UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1100 — Sannsynlighetsregning og statistisk modellering
Eksamensdag:	6. juni - 2025
Tid for eksamen:	15.00 – 19.00.
Oppgavesettet er på 4 sider.	
Vedlegg:	Formelsamling for STK1100/STK-FYS1110
Tillatte hjelpemidler:	Godkjent kalkulator Formelsamling for STK1100/STK-FYS1110

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Tabell over øvre persentiler for standard normalfordeling:

α	0.1	0.05	0.025	0.01	0.005	0.0025
z_α	0.1.282	1.645	1.960	2.326	2.576	2.807

Oppgave 1

Vi antar at trafikklysene i din by bare har to farger, rødt og grønt, og at de alltid virker slik at det er vekselsvis grønt lys i 40 sekunder og rødt lys i 20 sekunder.

(a) Du ankommer et kryss med et slikt trafikklys på et tilfeldig tidspunkt. Hva er sannsynligheten for at det er rødt lys? Hva er sannsynligheten for at det er grønt lys?

(b) Du passerer fire trafikklys på vei til jobb. Vi antar at de ligger langt fra hverandre og opererer uavhengig av hverandre. Hva er sannsynligheten for at det er "grønn bølge", dvs. at det er grønt lys når du passerer i alle fire kryssene? Hva er sannsynligheten for at du må stoppe minst tre ganger på rødt lys på vei til jobb?

Oppgave 2

Vi skal studere en situasjon der det utføres en sekvens av uavhengige, identiske forsøk, der resultatet i hvert forsøk er enten suksess (S) eller ikke (F). Sannsynligheten for suksess er den samme i hvert forsøk; $P(S) = p$. Hvis X er en tilfeldig variabel som beskriver hvor mange forsøk vi må gjøre frem til og med første suksess, vil X følge en geometrisk fordeling, gitt ved

(Fortsettes på side 2.)

$$P(X = x) = (1-p)^{x-1}p, \quad x = 1, 2, \ldots.$$

(a) Finn et uttrykk for den kumulative fordelingen til X, $F_X(x)$. Hvis $p = 1/3$, hva blir sannsynligheten for at første suksess skjer før 4. forsøk?

Noen ganger er man interessert i antall forsøk frem til man har opplevd r suksesser. Den geometriske modellen ovenfor er et spesialtilfelle der $r = 1$. Hvis Y er en tilfeldig variabel som beskriver hvor mange forsøk vi må gjøre frem til og med r-te suksess, vil Y følge en såkalt negativ binomisk fordeling, med

$$P(Y = y) = \binom{y-1}{r-1}(1-p)^{y-r}p^r, \quad y = r, r+1, r+2 \ldots.$$

Du skal ikke vise dette!

(b) Forklar hvorfor Y kan betraktes som summen av r uavhengige geometrisk fordelte variabler, alle med parameter p.
Bruk dette til å vise at $\mathsf{E}(Y) = \frac{r}{p}$.
Finn også $\mathsf{Var}(Y)$, samt den momentgenererende funksjonen til Y.

(c) Vi skal nå konsentrere oss om parameteren $\theta = 1/p$.
For en gitt r, bruk momentmetoden til å finne en estimator for θ basert på en observasjon av Y, og finn denne estimatorens forventning og standardfeil.
Er estimatoren forventningsrett?
Man observerer at det tok 11 forsøk før man hadde registrert 4 suksesser. Beregn et estimat for θ på bakgrunn av disse opplysningene, og medfølgende estimert standardfeil.

Oppgave 3

Anta en regresjonsmodell

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, ..., n$$

der $x_i \in \{0, 1\}$, dvs. antar enten verdien 0 eller verdien 1. Vi vil også anta at $\varepsilon_i \sim N(0, \sigma^2)$ og er uavhengige. Vi kan tenke på dette som en situasjon der vi har målinger fra to ulike grupper, den ene svarende til $x_i = 0$ og den andre svarende til $x_i = 1$.

(a) Hva blir forventningen til Y_i når $x_i = 0$ og når $x_i = 1$?

Argumentér for hvorfor β_1 måler forskjellen mellom de to gruppene.

La n_0 være antall observasjoner der $x_i = 0$ og n_1 antall observasjoner der $x_i = 1$.
En kan vise at

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n_0 n_1}{n}, \quad \text{der } \bar{x} = \frac{n_1}{n}$$

(Fortsettes på side 3.)

og

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{n_0 n_1}{n}[\bar{y}_1 - \bar{y}_0],$$

der $\bar{y}$ er gjennomsnittet av alle y_i-observasjonene, $\bar{y}_0$ er gjennomsnittet av y_i-observasjonene hvor $x_i = 0$ og $\bar{y}_1$ er gjennomsnittet av y_i-observasjonene hvor $x_i = 1$. Dette behøver du ikke å vise!

(b) Vis at minste kvadraters estimator for β_1 er $\hat{\beta}_1 = \overline{Y}_1 - \overline{Y}_0$ og utled en formel for $\hat{\beta}_0$, minste kvadraters estimator for β_0.

Diskutér om du synes disse estimatorene er rimelige.

Vi vil nå se på et konkret datasett som viser PEF ("Peak expiratory flow rate") målt for menn og kvinner. PEF er en metode for å måle lungekapasitet. Tabellen nedenfor viser målinger på 11 kvinner og 10 menn, totalt $n = 21$ observasjoner.

Kvinner	455	456	523	458	460	435	489	399	525	464	450
Menn	603	628	610	639	579	637	653	612	623	550	

Vi lar $x_i = 0$ for målinger fra kvinner og $x_i = 1$ for målinger fra menn. Vi vil anta at observasjonene følger den tidligere oppgitte regresjonsmodellen.
Vi har at

$$\bar{y}_0 = 464.91 \qquad \bar{y}_1 = 613.40$$
$$s_0^2 = 1327.70 \qquad s_1^2 = 934.49$$

der $\bar{y}_0$ er gjennomsnittet av observasjonene innenfor kvinnegruppen og $s_0^2 = \frac{1}{n_0-1}\sum_{i:x_i=0}(y_i - \bar{y}_0)^2$ er den empiriske variansen for de samme observasjonene. Notasjonen $\sum_{i:x_i=0}$ betyr her at vi kun summerer over de observasjoner der $x_i = 0$, dvs. observasjonene innenfor kvinnegruppen. Vi har tilsvarende $\bar{y}_1$ og s_1^2 for mannsgruppen.

(c) Vis at $\hat{\beta}_1$ er forventingsrett og at $\text{Var}[\hat{\beta}_1] = \frac{n}{n_0 n_1}\sigma^2$.

Beregn estimatet for β_1. Foreslå også et estimat for σ^2 og bruk dette til å gi et estimat på standardfeilen til $\hat{\beta}_1$.

Gir disse resultatene indikasjon på om det er forskjell på PEF-målinger mellom kvinner og menn?

(d) Argumentér for hvorfor $\hat{\beta}_1$ er normalfordelt.

Bruk dette til å utlede et eksakt $100(1-\alpha)\%$ konfidensintervall for β_1 når σ^2 er kjent.

Hva kan man si om konfidensintervallet hvis man setter inn et estimat for σ^2 i stedet?

(e) I denne oppgaven, anta at $\sigma = 33.7$.

Beregn både et 95% konfidensintervall og et 99% konfidensintervall for β_1.

Forklar forskjellene på disse intervallene, bl.a hvorfor det ene intervallet vil være større enn det andre.

Støtter disse intervallene opp dine betraktninger i punkt (*c*)?

(Fortsettes på side 4.)

Oppgave 4

Anta vi har en simultan sannsynlighetstetthet for (X, Y) gitt ved

$$f(x, y) = 2e^{-(x+y)}, \quad 0 \leq x \leq y.$$

(a) Finn den marginale sannsynlighetsfordelingen $f_X(x)$ for X og vis at

$$P(X < 1) = (1 - e^{-1})(1 + e^{-1}).$$

(b) Finn den betingede sannsynlighetsfordeling for $Y|X = x$.
Er X og Y uavhengige? Begrunn svaret.

(c) Beregn $P(Y < 1|X < 1)$.

SLUTT