

Oppgave 3 (35%)

Vi har et datasett med 3796 personer i USA. For disse personene har vi mål på variablene $yrsed_i$ (antall års utdanning inkludert grunnskole), $distmil_i$ (avstand i mil til nærmeste høyskole), samt $parentscoll_i$ som er en dummyvariabel som =1 dersom minst en av foreldrene har høyere utdanning.

Nedenfor følger et datasammendrag for alle 3796 observasjoner av disse variablene.

Summary Statistics, using the observations 1 - 3796					
Variable	Mean	Median	S.D.	Min	Max
parentscoll	0,251	0,000	0,434	0,000	1,00
yrsed	13,8	13,0	1,81	12,0	18,0
distmil	2,775	1,609	3,433	0,000	25,74

- a) Du skal nå teste hvorvidt personer i gjennomsnitt har mer enn 13 års utdanning. Bruk et 1% signifikansnivå.
- i) Hvilken av disse er da en passende hypotesetest?
 - 1) $H_0: \mu = 13$ $H_A: \mu < 13$
 - 2) $H_0: \mu = 13$ $H_A: \mu > 13$
 - 3) $H_0: \mu = 13$ $H_A: \mu \neq 13$
 - ii) Hva er testverdi?
 - iii) Hva er kritisk(e) verdi(er)?
 - iv) Hva er konklusjonen på testen?

For å estimere hvorvidt avstand til nærmeste høyskole og foreldrenes utdanning har en effekt på hvor lang utdanning personer har, så estimerer vi modellen under

(«Modell 1»: $yrsed_i = B_1 + B_2 distmil_i + B_3 parentscoll_i + u_i$ der u_i er et feilledd)

Model 1: OLS, using observations 1-3796
Dependent variable: yrsed

	Coefficient	Std. Error	t-ratio	p-value	
const	13,5856	0,0406787	334,0	<0,0001	***
distmil	-0,0268006	0,00820521	-3,266	0,0011	***
parentscoll	1,26574	0,0649366	19,49	<0,0001	***
Mean dependent var	13,82929	S.D. dependent var		1,813969	
Sum squared resid	11265,88	S.E. of regression		1,723420	
R-squared	0,097819	Adjusted R-squared		0,097343	
F(2, 3793)	205,6270	P-value(F)		1,64e-85	
Log-likelihood	-7450,994	Akaike criterion		14907,99	
Schwarz criterion	14926,71	Hannan-Quinn		14914,64	

- b) Gi en tolkning av de beregnede koeffisientene i Modell 1.

I tillegg beregner vi modellen «Modell 2» $yr_{sed}_i = B_1 + B_2 distmil_i + B_3 parentscoll_i + B_4 parentscoll_i \cdot distmil_i + u_i$ der u_i er et feilledd. I utskriften under er $pcolldistmil_i = parentscoll_i \cdot distmil_i$

Model 2: OLS, using observations 1-3796
Dependent variable: yrsed

	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-ratio</i>	<i>p-value</i>	
const	13,5900	0,0419545	323,9	<0,0001	***
distmil	-0,0282737	0,00888499	-3,182	0,0015	***
parentscoll	1,24353	0,0827905	15,02	<0,0001	***
pcolldistmil	0,0100223	0,0231754	0,4325	0,6654	
Mean dependent var	13,82929	S.D. dependent var		1,813969	
Sum squared resid	11265,33	S.E. of regression		1,723605	
R-squared	0,097863	Adjusted R-squared		0,097149	
F(3, 3792)	137,1176	P-value(F)		2,42e-84	
Log-likelihood	-7450,901	Akaike criterion		14909,80	
Schwarz criterion	14934,77	Hannan-Quinn		14918,68	

- Gi en tolkning av den beregnede koeffisienten b_4 i Modell 2.
- Beregn antall år utdanning ($\widehat{yr_{sed}}$) for person nr 116 i datasettet ved hjelp av Modell 2. Foreldrene til denne personen har ikke høyere utdanning, og avstanden til nærmeste høyskole er 8,045 mil.
- Hva blir anslagsfeilen for observasjon nr 116 dersom faktisk verdi er $yr_{sed}_{116} = 12$ år?