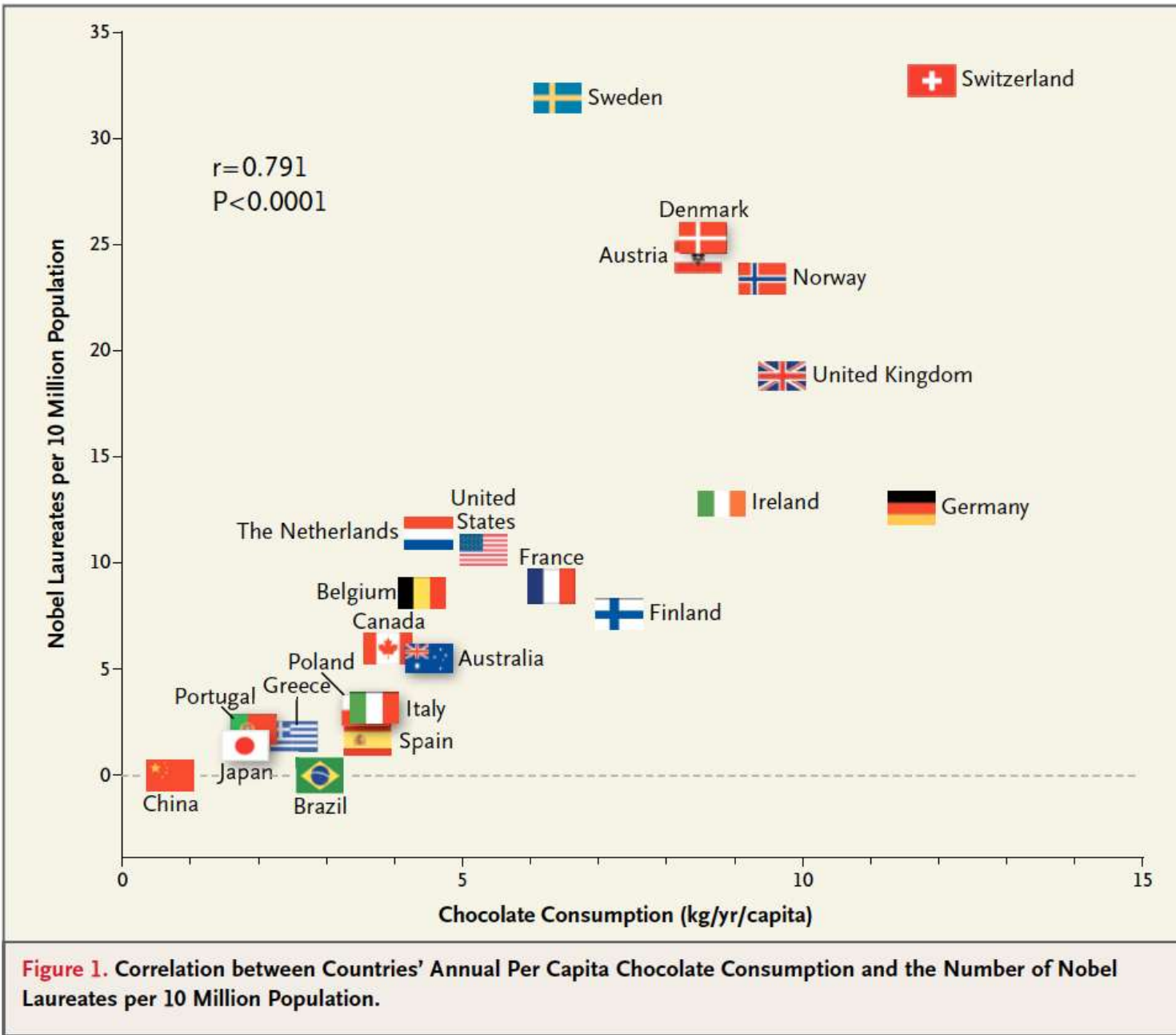


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

En annen måte å oppsummere lineære sammenhenger på: Beregn linja $y = ax + b$ som passer best til punktene (Ch 2.4 i boka)

Etter dette er det nyttig å **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.

Kommer:

STK1000 Uke 39 og utove. Studentene forventes å lese Ch 3.4 og Ch 4 i læreboka (MMC).

Essensen i statistisk analyse