UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1100 — Løsningsforslag
Eksamensdag:	6. juni - 2025
Tid for eksamen:	15.00 – 19.00.
Oppgavesettet er på 5 sider.	
Vedlegg:	Ingen
Tillatte hjelpemidler:	Godkjent kalkulator Formelsamling for STK1100/STK-FYS1110

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

Oppgave 1

(a) For hver periode på 60 sekunder, er det rødt lys i 20 sekunder og grønt lys i 40 sekunder. Vi har derfor at $P(\text{rødt}) = \frac{20}{60} = \frac{1}{3}$ og $P(\text{grønt}) = \frac{20}{60} = \frac{2}{3}$.

(b) $P(\text{grønt}) = \frac{2}{3}$ for hvert trafikklys. Vi har fått opplyst at alle trafikklysene er uavhengige av hverandre. Får derfor

$$\begin{aligned} P(\text{grønn bølge}) =& P(\text{grønt i første} \cap \text{grønt i andre} \cdots \cap \text{grønt i fjerde}) \\ =& P(\text{grønt i første}) \cdots P(\text{grønt i fjerde}) = (\frac{2}{3})^4 = 0.198 \end{aligned}$$

Neste spørsmål kan løses på mange måter ved å bruke kombinatorikk. En enkel løsning kan være å gjenkjenne et binomisk forsøk med $n = 4$, der rødt lys er suksess, og $P(\text{suksess}) = \frac{1}{3}$. Hvis X er antall røde lys på vei til jobb, er $X \sim \text{Bin}(4, \frac{1}{3})$. Vi finner da

$$P(X \geq 3) = P(X = 3) + P(X = 4) = \binom{4}{3}(\frac{1}{3})^3(1 - \frac{1}{3}) + \binom{4}{4}(\frac{1}{3})^4(1 - \frac{1}{3})^0 = \frac{1}{9}.$$

Oppgave 2

(a) Vi har

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

(Fortsettes på side 2.)

Ved å bruke formel for geometrisk rekke i formelsamlingen, finner vi

$$F_X(x) = P(X \leq x) = \sum_{k=1}^{x}(1-p)^{k-1}p = p\sum_{k=0}^{x-1}(1-p)^k = \frac{p(1-(1-p)^x)}{1-(1-p)} = 1-(1-p)^x.$$

Hvis $p = 1/3$, finner vi

$$F_X(3) = P(X \leq 3) = 1 - (1-\frac{1}{3})^3 = 1 - (\frac{2}{3})^3 = 1 - \frac{8}{27} = \frac{19}{27} = 0.704.$$

(b) Y kan betraktes som summen av r uavhengige X_i-er, der X_i-ene er geometrisk fordelt, fordi r suksesser fremkommer ved at man r ganger gjør forsøk frem til man får en suksess. Vi vet fra formelsamlingen at hvis X_i er geometrisk fordelt, er $\mathsf{E}(X_i) = 1/p$. Vi får derfor at $\mathsf{E}(Y) = \mathsf{E}(X_1 + X_2 \cdots X_r) = 1/p + 1/p \cdots + 1/p = r/p$.

Finner også på grunn av uavhengighet, $\mathsf{Var}(Y)$ som summen av variansene til X_i-ene, så vi får $\mathsf{Var}(Y) = r(1-p)/p^2$. Siden momentgenererende funksjon til en sum av uavhengige variable er lik produktet av momentgenererende funskjoner til variablene, finner vi

$$M_Y(t) = \prod_{i=1}^{r} M_{X_i}(t) = (\frac{pe^t}{1-(1-p)e^t})^r.$$

(c) Setter $\mathsf{E}(Y) = r/p = r\theta = Y$ og løser ut θ, som gir $\hat{\theta} = Y/r$. Regner ut $\mathsf{E}(\hat{\theta}) = \mathsf{E}(Y)/r = r\theta/r = \theta$, så estimatoren er forventningsrett. Videre finner vi

$$\mathsf{Var}(\hat{\theta}) = \mathsf{Var}(Y)/r^2 = r(1-p)/r^2p^2 = (1-p)/rp^2 = (\theta^2 - \theta)/r$$

slik at standard feil for $\hat{\theta}$ blir $\sqrt{(\theta^2-\theta)/r}$. Finner estimat $\hat{\theta} = 11/4 = 2.75$ og estimert standard feil som $\sqrt{(\hat{\theta}^2 - \hat{\theta})/r} = 1.1$.

Oppgave 3

(a) Vi har generelt at $E[Y_i] = \beta_0 + \beta_1 x_i$ som gir β_0 for $x_i = 0$ og $\beta_0 + \beta_1$ når $x_i = 1$ så β_1 blir da et mål på forventet forskjell mellom gruppene.

(b) Ved å sette inn i formler for $\hat{\beta}_1$, får vi

$$\begin{aligned}
\hat{\beta}_1 =& \bar{y}_1 - \bar{y}_0 \\
\bar{y} =& \frac{1}{n}[n_0\bar{y}_0 + n_1\bar{y}_1] \\
\hat{\beta}_0 =& \frac{1}{n}[n_0\bar{y}_0 + n_1\bar{y}_1] - (\bar{y}_1 - \bar{y}_0)\frac{n_1}{n} \\
=& \frac{n_0}{n}\bar{y}_0 + \frac{n_1}{n}\bar{y}_1 - \frac{n_1}{n}\bar{y}_1 + \frac{n_1}{n}\bar{y}_0 = \bar{y}_0
\end{aligned}$$

(Fortsettes på side 3.)

Dette medfører at $\hat{\beta}_0$ er gjennomsnitt av observasjoner fra gruppe 0, noe som er rimelig siden β_0 er forventningen i gruppe 0. Videre gjenspeiler $\hat{\beta}_1$ at β_1 er forskjell i forventinger i de to gruppene, som naturlig estimeres med forskjell i gjennomsnitt. Bevis for de deler som ikke kreves:

$$\begin{aligned}
\bar{x} =& \frac{1}{n} x_i = \frac{n_1}{n} \\
\sum_{i=1}^{n} (x_i - \bar{x})^2 =& \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = n_1 - \frac{n_1^2}{n} = \frac{n_0 n_1}{n} \\
\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) =& \sum_{i=1}^{n} (x_i - \bar{x}) y_i \\
=& \sum_{i:x_i=0} (0 - \frac{n_1}{n}) y_i + \sum_{i:x_i=1} (1 - \frac{n_1}{n}) y_i \\
=& - \frac{n_0 n_1}{n} \bar{y}_0 + \frac{n_0 n_1}{n} \bar{y}_1
\end{aligned}$$

(c) Vi har at

$$\begin{aligned}
E[\hat{\beta}_1] =& E[\bar{Y}_1 - \bar{Y}_0] = E[\bar{Y}_1] - E[\bar{Y}_0] \\
=& E[Y_i | x_i = 1] - E[Y_i | x_i = 0] \\
=& E[\beta_0 + \beta_1 * 1 + \varepsilon_i] - E[\beta_0 + \beta_1 * 0 + \varepsilon_i] = \beta_0 + \beta_1 - \beta_0 = \beta_1.
\end{aligned}$$

Vi har også at

$$\begin{aligned}
\text{Var}[\hat{\beta}_1] =& \text{Var}[\overline{Y}_1 - \overline{Y}_0] = \text{Var}[\overline{Y}_1] + \text{Var}[\overline{Y}_0] \\
=& \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_0} = \frac{n}{n_0 n_1} \sigma^2
\end{aligned}$$

siden variablene er uavhengige av hverandre. Vi får $\hat{\beta}_1 = 613.40 - 464.91 = 148.49$. Vi har at både s_0^2 og s_1^2 er forventningsrette estimatorer for σ^2. For estimering av σ^2 kan vi da bruke

$$\hat{\sigma}^2 = \frac{s_0^2 + s_1^2}{2} = 33.63^2,$$

eventuelt

$$\hat{\sigma}^2 = \frac{n_0 s_1^2 + n_1 s_1^2}{n} = 33.77^2$$

eller

$$\hat{\sigma}^2 = \frac{(n_0 - 1) s_1^2 + (n_1 - 1) s_1^2}{n - 2} = 33.79^2$$

(Fortsettes på side 4.)

der den andre er et fornuftig vektet gjennomsnitt mens den tredje svarer til hvordan man generelt estimerer varians i lineær regresjon.

Om vi setter inn ett av disse estimatene i uttrykket for $\sqrt{\text{Var}[\hat{\beta}_1]}$ ser vi at standardfeilen er liten i forhold til estimatet $\hat{\beta}_1$, noe som indikerer at det er forskjell mellom gruppene.

(d) Siden $\hat{\beta}_1$ er en lineærkombinasjoner av normalfordelte variable, så er den selv normalfordelt.

Vi har da

$$\begin{aligned}
\Pr(-z_{\alpha/2} < \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} < z_{\alpha/2}) &= 1-\alpha \\
&\Updownarrow \\
\Pr(-\sigma_{\hat{\beta}_1} z_{\alpha/2} < \hat{\beta}_1 - \beta_1 < z_{\alpha/2}\sigma_{\hat{\beta}_1}) &= 1-\alpha \\
&\Updownarrow \\
\Pr(-\hat{\beta}_1 - \sigma_{\hat{\beta}_1} z_{\alpha/2} < -\beta_1 < -\hat{\beta}_1 + z_{\alpha/2}\sigma_{\hat{\beta}_1}) &= 1-\alpha \\
&\Updownarrow \\
\Pr(\hat{\beta}_1 + \sigma_{\hat{\beta}_1} z_{\alpha/2} > \beta_1 > \hat{\beta}_1 - z_{\alpha/2}\sigma_{\hat{\beta}_1}) &= 1-\alpha
\end{aligned}$$

som gir et $100(1-\alpha)\%$ konfidensintervall $[\hat{\beta}_1 \pm z_{\alpha/2}\sigma_{\hat{\beta}_1}] = [\hat{\beta}_1 \pm z_{\alpha/2}\sigma\sqrt{n/(n_0 n_1)}]$.

Når vi setter inn et estimat for σ, så vil intervallet fremdeles være tilnærmet riktig (basert på store talls lov/sentralgrenseteoremet), men ikke lenger eksakt.

(e) Ved å bruke $z_{0.025} = 1.960$ og $z_{0.005} = 2.576$, får vi et (tilnærmet) 95% konfidensintervall $[119.63, 177.35]$ og et (tilnærmet) 99% konfidensintervall $[110.56, 186.42]$.

99% intervallet vil være større fordi vi lager et intervall som vi er sikrere vil dekke den sanne verdi. Større sikkerhet medfører imidlertid at intervall-lengden blir større.

Da begge intervallene indikerer at verdien av β_1 er veldig forskjellig fra 0, så indikerer det forskjell mellom de to gruppene.

Oppgave 4

(a) Vi har

$$f_X(x) = \int_x^\infty 2e^{-(x+y)}dy = 2e^{-x}[-e^{-y}]_x^\infty = 2e^{-2x}$$

for $0 < x < \infty$.

Vi har videre at

$$P(X < 1) = \int_0^1 2e^{-2x}dx = [-e^{-2x}]_0^1 = 1 - e^{-2} = (1-e^{-1})(1+e^{-1}).$$

(Fortsettes på side 5.)

(b) Det gir for $x \leq y \leq \infty$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{2e^{-(x+y)}}{2e^{-2x}} = e^{-(y-x)}.$$

Siden $f_{Y|X}(y|x)$ avhenger av x, så er X og Y *ikke* avhengige.

(c) Vi har

$$P(Y < 1|X < 1) = \frac{P(X < 1, Y < 1)}{P(X < 1)}$$

og

$$\begin{aligned} P(X < 1, Y < 1) &= \int_0^1 \int_x^1 2e^{-(x+y)} dy dx = 2\int_0^1 e^{-x}[-e^{-y}]_x^1 dx \\ &= 2\int_0^1 e^{-x}[e^{-x} - e^{-1}]dx = 2\int_0^1 [e^{-2x} - e^{-1-x}]dx \\ &= 2[e^{-1-x} - \frac{1}{2}e^{-2x}]_0^1 = 2[e^{-2} - \frac{1}{2}e^{-2} - e^{-1} + \frac{1}{2}] \\ &= e^{-2} - 2e^{-1} + 1 = (1 - e^{-1})^2 \end{aligned}$$

Dermed blir

$$P(Y < 1|X < 1) = \frac{1 - e^{-1}}{1 + e^{-1}}.$$

SLUTT