BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

TEMA 2: Enkel regresjon

1. Overblikk
2. De klassiske forutsetningene
3. Regresjonslinja
4. Ordinær minste kvadraters metode
5. Usikkerhet og modellpresisjon
6. Hvorfor OLS?

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

(1) Overblikk

Den multiple lineære regresjonsmodellen

→ $Y = B_1 + B_2 X_1 + \dots + B_K X_K + u$

Den enkle lineære regresjonsmodellen

→ $Y = \underbrace{B_1 + B_2 X}_{\text{"forklaring"}} + u$

variabelen vi ønsker å forklare

uforklart variasjonen i Y

→ Ikke-linearitet i variabel:

$$Y = B_1 + B_2 X^2 + u$$

$$Y = B_1 + B_2 \left(\frac{1}{X}\right) + u$$

$$Y = B_1 + B_2 (\ln X) + u$$

BI HANDELSHØYSKOLEN

Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo

Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500

www.bi.no
info@bi.no
Org. nr. 971 22 8865

→ Ikke-linearitet i parameter B_2:

$$Y = B_1 + B_2^2 X + u$$

$$Y = B_1 + e^{B_2} X + u$$

$$= B_1 + 2{,}714^{B_2} X + u$$

2) De klassiske regresjonsforutsetningene:

KF1: Linearitet i både parametre og variabler:

$$Y_i = B_1 + B_2 X_i + u_i$$

KF2: $Cov(X_i, u_i) = 0$. Forklaringsvariabel X_i og feilleddet skal være ukorrelerte. Tolkning: X_i øker forklaringskraften istedet for å redusere den

KF3: $E(u_i | X_i) = 0$ $\quad (\Rightarrow E(u_i) = 0)$

Tolkning: I gjennomsnitt skal forklaringsfeilen være lik null

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

kap.9

KF4: $Var(u_i) = \sigma^2$ for alle i. Tolking: Presisjonen til forklaringen vår er konstant (avhenger ikke av enhetene; i'ene)

kap.10

KF5: $Cov(u_i, u_j) = 0$ (feilleddene er ukorrelerte).

(3) Regresjonslinja

$$\rightarrow \quad Y_i = \underbrace{B_1 + B_2 X_i}_{\text{forklaringen}} + u_i \quad \text{(populasjon)}$$

$$\rightarrow \quad Y_i = \overbrace{b_1 + b_2 X_i} + e_i \quad \text{(utvalg)}$$

$E(Y_i | X_i) = B_1 + B_2 X_i$: Populasjons-regresjonsfunksjonen (Population Regression Function, PRF)

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

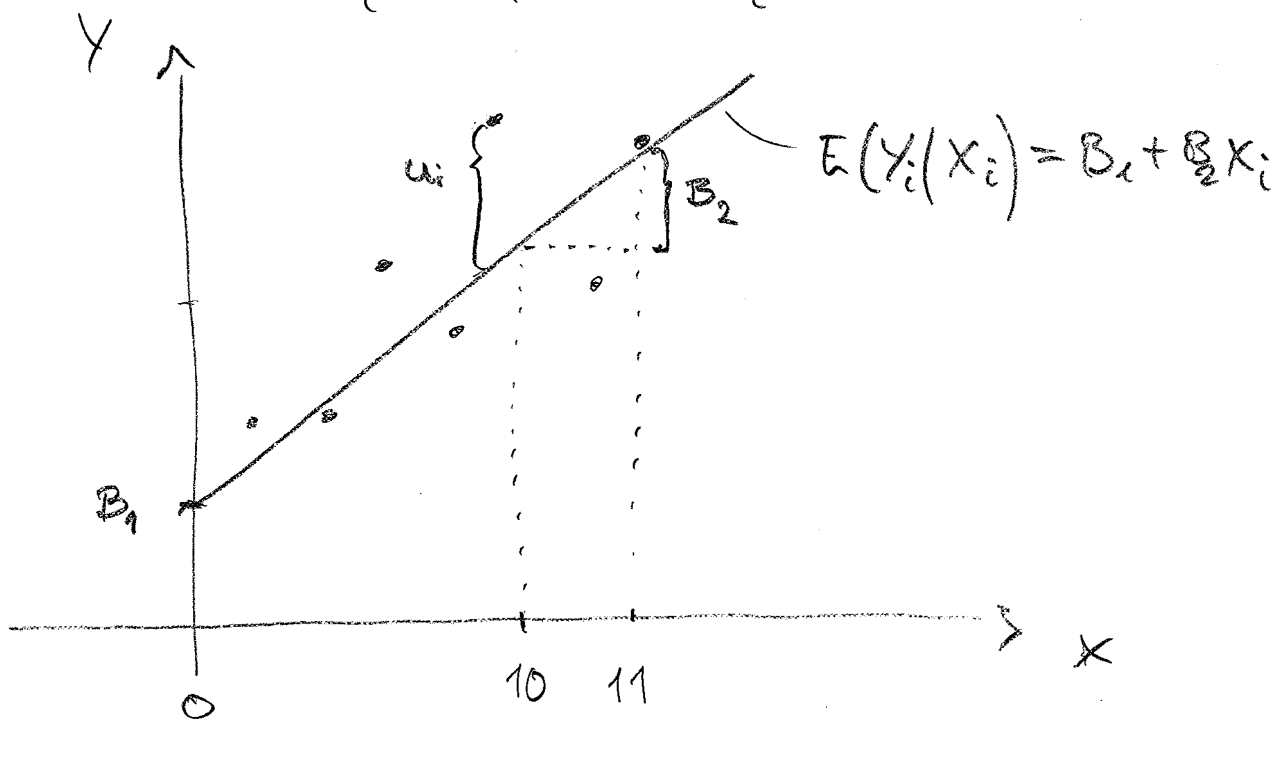
$\hat{Y}_i = b_1 + b_2 X_i$: Utvalgsregresjonsfunksjonen (Sample Regression Function, SRF)

→ b_1, b_2 er estimater av B_1, B_2

→ e_i er estimert forklaringsfeil/prediksjonsfeil; et estimat av u_i

→ Tolkning av B_1 og B_2 (b_1 og b_2 fortolkes på tilsvarende måte)

$$Y_i = B_1 + B_2 X_i + u_i$$



BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

B_1: Gjennomsnittlig verdi til Y_i når $X_i = 0$

MERK: Verdien til B_1 kan være negativ (eller økonomisk meningsløs).

Alternativ fortolkning: B_1 "plasserer" regresjonslinja slik at den går gjennom hovedtyngden til observasjonene

B_2: Den gjennomsnittlige endringen i Y_i gitt én enhets endring i X_i

Eksempel: $B_1 + B_2 \cdot 11 - (B_1 + B_2 \cdot 10)$

$$= B_2$$

Eksempel (tabell 2,5, s. 38):

Y_i = timelønn i USD for person i

X_i = ant. år med skolegang for person i

$b_1 = -0{,}0144 \qquad b_2 = 0{,}7241$

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

b_1 : plasserer regresjonslinja slik at den går gjennom hovedtyngden til observasjonene

b_2 : et år mer med skolegang gir i gjennomsnitt 0,72 USD mer i timelønn

(4) Minste kvadratersmetode

→ lineær / ordinær minste kvadratersmetode

→ Ordinary Least Squares (OLS)

Hvilke verdier skal vi tillegge b_1 og b_2?

↳ Vanligste framgangsmåte: OLS

Gitt $Y_i = b_1 + b_2 X_i + e_i$

$$\Updownarrow$$

$$e_i = Y_i - b_1 - b_2 X_i$$

$$\Downarrow$$

$$e_i^2 = (Y_i - b_1 - b_2 X_i)^2$$

$$\sum e_i^2 = \sum (Y_i - b_1 - b_2 X_i)^2$$

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

* OLS estimering består i å velge b_1 og b_2 slik at $\sum e_i^2$ minimeres

Førsteordensbetingelsene

$$\frac{\partial(\sum e_i^2)}{\partial b_1} = 0$$

$$\frac{\partial(\sum e_i^2)}{\partial b_2} = 0$$

Appendix 2A:

$$\frac{\partial(\sum e_i^2)}{\partial b_1} = (-2)\cdot \sum (Y_i - b_1 - b_2 X_i)$$

$$\frac{\partial(\sum e_i^2)}{\partial b_2} = (-2)\cdot \sum (Y_i - b_1 - b_2 X_i) X_i$$

$$b_2 = \frac{S_{XY}}{S_X^2}$$

utvalgskovariansen mellom X og Y

utvalgsvariansen til X

BI HANDELSHØYSKOLEN

Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo

Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500

www.bi.no
info@bi.no
Org. nr. 971 22 8865

$$b_1 = \bar{Y} - b_2 \bar{X}$$

gjennomsnitt

→ Nyttig: OLS estimatet b_2 kan skrives på mange måter

$$b_2 = \frac{S_{XY}}{S_X^2} = \frac{\dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\dfrac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$= \frac{\sum (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y})}{\sum (x_i^2 - x_i \bar{x} - \bar{x} x_i + \bar{x}^2)}$$

$$= \frac{\sum (x_i y_i) - n \bar{x}\bar{y}}{\sum (x_i^2) - n \bar{x}^2}$$

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

→ Andre egenskaper ved OLS estimering:

* $\bar{e} = \frac{1}{n} \sum e_i = 0$
* $\sum (e_i x_i) = 0$
* $\sum (e_i \hat{Y}_i) = 0$ $\qquad \hat{Y}_i = b_1 + b_2 X_i$,

$$\Rightarrow \sum \left[e_i (b_1 + b_2 X_i) \right] = 0$$

Eksempel (tabell 2.4, s. 37):

Y_i = S.A.T.
X_i = familieinntekt

$Y_i = \{410, 420, 440, 490, 530, 530, 550, 540, 570, 590\}$

$X_i = \{5000, 15000, 25000, 35000, 45000, 55000, 65000, 75000, 90000, 15000\}$

$n = 10$ $\qquad \bar{X} = \frac{1}{10} \cdot 560\,000$ $\qquad (\sum X_i = 560\,000)$

$= 56\,000$

$\bar{Y} = \frac{1}{10} \cdot 5070$ $\qquad (\sum Y_i = 5070)$

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

$$\Sigma(x_i y_i) = 305\ 550\ 000$$

$$\Sigma(x_i^2) = 47\ 600\ 000\ 000$$

$$\bar{x}^2 = (56\ 000)^2$$

$$b_2 = \frac{\Sigma(y_i x_i) - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2} = \frac{21\ 630\ 000}{16\ 240\ 000\ 000}$$

$$\approx 0{,}0013$$

$$b_1 = \bar{y} - b_2 \cdot \bar{x} = 507 - 0{,}0013 \cdot 56\ 000$$

$$= 434{,}2$$

* Oppgavesett 2:

1. $$S_y = \frac{1}{n-1} \cdot \underbrace{\Sigma(y_i - \bar{y}^2)}_{TSS}$$

$$\Downarrow$$

$$39{,}22 = \frac{1}{37-1} \cdot TSS$$

$$\Downarrow$$

$$36 \cdot 39{,}22 \approx 1412.$$

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

2. B

3. $b_2 = \frac{S_{xy}}{S_x^2}$ + . C

4. $Y_i = \alpha + \beta X_i + u_i$

$\alpha \uparrow B_1$, $\beta \uparrow B_2$

$\hat{\alpha} = b_1, \quad \hat{\beta} = b_2$

C.

5. $b_2 = \hat{\beta} = \frac{S_{xy}}{S_x^2}$ ÷

C.

6. $\hat{\beta} = \frac{\Sigma(x_i y_i)}{\Sigma x_i^2} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$

$= b_2$

C.

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

7. Parametrene: α og β

$$Y_t = \alpha + \frac{\beta}{X} + u_t$$

$$= \underline{\alpha + \beta \cdot X_t^* + u_t}$$

$$X^* = \frac{1}{X_t}$$

(5) Usikkerhet og modellpresisjon

$$Y_i = \underbrace{B_1 + B_2 X_i}_{\text{forklaring}} + u$$

forklaringsfeil

$$Y_i = \overbrace{b_1 + b_2 X_i} + e_i$$

→ $E(u^2) = \sigma^2$: Et mål på presisjonen til forklaringen vår.

→ $\frac{1}{n-2} \sum e_i^2$: Vårt empiriske estimat på σ^2

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

→ Notasjon:

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum e_i^2$$

ant. estimerte parametre

$$\hat{\sigma} = \sqrt{\frac{1}{n-k} \sum e_i^2}$$

: regresjonens standardfeil (Standard Error of Regression; SER)

EViews: S.E. of regression

→ $\sum e_i^2$: Kvadratsummen til residualene (Sum of Squared Errors; SSE)

EViews: Sum squared resid

→ b_1 og b_2 er estimater av B_1 og B_2.

Mål på presisjonen til b_1 og b_2:

$\left.\begin{array}{l} se(b_1) \\ se(b_2) \end{array}\right\}$ standardfeilene til b_1 og b_2

BI HANDELSHØYSKOLEN
Nydalsveien 37, 0484 Oslo
Postadresse 0442 Oslo
Telefon 06600
Telefaks 21 04 80 00
Studieinfo. tlf. 810 00 500
www.bi.no
info@bi.no
Org. nr. 971 22 8865

$$se(b_1) = \sqrt{var(b_1)}$$

$$= \sqrt{\left[\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}\right] \cdot \hat{\sigma}^2}$$

SER

$$se(b_2) = \sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}}$$

EViews: Standardfeilene til estimatene står i kolonnen kalt "Std. Error"

⑥ Hvorfor OLS?

Gitt KF1–KF5 så er OLS "BLUE":

→ Best
→ Linear

"Best" (minst varians) i klassen av lineære estimatorer

→ Unbiased
→ Estimator

b_1 og b_2 forventningsrette estimatorer av β_1 og β_2