UNIVERSITETET I OSLO

DET MEDISINSKE FAKULTET

Eksamen
ERN4000 STATISTIKK
Fredag 5. september 2025 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.

Datafilen titanicExam2025.dta brukes i alle oppgavene

Datafilen inneholder informasjon for 236 personer som reiste med Titanic i 1912.

Variabelnavn	**Beskrivelse**
Survived	Survival (overlevelse): 0 = no (dead); 1 = yes (alive)
Pclass	Passenger class (sosial klasse): 1 = 1st (upper), 2 = 2nd (middle), 3 = 3rd (lower)
Age	Age (alder) in years (fractional if less than 1 year)
Fare	Fare (billet pris) (British pound)
Sex	Sex (kjønn): 1 = female, 2 = male
Embark	Port of embarkation (avreisested): 1 = C (Cherbourg), 2 = Q (Queenstown), 3 = S (Southampton)

Merk:

- Vi antar at personene er uavhengige av hverandre. Med andre ord, i oppgavene ser vi bort i fra ethvert slektskap mellom personene.
- Datafilen inkluderer både voksne og barn, og i oppgavene skiller vi ikke mellom disse.
- Noen av variablene har manglende verdier.

Oppgave 1

Datafil titanicExam2025.dta.

a. Gi en beskrivelse av alder og billettpris. Bruk det du synes er mest hensiktsmessig av beskrivende statistikk.
b. Hvor mange menn og kvinner er det i datafilen? Angi også prosent.
c. Hva var gjennomsnittsalderen for henholdsvis menn og kvinner?
d. Sammenlign alderen for menn og kvinner ved hjelp av en statistisk test. Begrunn valget av test. Oppgi antall observasjoner, estimat, 95% konfidensintervall og p-verdi. Kommenter kort det du finner.
e. For hvor mange av mennene og kvinnene er det informasjon om billettpris? Sammenlign billettprisen for menn og kvinner ved hjelp av en ikke-parametrisk test, og beregn median billettpris for menn og kvinner. Kommenter kort det du finner.

Oppgave 2

Vi fortsetter med datafilen titanicExam2025.dta.

a. Lag et spredningsdiagram for alder og billettpris.
b. Beregn Pearsons korrelasjonskoeffisient med 95% konfidensintervall for sammenhengen mellom alder og billettpris.
c. Sammenhengen mellom alder og billettpris kunne også vært studert med lineær regresjonsanalyse der billettpris er avhengig variabel. Utfør en slik regresjonsanalyse og oppgi regresjonskoeffisienten med 95% konfidensintervall. Er forutsetningene til stede for å gjøre denne analysen?

Oppgave 3

Vi fortsetter med datafilen titanicExam2025.dta.

a. Hvor mange overlevde? Angi også prosent.
b. Hvor stor andel av mennene overlevde, og hvor stor andel av kvinnene overlevde?
c. Sammenlign overlevelsen for menn og kvinner ved hjelp av en statistisk test. Begrunn valget av test.
d. Beregn relativ risiko for å overleve for kvinner sammenlignet med menn, med 95% konfidensintervall. Kommenter kort det du finner.
e. Hvor mange personer var det i hver sosial klasse? Og hvor stor andel overlevde i hver sosial klasse? Undersøk om det er sammenheng mellom sosial klasse og overlevelse. Begrunn valget av test.

Oppgave 4

Vi fortsetter med datafilen titanicExam2025.dta.

a. Hvor stor andel av personene i hver av de tre sosiale klassene har informasjon om alder?
f. Er det noen forskjell i alder mellom de tre sosiale klassene? Utfør en parametrisk test og sjekk at forutsetningene er til stede. Dersom du finner en forskjell, foreta en parvis sammenligning av gruppene. Kommenter kort det du finner.

Oppgave 5

Vi fortsetter med datafilen titanicExam2025.dta.

a. Hvor mange personer gikk om bord fra hvert utreisested og for hvor mange personer mangler det informasjon om utreisested?
b. Er det noen sammenheng mellom sosial klasse og utreisested? Begrunn valget av test. Kommenter kort det du finner.