Oppgavesettet består av 11 delspørsmål som alle teller likt ved bedømmelse av besvarelsen. For alle spørsmål skal alle "underspørsmål" besvares. Les oppgavene nøye. Uklarheter kan oppklares med foreleser rundt kl. 17. Alle oppgaver **skal** besvares med et passende nivå forklaring/utledning uten at man trenger å overdrive. Det vil derfor **ikke** være nok å alene referere til at kalkulatorer gir dere et svar. Kravet om utledning er naturligvis høyere på oppgaver der dere er gitt svaret og blir bedt om å vise hvordan man kommer frem til svaret. Men, ikke overdriv her heller. Der kilder ikke er oppgitt i oppgavesettet så er tallene i oppgavesettet fiktive.

OPPGAVE 1

En jeger skal på jakt i et område der det både er rype og hare. Jegeren skal prøve å jakte på begge dyreartene og hun anslår at sannsynligheten for at hun skyter minst én rype er 0.65 og sannsynligheten for at hun skyter minst én hare er 0.40. Sannsynligheten for at hun skyter minst ett dyr (enten rype eller hare) anslår hun til 0.88. Kall det at hun skyter minst én rype for hendelse R og at hun skyter minst én hare for hendelse H. Oppgavene under kan løses ved formell regning eller ved hjelp av resonnementer knyttet til Venn-diagrammet.

a) Vis at $P(R \cap H) = 0.17$ og forklar med ord hva denne sannsynligheten forteller oss.
Lag et Venn-diagram som viser sammenhengen mellom R og H.
Vis i Venn-diagrammet hvor vi finner $R \cap H$, $R \cap \bar{H}$ og $\bar{R} \cap H$.
Finn sannsynligheten for at hun ikke skyter noen ryper, men skyter minst én hare.

Fra oppgaveteksten har vi at

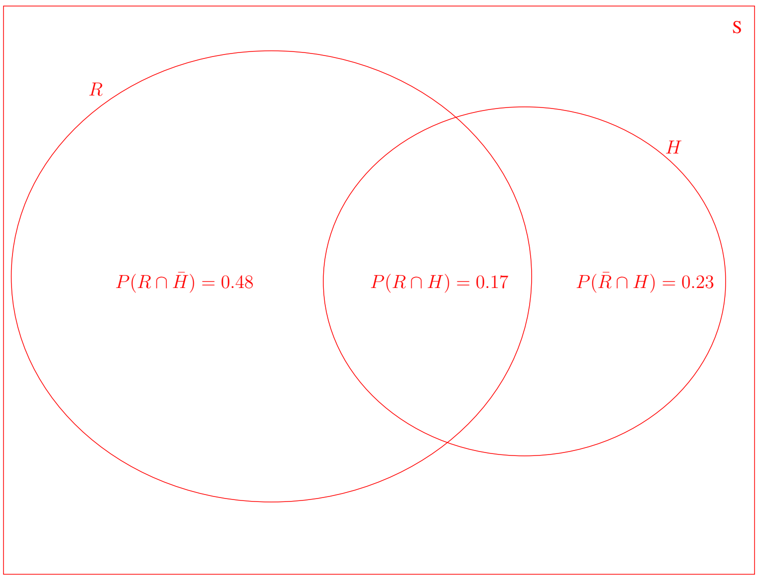
$$P(R) = 0.65 \quad , \quad P(H) = 0.40 \quad , \quad P(R \cup H) = 0.88.$$

Vi finner da $P(R \cap H)$ som

$$P(R \cap H) = P(R) + P(H) - P(R \cup H) = 0.65 + 0.40 - 0.88 = 0.17.$$

$P(R \cap H)$ beskriver sannsynligheten for at hun skyter både minst én rype og minst én hare.

Vi får da at det overlappende området i Venn-diagrammet skal være 0.17 stort. Dermed blir den resterende delen av R-sirkelen 0.48 stor, mens den resterende H-sirkelen blir 0.23 stor. Området utenfor sirklene er $P(\overline{R \cup H}) = 1 - P(R \cup H) = 1 - 0.88 = 0.12$. Vi får da Venn-digrammet



I det siste delspørsmålet skal man finne $P(\bar{R} \cap H)$. Man kan bruke Venn-diagrammet til å løse denne oppgaven. Med regning blir det med å snu på setning om total sannsynlighet

$$
\begin{aligned}
P(H) &= P(\bar{R} \cap H) + P(R \cap H) \\
&\Rightarrow \quad P(\bar{R} \cap H) = P(H) - P(R \cap H) = 0.40 - 0.17 = 0.23.
\end{aligned}
$$

b) Regn ut sannsynligheten for at hun skyter minst én hare når vi vet at hun skjøt minst én rype.
Er hendelsene R og H avhengige hendelser? Begrunn svaret.
Dersom hun forteller at hun har skutt minst ett dyr, men hun sier ikke noe om hvor mange dyr eller hvilke dyr hun har skutt, hva er sannsynligheten for at hun har skutt minst én rype?

Vi skal først finne $P(H|R)$, og vi får

$$
P(H|R) = \frac{P(R \cap H)}{P(R)} = \frac{0.17}{0.65} = 0.2615.
$$

Vi har altså at $P(H|R) = 0.2615$ som ikke er lik $P(H) = 0.40$, og vi kan da konkludere med at R og H er avhengige hendelser siden $P(H|R) \neq P(H)$.

Opplysningen vi får tilsvarer $R \cup H$. Vi kan da sette opp en formell utregning som gir

$$
\begin{aligned}
P(R|R \cup H) &= \frac{P(R \cap (R \cup H))}{P(R \cup H)} \\
&= \frac{P(R)}{P(R \cup H)} = \frac{0.65}{0.88} = 0.7386.
\end{aligned}
$$

Vi har her forenklet telleren siden det er nok at R skjer for at både R og $R \cup H$ skal skje, og vi kan dermed forenkle med at $P(R \cap (R \cup H)) = P(R)$.

Fra Venn-diagrammet ser vi at opplysningen vi får tilsvarer at vi befinner oss i området innenfor én eller begge sirklene, dvs $R \cup H$. Dette området tilsvarer det begrensede utfallsrommet og er $P(R \cup H) = 0.88$ stort. Sannsynligheten for at R skal skje på det begrensede utfallsrommet er da $P(R)/P(R \cup H) = 0.7386$, som før.

OPPGAVE 2

I Norge planlegges det å la kommunene bestemme om det skal være lovlige snøscooterløyper i de enkelte kommunene. I den forbindelse ønsker vi å finne ut hva folk synes om dette. I en meningsmåling blir 100 personer spurt om de er motstandere av at kommunene skal kunne tillate slike snøscooterløyper. Resultatet ble at 59 svarte "Ja", mens 41 svarte "Nei".

a) Kall antall personer som svarte "Ja" for X (de er motstandere/negative til snøscooterløyper). Hva slags fordeling X vil ha?
Hva er den naturlige estimatoren for andelen personer i Norge som er motstandere av snøscooterløyper? Forklar og grunngi svarene og regn ut estimatoren.

X vil være binomisk fordelt siden kravene til en binomisk forsøksrekke er oppfylt, dvs

i) Hvert forsøk/person har to mulige utfall/svar,

ii) sannsynligheten p for de mulige utfallene er antatt konstant for alle forsøk/personer, og

iii) hvert forsøk/person er uavhengig.

Den naturlige estimatoren $\hat{p}$ for andelen som er motstandere av snøscooterløyper i Norge p er antall personer som svarer "Ja" delt på antall som ble spurt, dvs med $n = 100$ lik antall personer som ble spurt i meningsmålingen får vi

$$\hat{p} = \frac{X}{n} = \frac{59}{100} = 0.59.$$

b) Det hevdes av enkelte at "mer enn 50% av Norges befolkning" er motstandere av slike snøscooterløyper. Bruk hypotesetesting til å undersøke om svarene fra meningsmålingen gir grunnlag for å hevde dette på 5% signifikansnivå.
Angi hypoteser, sjekk eventuelle nødvendige antagelser og angi en konklusjon av testen.

Siden påstanden som skal undersøkes er at flertallet av Norges befolkning er mot slike snøscooterløyper, så tilsvarer dette alternativhypotesen. Vi får da at hypotesene blir

$$\begin{aligned} &H_0: \quad p \leq 0.5 \quad \text{dvs at mindre enn eller lik 50\% er motstandere av snøscooterløyper} \\ &H_1: \quad p > 0.5 \quad \text{dvs at mer enn 50\% er motstandere av snøscooterløyper.} \end{aligned}$$

Vi skal gjøre hypotesetesting på en binomisk sannsynlighet, og vi bruker da normaltilnærming. Vi sjekker at denne tilnærmingen er gyldig, og kravet er da at variansen til X er større enn fem. Vi finner $\mathbf{Var}(X) = n \cdot p \cdot (1-p) = 100 \cdot 0.5 \cdot 0.5 = 25$, og siden $\mathbf{Var}(X) > 5$, så kan vi bruke normaltilnærmingen. Vi finner da testobservatoren som

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.59 - 0.50}{\sqrt{\frac{0.50(1-0.50)}{100}}} = \frac{0.09}{0.05} = 1.8.$$

Siden vi har en ensidig test, hvor vi forkaster for høye X-verdier, finner vi at forkastningsterskelen er gitt ved $z_{0.05} = 1.645$. Siden testobservatoren Z er større enn forkastningsterskelen $z_{0.05}$, så kan vi forkaste nullhypotesen, og vi kan konkludere med at mer enn 50% av Norges befolkning er mot snøscooterløyper.

Oppgaven kan også løses uten å gå via utregning av en Z-verdi, og da heller regne ut p-verdien direkte. Vi finner da nøyaktig p-verdi som (dette må gjøres på kalkulator)

$$\begin{aligned}
\text{p-verdi} &= P(X \geq 59) = 1 - P(X \leq 58) \\
&= 1 - \sum_{x=0}^{58} \binom{100}{x} 0.5^x (1-0.5)^{100-x} \\
&= 1 - \sum_{x=0}^{58} \binom{100}{x} 0.5^{100} = 1 - 0.9557 = 0.0443.
\end{aligned}$$

Eller vi kan finne dette tilnærmet som

$$\begin{aligned}
\text{p-verdi} &= P(X \geq 59) = 1 - P(X \leq 58) \\
&\approx 1 - G\left(\frac{58 + 0.5 - n \cdot p_0}{\sqrt{n \cdot p_0(1-p_0)}}\right) \\
&= 1 - G\left(\frac{58 + 0.5 - 50}{\sqrt{25}}\right) = 1 - G(1.7) = 1 - 0.9554 = 0.0446.
\end{aligned}$$

Vi ser at både eksakt og tilnærmet p-verdi er mindre enn 0.05, og vi kan da på samme måte som med Z-testobservatoren forkaste nullhypotesen.

Blant de som ønsker snøscooterløyper hevdes det at det hovedsakelig er folk som bor i de store byene som er motstandere av snøscooterløyper. De samme meningsmålingsdataene deles derfor opp etter om den som ble spurt bor i en stor by eller ikke. Av de 41 som ønsker snøscooterløyper bodde 12 i en stor by. Av de som er motstandere av snøscooterløyper bodde 48 i en stor by.

c) Bruk denne oppdelingen til å avgjøre om det er forskjell mellom de som bor i en stor by og de som ikke bor i en stor by med tanke på om de ønsker snøscooterløyper eller ikke. Sett opp en kjikvadrattest og bruk 5% signifikansnivå. Angi hypoteser og gi en konklusjon.

Vi bruker en kjikvadrattest med en 2x2-krysstabell. De to kategoriene er positiv/negativ til snøscooterløyper og bor/bor ikke i en stor by. Fra oppgaven får vi da

	Bor i stor by	Bor ikke i stor by	Radsum R_i
Motstander av snøscooterløyper	48	11	59
Tilhenger av snøscooterløyper	12	29	41
Kolonnesum K_j	60	40	100

Nullhypotesen tilsvarer at kolonnene og radene er uavhengige, dvs. at det ikke er noen sammenheng mellom holdning til snøscooterløyper og hvor man bor. Alternativhypotesen er at det er en sammenheng mellom disse to. Vi får da

H_0 : Det er ingen sammenheng mellom holdning til snøscooterløyper og hvor man bor

H_1 : Det er en sammenheng mellom holdning til snøscooterløyper og hvor man bor.

Vi finner at radsummene er $R_1 = 59$ og $R_2 = 41$, og kolonnesummene er $K_1 = 60$ og $K_2 = 40$. Vi finner da forventet antall, under nullhypotesen, for hver av cellene som $E_{\text{rad } i,\text{kolonne } j} = \frac{R_i \cdot K_j}{n}$, som gir

$$E_{1,1} = \frac{R_1 \cdot K_1}{n} = \frac{59 \cdot 60}{100} = \frac{3540}{100} = 35.4.$$

Tilsvarende får vi

$$E_{1,2} = \frac{59 \cdot 40}{100} = 23.6 \quad , \quad E_{2,1} = \frac{41 \cdot 60}{100} = 24.6 \quad , \quad E_{2,2} = \frac{41 \cdot 40}{100} = 16.4.$$

Vi finner da de skalerte, kvadratiske avvikene for hver av cellene som $(X_{i,j} - E_{i,j})^2/E_{i,j}$, som gir

$$\frac{(X_{1,1} - E_{1,1})^2}{E_{1,1}} = \frac{(48 - 35.4)^2}{35.4} = \frac{(12.4)^2}{35.4} = 4.48475$$
$$\frac{(X_{1,2} - E_{1,2})^2}{E_{1,2}} = \frac{(11 - 23.6)^2}{23.6} = \frac{(-12.4)^2}{23.6} = 6.7271$$
$$\frac{(X_{2,1} - E_{2,1})^2}{E_{2,1}} = \frac{(12 - 24.6)^2}{24.6} = \frac{(-12.4)^2}{24.6} = 6.4537$$
$$\frac{(X_{2,2} - E_{2,2})^2}{E_{2,2}} = \frac{(29 - 16.4)^2}{16.4} = \frac{(12.4)^2}{16.4} = 9.6805.$$

Vi får da at testobservatoren er lik

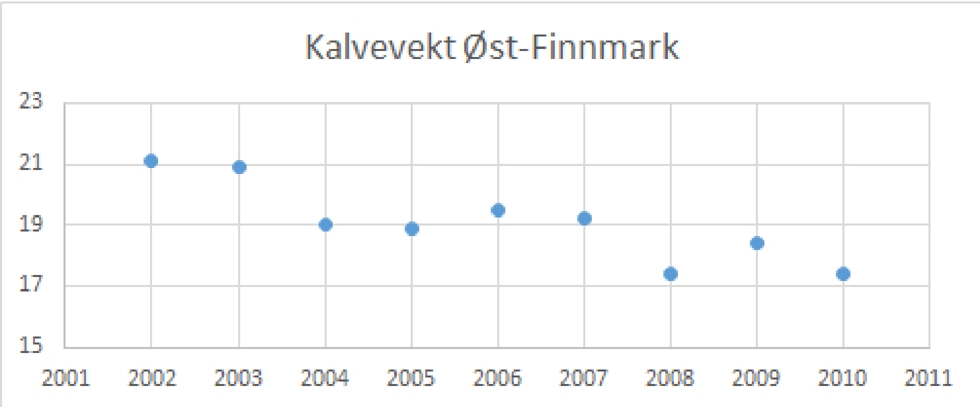
$$Q = \sum_{\text{over alle celler}} \frac{(X_{i,j} - E_{i,j})^2}{E_{i,j}} = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(X_{i,j} - E_{i,j})^2}{E_{i,j}}$$
$$= 4.48475 + 6.7271 + 6.4537 + 9.6805 = 27.346.$$

Vi finner antall frihetsgrader som (antall rader - 1)(antall kolonner -1), som gir $(2-1)(2-1) = 1$ antall frihetsgrader. Vi finner da at forkastningsterskelen for 1 frihetsgrad i chikvadratfordeligen er lik 3.84. Siden testobservatoren Q er større enn forkastningsterskelen kan vi forkaste nullhypotesen, og vi kan fastslå at det er sammenheng mellom holdning til snøscooterløyper og om man bor i en stor by eller ikke.

OPPGAVE 3

I denne oppgaven skal vi se på slaktevekter av tamrein. Dataene er tatt fra Landbruksdirektoratets sider om reindrift, mer spesifikt fra rapporten "Ressursregnskap for reindriftsnæringen" fra mai 2012 (`http://www.reindrift.no/asset/4922/1/4922_1.pdf`). Fra rapporten har vi at slaktevekten i antall kilo av kalver for området "Øst-Finnmark" er

Årstall (x_i)	2002	2003	2004	2005	2006	2007	2008	2009	2010
Slaktevekt kalv (y_i)	21.1	20.9	19.0	18.9	19.5	19.2	17.4	18.4	17.4



Vi betegner en regresjonslinje ved $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$, og videre i oppgaven kan du bruke at (der $i = 1, 2, \ldots, 9$ tilsvarer årene fra 2002 til 2010)

$$\sum_{i=1}^{9}(x_i - \bar{x})(y_i - \bar{y}) = -25.2 \quad , \quad \sum_{i=1}^{9} x_i = 18054 \quad ,$$

$$SS_E = \sum_{i=1}^{9}(y_i - \hat{y}_i)^2 = 3.14489 \quad , \quad \sum_{i=1}^{9}(x_i - \bar{x}_i)^2 = 60 \quad , \quad \sum_{i=1}^{9} y_i = 171.8.$$

a) Anta at dataene kan beskrives godt av en lineær regresjonsmodell og bestem regresjonslinjen $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$. Tegn inn regresjonslinjen i koordinatsystemet sammen med datapunktene. Dette må gjøres på vanlige svarark, dvs. dere kan **ikke** tegne inn på oppgavearket. Vis hvordan du regner deg frem til hvor regresjonslinjen skal tegnes inn (det holder ikke å henvise til at du leser av fra kalkulator).
Gi en tolkning av de to parameterne som trengs for å beskrive regresjonslinjen. Gir begge parameterne meningsfull informasjon om man ønsker å se på utviklingen av slaktevekten?

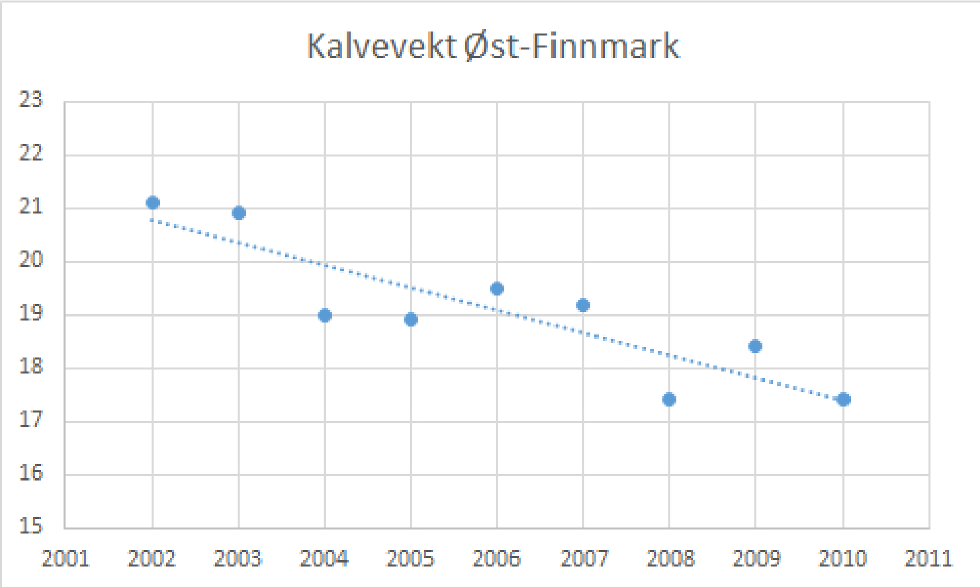
Vi finner først det estimerte stigningstallet $\hat{\beta}$ som

$$\hat{\beta} = \frac{\sum_{i=1}^{9}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{9}(x_i - \bar{x}_i)^2} = \frac{-25.2}{60} = -0.42.$$

Vi finner så $\bar{x} = \sum x_i/n = 18054/9 = 2006$ og $\bar{y} = \sum y_i/n = 171.8/9 = 19.089$, og bruker disse til å finne $\hat{\alpha}$ som

$$\bar{y} = \hat{\alpha} + \hat{\beta} \cdot \bar{x}$$
$$\Rightarrow \quad \hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x} = 19.089 - (-0.42) \cdot 2006 = 861.6089,$$

som gir regresjonslinjen $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x = 861.6089 - 0.42 \cdot x$.



Vi finner to punkter som ligger på regresjonslinjen og trekke så en rett linjen gjennom disse punktene. Vi får da

$$\hat{y}(2002) = 861.6089 - 0.42 \cdot 2002 = 20.7689$$
$$\hat{y}(2010) = 861.6089 - 0.42 \cdot 2010 = 17.4089.$$

Det estimerte stigningstallet $\hat{\beta}$ forteller oss hvor mye slaktevekten har gått ned per år, og dette er meningsfull informasjon. Den estimerte skjæringsverdien med andreaksen $\hat{\alpha}$ forteller oss hva slaktevekten, i følge en slik lineær modell, ville vært i år null. Dette er ikke relevant i og med at vi da vil bruke modellen til å predikere verdier ekstremt langt utenfor området vi har data for, og i tillegg er en vekt på 861 kg på en reinkalv tull.

b) Om utviklingen av slaktevekten fortsetter å følge den lineære modellen regnet ut i oppgave a), når vil slaktevekten være lik 5 kg? Du skal regne ut svaret og du kan angi svaret i desimalår, f.eks. år 2017.74.
Diskuter også om en slik bruk av en lineær regresjonsmodell er meningsfull.

Vi har regresjonslinjen

$$\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$$

og når vi er ute etter året så løser vi ut x alene og setter inn $\hat{y} = 5$, som gir

$$\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$$
$$\hat{y} - \hat{\alpha} = \hat{\beta} \cdot x$$
$$\frac{\hat{y} - \hat{\alpha}}{\hat{\beta}} = x$$
$$x = \frac{\hat{y} - \hat{\alpha}}{\hat{\beta}} = \frac{5 - 861.61}{-0.42} = 2039,547619 \approx 2039.5.$$

I følge den lineære regresjonsmodellen vil altså gjennomsnittlig kalveslaktevekt være fem kg i år 2039.5.

Rent teoretisk så vet vi at man skal være veldig forsiktig med å bruke lineære, og andre, regresjonsmodeller til å gjøre prediksjoner for verdier av den uavhengige variabelen som er utenfor området vi har data for. I vårt tilfelle har vi data for årene 2002 til 2010, og da er det svært tvilsomt å hevde at man kan bruke en lineær regresjonsmodell til å si noe om hva som skjer 30 år frem i tid. Biologisk sett så er det vel også meningsløst å snakke om slaktevekter på fem kg, da ingen kalver vil kunne overleve vinteren med så lav vekt. Til slutt, så er det vel gode sjanser for å lovgivende myndigheter griper inn lenge før vi kommer ned på slike lave slaktevekter. Det er derfor av mange grunner meningsløst å bruke en lineær regresjonsmodell til å predikere år for når gjennomsnittlig slaktevekt blir fem kg.

c) Vis at et 95 % konfidensintervall for stigningstallet i regresjonslinjen er gitt ved nedre og øvre grense $[-0.6246\ ,\ -0.2154]$.
Bruk dette intervallet til å avgjøre om det har vært en signifikant nedgang/økning i slaktevekten over denne tidsperioden. Bruk 5% signifikansnivå.

Vi har at et 95% konfidensintervall for stigningstallet β er gitt ved

$$\hat{\beta} \pm t_{\alpha/2}^{(n-2)} \cdot \mathbf{SE}(\hat{\beta}),$$

der $\alpha = 0.05$ i vårt tilfelle og antall frihetsgrader er $n-2 = 9-2 = 7$. Vi har da at $t_{\alpha/2}^{(n-2)} = 2.365$.

For å finne $\mathbf{SE}(\hat{\beta}) = \sqrt{\mathbf{Var}(\hat{\beta})}$ må vi først finne

$$\mathbf{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{9}(x_i - \bar{x})^2},$$

der σ^2 er den sanne variansen til slaktevektene y_i. Denne kjenner vi ikke, og vi må da estimere den via

$$s^2 = \frac{SS_E}{n-2} = \frac{3.14489}{7} = 0.44927.$$

Vi får da

$$\mathbf{Var}(\hat{\beta}) = \frac{s^2}{\sum_{i=1}^{9}(x_i - \bar{x})^2} = \frac{0.44927}{60} = 0.0074878.$$

Vi får da $\mathbf{SE}(\hat{\beta}) = \sqrt{\mathbf{Var}(\hat{\beta})} = \sqrt{0.0074878} = 0.08653$, som gir et 95 % konfidensintervall for stigningstallet β gitt ved

$$\hat{\beta} \pm t_{\alpha/2}^{(n-2)} \cdot \mathbf{SE}(\hat{\beta}) = -0.42 \pm 2.365 \cdot 0.08653 = -0.42 \pm 0.20464,$$

som gir det oppgitte konfidensintervallet $[-0.6246\ ,\ -0.2154]$.

Siden 95%-konfidensintervallet for stigningstallet for Øst-Finnmark ikke inneholder null kan vi på 5 % signifikansnivå hevde at det har vært et signfikant nedgang (siden konfidensintervallet kun inneholder negative verdier av stigningstallet) i slaktevekten.

OPPGAVE 4

Vi skal i denne oppgaven se på hvor mye bensin vi bruker med ulike typer bensin. Fra før av vet vi at vår bil bruker i gjennomsnitt $\mu_V = 0.5$ liter bensin per mil (1 mil tilsvarer 10 kilometer) med "vanlig" bensin. Anta at bensinforbruket er normalfordelt med et standardavvik på $\sigma = 0.05$ liter per mil. Dette standardavviket er knyttet til at vi kjører tur-retur til arbeid, og videre i oppgaven er én slik kjøring tur-retur til arbeid å betrakte som én måling av bensinforbruket. Vi antar at bensinforbruket de ulike dagene er uavhengig av hverandre.

a) Hva er sannsynligheten for at vi bruker mindre enn 0.40 liter per mil med vanlig bensin? Hva er sannsynligheten for at vi i løpet av fire arbeidsdager i gjennomsnitt bruker mindre enn 0.45 liter per mil?

Vi kaller målt bensinforbruket for X_V for vanlig bensin og vi kaller gjennomsnittlig bensinforbruk med vanlig bensin for $\mu_V = 0.50$. Vi skal finne $P(X_V < 0.40)$. Vi transformerer problemet til en standard normalfordelt variabel og får da

$$P(X_V < 0.40) = G\left(\frac{0.40 - 0.50}{0.05}\right) = G\left(\frac{-0.10}{0.05}\right) = G(-2) = 0.0228.$$

Vi vet at et gjennomsnitt $\bar{X}$ av n uavhengige normalfordelte variabler vil være fordelt med samme forventningsverdi som enkeltmålingene, men med et standardavvik som er lik standardavviket for en enkeltmåling delt på roten av antall målinger. Det tilsvarer $\bar{X}_V \sim \mathcal{N}(\mu_V, \sigma/\sqrt{n})$. Vi finner da $P(\bar{X}_V < 0.45)$ som

$$P(\bar{X} < 0.45) = G\left(\frac{0.45 - 0.50}{0.05/\sqrt{4}}\right) = G\left(\frac{-0.05}{0.05/2}\right) = G(-2) = 0.0228.$$

Et drivstoffselskap selger bensinen "B-power" som de hevder gjør at man bruker mindre bensin enn med vanlig bensin. Vi ønsker å undersøke om vi bruker mindre bensin når vi bruker "B-power" i vår bil. Vi måler derfor nøyaktig X_i, som er hvor mye bensin vi bruker per mil når vi kjører tur-retur arbeid hver dag i to uker med "B-power" og vi får da

Dag i	1	2	3	4	5	6	7	8	9	10	Sum
X_i	0.372	0.421	0.527	0.453	0.548	0.410	0.473	0.517	0.438	0.488	4.647.

b) Bruk de 10 målingene i tabellen over til å teste om vi på 5% signifikansnivå kan hevde at vi faktisk bruker mindre bensin når vi bruker "B-power". Du kan **ikke** anta at standardavviket til bensinforbruket med "B-power" er kjent.
Forklar ditt valg av ensidige eller tosidige hypoteser i denne oppgave, angi hypotesene og angi en konklusjon.
Du kan i denne deloppgaven anta at målingene er normalfordelte og bruke at

$$\sum_{i=1}^{10}(X_i - 0.4647)^2 = 0.0285121.$$

Siden vi må estimere standardavviket fra dataene så blir det en t-test vi skal utføre. Det blir en ensidig test, der vi forkaster for lave verdier av testobservatoren T, siden vi kun vil undersøke om vi bruker mindre bensin med "B-power". Vi får da at hypotesene blir

$$H_0 : \mu_B \geq \mu_V$$
$$H_1 : \mu_B < \mu_V$$

der μ_B er gjennomsnittlig bensinforbruk med "B-power". Fra oppgaven har vi at det empiriske gjennomsnittet er lik $\bar{X} = 4.647/10 = 0.4647$ og det empiriske standardavviket er lik

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{10}(X_i - \bar{X})^2} = \sqrt{\frac{1}{9}\sum_{i=1}^{10}(X_i - 0.4647)^2} = \sqrt{\frac{1}{9}0.0285121} = 0.05629.$$

Testobservatoren blir da

$$T = \frac{\bar{X} - \mu_V}{s/\sqrt{n}} = \frac{0.4647 - 0.50}{0.05629/\sqrt{10}} = -1.9831.$$

Med 9 frihetsgrader og 5% signifikansnivå blir forkastningsterskelen for den ensidige t-testen lik $t_{0.05}^{(9)} = -1.833$. Siden testobservatoren er mer negativ en forkastningsterskelen, dvs. $T = -1.9831 < t_{0.05}^{(9)} = -1.833$, så kan vi forkaste nullhypotesen, og vi kan dermed hevde at bilen vår faktisk bruker mindre bensin med "B-power" enn med vanlig bensin.

c) I oppgave b) kunne vi anta at målingene var normalfordelte. Vi skal i denne oppgaven se hvordan vi kan teste samme hypoteser som i b), men når vi **ikke** kan anta at målingene er normalfordelte. Vi skal da bruke en fremgangsmåte som er veldig lik som for "Paret fortegnstest" i læreboken. Vi kan da hevde at hvis det faktisk er sånn at vi i gjennomsnitt bruker like mye bensin med "B-power" som med vanlig bensin, så er det 50% sannsynlighet for at vi en tilfeldig dag bruker mindre enn 0.5 liter per mil også med "B-power".
Bruk dette til å lage en test der vi kun ser på om vi for de ulike dagene bruker mer eller mindre bensin enn 0.5 liter per mil. Bruk samme signifikansnivå som i b).
Sammenlign konklusjonen denne testen gir med konklusjonen fra oppgave b).

Vi ser her på kun antall dager da vi brukte under 0.50 liter per mil med "B-power". Vi kaller dette antallet U og finner at $U = 7$. Vi har at U vil være binomisk fordelt (to mulige utfall med lik sannsynlighet for hver uavhengige dag) med $p = 0.50$ og $n = 10$ under nullhypotesen. Som for "Paret fortegnstest" finner vi da p-verdien som

$$\text{p-verdi} = P(U \geq 7) = 1 - P(U \leq 6) \overset{\text{Tabell E.1}}{=} 1 - 0.828 = 0.172.$$

Det betyr at kun ved tilfeldighet, dvs. i følge nullhypotesen, kan vi oppleve å bruke under 0.50 liter per mil for syv eller flere dager i løpet av 10 dager med 17.2% sannsynlighet. Vi kan dermed ikke forkaste nullhypotesen gjennom denne ikkeparametriske testen. Vi ser at vi får ulike konklusjon i b) og c), noe som viser det kjente faktum at ikkeparametriske tester har lavere styrke enn parametriske tester (gitt at forutsetningene i den parametriske testen er oppfylt).