

Lineær regresjon

Regresjonsanalyse: Analyse av smh mellom to eller flere variabler; der den ene er definert som respons og en eller flere andre er definert som forklaringsvariabler.

Respons

y

Outcome

Endepunkt

Avhengig variabel

Output-variabel

Endogen variabel

...

Dependent variable

Forklaringsvariabel

x

Prediktor

Kovariat

Uavhengig variabel

Eksogen variabel

...

Independent variable

Residualer

Bruktbilpris =

Lineær funksjon

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Fortsatt lineær

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

Fortsatt lineær

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 x_3 + \beta_3 x_4^2 + \beta_4 \ln(x_5) + \varepsilon$$

Ikke-lineær regresjon

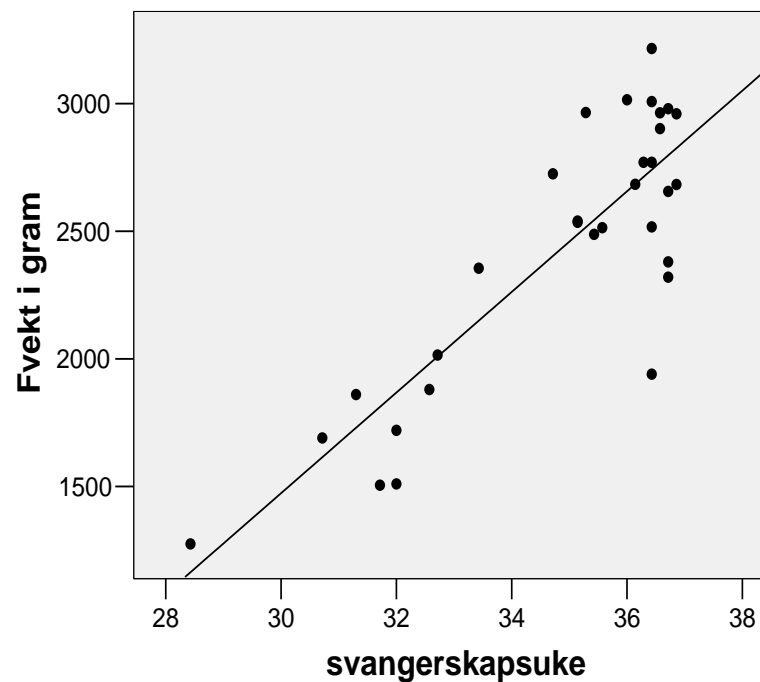
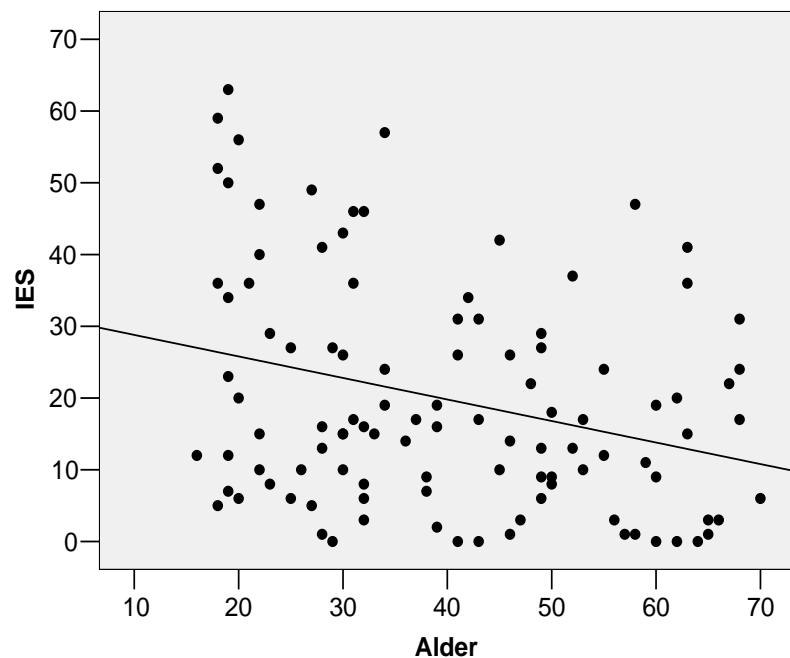
$$y = \beta_0 + \frac{\beta_1 x_1 + \beta_2 x_2 x_3}{\beta_3 x_4^2} \cdot \beta_4 \ln(x_5) + \varepsilon$$

Ulike typer data* krever ulike regresjonsmodeller

Respons-variabel*	Kategorisk (syk/frisk)	Kontinuerlig	Sensurerte data (levetider)
Regresjons-analyse	Logistisk regresjon	Lineær regresjon	Cox-regresjon
Effektmål	OR	Stigningstallet B (SPSS-notasjon), som angir hvor mange enheters økning som forventes i responsen per enhets økning i forklaringsvariabelen.	RR

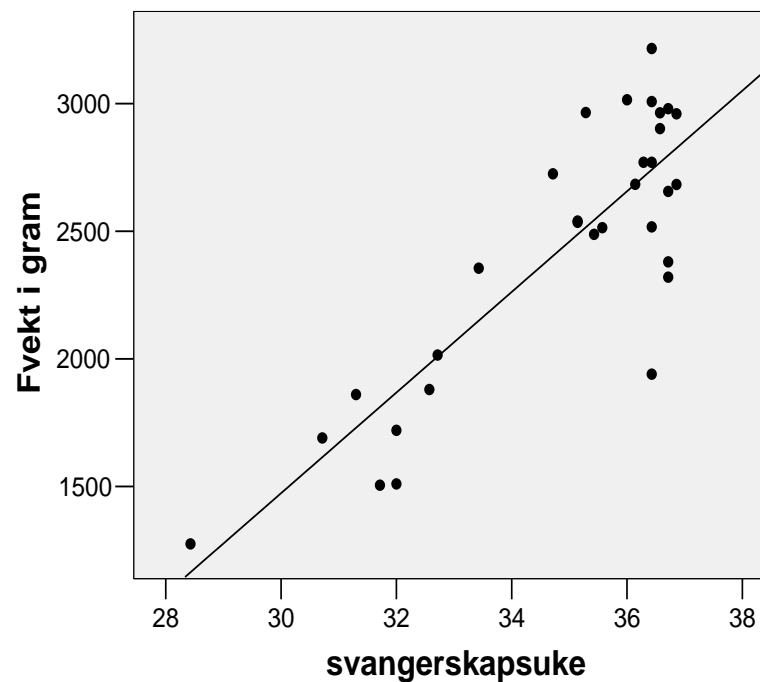
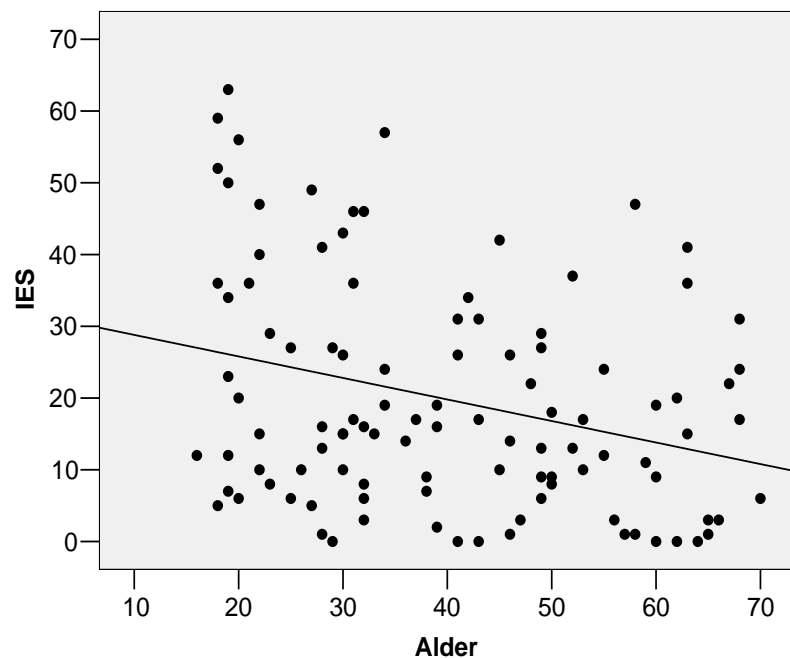
* OBS: Forklaringsvariablene kan være hvilken som helst type variabler: Binære, kategoriske (ordinale eller nominale) og kontinuerlige. Det er imidlertid enklest når de er binære eller kontinuerlige.

Effektmål: Tallfesting av det vi ønsker å studere: Kjenn din(e) måleskala(er)!



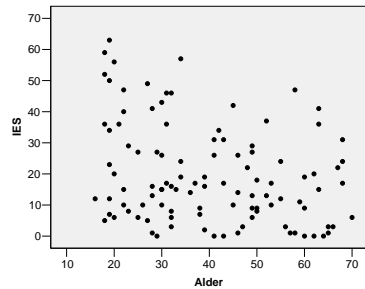
Regresjonskoeffisienten B angir hvor mange enheter fødselsvekten eller stresscoren IES forventes å endre seg når forklaringsvariabelen endres med en enhet.
IES: Impact of Event Scale.

Gjennomsnittlig enheters reduksjon i IES ved å øke alderen fra 20 til 60 år.



Mellom svangerskapsuke 28 og 38 øker barnets fødselsvekt med gjennomsnittlig gram i uka.

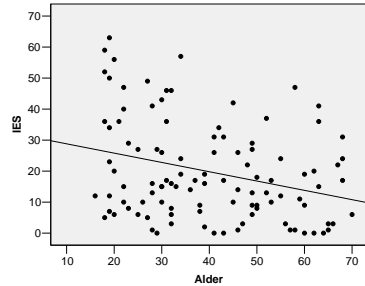
Tolkning av effektmål. Responsvariabel: Impact of Event Scale, IES.



Kont. f.var

Eks: alder

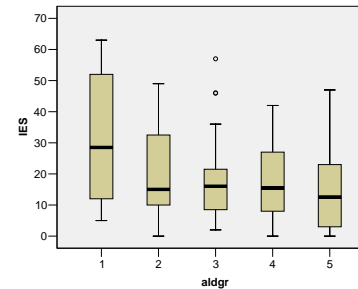
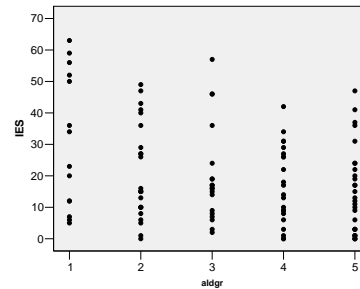
Med regresjonslinje:



B= -0.3

95%KI [-0.5 , -0.1]

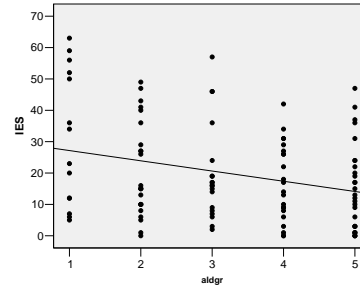
p=0.002



Ordinal f.var

Eks: 10 års aldersgrupper

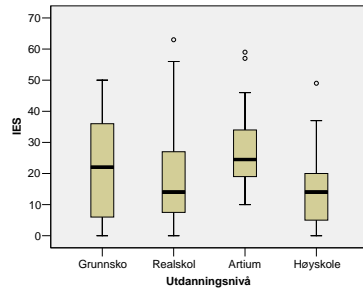
Med regresjonslinje:



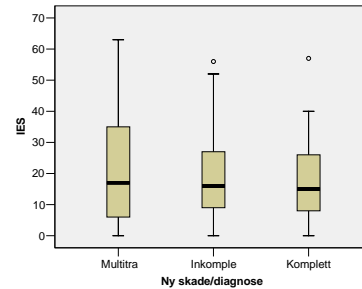
B=-3.3

95% KI [-5,4 , -1,1]

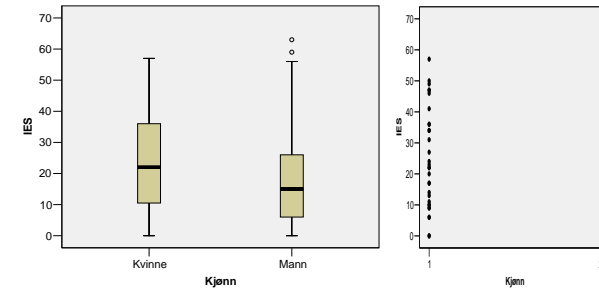
p=0.003



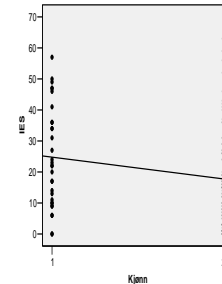
”Ordinal” f.var
Eks: utdanning



Nominal f.var
Eks: skadetype



Binær f.var
Eks: Kjønn
Med regresjonslinje:



Nødvendig å lage dummy-variabler.

B= -7.0
95%KI [-13.6 , -0.4]
p=0.04

Bivariat, univariat, multippel eller multivariat?

Bivariate analyser er analyse av smh mellom to variabler

Regresjonsanalyser med én responsvariabel og én forklaringsvariabel kalles ofte univariate analyser, eller «enkle» regresjonsanalyser

Bivariate analyser og univariate (enkle) regresjonsanalyser gir identiske/svært like resultater (p-verdier) selv om effektmålene er forskjellige:

- **Pearson korrelasjon gir identiske p -verdier som lineær-regresjon med én kontinuerlig forklaringsvariabel;**
- **Pearsons χ^2 gir identiske p -verdier som logistisk regresjon med én binær forklaringsvariabel.**

Resultatene fra enkel, eller univariat regresjonsanalyse kalles ujusterte (crude, unadjusted):

Ujusterte estimat, ujusterte KI, ujusterte p -verdier.

Regresjonsanalyse med én respons-variabel og flere forklaringsvariabler kalles ofte multippel eller multivariabel analyse.

Resultatene fra multippel regresjonsanalyse kalles justerte/korrigerte (adjusted/corrected): Justerte estimat, justerte KI og justerte p -verdier.

Begrepet *Multivariat analyse* brukes ofte om regresjonsanalyser der man ser på flere (respons)variabler simultant (ofte repeterte målinger) og evt også en eller flere forklaringsvariabler. Ikke aktuelt i STK1000.

Univariate analyser (enkle analyser, som i Ch 10) er første steg i en regresjonsanalyse.

	Univariat analyse Ujusterte estimat, KI, p -verdier			Multippel analyse Justerte estimat, KI, p -verdier		
	β	95%KI for β	p	β	95%KI for β	p
Forklarings-variabel Eller prediktor						

For å gjøre en univariat regresjonsanalyse (enkel regresjonsanalyse), må vi vite prinsippene for regresjon og hvordan en regresjonsanalyse gjennomføres og tolkes.

H_0 "ingen effekt" tilsvarer $\beta_1 =$.

Mao: Dersom KI for β_1 ikke inneholder 0, er det en signifikant effekt (av forklaringsvariabelen eller prediktoren på responsvariabelen).

Multippel regresjon (Ch 11)

**Hvorfor en regresjonsanalyse er som en blodprøve:
Hva er egentlig problemstillingen?**

«To predict or to explain»

**Skal vi forklare; altså estimere sammenhengen
mellom responsvariabelen og det vi antar er en
forklaringsvariabel, så riktig, «unbiased» som mulig?
eller**

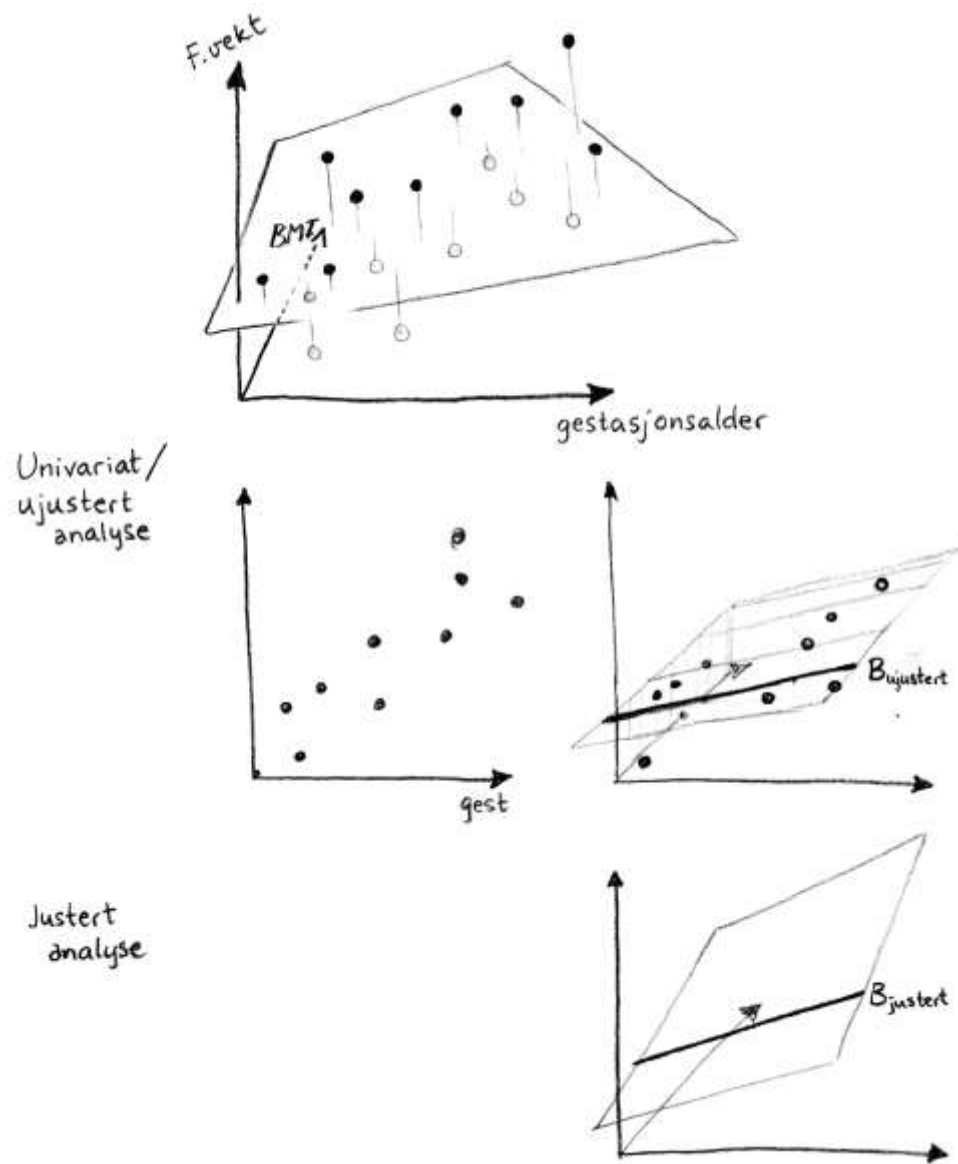
**Skal vi predikere; altså forutsi fremtidige verdier av
responsvariabelen så godt som mulig?**

**Først: Hvorfor multippel regresjon er som å slå en
sponplate gjennom en myggsverm eller en
fiskestim**

Eksempel:

Responsvariabel: Fødselsvekt

**Kovariater (enten forklaringsvariabler eller
prediktorer): Svangerskapslengde («gestasjonsalder»)
og BMI, altså $\{(vekt \text{ i kg})/(\text{høyde i m})^2\}$**



For at dette planet skal passe best mulig til både fødselsvekt og begge de to forklaringsvariablene (nå må vi ikke bare tilpasse modellen utelukkende etter svangerskapslengden, men også til BMI), kan det hende at planet "vipper" litt i forhold til den første linja vi tilpasset, og brattheten til den første linja justeres litt.

Dette kaller vi "justert effekt-estimat".

Denne nye β -en viser altså "effekten av BMI, korrigert for gestasjonsalder".

Tilsvarende vil den nye β -en for effekten av gestasjonsalder være justert for BMI.

Eksempel fra Pubmed: [Epidemiology](#). 2004 Sep;15(5):557-64.

The relation between birth weight and intima-media thickness in middle-aged adults.

[Tilling K](#), [Smith GD](#), [Chambless L](#), [Rose K](#), [Stevens J](#), [Lawlor D](#), [Szklo M](#).

Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol, UK. kate.tilling@bristol.ac.uk

BACKGROUND: Birth weight has been found to be inversely associated with the risk of coronary heart disease and stroke, although the mechanisms for this association remain unclear. Here, we investigate the relation between reported birth weight and atherosclerosis in middle age.

METHODS: We included the 9817 participants (age 44-65) in the Atherosclerosis Risk in Communities (ARIC) study who were neither a twin nor born prematurely. Carotid atherosclerosis was assessed as intima-media thickness measured by B-mode ultrasound. We studied the association with recalled exact birth weight, and for those unable to recall exact birth weight, with recalled birth weight category.

RESULTS: Mean intima-media thickness (+/- standard deviation) was 0.73 +/- 0.17 mm. Mean birth weight for the 4635 participants recalling exact birth weight was 3.49 +/- 0.71 kg. A further 4946 participants recalled birth weight category, with 4730 (96%) reporting "medium" birth weight. In univariate analysis, birth weight and intima-media thickness were positively related. However, adjustment for confounding factors reduced the association to only a 0.004 mm higher intima-media thickness (95% CI = - 0.003 to 0.011) mm per 1 kg of birth weight. The same pattern of univariate positive relationship and attenuation with adjustment was seen for birth weight category and intima-media thickness. There was no evidence of interaction between adult body mass index (BMI) and birth weight, or of interaction between category of adult BMI and birth weight category. An inverse relation between birth weight category and intima-media thickness was seen only for those in the lowest category of adult BMI (BMI <25 kg/m).

CONCLUSIONS: We found no evidence of a clinically significant relation between birth weight and carotid atherosclerosis.