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 9: Multikolinearitet

1. Eksakt multikolinearitet
2. Imperfekt multikolinearitet
3. Hvordan håndterer vi (sterk) imperfekt multikolinearitet
4. Oppgavesett 9

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) Eksakt multikolinearitet

"Kolinearitet" mellom X_1 og X_2 oppstår når

$$X_1 = a + b X_2$$

Dvs.: X_1 kan skrives som en <u>eksakt</u> lineær funksjon av X_2

<u>Når skjer dette?</u>

En vanlig situasjon: X_1 og X_2 er konstante

→ Eksempel: $X_1 = 1 \quad X_2 = 2$

Vi forsøker å estimere

$$pris = B_1 X_1 + B_2 X_2 + u$$

En annen vanlig situasjon: "Dummyfellen"

→ Eksempel: $X_1 = 1$ eller 2 ("dummy-trap")

$$X_2 = \begin{cases} 1 & \text{hvis } X_1 = 1 \\ 0 & \text{ellers} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{hvis } X_1 = 2 \\ 0 & \text{ellers} \end{cases}$$

$\Rightarrow X_1 = X_2 + 2 \cdot X_3$: X_1 er en eksakt lineær kombinasjon av X_2 og X_3

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

→ Eksempel: $X_1 = 1$ eller 2

$$X_2 = \begin{cases} 0,5 & \text{hvis } X_1 = 1 \\ 0 & \text{ellers} \end{cases}$$

$$X_3 = \begin{cases} 4 & \text{hvis } X_1 = 2 \\ 0 & \text{ellers} \end{cases}$$

$$\Rightarrow X_1 = 2 \cdot X_2 + \frac{1}{2} X_3$$

$$1 = \underbrace{0,5}_{1} \qquad 0$$

$$2 = 0 \qquad 2$$

② Imperfekt multikolinearitet

Ikorthet: X_1 og X_2 er korrelerte

Hvis X_1 og X_2 er korrelerte så vil R^2 til

$$X_1 = a_1 + a_2 X_2 + e$$

være større enn 0 (husk: $R^2 = (r)^2$

korrelasjon mellom X_1 og X_2

↳ Jo høyere R^2, jo mer/sterkere kolinearitet

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

X_1, X_2, X_3 er imperfekt multikolineære

hvis R^2 til

$$X_1 = a_1 + a_2 X_2 + a_3 X_3 + e$$

er ~~større enn~~ 0 og mindre enn 1

↳ Jo høyere R^2, jo mer/sterkere multikolinearitet

Når er imperfekt multikolinearitet et problem?

↳ Når R^2 er "høy"

Hva er problemet?

↳ Var(b): Blir veldig høye (dvs.: b er et veldig upresist estimat)

Variansinflasjonsfaktoren (VIF, "Variance Inflation Factor")

Betrakt:

$$Y = B_1 + B_2 X_2 + B_3 X_3 + u$$

Det kan vises at:

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

$$Var(b_2) = \frac{\sigma^2}{\sum (X_2 - \bar{X}_2)^2} \cdot \frac{1}{(1 - R_2^2)}$$

$$Var(b_3) = \frac{\sigma^2}{\sum (X_3 - \bar{X}_3)^2} \cdot \frac{1}{(1 - R_3^2)}$$

VIF

hvor R_2^2 er R^2 til

$$X_2 = c_1 + c_2 X_3 + v$$

og R_3^2 er R^2 til

$$X_3 = d_1 + d_2 X_2 + w$$

I ord:

- $\sigma^2 \nearrow \Rightarrow Var(b)$ går ned
- Variasjon i X går opp, så går $Var(b)$ ned
- Jo høyere R_j^2 (sterkere multikolinearitet), jo større $Var(b)$

Generelt:

$$Y = B_1 + B_2 X_2 + \dots + B_K X_K + u$$

så er

VIF

$$Var(b_j) = \frac{\sigma^2}{\sum (X_j - \bar{X}_j)^2} \cdot \frac{1}{(1 - R_j^2)}$$

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

hvor R_j^2 er R^2 til en regresjon med X_j som venstreside variabel og de andre X'ene som høyreside variabler (inkludert en konstant)

<u>Teoretiske</u> konsekvenser av sterk imperfekt multikolinearitet: Få eller ingen

<u>Praktiske</u> konsekvenser:

→ Upresise estimater

→ Estimater og resultater kan endre seg vesentlig ved små endringer i modellen (f.eks. inklusjon/utelating av bare én variabel, små transformasjoner, osv.)

→ Estimatene kan ha "uriktige" fortegn

Eksempel: Huspriser 2 i USA

(1) $pris = B_1 + B_2$ Vurdering $+ B_3$ BOA $+ B_4$ Tomt $+ B_5$ Badrom $+ u$

→ b_3 negativ!

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

Bedre modell? → (2) $\frac{\text{pris}}{\text{BOA}} = B_1 + B_2 \frac{\text{vurdering}}{\text{BOA}} + B_3 \text{Tomt} + B_5 \text{Baderom} + u$

"Moral"/lærdom: Med mindre det er essensielt å studere innvirkninger av BOA på pris, så er det en god idé å studere innvirkningen av $\frac{\text{vurdering}}{\text{BOA}}$ på $\frac{\text{pris}}{\text{BOA}}$ i stedet for (1)

(3) Hvordan håndterer vi (sterk) imperfekt multikolinearitet?

→ Kan vi transformere variablene?

→ Kan vi utelate variabler?
  ↳ Helst ikke! : Kan føre til utelatingsskjevhet

→ Skaffe nye/mer data?
  ↳ Ofte upraktiske

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

"Moderne" løsninger: Automatiske modellseleksjonsalgoritmer

- → Forover regresjon (FR)
- → Bakover eliminasjon (BE)
- → Kompliserte varianter av FR og BE: "General-to-specific"

BI NORWEGIAN SCHOOL OF MANAGEMENT
Nydalsveien 37, 0484 Oslo
Mailing address NO-0442 Oslo
Telephone [+47] 06600
Telefax [+47] 21 04 80 00
Call Center [+47] 810 00 500
www.bi.edu
info@bi.no
Org. no. 971 22 8865

4 Oppgavesett 9

→ Eksakt multikolinearitet mellom 2 eller flere forklaringsvariabler: Forklarings-variablene kan skrives som en eksakt lineær kombinasjon av hverandre

↳ Problem: Estimering umulig

→ Imperfekt multikolinearitet mellom forklaringsvariabler: At de er korrelerte

↳ Problem: Høy Var(b) (Høy se(b)) } "upresise estimater"

1. A) Feil
B) Feil
C) Riktig
D) Feil

2. Tolk "lineær avhengighet" som "eksakt

BI NORWEGIAN SCHOOL OF MANAGEMENT
Nydalsveien 37, 0484 Oslo
Mailing address NO-0442 Oslo
Telephone [+47] 06600
Telefax [+47] 21 04 80 00
Call Center [+47] 810 00 500
www.bi.edu
info@bi.no
Org. no. 971 22 8865

lineær avhengighet"

C) Riktig

D) Kan være riktig i visse tilfeller, men ikke alltid (se eksempel 3 i del 1 i notatene)

3.) A) → $t = \frac{b_i}{se(b_i)}$

$\rightarrow se(b_i) = \sqrt{Var(b_i)}$ VIF

$\rightarrow Var(b_i) = \frac{\sigma^2}{\sum(X_i - \bar{X})^2} \cdot \frac{1}{(1 - R_i^2)}$

hvor R_i^2 er R^2 til en regresjon hvor X_i er venstre side variabel og de andre forklaringsvariablene utgjør høyre side variablene

⇒ A) Feil

B) Riktig

C) Feil

D) Feil: Ekskludering av forklaringsvariabler skal helst unngås

BI NORWEGIAN SCHOOL OF MANAGEMENT
Nydalsveien 37, 0484 Oslo
Mailing address NO-0442 Oslo
Telephone [+47] 06600
Telefax [+47] 21 04 80 00
Call Center [+47] 810 00 500
www.bi.edu
info@bi.no
Org. no. 971 22 8865

4. $t = \dfrac{b}{se(b)}$: Høy $Var(b)$ $\Rightarrow$ lav t-verdi
$\left(se(b) = \sqrt{Var(b)} \right)$

$\Rightarrow$ Generelt: Høy P-verdi

A) Riktig

B) } GALT
C)
D)

6. Husk: $VIF = \dfrac{1}{1-R_2^2}$ hvor R_2^2 er R^2 fra

$$\ln(\text{Vurdering}) = \gamma_1 + \gamma_2 \ln(\text{Tomt}) + \gamma_3 \ln(FF) + \gamma_4 \text{Baderom} + v$$

$$VIF = \frac{1}{1-R_2^2} = 6{,}247$$

$$\Updownarrow$$

$$1 = 6{,}247 \cdot (1 - R_2^2)$$

$$\Updownarrow$$

$$R_2^2 = \frac{5{,}247}{6{,}247} = 0{,}84 = 84\%$$

c)

BI NORWEGIAN SCHOOL OF MANAGEMENT
Nydalsveien 37, 0484 Oslo
Mailing address NO-0442 Oslo
Telephone [+47] 06600
Telefax [+47] 21 04 80 00
Call Center [+47] 810 00 500
www.bi.edu
info@bi.no
Org. no. 971 22 8865

7. $Var(\hat{\beta}_2)$: $c = \dfrac{1}{1-R_2^2}$ hvor R_2^2 er R^2

til $x_2 = c_1 + c_2 x_3 + v$

$Var(\hat{\beta}_3)$: $c = \dfrac{1}{1-R_3^2}$ hvor R_3^2 er R^2

til $x_3 = d_1 + d_2 x_2 + w$

$\Rightarrow R_2^2 = R_3^2$

A) $r_{x_2 x_3} < 0 \Rightarrow R^2 > 0 \Rightarrow \underline{\underline{c > 1}}$

RIKTIG

B) $r_{x_2 x_3} > 0 \Rightarrow R^2 > 0 \Rightarrow \underline{\underline{c > 1}}$

RIKTIG

C) $r_{x_2 x_3} = 0 \Rightarrow R^2 = 0 \Rightarrow \underline{\underline{c = 1}}$

GALT

D) RIKTIG

BI NORWEGIAN SCHOOL OF MANAGEMENT
Nydalsveien 37, 0484 Oslo
Mailing address NO-0442 Oslo
Telephone [+47] 06600
Telefax [+47] 21 04 80 00
Call Center [+47] 810 00 500
www.bi.edu
info@bi.no
Org. no. 971 22 8865

8 A) $r_{x_2x_3} = 0 \Rightarrow R^2 = 0 \Rightarrow \underline{\underline{c = 1}}$

GALT

B) $\lambda_2 \neq 0 \Rightarrow r_{x_2x_3} \neq 0$ (Husk: ...)

$\Rightarrow R_2^2 > 0 \Rightarrow \underline{\underline{c > 1}}$

RIKTIG

C) GALT

D) $R^2 = 0 \Rightarrow \underline{\underline{c = 1}}$. GALT