UNIVERSITETET I OSLO
ØKONOMISK INSTITUTT

Eksamen i: ECON2130 - Statistikk 1, høsten 2003

Eksamensdag: Mandag 8. desember 2003

Tid for eksamen: kl. 09:00 – 12:00

Oppgavesettet er på X sider

Tillatte hjelpemidler:

- Alle trykte og skrevne hjelpemidler, samt lommekalkulator er tillatt

Eksamen blir vurdert etter ECTS-skalaen. A-F, der A er beste karakter og E er dårligste ståkarakter. F er ikke bestått.

Oppgavesettet inneholder 4 oppgaver med underspørsmål. Les hele oppgaveteksten til hver enkelt av oppgavene 1, 2, 3 og 4 før du begynner å løse enkeltoppgaven og svare på delspørsmålene. Ikke bruk tid på å skrive deler av oppgaveteksten i din besvarelse.

Oppgave 1

En vanlig kortstokk inneholder 52 kort, som består av 13 kort i 4 ulike farger: Spar, hjerter, ruter og kløver. Du blir tildelt 5 kort fra kortstokken. Kortene ligger med baksiden opp slik at du ikke ser kortenes farge og verdi. Svar på delspørsmålene under punkt a og b, og begrunn kort beregningsmetoden din. Hvis formlene du bruker inneholder brøker og/eller binomialkoeffisienter, så trenger du ikke regne ut deres desimalverdier.

a) Før du tar opp kortene du er blitt tildelt, hva er sannsynligheten for at det tredje kortet er et ess i en hvilken som helst farge? Hva er sannsynligheten for at det er ess i hjerter? Hva er sannsynligheten for at alle fem kortene du er blitt tildelt har samme farge, enten spar, hjerter, ruter eller kløver?

b) Du tar opp de to første kortene blant de fem du har fått tildelt, og ser at de er hjerter 3 og hjerter knekt. Før du tar opp og ser de resterende tre kortene (dvs betingete på at du kjenner fargen og verdiene av de 2 første kortene av de 5 du er blitt tildelt) hva er sannsynligheten for at det tredje kortet er et ess, og hva er sannsynligheten for at det er ess i hjerter? Hva er nå sannsynligheten for at alle fem kort er i fargen hjerter?

Oppgave 2

X er en stokastisk variabel med en ukjent sannsynlighetsfordeling. Vi har informasjon om X bare i form av observasjoner av hendelser eller verdier x. Ved å analysere observasjonene kan vi foreslå en passende sannsynlighetsfordeling.

a) Forklar generelt og kort begrepene parameter, estimator, punktestimat og konfidensintervall. Hvilke av disse størrelsene vil du betrakte som usikre, og hvorfor?

b) Forklar hvorfor du kan bruke gjennomsnittet av observerte verdier x som en estimator for variabelens forventningsverdi E(X), og hvorfor du kan bruke relativ hyppighet av forekomsten x som en estimator for sannsynligheten P(X =x)?

c) La oss anta at det har vært 3 rentenedsettelser fra Norges Bank i løpet av en 12 måneders periode. Du ønsker å anslå sannsynligheten for nye rentenedsettelser de kommende 12 måneder. Først vil du bruke en Poisson modell. På grunnlag av de observerte 3 rentenedsettelsene, hva estimerer du parameteren λ i Poissonfordelingen til å være? Hva er tidsenheten din? Vis at sannsynligheten for minst en rentenedsettelse i løpet av de kommende 12 måneder i følge Poissonmodellen er lik 0.95.

d) Hovedstyret i Norges Bank møtes omtrent hver 6. uke, la oss si 8 ganger i løpet av 12 måneder. På hvert møte bestemmes det hvorvidt Norges Bank skal justere renta eller la den forbli uendret. På grunnlag av denne informasjonen samt de observerte 3 rentenedsettelsene de siste 12 måneder, bestemmer du deg for å bruke en binomisk modell. Hva blir ditt estimat på sannsynligheten for at Norges bank setter ned renten etter et møte? Vis at sannsynligheten for minst en rentenedsettelse i løpet av de kommende 12 måneder / 8 møter i følge den binomiske modellen er lik 0.977.

e) Drøft kort hvor godt de to sannsynlighetsmodellene egner seg til å anslå sannsynligheten for rentenedsettelser.

Oppgave 3

I en befolkningsprognose har man analysert den mulige befolkningsutviklingen fram til 2050. Framtiden er usikker - derfor er befolkningsprognosen basert på mange stokastiske variabler. I denne oppgaven er vi interessert i en av disse stokastiske variablene, nemlig fødselsraten, definert som forholdet mellom antall fødsler i et år og folkemengden ved årets begynnelse. Vi skriver fødselsraten som B.

Man har simulert 5000 mulige utviklingsbaner for denne variabelen B for hvert år mellom 2000 og 2050 (51 år). Dermed varierer B i hver bane over tid. For hvert år mellom 2000 og 2050 kan vi beregne gjennomsnittet av B over de 5000 simuleringer. Dette gjennomsnittet viser seg å være nokså konstant over tid, på et nivå rundt dagens verdi på 13 promille. Men på grunn av variansen i B er de 5000 utviklingsbanene svært forskjellige.

Forskerne er interessert i sannsynligheten for en ny babyboom. De definerer en slik begivenhet for en gitt simuleringsbane som en situasjon der fødselsraten B for minst ett år er større enn 18 promille i denne banen (som var omtrent lik verdien observert på 1950- og 1960-tallet). Det viser seg at dette var tilfelle i 992 av de 5000 simulerte utviklingsbaner for B. Besvar følgende spørsmål kort og presist.

a) Skriv sannsynligheten for en ny babyboom som p. Oppfatt hver simuleringsbane som et delforsøk. Hva må du anta om simuleringsopplegget for at p kan oppfattes som en parameter i en binomisk fordeling?

b) Anta at p er parameter i en binomisk fordeling. Gi en estimator for p. Er den forventningsrett? Begrunn ditt svar.

c) Estimer p, og gi et 95 prosent konfidensintervall for p. Er intervallet eksakt eller tilnærmet? Hvorfor?

d) En demograf hevder at sannsynligheten for en ny babyboom umulig kan være høyere enn 15%. Du skal teste nullhypotesen $p \leq 0{,}15$ mot den alternative hypotesen $p > 0{,}15$ ved hjelp av dette datamaterialet. Velg signifikansnivå lik 5 prosent. Hva er den kritiske verdien for p-estimatet? Må du forkaste nullhypotesen?

e) Demografen tror ikke at det er riktig å teste nullhypotesen $p \leq 0{,}15$. Han ønsker heller å teste nullhypotesen $p \geq 0{,}15$ mot den alternative hypotesen $p < 0{,}15$. Hva ville du foretrekke: å teste nullhypotese og alternativ hypotese som i oppgave d, eller teste det som demografen foretrekker? Du behøver ikke å gjennomføre testen som demografen foreslår.

Oppgave 4

X og Y er to stokastiske variabler, og du har observerte N par av verdier (x_1,y_1), (x_2,y_2), ..., (x_N,y_N).

a) Hva betyr det at de to variablene X og Y samvarierer, og hvordan kan du beskrive samvariasjon? Hva uttrykker regresjonsmodellen $Y_i = a + b\,X_i + u_i$ av samvariasjon og sammenheng mellom X og Y? Gjør rede for og begrunn egenskapene til feilleddet u.

b) Tegn et koordinatsystem med horisontal X-akse og vertikal Y-akse, samt en tilfeldig sky av datapunkter (x_i,y_i) som uttrykker en lineær sammenheng mellom de to variablene X og Y. La datapunktene ha god spredning både i X- og Y-retning, slik at sammenhengen ikke er eksakt. Regresjonslinjen $E(Y_i|X_i) = a + b\,X_i$ bestemmes slik at den minimerer summen av de kvadrerte avstandene fra datapunktene loddrett ned på eller loddrett opp til regresjonslinjen. Tegn inn linjen der du tror den må ligge omtrentlig. Tegn også inn en tenkt linje for den motsatte regresjonen $X_i = c + d\,Y_i + v_i$ Begrunn med ord hvor du plasserer regresjonslinjene.

c) La X = sparing og la Y = inntekt. Hva vil du som økonom forvente når det gjelder samvariasjonen mellom X og Y? Du ønsker å estimere en lineær modell for sammenhengen mellom X og Y. Gjør rede for hvordan du vil avgjøre hvilken av de to regresjonsmodellene i a) og b) du vil estimere.