UNIVERSITETET I OSLO

DET MEDISINSKE FAKULTET

Eksamen
ERN4000 STATISTIKK
Fredag 5. september 2026 kl. 09:00

Hjelpemidler

- Stata (på den stasjonære maskinen)
- Egen kalkulator med exp og ln, og uten trådløs kommunikasjon
- Stata Cheat Sheet
- Lærerbok: Aalen OO (red), Frigessi A, Moger TA, Scheel I, Skovlund E, Veierød MB. Statistiske metoder i medisin og helsefag, 2. utgave. Gyldendal Akademisk 2018.
  eller
  Aalen OO (red), Frigessi A, Moger TA, Scheel I, Skovlund E, Veierød MB. Statistiske metoder i medisin og helsefag. Gyldendal Akademisk 2006.

Kort forklaring til oppgavene

I oppgavene skal du foreta statistisk analyse av noen problemstillinger. I presentasjonen av resultatene er det tre statistiske mål som er viktige å få frem: estimat (f.eks. gjennomsnitt, andel, odds ratio, regresjonskoeffisient), konfidensintervall og p-verdi.

Oppgave 1

I datafilen data2026.dta finner du resultater fra en randomisert studie med 304 personer der man sammenlignet to intervensjoner, kostomlegging og et program for fysisk aktivitet, og undersøkte vektendring i løpet av 18 måneder.

Variabelnavn	**Beskrivelse**
intervensjon	Intervensjonsgruppe: 0 = Kostomlegging, 1 = Fysisk aktivitet
alder	Alder (år) ved inklusjon
kjønn	Kjønn: 0 = Kvinne, 1 = Mann
status	Samlivsform: 0 = Enslig, 1 = Samboende
høyde	Høyde (cm)
vekt0	Vekt (kg) ved inklusjon
vekt1	Vekt (kg) etter 18 måneder
vektendring	vekt0 minus vekt1
vektendring2	kategorisk variabel: 1 = vektendring ≥5 kg, 2 = vektendring <5 kg

a) Undersøk om kjønn, samlivsform, alder, høyde og vekt er signifikant forskjellig i de to intervensjonsgruppene ved inklusjon. Oppgi resultatene i en tabell med samme oppsett som nedenfor og angi i fotnote til p-verdiene hvilke tester som er brukt for hver variabel. Skriv en kort tekst som begrunner valget av tester. Oppgi hvilke Stata kommandoer som ble brukt for å fylle inn i tabellen og for å velge tester.

Variabel	Kost (n=150) n (%)	Fysisk aktivitet (n=154) n (%)	p-verdi
Kjønn			
Kvinne			
Mann			
Samlivsform			
Enslig			
Samboende			
	Gjennomsnitt (SD)	Gjennomsnitt (SD)	
Alder			
Høyde			
Vekt			

SD = standard avvik

b) Estimer forskjellen i vektendring mellom de to gruppene med 95% konfidensintervall. Er det signifikant forskjellig vektendring i de to gruppene? Begrunn valget av test.

Oppgave 2

I denne oppgaven skal du bruke datafilen kostgruppen.dta. Denne filen inkluderer bare de 150 personene som ble randomisert til gruppen som fikk kostomlegging. Variablene er de samme som i oppgave 1.

a) Undersøk ved hjelp av lineær regresjon om det er signifikant sammenheng mellom vektendring og kjønn. Angi også den estimerte regresjonskoeffisienten med 95% konfidensintervall. Hvordan fortolkes regresjonskoeffisienten for kjønn?

b) Undersøk ved hjelp av lineær regresjon om det er signifikant sammenheng mellom vektendring og alder. Angi også den estimerte regresjonskoeffisienten med 95% konfidensintervall.

c) Forskerne utførte en multippel regresjonsanalyse med vektendring som avhengig variabel, og alder og kjønn som uavhengige variabler. Utfør en slik multippel regresjonsanalyse og oppgi R^2. Hva forteller R^2 deg?

d) Lag en tabell som angitt nedenfor som viser estimerte regresjonskoeffisienter, konfidensintervaller og p-verdier for alder og kjønn. Både de ujusterte og de justerte resultatene.

	Ujustert			Justert		
Variabel	Regresjons-koeffisient	95% KI	p-verdi	Regresjons-koeffisient	95% KI	p-verdi
Kjønn*						
Alder						

*Menn i forhold til kvinner

e) Var forutsetningene tilstede for å gjøre analysen i d? Begrunn svaret.

Oppgave 3

Vi skal nå bruke datafilen data2026.dta som vi brukte i oppgave 1 (alle 304 personene).

a) Lag en 2x2 tabell med variabelen intervensjon i kolonner og variabelen vektendring2 i rader for kvinner og en tilsvarende tabell for menn. Prosentuer tabellene på en hensiktsmessig måte utfra studiens design. Hva ser du fra disse tabellene?

   I Stata kan du bruke
   `bysort var1: tabulate var2 var3, option`
   der du setter inn oppgavens variabelnavn for `var1`, `var2` og `var3`, og bruker `row` eller `column` som `option`.

b) For kvinner skal du beregne relativ risiko (RR) for en vektendring på ≥5 kg for kostbehandling sammenlignet med fysisk aktivitet. Beregn også 95% konfidensintervall. For menn er RR=1.79 og 95% KI er (1.14, 2.81). Kommenter resultatene for begge kjønn.

Oppgave 4

Vi fortsetter med datafilen data2026.dta (alle 304 personene) som vi brukte i oppgave 1 og 3.

I en randomisert studie med en kontinuerlig variabel målt ved start og ved studieslutt, er det vanlig å bruke en regresjonsmodell der verdien ved studieslutt er avhengig variabel og verdien ved studiestart er uavhengig variabel sammen med gruppe. Vekt1 er da vår avhengige variabel med vekt0 og intervensjon som uavhengige variabler. I tillegg skal vi her ta med kjønn som uavhengig variabel og interaksjon mellom intervensjon og kjønn. Utfør en slik multippel lineær regresjonsanalyse. Merk at siden vi har med et interaksjonsledd skal det være

prefiks `i.` for både kjønn (`i.kjønn`) og intervensjon (`i.intervensjon`) i Stata. Er det signifikant interaksjon mellom intervensjon og kjønn? Forklar hva det betyr. Dere skal ikke tolke regresjonskoeffisientene.