Oppgave 1 Spillet Tenzies

En regulær terning har 6 sider. Når man triller terningen er sannsynligheten for hvert utfall, x, lik $1/6 = p$, $x = 1, \ldots, 6$. I spillet Tenzies triller man 10 terninger i gjentatte runder inntil alle har vist tallet '6'. I praksis legger man til side terninger som allerede har vist '6', og kun de som ikke har vist '6' blir trillet i neste runde.

a) La Y være antall terninger som viser '6' i første runde.

Hva er sannsynlighetsfordelingen til Y?

Regn ut $P(Y > 0)$.

Regn ut $P(Y = 1|Y > 0)$.

b) La Z_i være antall runder inntil terning $i = 1, \ldots, 10$ viser '6'.

Forklar hvorfor hver Z_i, $i = 1, \ldots, 10$ er geometrisk fordelt. Hva blir formelen for punktsannsynlighetene til hver Z_i?

Bruk at for $0 < q < 1$, så er $1 + q + q^2 + \ldots + q^{z-1} = \sum_{i=1}^{z} q^{i-1} = \frac{1-q^z}{1-q}$, til å regne ut den kumulative fordelingsfunksjonen til hver Z_i.

Finn verdien z som splitter sannsynlighetsfordelingen til Z_i i to like deler (teoretisk median).

c) La W være antall runder inntil man oppnår Tenzies.

Finn et uttrykk for den kumulative fordelingsfunksjonen til W.

Dersom man oppnår Tenzies på 10 eller færre runder så er man konge. Hva er sannsynligheten for å bli konge?

Oppgave 2 Støymålinger

I en prosjektoppgave skal Jan Thomas analysere data av støymålinger.

a) Vi antar at en støymåling, Y, er normalfordelt med forventning $\mu = 40$ dB og standardavvik $\sigma = 10$ dB.

Hva er sannsynligheten for at en støymåling er høyere enn 60 dB?

Hva er sannsynligheten for at en støymåling ligger mellom 50 dB og 60 dB?

La c være en grense for kritisk støy bestemt slik at $P(Y > c) = 0.01$. Hva er c?

Jan Thomas har fått tilgang til data med $n = 30$ målinger fra et sted i Trondheim. Av hensyn til bomiljøet skal støynivået helst ikke overskride 50 dB.

b) Jan Thomas antar at støymålingene er uavhengige og identisk fordelte. Han lar Y_i være måling i og setter at $Y_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$, der μ og σ^2 er ukjente.

I tallmaterialet er $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 57.0$ og $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = 11.9^2$.

Finn et 95 % konfidensintervall for μ.

Jan Thomas får kontakt med firmaet som har samlet inn støymålingene, og han får oversendt data som vist i Tabell 1. Firmaet oppgir at målingene er gjennomførte i løpet av to uker i November 2022.

Tabell 1: Måling av støy (dB) på ulike dager og hele klokketimer.

	Mandag			Tirsdag			Onsdag			Torsdag			Fredag		
Tid (kl time)	8	12	16	8	13	15	9	12	17	8	10	16	9	13	17
Støy (dB)	68	41	48	74	32	60	71	38	51	64	46	60	58	49	60
	Mandag			Tirsdag			Onsdag			Torsdag			Fredag		
Tid (kl time)	10	13	16	8	12	16	8	12	17	9	12	15	8	12	16
Støy (dB)	61	51	58	71	39	63	66	44	59	70	52	65	80	45	65

Basert på denne informasjonen, ser Jan Thomas at han må ta hensyn til klokketiden t_i på døgnet målingene er gjort. For klokketidene som data er samlet inn på, foreslår han en regresjonsmodell:

$$Y_i = \beta_0 + \beta_1 \left[\cos\left(\frac{t_i - 8}{4}\pi\right) + 1\right] + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

der støyleddene $\epsilon_1, \ldots, \epsilon_n$ er uavhengige og identisk normalfordelte med forventning 0 og varians σ^2. Det oppgis at $\sum_{i=1}^{30} x_i = 38.5$, $\sum_{i=1}^{30} y_i = 1709$, $\sum_{i=1}^{30}(x_i - \bar{x})^2 = 20.1$ og $\sum_{i=1}^{30}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{30}(x_i - \bar{x})y_i = 227.7$, der $x_i = \cos\left(\frac{t_i - 8}{4}\pi\right) + 1$.

c) Hva er tolkningen av β_0 i ligning (1)?

Jan Thomas vil undersøke om det er grunn til å konkludere med at forventet støynivået under normalforhold (definert som kl 12 på dagen) er under 50 dB ved å utføre en hypotesetest for β_0. Formuler hypotesetesten.

Variansen til minstekvadratsumsestimatoren for β_0 er gitt ved $\sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2})$. Det oppgis at estimatet for σ^2 er 7.32^2. Gjennomfør hypotesetesten når signifikansnivået er satt til $\alpha = 0.05$.

Oppgave 3 Bilverkstedet

Et verksted klassifiserer feil i det elektriske anlegget til en bil til å være av 2 typer, A-feil og B-feil, avhengig av hvilken feilkode som blir gitt i feilsøkingen. Tidene målt i timer som går med til å finne en A-feil, X, og en B-feil, Y, er gitt ved sannsynsynlighetstetthetene $f_X(x) = \frac{1}{\beta}e^{-x/\beta}$, $x > 0$, $\beta > 0$ og $f_Y(y) = \frac{y}{\beta^2}e^{-y/\beta}$, $y > 0$, $\beta > 0$. Parameteren β er den samme i begge fordelingene.

a) Anta at $\beta = 1$ og finn sannsynligheten for at X er mindre enn 1.

Den momentgenererende funksjonen til X er gitt ved $M_X(t) = \frac{1}{1-\beta t}, t < \frac{1}{\beta}$. Bruk denne til å finne $E(X)$ og $Var(X)$ uttrykt ved β.

I resten av oppgaven skal vi anta at β er ukjent.

b) Vis at $Z = \frac{2Y}{\beta}$ er χ^2-fordelt (kjikvadratfordelt) med 4 frihetsgrader og bruk dette til å finne $E(Y)$ og $Var(Y)$ uttrykt ved β.

c) Anta at vi har m_1 tider $X_1, \ldots, X_{m_1}$ for feilsøking etter A-feil og m_2 tider $Y_1, \ldots, Y_{m_2}$ for feilsøking etter B-feil. Alle tidene antas å være uavhengige. Vis at sannsynlighetsmaksimeringsestimatoren for β er gitt ved $\hat{\beta} = \frac{\sum_{i=1}^{m_1} X_i + \sum_{j=1}^{m_2} Y_j}{m_1 + 2m_2}$

Finn forventning og varians til denne estimatoren.

d) En kunde har ventet i mange timer på at feilen skal bli funnet og mistenker at β er større enn 1 som er den verdien verkstedet hevder er riktig. Anta vi har $m_1 = m_2 = 100$ og at $\bar{x} = 1.43$ og $\bar{y} = 2.80$.

Forklar hvorfor $\hat{\beta}$ tilnærmet er normalfordelt.

Formuler en hypotesetest for denne situasjonen basert på mistanken til kunden. Foreslå en egnet testobservator (teststatistikk) og regn ut p-verdien til denne hypotesetesten med de gitte dataene. Kommenter resultatet.