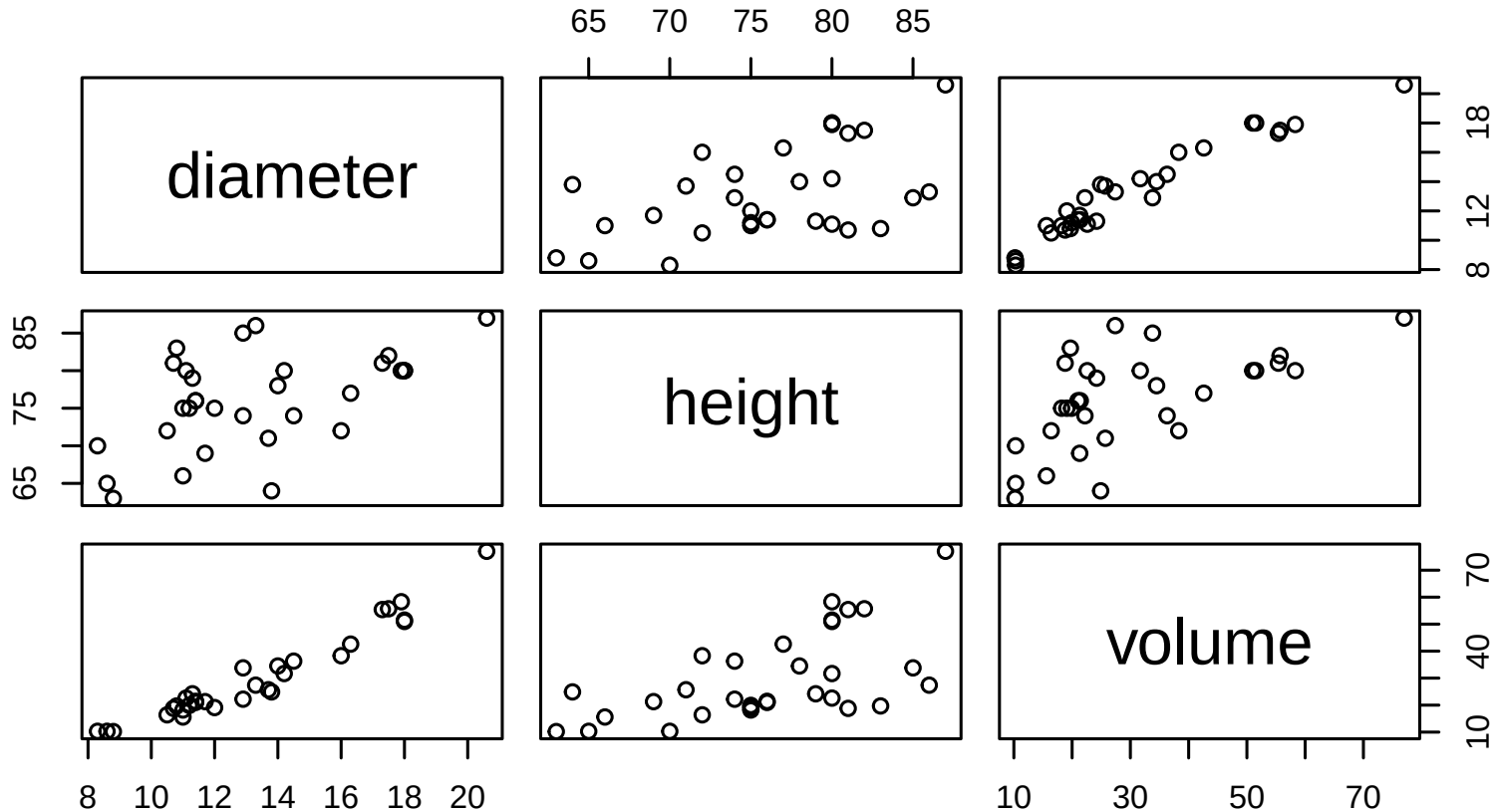
STK1110 Høsten 2026

Multippel Lineær Regresjon

Tilsvarer Avsnitt 12.7, 12.8 og 12.9

Flere forklaringsvariable

- Trær: Volum, høyde, diameter



- Naturlig å tenke seg at volum er avhengig av høyde og diameter

Multippel lineær regresjon

- En multippel lineær regresjonsmodell basert på k forklaringsvariabler er gitt ved

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i, \quad \varepsilon_i \overset{uif}{\sim} N(0, \sigma^2), \quad i = 1, \ldots, n$$

  der verdiene x_{ij} av forklaringsvariablene anses som gitt.
- Da er

$$\mathrm{E}(Y_i | x_{i1}, \ldots, x_{ik}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}.$$

- Det betyr at β_0 er forventet respons når $x_{i1} = \ldots = x_{ik} = 0$ og β_j er forventet endring i responsen når x_j øker med 1, mens alle andre forlaringsvariabler holdes fast.
- Her tar vi for oss resultater knyttet til multippel lineær regresjon, men viser ikke disse før vi kommer til Avsnitt 12.9.

Minste kvadraters estimatene (MKE)

- MKE $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ er de verdiene $b_0, b_1, \ldots, b_k$ som minimerer

$$g(b_0, b_1, \ldots, b_k) = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_{i1} + \ldots + b_k x_{ik}))^2.$$

- MKE finnes ved å løse ligningene, $j = 0, 1, 2, \ldots, k$,

$$\frac{\partial g(b_0, b_1, \ldots, b_k)}{\partial b_j} = 0.$$

- Disse ligningene blir

$$\begin{aligned}
\frac{\partial g(b_0,b_1,\ldots,b_k)}{\partial b_0} &= -2 \textstyle\sum_{i=1}^{n} (y_i - (b_0 + b_1 x_{i1} + \ldots + b_k x_{ik})) &&= 0 \\
\frac{\partial g(b_0,b_1,\ldots,b_k)}{\partial b_1} &= -2 \textstyle\sum_{i=1}^{n} x_{i1} (y_i - (b_0 + b_1 x_{i1} + \ldots + b_k x_{ik})) &&= 0 \\
&\vdots &&\vdots \\
\frac{\partial g(b_0,b_1,\ldots,b_k)}{\partial b_k} &= -2 \textstyle\sum_{i=1}^{n} x_{ik} (y_i - (b_0 + b_1 x_{i1} + \ldots + b_k x_{ik})) &&= 0
\end{aligned}$$

MKE forts., tilpassede verdier og residualer

- Man finner dermed MKE ved å løse $k+1$ lineære ligninger med $k+1$ ukjente
- Eksplisitte løsninger på matriseform utledes i Avsnitt 12.9

- De tilsvarende minste kvadraters estimatorene får en ved å erstatte y_i-ene med Y_i-ene i ligningene for MKE.

- Minste kvadraters estimatorer oppfyller $\hat{\beta}_j \sim N(\beta_j, c_{jj}\sigma^2)$, der c_{jj} er konstanter vi vil finne uttrykk for i Avsnitt 12.9.

- De tilpassede verdiene og tilhørende residualene er nå gitt ved:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_k x_{ik} \quad \text{og} \quad e_i = y_i - \hat{y}_i, \quad i = 1, \ldots, n.$$

Kvadratsummer og estimat for variansen

- Feil- eller residualkvadratsummen er gitt ved

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

som gir oss et estimat for σ^2 ved

$$s^2 = \frac{SSE}{n - (k+1)} = MSE.$$

- Den tilsvarende estimatoren S^2 får vi ved å erstatte y_i med Y_i og $\hat{y}_i$ med $\hat{Y}_i$, og S^2 er forventningsrett for σ^2.
- De andre kvadratsummene er som før definert som $SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ og $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$, og vi har fortsatt $SST = SSR + SSE$.

Forklart variasjon

- Et mål for hvor mye av variasjonen i y_i-ene som er forklart av x_{ij}-ene er

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

- Denne øker imidlertid med antall forklaringsvariabler, selv når en inkluderer variabler som har helt marginal betydning for y_i-ene.
- Et bedre mål er derfor justert R^2, gitt ved

$$R^2_{adj} = 1 - \frac{SSE/(n-(k+1))}{SST/(n-1)}.$$

Konfidensintervall for og hypotesetester angående β_j

- La $S_{\hat{\beta}_j} = \sqrt{c_{jj}}S$ være en estimator for standardavviket til $\hat{\beta}_j$.
- Vi har $\frac{\hat{\beta}_j - \beta_j}{S_{\hat{\beta}_j}} \sim t_{n-(k+1)}$.
- Det betyr at et $100 \cdot (1-\alpha)\%$ konfidensintervall for β_j er gitt ved

$$\hat{\beta}_j \pm t_{\alpha/2, n-(k+1)} s_{\hat{\beta}_j}.$$

Hypotesetester angående β_j

- Vi kan vi bruke observatoren $T = \frac{\hat{\beta}_j - \beta_{j0}}{S_{\hat{\beta}_j}}$ til å teste hypoteser
- Tester med signifikansnivå α:
  1. $H_0 : \beta_j \leq \beta_{j0}$ mot $H_a : \beta_j > \beta_{j0}$
     - Forkast H_0 dersom $T \geq t_{\alpha, n-(k+1)}$
     - Med observert $T = t$ blir p-verdien $\mathrm{P}(T_{n-k-1} > t)$ der T_{n-k-1} er t_{n-k-1}-fordelt
  2. $H_0 : \beta_j \geq \beta_{j0}$ mot $H_a : \beta_j < \beta_{j0}$
     - Forkast H_0 dersom $T \leq -t_{\alpha, n-(k+1)}$
     - Med observert $T = t$ blir p-verdien $\mathrm{P}(T_{n-k-1} < t)$
  3. $H_0 : \beta_j = \beta_{j0}$ mot $H_a : \beta_j \neq \beta_{j0}$
     - Forkast H_0 dersom $T \leq -t_{\alpha/2, n-(k+1)}$ eller $T \geq t_{\alpha/2, n-(k+1)}$
     - Med observert $T = t$ fås p-verdi $\mathrm{P}(T_{n-k-1} > |t|) + \mathrm{P}(T_{n-k-1} < -|t|)$

F-tester for β_j-ene

- Anta vi nå vil teste

$$H_0 : \beta_j = 0, \quad j = 1, \ldots, k \quad \text{mot} \quad H_a : \text{ minst én } \beta_j \neq 0.$$

- Testobservator: $F = \frac{SSR/k}{SSE/(n-(k+1))} = \frac{MSR}{MSE}$
- Under H_0 er $F \sim F_{k,n-(k+1)}$
- Vi forkaster da H_0 ved signifikansnivå α dersom $F \geq F_{\alpha,k,n-(k+1)}$.
- Variansanalysetabell:

Kilde til var.	**Df**	**SS**	**MS**	**f**
Regresjon	k	SSR	$MSR = \frac{SSR}{k}$	$\frac{MSR}{MSE} = \frac{SSR/k}{SSE/(n-(k+1))}$
Feil	$n-(k+1)$	SSE	$MSE = \frac{SSE}{n-(k+1)}$	
Total	$n-1$	SST		

F-tester for β_j-ene (forts.)

- Vi kan også være interessert i å teste om noen av, men ikke alle, β_j-ene er 0, f.eks.

  $H_0 : \beta_{m+1} = \ldots = \beta_k = 0$ mot $H_a :$ minst én $\beta_{m+1}, \ldots, \beta_k \neq 0$.

- La $\beta_0^*, \ldots, \beta_m^*$ være minste kvadraters estimatorer for $\beta_0, \ldots, \beta_m$ i den reduserte modellen

  $$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im} + \varepsilon_i.$$

- Videre, la $Y_i^* = \beta_0^* + \beta_1^* x_{i1} + \ldots + \beta_m^* x_{im}$, og sett $SSE_0 = \sum_{i=1}^n (Y_i - Y_i^*)^2$.
- Da er $F = \frac{(SSE_0 - SSE)/(k-m)}{SSE/(n-(k+1))} \sim F_{k-m,n-(k+1)}$.
- Vi forkaster da H_0 ved signifikansnivå α dersom $F \geq F_{\alpha,k-m,n-(k+1)}$.

Sjekk av modellantakelser

- I den multiple lineære regresjonsmodellen antar vi fortsatt at
  1. Sammenhengen er lineær.
  2. Variansen er lik for alle verdier av x_{ij}-ene.
  3. Feilleddene ε_i er normalfordelt.
  4. Feilleddene ε_i er uavhengige.
- De tre første antakelsene kan vi sjekke ved hjelp av samme plott som for enkle lineære regresjonsmodeller.

Konfidensinterv. for $\mu_{Y|x_1^*,\ldots,x_k^*}$ og prediksjonsinterv. for Y^*

- Anta at vi for gitte verdier $x_1^*, \ldots, x_k^*$ er interessert i å estimere $\mu_{Y|x_1^*,\ldots,x_k^*} = \mathrm{E}(Y|x_1^*, \ldots, x_k^*) = \beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^*$.
- Denne kan estimeres med

$$\hat{Y} = \hat{\mu}_{Y|x_1^*,\ldots,x_k^*} = \hat{\beta}_0 + \hat{\beta}_1 x_1^* + \ldots + \hat{\beta}_k x_k^*,$$

- La $s_{\hat{Y}}$ være estimatet for standardavviket $\sigma_{\hat{Y}}$ til $\hat{Y}$. Da er et $100 \cdot (1-\alpha)\%$ konfidensintervall for $\mu_{Y|x_1^*,\ldots,x_k^*}$ gitt ved:

$$\hat{y} \pm t_{\alpha/2,n-(k+1)} s_{\hat{Y}}.$$

- Videre kan det vises at $\mathrm{V}(Y - \hat{Y}) = \sigma^2 + \sigma^2_{\hat{Y}}$, slik at et $100 \cdot (1-\alpha)\%$ prediksjonsintervall for en ny verdi Y^* med tilsvarende kovariatverdier $x_1^*, \ldots, x_k^*$ er

$$\hat{y} \pm t_{\alpha/2,n-(k+1)} \sqrt{s^2 + s^2_{\hat{Y}}}.$$

Polynomisk regresjon

- Anta at vi nå bare har én forklaringsvariabel x, men at sammenhengen mellom x og y beskrives bedre av en kvadratisk enn av en lineær funksjon, dvs.

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

- Da er $\mathrm{E}(Y|x) = \beta_0 + \beta_1 x + \beta_2 x^2$, som er en kvadratisk funksjon.
- Vi kan imidlertid formulere dette som en multippel lineær regresjonsmodell med $x_{i1} = x_i$ og $x_{i2} = x_i^2$.
- Da kan vi bruke metodene for multippel lineær regresjon.
- Imidlertid blir tolkningen av β_1 og β_2 annerledes.
- Dette kan umiddelbart generaliseres til 3. og høyere grads polynomer (men nytten av dette er ofte begrenset)

Interaksjonseffekter

- Anta at vi nå at vi har to forklaringsvariabler og modellen

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

- Da er $\mathrm{E}(Y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, hvilket betyr at effekten av x_2 er den samme for alle verdier av x_1 og omvendt.
- Hvis derimot modellen er

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i,$$

  så er $\mathrm{E}(Y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_{i1} x_{i2}$, slik at effekten av x_2 avhenger av verdien til x_1.
- Dette kalles en interaksjonseffekt mellom to forklaringsvariabler.
- En kan også ha (flere) interaksjonseffekter i modeller med mer enn to kovariater, og en interaksjonen kan involvere mer enn to variabler.

Kategoriske forklaringsvariabler

- Vi ønsker nå å modellere sammenhengen mellom responsvariabelen Y og de to forklaringsvariablene x_1 og x_2, der x_1 er kategorisk, mens x_2 er numerisk.
- Så langt har vi sett på numeriske forklaringsvariabler, men en kan gjerne også inkludere kategoriske forklaringsvariabler i en lineær regresjonsmodell.
- Anta først at x_1 kan tilhøre én av to mulige kategorier, slik at

$$x_1 = \begin{cases} 0, & \text{kategori 1} \\ 1, & \text{kategori 2} \end{cases}.$$

- Vi kan da spesifisere modellen som

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i.$$

Kategoriske forklaringsvariabler (forts.)

- Når $x_1 = 0$, er da

$$Y_i = \beta_0 + \beta_2 x_{i2} + \varepsilon_i \text{ og } \mathrm{E}(Y|x_1, x_2) = \beta_0 + \beta_2 x_2,$$

mens når $x_1 = 1$, er

$$Y_i = \beta_0 + \beta_1 + \beta_2 x_{i2} + \varepsilon_i \text{ og } \mathrm{E}(Y|x_1, x_2) = \beta_0 + \beta_1 + \beta_2 x_2.$$

- Fortolkningen av β_1 er dermed forventet endring i responsen når x_1 er i kategori 2, sammenlignet med når x_1 er i kategori 1.

Regresjonsmodellen består altså av to parallelle linjer,
en for hver av de to kategoriene til x_1.

Kategoriske variable og interaksjon

Vi kan utvide modellen med den kategoriske x_1 og den numeriske x_2 til

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i.$$

- Når $x_1 = 0$, blir som foran

$$\mathrm{E}(Y|x_1, x_2) = \beta_0 + \beta_2 x_2,$$

mens når nå $x_1 = 1$, blir

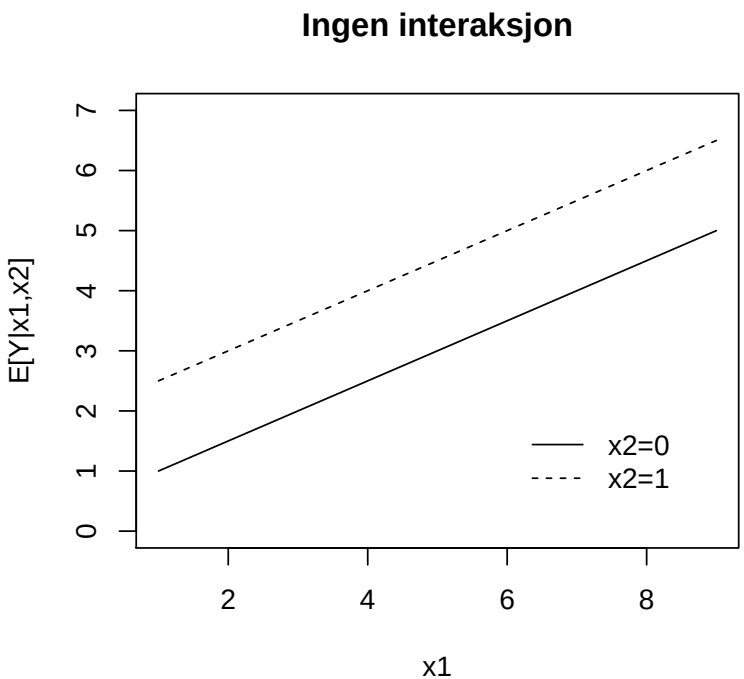
$$\mathrm{E}(Y|x_1, x_2) = \beta_0 + \beta_1 + (\beta_2 + \beta_3) x_2.$$

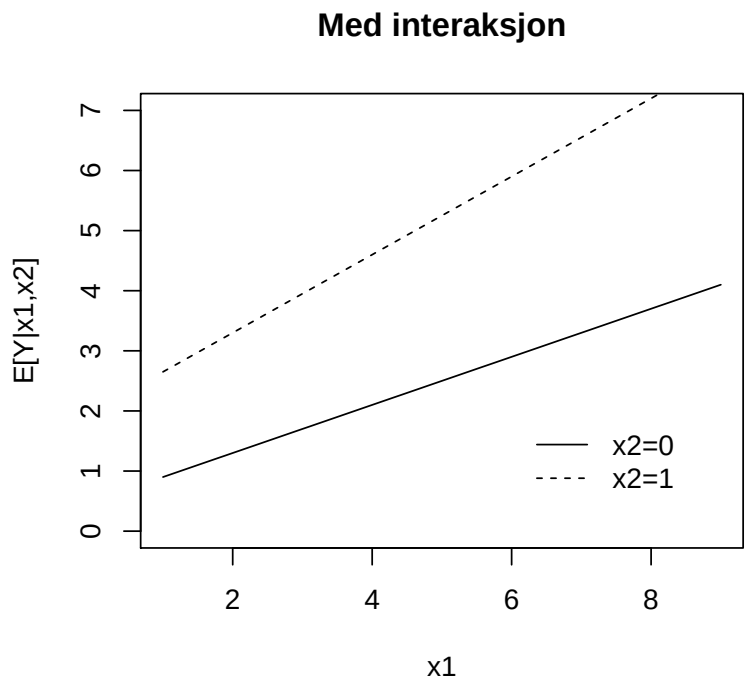
Dette generer en modell med både ulike konstantledd β_0 og $\beta_0 + \beta_1$ og ulike stigningskoeffisienter β_2 og $\beta_2 + \beta_3$ avhengig av om $x_1 = 0$ eller $x_1 = 1$.

Illustrasjon interaksjon

Anta at x_2 er binær, altså antar verdiene 0 eller 1. Da vil altså

- en modell uten interaksjon kunne angis som to parallelle regresjonslinjer mhp x_1
- mens i en modell med interaksjon så vil disse regresjonslinjene ha ulike stigningskoeffisenter





Kategoriske forklaringsvariabler (forts.)

- Generelt, dersom en forklaringsvariabel kan tilhøre en av p kategorier, lager en $p-1$ nye forklaringsvariabler med 0-er og 1-ere, som en bruker i modellen.
- Anta at vi vil lage en regresjonsmodell for Y med en kategorisk forklaringsvariabel x_1, med 3 mulige kategorier og en numerisk forklaringsvariabel x_2.
- Vi lager da de nye forklaringsvariablene

$$x_{11} = \begin{cases} 0, & \text{kategori 1 eller 3} \\ 1, & \text{kategori 2} \end{cases} \quad \text{og } x_{12} = \begin{cases} 0, & \text{kategori 1 eller 2} \\ 1, & \text{kategori 3} \end{cases}.$$

- Dette gjør vi for å ta hensyn til at til at det ikke må være noen ordning mellom kategoriene (2 er ikke "større" enn 1 og 3 er ikke "større" enn 2 i denne sammenhengen, nødvendigvis).

Kategoriske forklaringsvariabler (forts.)

- Regresjonsmodellen blir da:

$$Y_i = \beta_0 + \beta_1 x_{i,11} + \beta_2 x_{i,12} + \beta_3 x_{i2} + \varepsilon_i \text{ med}$$

med

$$\mathrm{E}(Y|x_1, x_2) = \begin{cases} \beta_0 + \beta_3 x_2, & \text{kategori 1} \\ \beta_0 + \beta_1 + \beta_3 x_2, & \text{kategori 2} \\ \beta_0 + \beta_2 + \beta_3 x_2, & \text{kategori 3} \end{cases}.$$

- Fortolkningen av β_1 er dermed forventet endring i responsen når x_1 er i kategori 2, sammenlignet med når x_1 er i kategori 1, og fortolkningen av β_2 er forventet endring i responsen når x_1 er i kategori 3, sammenlignet med når x_1 er i kategori 1.
- Omkodingen av kategoriske variabler til $p - 1$ $0 - 1$-variabler, som representerer de forskjellige kategoriene, kan gjøres med funksjonen `factor()` i R.

Enveis variansanalyse og multippel lineær regresjon

- Enveis variansanalyse er et spesialtilfelle av multippel lineær regresjon.
- I enveis variansanalyse antar vi at vi har I grupper med observasjoner $X_{i1}, \ldots, X_{iJ_i} \overset{uif}{\sim} N(\mu_i, \sigma^2)$ fra hver gruppe i, og X_{ij}-ene er uavhengige på tvers av grupper også.
- Her: "Tradisjonell" parametrisering av enveis ANOVA.
- La $n = \sum_{i=1}^{I} J_i$, $\bar{\mu} = \frac{1}{I}\sum_{i=1}^{I} \mu_i$ og $\alpha_i = \mu_i - \bar{\mu}$.
- Da har vi

$$X_{ij} = \bar{\mu} + \alpha_i + \varepsilon_i, \quad \varepsilon_{ij} \sim N(0, \sigma^2), \quad \sum_{i=1}^{I} \alpha_i = \sum_{i=1}^{I} (\mu_i - \bar{\mu}) = 0.$$

- Det betyr at $\alpha_I = -\sum_{i=1}^{I-1} \alpha_i$, og vi kan skrive

$$X_{ij} = \begin{cases} \bar{\mu} + \alpha_i + \varepsilon_{ij}, & i = 1, \ldots, I-1 \\ \bar{\mu} - \sum_{k=1}^{I-1} \alpha_k + \varepsilon_{ij}, & i = I \end{cases} \qquad j = 1, \ldots, J_i.$$

Enveis ANOVA og multippel lineær regresjon (forts.)

- Renummerer nå responsen X_{ij} til

$$Y_1 = X_{1,1}, \ldots, Y_{J_1} = X_{1,J_1}, Y_{J_1+1} = X_{2,1}, \ldots, Y_n = X_{I,J_I}$$

og definerer forklaringsvariablene $x_{k,i}$ ved

$$x_{k,i} = \begin{cases} 1, & \text{hvis individ } k \text{ tilhører gruppe } i,\ i = 1, \ldots, I-1 \\ -1, & \text{hvis individ } k \text{ tilhører gruppe } I \\ 0, & \text{ellers} \end{cases}.$$

- Da kan vi skrive

$$\begin{aligned} Y_k &= \bar{\mu} + \alpha_1 x_{k,1} + \ldots + \alpha_{I-1} x_{k,I-1} + \varepsilon_k \\ &= \beta_0 + \beta_1 x_{k,1} + \ldots + \beta_{I-1} x_{k,I-1} + \varepsilon_k \end{aligned}$$

med $\beta_0 = \bar{\mu}$ og $\beta_i = \alpha_i = \mu_i - \bar{\mu}$, $i = 1, \ldots, I-1$.

Sammenheng mellom enveis variansanalyse og multippel lineær regresjon (forts.)

- I R brukes (default) en annen parametrisering med

$$\alpha_1 = 0, \quad x_{k,i} = \begin{cases} 1, & \text{hvis individ } k \text{ tilhører gruppe } i, \\ 0, & \text{ellers} \end{cases} \quad i = 2, \ldots, I$$

- Da er $\beta_0 = \mu_1$ og $\beta_i = \mu_i - \mu_1$, $i = 2, \ldots, I$.

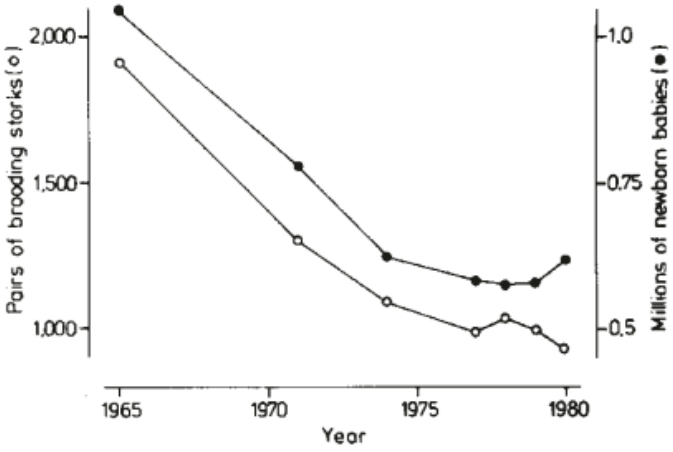
Eksempel

Fettopptak i rotter.

Konfundering: eksempel

A new parameter for sex education

SIR—There is concern in West Germany over the falling birth rate. The accompanying graph[1,2] might suggest a solution that every child knows makes sense.



HELMUT SIES
Institut für Physiologische Chemie 1, Universität Düsseldorf, Moorenstrasse 5, D-4000 Düsseldorf, FRG

1. *Fachserie Gebiet und Bevölkerung* (Statistisches Bundesamt, Kohlhammer, Stuttgart, 1984).
2. Bauer, S. & Thielcke, G. *Die Vogelwarte* **31**, 183–191 (1982).

Nature 1988, vol. 332, s. 495.

Figuren til venstre viser sammenhengen mellom antall par hekkende storker og antall fødsler i Vest-Tyskland i perioden 1965-1980.

Vi ser en sammenheng mellom slik at antall fødsler er høyt når antall storker også er høyt.

Men det er neppe snakk om en årsakssammenheng.

De neste slidene illustrerer hvordan en utelatt variabel kan gi et skjevt bilde av sammenhenger.

Dette refereres til som konfunderende eller sammenblandende (evt. "lurkende") variable

Konfundering

- Anta at den sanne modellen er

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

- I stedet tilpasser vi modellen

$$Y_i = \gamma_0 + \gamma_1 x_{i1} + \varepsilon_i^*, \quad \varepsilon_i^* \sim N(0, (\sigma^*)^2).$$

- Minste kvadraters estimator for γ_1 er $\hat{\gamma}_1 = \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1) Y_i}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}$, og man kan vise at

$$\hat{\gamma}_1 = \hat{\beta}_1 + \hat{\beta}_2 r_{12} \sqrt{\frac{S_{22}}{S_{11}}},$$

der $S_{11} = \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2$, $S_{22} = \sum_{i=1}^n (x_{i2} - \bar{x}_2)^2$ og $r_{12} = \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{i,1} - \bar{x}_1)^2 \sum_{i=1}^n (x_{i,2} - \bar{x}_2)^2}}$.

Konfundering (forts.)

- Her er r_{12} den empiriske korrelasjonen mellom de to forklaringsvariablene x_1 og x_2.
- Videre er

$$\mathsf{E}(\hat{\gamma}_1) = \beta_1 + \beta_2 r_{12} \sqrt{\frac{S_{22}}{S_{11}}}.$$

- Det betyr at hvis forklaringsvariablene er korrelert ($r_{12} \neq 0$), vil $\hat{\gamma}_1$ ikke estimere effekten av x_1 alene, men en størrelse som avhenger av både x_1 og x_2.
- Vi sier da at effekten av den første forklaringsvariabelen er konfundert av den andre forklaringsvariabelen.

Eksempel

Karakterer for studenter i "Computer Science".

Avsnitt 12.9: Multippel lineær regresjon på matriseform

- Den multiple lineære regresjonsmodellen kan skrives, samlet for alle observasjoner, som

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_{11} + \ldots + \beta_k x_{1k} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \ldots + \beta_k x_{nk} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad \varepsilon_i \overset{uif}{\sim} N(0, \sigma^2).$$

- Dette kan også skrives på matriseform.
- La

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \text{ og } \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix}$$

  der $\boldsymbol{X}$ kalles modell- eller designmatrisen.
- Modellen over kan da skrives som

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Minste kvadraters estimator

Minste kvadraters estimater er de verdiene $b_0, b_1, \ldots, b_k$ som minimerer

$$g(b_0, b_1, \ldots, b_k) = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_{i1} + \ldots + b_k x_{ik}))^2.$$

Hvis vi deriverer m.h.p. $b_0, b_1, \ldots, b_k$ og setter lik 0, får vi de såkalte normallikningene

$$\begin{aligned}
b_0 n + b_1 \textstyle\sum_{i=1}^{n} x_{i1} + \ldots + b_k \sum_{i=1}^{n} x_{ik} &= \sum_{i=1}^{n} y_i \\
b_0 \textstyle\sum_{i=1}^{n} x_{i1} + b_1 \sum_{i=1}^{n} x_{i1}^2 + \ldots + b_k \sum_{i=1}^{n} x_{i1} x_{ik} &= \sum_{i=1}^{n} x_{i1} y_i \\
&\vdots \\
b_0 \textstyle\sum_{i=1}^{n} x_{ik} + b_1 \sum_{i=1}^{n} x_{i1} x_{ik} + \ldots + b_k \sum_{i=1}^{n} x_{ik}^2 &= \sum_{i=1}^{n} x_{ik} y_i.
\end{aligned}$$

Minste kvadraters estimator (forts.)

- Dette kan skrives som

$$\begin{pmatrix} \sum_{i=1}^n 1 & \sum_{i=1}^n x_{i1} & \cdots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{i1}x_{ik} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{pmatrix}$$

- På matriseform blir dette

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}^T\boldsymbol{Y},$$

der $\boldsymbol{X}^T$ er den transponerte til $\boldsymbol{X}$.

- Minste kvadraters estimator $\hat{\boldsymbol{\beta}}$ blir da gitt ved (så sant $\boldsymbol{X}^T\boldsymbol{X}$ er inverterbar):

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}.$$

Alternativ utledning av normallikningene på matriseform

- Vi har kvadratsummen
  $g(b_0, b_1, \ldots, b_k) = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_{i1} + \ldots + b_k x_{ik}))^2$
- Med matriseuttrykk kan denne skrives om til
  $g(\boldsymbol{b}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) = \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{b}^T\boldsymbol{X}^T\boldsymbol{y} + \boldsymbol{b}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{b}$
- Videre har vi lov til skrive (siden $X^T\boldsymbol{X}$ er symmetrisk)

$$\frac{\partial g(\boldsymbol{b})}{\partial \boldsymbol{b}^T} = -2\boldsymbol{X}^T\boldsymbol{y} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{b}$$

- og hvis vi setter til null får vi normal-ligningene på matriseform

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}^T\boldsymbol{y}$$

som (igjen) gir $\boldsymbol{b} = \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Hattmatrisen

- Predikerte verdier er da gitt ved:

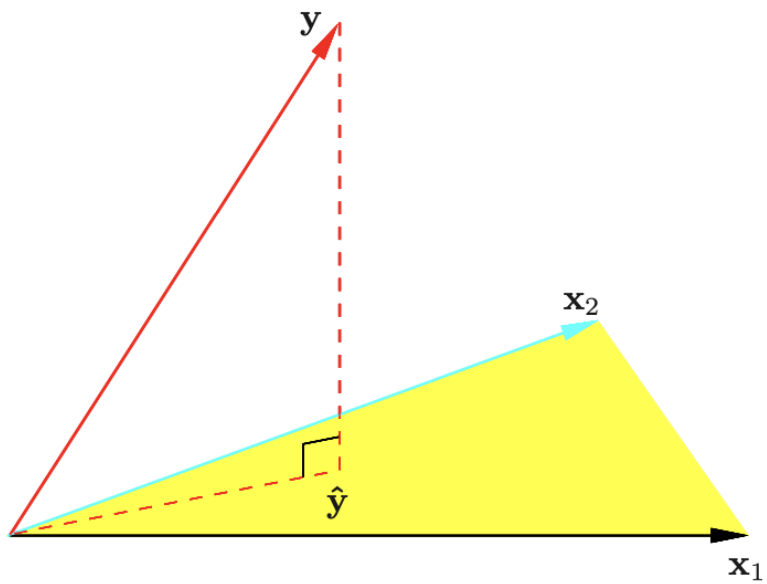
$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta} = \underbrace{\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T}_{\boldsymbol{H}}\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Y},$$

  der $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ kalles hattmatrisen.

- Merk

$$\boldsymbol{H}^2 = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \boldsymbol{H}$$

  dvs $\boldsymbol{H}$ er en Projeksjonsmatrise



Residualer og estimerte (predikerte) verdier

De estimerte verdiene $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$ kan skrives ved følgende matriseuttrykk

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \underbrace{\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T}_{\boldsymbol{H}}\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Y},$$

- Vektoren av residualer $e_i = y_i - \hat{y}_i$ er da gitt ved

$$\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{H}\boldsymbol{Y} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y},$$

  der $\boldsymbol{I}$ er $n \times n$-identitetsmatrisen.
- Vi har $\boldsymbol{X}^T\boldsymbol{e} = \boldsymbol{0}$, der $\boldsymbol{0}$ er en $k \times 1$ vektor av nuller.
- Det betyr at $\boldsymbol{e}$ er ortogonal til alle kolonnene til $\boldsymbol{X}$.
- Skalarproduktet mellom $\boldsymbol{e}$ og første kolonne i $\boldsymbol{X}$ er summen av residualene i $\boldsymbol{e}$, og denne er altså lik 0.

Kvadratsummer

- Residualkvadratsummen er gitt ved

$$SSE = \boldsymbol{e}^T \boldsymbol{e} = (\boldsymbol{Y} - \hat{\boldsymbol{Y}})^T (\boldsymbol{Y} - \hat{\boldsymbol{Y}}) = \|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\|^2,$$

  der $\|\mathbf{u}\|$ er lengden av vektoren $\mathbf{u}$.
- Regresjonskvadratsummen er

$$SSR = (\hat{\boldsymbol{Y}} - \bar{\boldsymbol{Y}})^T (\hat{\boldsymbol{Y}} - \bar{\boldsymbol{Y}}) = \|\hat{\boldsymbol{Y}} - \bar{\boldsymbol{Y}}\|^2$$

- Den totale kvadratsummen er gitt ved

$$SST = (\boldsymbol{Y} - \bar{\boldsymbol{Y}})^T (\boldsymbol{Y} - \bar{\boldsymbol{Y}}) = \|\boldsymbol{Y} - \bar{\boldsymbol{Y}}\|^2 = SSE + SSR.$$

  som altså er Pytagoras setning!

Forventningsvektor og kovariansmatrise

- La $\mathbf{U}$ være en n-dimensjonal stokastisk vektor, dvs. $\mathbf{U} = (U_1, \ldots, U_n)^T$ med $\mathrm{E}(U_i) = \mu_i$ og $\mathrm{Cov}(U_i, U_j) = \sigma_{ij}$, $i, j = 1, \ldots, n$.
- Da er forventningsvektoren til $\mathbf{U}$

$$\mathrm{E}(\mathbf{U}) = \begin{pmatrix} \mathrm{E}(U_1) \\ \vdots \\ \mathrm{E}(U_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \boldsymbol{\mu}$$

og kovariansmatrisen er gitt ved

$$\begin{aligned} \mathrm{Cov}(\mathbf{U}) &= \begin{pmatrix} \mathrm{E}((U_1 - \mu_1)^2) & \ldots & \mathrm{E}((U_1 - \mu_1)(U_n - \mu_n)) \\ \vdots & & \vdots \\ \mathrm{E}((U_n - \mu_n)(U_1 - \mu_1)) & \ldots & \mathrm{E}((U_n - \mu_n)^2) \end{pmatrix} \\ &\mathrm{E}\left(\begin{pmatrix} U_1 - \mu_1 \\ \vdots \\ U_n - \mu_n \end{pmatrix} (U_1 - \mu_1 \ldots U_n - \mu_n) \right) \\ &= \mathrm{E}((\mathbf{U} - \boldsymbol{\mu})(\mathbf{U} - \boldsymbol{\mu})^T) \\ &= \boldsymbol{\Sigma}. \end{aligned}$$

Egenskaper ved minste kvadraters estimator

- La nå $\mathbf{V} = \mathbf{c} + \mathbf{AU}$, der $\mathbf{c}$ er en m-dimensjonal vektor med konstanter og $\mathbf{A}$ er en $m \times n$-matrise med konstanter.
- Da er

$$\begin{aligned} \mathrm{E}(\mathbf{V}) &= \mathbf{c} + \mathbf{A}\mathrm{E}(\mathbf{U}) = \mathbf{c} + \mathbf{A}\boldsymbol{\mu} \\ \mathrm{Cov}(\mathbf{V}) &= \mathbf{A}\mathrm{Cov}(\mathbf{U})\mathbf{A}^T = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T. \end{aligned}$$

- Dette kan vi bruke til å finne egenskapene til minste kvadraters estimator.
- Vi har:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{og} \quad \mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}.$$

- Da er

$$\begin{aligned} \mathrm{E}(\boldsymbol{Y}) &= \mathrm{E}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{X}\boldsymbol{\beta} + \mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{X}\boldsymbol{\beta} \\ \mathrm{Cov}(\boldsymbol{Y}) &= \mathrm{Cov}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}. \end{aligned}$$

Egenskaper ved minste kvadraters estimator (forts.)

- Videre er $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, slik at

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{\beta} \\
\mathrm{Cov}(\hat{\boldsymbol{\beta}}) &= \mathrm{Cov}((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}) \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathrm{Cov}(\boldsymbol{Y})((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)^T \\
&= \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \sigma^2\mathbf{C},
\end{aligned}$$

  der $\mathbf{C} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

- Da $\hat{\beta}_j$-ene er lineærkombinasjoner av Y_i-ene, som er normalfordelt, må $\hat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$, $j = 0, 1, \ldots, k$, der c_{jj}-ene er diagonalelementene i $\mathbf{C}$.

- Det betyr at $\frac{\hat{\beta}_j - \beta_j}{S\sqrt{c_{jj}}} \sim t_{n-(k+1)}$, som gir oss konfidensintervallene og hypotesetestene presentert i Avsnitt 12.7.

Hattmatrisen, residualer og innflytelse

- Hattmatrisen har egenskapene $\boldsymbol{H}^T = \boldsymbol{H}$ og $\boldsymbol{HH} = \boldsymbol{H}$.
- De predikerte verdiene er gitt ved $\hat{\boldsymbol{Y}} = \boldsymbol{HY}$, slik at

$$\text{Cov}(\hat{\boldsymbol{Y}}) = \text{Cov}(\boldsymbol{HY}) = \sigma^2 \boldsymbol{H}.$$

- Videre er også $(I - H)^T = (I - H)$ og $(I - H)(I - H) = (I - H)$
- og siden residualene gitt ved $\boldsymbol{e} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$, blir

$$\text{Cov}(\boldsymbol{e}) = \text{Cov}((\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}) = \sigma^2(\boldsymbol{I} - \boldsymbol{H}),$$

som vi kan bruke til å finne de standardiserte residualene

$$e_i^* = \frac{e_i}{s\sqrt{1 - h_{ii}}}$$

Hattmatrisen, residualer og innflytelse (forts.)

- Hattmatrisen kan også brukes til å kvantifisere hvor stor innflytelse hver av observasjonene har på den tilpassede regresjonsmodellen.
- Siden $\hat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, er

$$\hat{y}_i = h_{i1}y_1 + h_{i2}y_2 + \ldots + h_{in}y_n,$$

  der h_{ij} er element (i, j) i $\boldsymbol{H}$.
- Det betyr at $\frac{\partial \hat{y}_i}{\partial y_i} = h_{ii}$, altså gir diagonalelementene i $\boldsymbol{H}$ informasjon om hvor stor betydning hver observasjon y_i har på sin egen prediksjon $\hat{y}_i$
- Dette kalles "leverage" på engelsk.

Eksempel

Fra Oppg. 12.137 i boka.

Hattmatrisen, residualer og innflytelse (forts.)

- Standardiserte residualer:
  - Stor residual: $|e_i^*| > 2$
  - Veldig stor residual: $|e_i^*| > 3$
  - Stort avvik mellom y_i og $\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$
- Kan vise:

$$0 \leq h_{ii} \leq 1 \text{ og } \sum_{i=1}^{n} h_{ii} = k + 1$$

- Gjennomsnittlig: $(k+1)/n$
  - Høy innflytelse hvis $h_{ii} > 2(k+1)/n$
  - Veldig høy innflytelse hvis $h_{ii} > 3(k+1)/n$
  - $\boldsymbol{x}_i$ ekstrem (outlier)

Oppsummering

Vi oppsummerer multippel lineær regresjon med utvalgte eksamensoppgaver på tavla.

Eksempel

Oppgave 3, eksamen 2025.

Eksempel

Oppgave 3 c) og d), eksamen 2019.