

Mappeinnlevering 2

Fakultet for biovitenskap, fiskeri og økonomi.

Kandidatnummer 6, SOK-2009, Høst 2023

08-11-2023

Table of contents

Oppgave 1:	3
a) Kjøre en enkel lineær regresjonsanalyse	3
b) Forklaring av resultater	3
c) Bryter modellen med antakelsene	5

Oppgave 2:	5
a) Kjør en multippel lineær regresjonsanalyse med minst to uavhengige variabler. Velg selv om du tilføyer en eller flere variabler til din tidligere analyse, eller om du lager en helt ny. Forklar hvorfor du har valgt denne kombinasjonen av variabler.	5
b) Vis og forklar resultatene dine. Bruk grafer, tabeller, og output til å forklare din modell og hva modellen kan fortelle oss.	5
c) Test hvorvidt modellen din bryter med antakelsene til multippel lineær regresjon. Vis og forklar hvordan du testet/undersøkte	5

Appendiks	5
------------------	----------

Figurliste

1	Lineær regresjon mellom kvinner og kvinners mødre sin utdanning	4
---	---	---

Oppgave 1:

a) Kjøre en enkel lineær regresjonsanalyse

Datasettet vi har fått inneholder mye data på inntekt, arbeidserfaring og utdanningsnivå, spesielt for kvinner. Jeg har lyst å se om det kan finnes en korrelasjon mellom utdanning og slekt, og vil da se om det er en større korrelasjon om at barn som har utdannede foreldre også tar utdanning selv.

Derfor tar jeg tar den avhengige variabelen “educ” som direkte oversatt er “Kvinnens utdanningsnivå i år” for 1975 og den uavhengige variabelen “motheduc” som også direkte oversatt er “Kvinnens mor sitt utdanningsnivå i år” og sjekker om det er en korrelasjon.

b) Forklaring av resultater

Når jeg kjører en enkel lineær regresjonsanalyse så får jeg ut [Table 1](#).

	Avhengig variabel	
	Kvinner utdanning	
Intercept	9.55981***	(0.21898)
Mødres utdanning	0.29478***	(0.02224)
Observations	753	
R ²	0.1895	
Adjusted R ²	0.1884	
Residual Std. Error	2.054 (df = 751)	
F Statistic	175.6*** (df = 1; 751)	

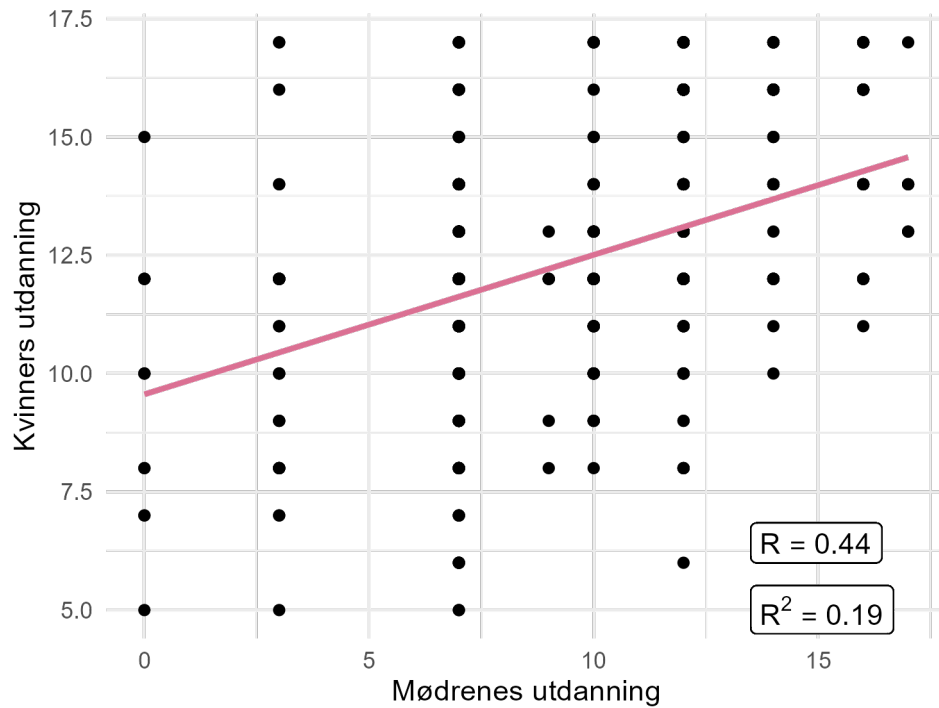
Tabell 1: Lineær regresjonsanalyseresultat for kvinners utdanning og deres mødre.

R^2 forklarer at 18.95% av kvinners utdanning kan forklares av den uavhengige variabelen som er kvinnens mor sin utdanning. Man kan også se at for hver “x” eller utdanning for kvinner så går “y” eller morens utdanning 0.294 mot høyere utdanning. Intercepten er på 9.55 som sier hva verdien på kvinners utdanning er når motheduc er 0.

Std. Error eller standardfeilen forklarer hvor “sikker” man er på det man har estimert, om det er et høyt tall er man mer usikker på om estimatet er riktig utifra det utvalget man har tatt ifra en populasjon. En lavere standardfeil vil gi mindre usikkerhet.

Degrees of freedom er antallet uavhengige variabler som er “fri” til å variere ved tilfeldig trekning.

Tre stjerner eller “***” forteller meg tilslutt at resultatet er signifikant, som vil si at det er lite sannsynlighet for at resultatet har oppstått tilfeldig



Figur 1: Lineær regresjon mellom kvinner og kvinners mødre sin utdanning

I [Figur 1](#) kan man se forholdet mellom den avhengige variabelen på y-aksen som er Kvinner utdanning mot den uavhengige variabelen mødrenes utdanning som er på x-aksen, og at korrelasjonen er positiv. Det vi tar ut av koeffisientene er at det er en “svak” mot “moderat” korrelasjon for verdien til korrelasjonskoeffisienten R som er på 0.44. Og som sagt tidligere sier R^2 at omtrent 19% av kvinners utdanning kan forklares av moren sin utdanning.

c) Bryter modellen med antakelsene

Oppgave 2:

- a) Kjør en multippel lineær regresjonsanalyse med minst to uavhengige variabler. Velg selv om du tilføyer en eller flere variabler til din tidligere analyse, eller om du lager en helt ny. Forklar hvorfor du har valgt denne kombinasjonen av variabler.
- b) Vis og forklar resultatene dine. Bruk grafer, tabeller, og output til å forklare din modell og hva modellen kan fortelle oss.
- c) Test hvorvidt modellen din bryter med antakelsene til multippel lineær regresjon. Vis og forklar hvordan du testet/undersøkte

Appendiks

Bruk av KI: ChatGPT 4 inkludert advanced data analysis.

[Noe hjelp med hypotesen](#)