# Løsning øving 4

## Oppgave 1

Basert på 5 målinger med snitt $\bar{X} = 0,22$ skal vi konstruere et 95 %-konfidensintervall for alkoholnivået $\mu$ i blodet til bilføreren, når hver uavhengige målinger har et kjent standardavvik på $\sigma = 0,044$.

Gjennomsnittet $\bar{X}$ vil være normalfordelt som $\bar{X} \sim N\left(\mu, \frac{0{,}044^2}{5}\right)$. Et 95 %-konfidensintervall er da gitt ved (fra formelarket)

$$\bar{X} \pm u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}},$$

der $\frac{\alpha}{2}$-kvantilen her er $u_{\frac{0{,}05}{2}} = u_{0{,}025} = 1,96$. På intervallform blir dette

$$\begin{aligned} & \left[\bar{X} - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \\ = & \left[0,22 - 1,96 \frac{0,044}{\sqrt{5}}, 0,22 + 1,96 \frac{0,044}{\sqrt{5}}\right] \\ = & \underline{\underline{[0,18, 0,26]}} \end{aligned}$$

## Oppgave 2

Når man utfører en hypotesetest, er alltid nullhypotesen den "konservative"/"tilbakeholdne"/"status quo"-situasjonen. Ettersom det er promillegrensa på 0,20 som bestemmer hvorvidt bilføreren skal dømmes for promillekjøring, vil en naturlig nullhypotese her være

$H_0$: Bilføreren har en promille på $\mu = 0,20$.

## Oppgave 3

Vi skal nå gjennomføre følgende hypotesetest med 5 % signifikansnivå:

$H_0$: Bilføreren har lovlig promille, $\mu = 0,20$.

vs. mothypotesen

$H_1$: Bilføreren har for høy promille, $\mu > 0,20$.

Når vi altså skal teste hypotesen $\mu = \mu_0 = 0,20$ mot $\mu > 0,20$, bruker vi testobservatoren

$$U_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{0,22 - 0,20}{\frac{0{,}044}{\sqrt{5}}} = \underline{1,02}$$

Vi sammenlikner nå verdien på testobservatoren $U_0$ med verdien på kvantilen $u_{0{,}05}$, ettersom vi bruker 5 % signifikansnivå. Denne er

$$u_{0{,}05} = 1,645$$

Ettersom $U_0 = 1,02 < u_{0{,}05} = 1,645$, vil vi konkludere med at vi ikke skal forkaste nullhypotesen - altså vi kan ikke dømme bilføreren for promillekjøring.

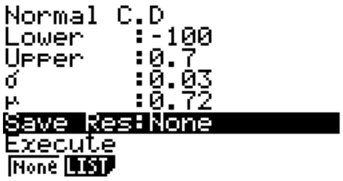
## Oppgave 4

Bensinforbruket $X$ for en bil er normalfordelt med $\mu = 0,72$l/mil og $\sigma = 0,03$l/mil, dvs. $X \sim N\left(\mu, \sigma^2\right) = N\left(0,72, 0,03^2\right)$. Vi skal bestemme sannsynligheten for at en tilfeldig bil av denne typen har et bensinforbruk større enn $0,70\,$l/mil.

Vi får at:

$$P\left(X > 0,70\right) = 1 - P\left(X \leq 0,70\right),$$

der vi finner den kumulative sannsynligheten fra kalkis:




Vi får at $P\left(X \leq 0,70\right) = 0,252$, slik at

$$\begin{aligned} P\left(X > 0,70\right) &= 1 - P\left(X \leq 0,70\right) \\ &= 1 - 0,252 \\ &= 0,748 \\ &\approx \underline{\underline{0,75}} \end{aligned}$$

## Oppgave 5

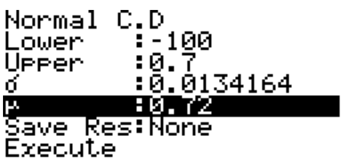
b) Vi skal nå finne sannsynligheten for at det gjennomsnittlige bensinforbruket for 5 slike biler blir lavere enn 0,70 liter per mil. Vi vet at snittet av 5 normalfordelte variable $X_1, , , , , X_5$ med samme $\mu$ og $\sigma^2$ er

$$\bar{X} = \frac{X_1 + X_2 +, , , + X_5}{5} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \sim N\left(0,72, \frac{0,03^2}{5}\right)$$

slik at vi skal bestemme

$$P\left(\bar{X} \leq 0,70\right)$$

Ettersom vi vet fordelingen til $\bar{X}$, finner vi må den kumulative sannsynligheten via kalkis (standardavviket blir $\sqrt{\frac{0,03^2}{5}}$):




$$P\left(\bar{X} \leq 0,70\right) = \underline{\underline{0,068}}$$

## Oppgave 6

Vi skal finne sannsynligheten for at minst 3 av de 5 bilene har et bensinforbruk lavere enn 0.70 liter/mil. Her kan vi bruke en binomisk modell: vi har 2 mulige utfall (har/har ikke lavere forbruk enn 0,70 l/mil), og vi gjør 5 uavhengige "delforsøk" (måler på 5 uavhengige biler).

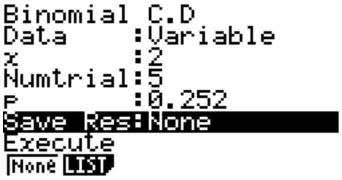
Sannsynligheten for at én bil har **lavere** forbruk enn $0,70\,\mathrm{l/mil}$ fant vi tidligere, lik $p = 0,252$. Hvis $X$ angir antall biler med forbruk lavere enn $0,70\,\mathrm{l/mil}$, er altså

$$X \sim bin\,(n, p) \sim bin\,(5, 0, 252)\,,$$

slik at sannsynligheten for at minst 3 av 5 biler har lavere forbruk enn 0.70 l/mil blir

$$P\,(X \geq 3) = 1 - P\,(X \leq 2)\,,$$

og vi kan finne den kumulative binomiske fordelingen fra kalkis:




Får at

$$P\,(X \leq 2) = 0,894,$$

slik at

$$\begin{aligned} P\,(X \geq 3) &= 1 - P\,(X \leq 2) \\ &= 1 - 0,894 \\ &= 0,106 \\ &\approx \underline{\underline{0,11}} \end{aligned}$$

### Oppgave 7

a) Vi har gitt følgende sett med målinger av motoreffekt (i HK):

147,3 152,9 148,1 149,0 150,6

For et 90 % konfidensintervall for den målte motoreffekten er $\alpha = 0,10$, slik at formelarket gir:

$$\bar{X} \pm u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \bar{X} \pm u_{\frac{0,10}{2}} \frac{\sigma}{\sqrt{n}} = \bar{X} \pm u_{0,05} \frac{\sigma}{\sqrt{n}},$$

der $X$ er det estimerte gjennomsnittet av målingene, $u_{0,05} = 1,645$ (fra kvantiltabellen) og $\sigma = 3,0$ er det kjente standardavviket for målingene. Gjennomsnittet blir

$$\bar{X} = \frac{147,3 + 152,9 + 148,1 + 149,0 + 150,6}{5} = \underline{149,6}$$

Konfidensintervallet blir da

$$\bar{X} \pm u_{0,05} \frac{\sigma}{\sqrt{n}} = 149,6 \pm 1,645 \cdot \frac{3,0}{\sqrt{5}} = 149,6 \pm 2,21$$

Altså: et 90 % konfidensintervall for målt motoreffekt blir

$$[149,6 - 2,21, 149,6 + 2,21] = \underline{\underline{[147,4, 151,8]}}$$

b) Sammenhengen mellom utvalgsstørrelsen $n$ og feilmarginen $d$ for et konfidensintervall er gitt ved (fra formelarket)

$$n = \left(\frac{u_{\frac{\alpha}{2}} \sigma}{d}\right)^2,$$

Med de oppgitte dataene får vi følgende anslag for antallet målinger $n$ vi må gjennomføre for at feilmarginen skal bli $d = 1,0$ (HK):

$$\begin{aligned} n &= \left(\frac{u_{\frac{\alpha}{2}}\sigma}{d}\right)^2 \\ &= \left(\frac{u_{0,05}\sigma}{d}\right)^2 \\ &= \left(\frac{1,645 \cdot 3,0}{1,0}\right)^2 \\ &= 24,3 \\ &\approx \underline{\underline{24}} \text{ (vanlig å opphøye oppover)} \end{aligned}$$

Ved å gjøre 24 målinger, får vi en feilmargin i 90 %-konfidensintervallet på 1.0 HK.

c) Vi skal gjennomføre en hypotesetest med signifikansnivå 5 % for å undersøke om dataene gir grunnlag for å hevde at motoreffekten er lavere enn de annonserte 150 HK. Vi formulerer to hypoteser:

$$\begin{aligned} H_0 &: \mu = 150 \\ H_1 &: \mu < 150 \end{aligned}$$

Vi kan bruke hypotesetesten for kjent standardavvik (vi har $\sigma = 3,0$), og testobservatoren er gitt på formelarket:

$$U_0 = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}},$$

der $\bar{X}$ er gjennomsnittet av effektmålingene, $\mu$ er den teoretiske forventningsverdien (som ved nullhypotesen er $\mu = 150$) og $n = 5$. Vi får følgende verdi ved denne observatoren:

$$\begin{aligned} U_0 &= \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \\ &= \frac{149,6 - 150}{\frac{3,0}{\sqrt{5}}} \\ &= \underline{-0,298} \end{aligned}$$

Vi vet at vi forkaster nullhypotesen dersom absoluttverdien av $U_0$ er større en $u_\alpha$, altså dersom $|U_0| < u_\alpha$, som med 5 % signifikansnivå er $u_{0,05} = 1,645$. Vi ser at dette ikke er tilfellet her - derfor beholder vi nullhypotesen. Dette er i tråd med konfidensintervallet vi fant tidligere - dette inneholder verdien 150 HK, og støtter konklusjonen om at vi IKKE kan hevde at motoreffekten er lavere enn dette.

Dataene gir **ikke grunnlag** or å hevde at motoreffekten er lavere enn 150 HK med 5 % signifikansnivå.

d) Vi skal nå gjennomføre en hypotesetest med ukjent standardavvik $\sigma$. Vi bruker da en $t$-test, med testobservator

$$T_0 = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}},$$

der estimatoren $S^2$ er den empiriske variansen:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2,$$

som her blir

$$S^2 = \frac{1}{5-1}\sum_{i=1}^{n}(X_i - 149,6)^2 = \underline{4,947}$$

Det estimerte standardavviket blir altså

$$S = \sqrt{S^2} = \sqrt{4,947} = \underline{2,22}$$

Testobervatoren blir

$$\begin{aligned} T_0 &= \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \\ &= \frac{149,6 - 150}{\frac{2,22}{\sqrt{5}}} \\ &= \underline{-0,403} \end{aligned}$$

Vi forkaster nullhypotesen dersom $|T_0| > t_{\alpha,n-1}$, der $t_{a,n-1}$ er en kvantil i $t$-fordelingen. Med valgt signifikansnivå $\alpha = 0,05$ og $n = 5$, finner vi den aktuelle kvantilen i kvantiltabellen for $t$-fordelingen på formelarket:

$$\begin{aligned} t_{a,n-1} &= t_{0,05,4} \\ &= \underline{2,132}\ (\alpha = 0,05,\ 4 \text{ frihetsgrader}) \end{aligned}$$

Vi ser at vi ikke har grunnlag for å forkaste nullhypotesen, ettersom $|T_0| = 0,403 < t_{a,n-1} = 2,132$ (og ikke større, slik kravet tilsier).

Hypotesetesten med ukjent $\sigma$ og 5 % signifikansnivå gir **ikke grunnlag** for å hevde at motoreffekten er lavere enn 150 HK.

**Oppgave 8**

a) Vi har gitt følgende ti målinger av bruddstyrken til vaieren fra laboratorium nr. 1:

2,2 2,3 2,5 2,1 2,3 2,6 2,2 2,1 2,0 2,3

Gjennomsnittet av målingene blir:

$$\bar{X} = \sum_{i=1}^{10} X_i = \underline{2,26}$$

Målingene fra laboratium nr. 2:

2,8, 3,0, 2,6, 2,6, 2,4

Gjennomsnittet av målingene blir:

$$\bar{Y} = \sum_{i=1}^{5} Y_i = \underline{2,68}$$

b) Vi har gitt to forskjellige estimatorer for bruddstyrken:

$$\hat{u}_1 = \frac{2}{3}\bar{X} + \frac{1}{3}\bar{Y} \text{ og } \hat{u}_2 = \frac{1}{2}\bar{X} + \frac{1}{2}\bar{Y}$$

Vi sjekker om begge er forventningsrette - dvs. at begge estimatorene har forventningsverdi lik $\mu$, som er den ukjente forventningsverdien for bruddstyrken. Her bruker vi at begge seriene med målinger har samme fordeling, dvs. $N\left(\mu, \sigma^2\right)$. Vi bruke regnereglene for forventningsverdier[1] til å sjekke dette:

$$\begin{aligned}
E\left(\hat{u}_1\right) &= E\left(\frac{2}{3}\bar{X} + \frac{1}{3}\bar{Y}\right) \\
&= E\left(\frac{2}{3}\bar{X}\right) + E\left(\frac{1}{3}\bar{Y}\right) \\
&= \frac{2}{3}E\left(\bar{X}\right) + \frac{1}{3}E\left(\bar{Y}\right) \\
&= \frac{2}{3}\cdot\mu + \frac{1}{3}\cdot\mu \\
&= \underline{\underline{\mu}}
\end{aligned}$$

$$\begin{aligned}
E\left(\hat{u}_2\right) &= E\left(\frac{1}{2}\bar{X} + \frac{1}{2}\bar{Y}\right) \\
&= E\left(\frac{1}{2}\bar{X}\right) + E\left(\frac{1}{2}\bar{Y}\right) \\
&= \frac{1}{2}E\left(\bar{X}\right) + \frac{1}{2}E\left(\bar{Y}\right) \\
&= \frac{1}{2}\cdot\mu + \frac{1}{2}\cdot\mu \\
&= \underline{\underline{\mu}}
\end{aligned}$$

Begge estimatorene er altså forventingsrette - dvs. jo flere målinger vi gjør ved hjelp av de, desto nærmere vil vi komme den "sanne" verdien til $\mu$ (som ingen kjenner i utgangspunktet).

c) Ettersom den estimerte bruddstyrken ved metode 2 er $\bar{Y} = 2,68$, blir den estimerte bruddstyrken med de to metodene lik

$$\begin{aligned}
\hat{u}_1 &= \frac{2}{3}\bar{X} + \frac{1}{3}\bar{Y} \\
&= \frac{2}{3}\cdot 2,26 + \frac{1}{3}\cdot 2,68 \\
&= \underline{\underline{2,40}}
\end{aligned}$$

$$\begin{aligned}
\hat{u}_2 &= \frac{1}{2}\bar{X} + \frac{1}{2}\bar{Y} \\
&= \frac{1}{2}\cdot 2,26 + \frac{1}{2}\cdot 2,68 \\
&= \underline{\underline{2,47}}
\end{aligned}$$

Standardavviket med de to metodene finner vi via regnereglene for variansen. Her kommer vi til å få bruk for variansen til et gjennomsnitt:

$$Var\left(\bar{X}\right) = Var\left(\frac{X_1 + X_2 +,,, + X_n}{n}\right) = \frac{1}{n^2}Var\left(X_1 +,,, X_n\right) = \frac{1}{n^2}\cdot n\sigma^2 = \underline{\frac{\sigma^2}{n}}$$

Metode 1:

[1] Husk at forventningsverdien til et gjennomsnitt er lik forventningen til hver av målingene når disse har samme forventning; $E\left(\bar{X}\right) = \mu$

$$
\begin{aligned}
Var\left(\hat{u}_1\right) &= Var\left(\frac{2}{3}\bar{X}+\frac{1}{3}\bar{Y}\right) \\
&= Var\left(\frac{2}{3}\bar{X}\right)+Var\left(\frac{1}{3}\bar{Y}\right) \quad \text{(fordi } \bar{X} \text{ og } \bar{Y} \text{ er uavh,)} \\
&= \left(\frac{2}{3}\right)^2 Var\left(\bar{X}\right)+\left(\frac{1}{3}\right)^2 Var\left(\bar{Y}\right) \quad \text{(regel: } Var\left(aX\right)=a^2 Var\left(X\right)\text{)} \\
&= \left(\frac{2}{3}\right)^2 \frac{\sigma^2}{10}+\left(\frac{1}{3}\right)^2 \frac{\sigma^2}{5} \text{ (har hhv, 10 og 5 målinger for } \bar{X} \text{ og } \bar{Y}\text{)} \\
&= \left(\frac{2}{3}\right)^2 \frac{(0,3)^2}{10}+\left(\frac{1}{3}\right)^2 \frac{(0,3)^2}{5} \\
&= \underline{0,0060}
\end{aligned}
$$

Da er standardavviket for bruddstyrken ved metode 1 lik

$$SD\left(\hat{u}_1\right)=\sqrt{Var\left(\hat{u}_1\right)}=\sqrt{0,0060}=\underline{\underline{0,077}}$$

Metode 2:

$$
\begin{aligned}
Var\left(\hat{u}_2\right) &= Var\left(\frac{1}{2}\bar{X}+\frac{1}{2}\bar{Y}\right) \\
&= Var\left(\frac{1}{2}\bar{X}\right)+Var\left(\frac{1}{2}\bar{Y}\right) \\
&= \left(\frac{1}{2}\right)^2 Var\left(\bar{X}\right)+\left(\frac{1}{2}\right)^2 Var\left(\bar{Y}\right) \\
&= \left(\frac{1}{2}\right)^2 \frac{\sigma^2}{10}+\left(\frac{1}{2}\right)^2 \frac{\sigma^2}{5} \\
&= \left(\frac{1}{2}\right)^2 \frac{(0,3)^2}{10}+\left(\frac{1}{2}\right)^2 \frac{(0,3)^2}{5} \\
&= \underline{0,00675}
\end{aligned}
$$

Da er standardavviket for bruddstyrken ved metode 2 lik

$$SD\left(\hat{u}_2\right)=\sqrt{Var\left(\hat{u}_2\right)}=\sqrt{0,00675}=\underline{\underline{0,082}}$$

Hvilken metode er "best"? Vi ser at estimatoren $\hat{u}_1$ har den minste variansen - dvs. når vi beregner gjennomsnittlig bruddstyrke ved hjelp av denne, blir spriket i dataene minst. På grunn av lavere varians er $\hat{u}_1$ den "beste" estimatoren.