(versjon 02)

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

LØSNING: Oppgavesett nr. 5

"*MAT110 Statistikk 1*", 2020

Oppgave 1: (etterspørsel)

a) Populasjonsmengden er alle potensielle kunder hos McDonalds i Molde i hverdagene fra 16:00 til 18:00.

b) Slik gjennomføres datainnsamlingen:

- observere påfølgende kunder mellom 16:00 og 18:00 på en gitt hverdag.
- Men hvilke hverdager? For at vi skal kunne få et *tilfeldig* utvalg må vi velge 100 hverdager, hvor hver hverdag har like stor sjanse for å bli valgt.
- I tillegg må vi luke ut spesielle dager hvor vi *vet* at etterspørselen enten er merkbart høyere eller lavere enn vanlig. Fridager eller kampanjedager er slike dager. Disse dagene er ikke representative for de dagene vi studerer.

Rent praktisk:

- Første ankommende kunde teller vi ikke, siden vi ikke har en tidligere kunde å regne IAT fra.
- Vi må notere ankomsttiden for neste ankommende gruppe. Vi kan da beregne IAT-tiden x_i for denne gruppen. Deretter må vi registrere hva *hele* gruppen bestilte, og beregne den totale bearbedingstiden y_i for alle ordrene.

c) Regner relativ frekvens f_r for intervallet $[0.0\,,\,0.2)$: [1]

$$\underline{f_r(n_1)} \;=\; \frac{n_1}{n} \;=\; \frac{34}{100} \;=\; \underline{0.34} \tag{1}$$

Tilsvarende for alle andre intervall.

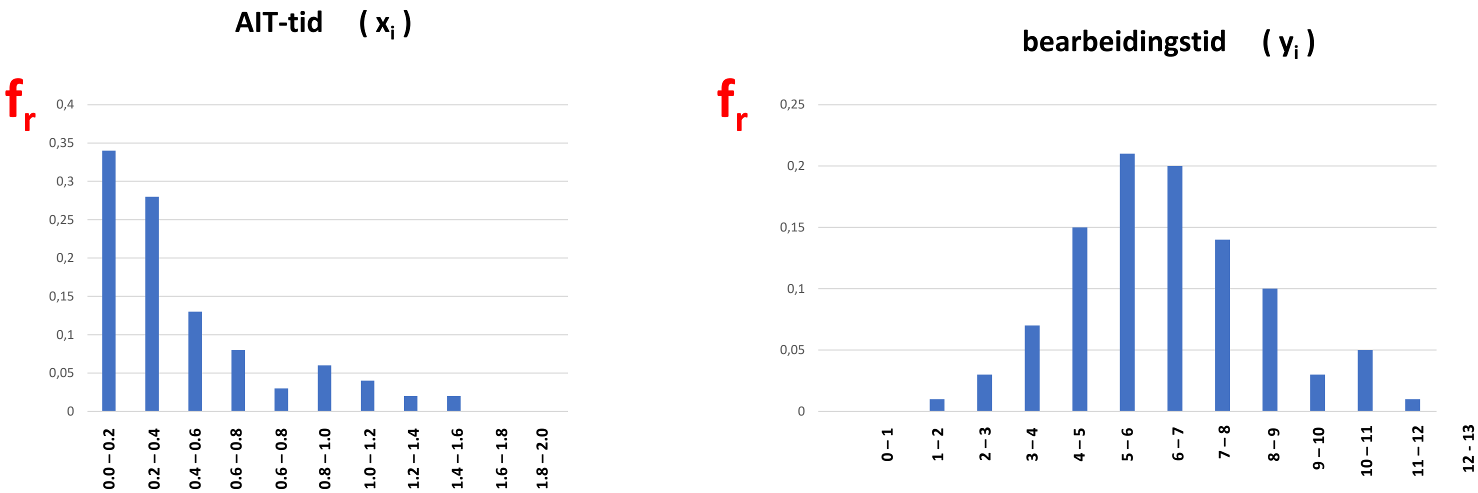
Intervall	**Relativ frekvens** f_r
$[0.0,\ 0.2)$	0.34
$[0.2,\ 0.4)$	0.28
$[0.4,\ 0.6)$	0.13
$[0.6,\ 0.8)$	0.08
$[0,8,\ 1.0)$	0.03
$[1.0,\ 1.2)$	0.06
$[1.2,\ 1.4)$	0.04
$[1.4,\ 1.6)$	0.02
$[1.6,\ 1.8)$	0.02
$[1.8,\ 2.0)$	0.00
$[2.0,\ 2.2)$	0.00
$[2.2,\ 2.4)$	0.00
$[2.4,\ 2.6)$	0.00
$[2.6,\ 2.8)$	0.00
$[2.8,\ 3.0]$	0.00

Tabell 1: f_r for x_i-ene (AIT).

Intervall	**Relativ frekvens** f_r
$[0,\ 1)$	0.00
$[1,\ 2)$	0.01
$[2,\ 3)$	0.03
$[3,\ 4)$	0.07
$[4,\ 5)$	0.15
$[5,\ 6)$	0.21
$[6,\ 7)$	0.20
$[7,\ 8)$	0.14
$[8,\ 9)$	0.10
$[9,\ 10)$	0.03
$[10,\ 11)$	0.05
$[11,\ 12)$	0.01
$[12,\ 13)$	0.00

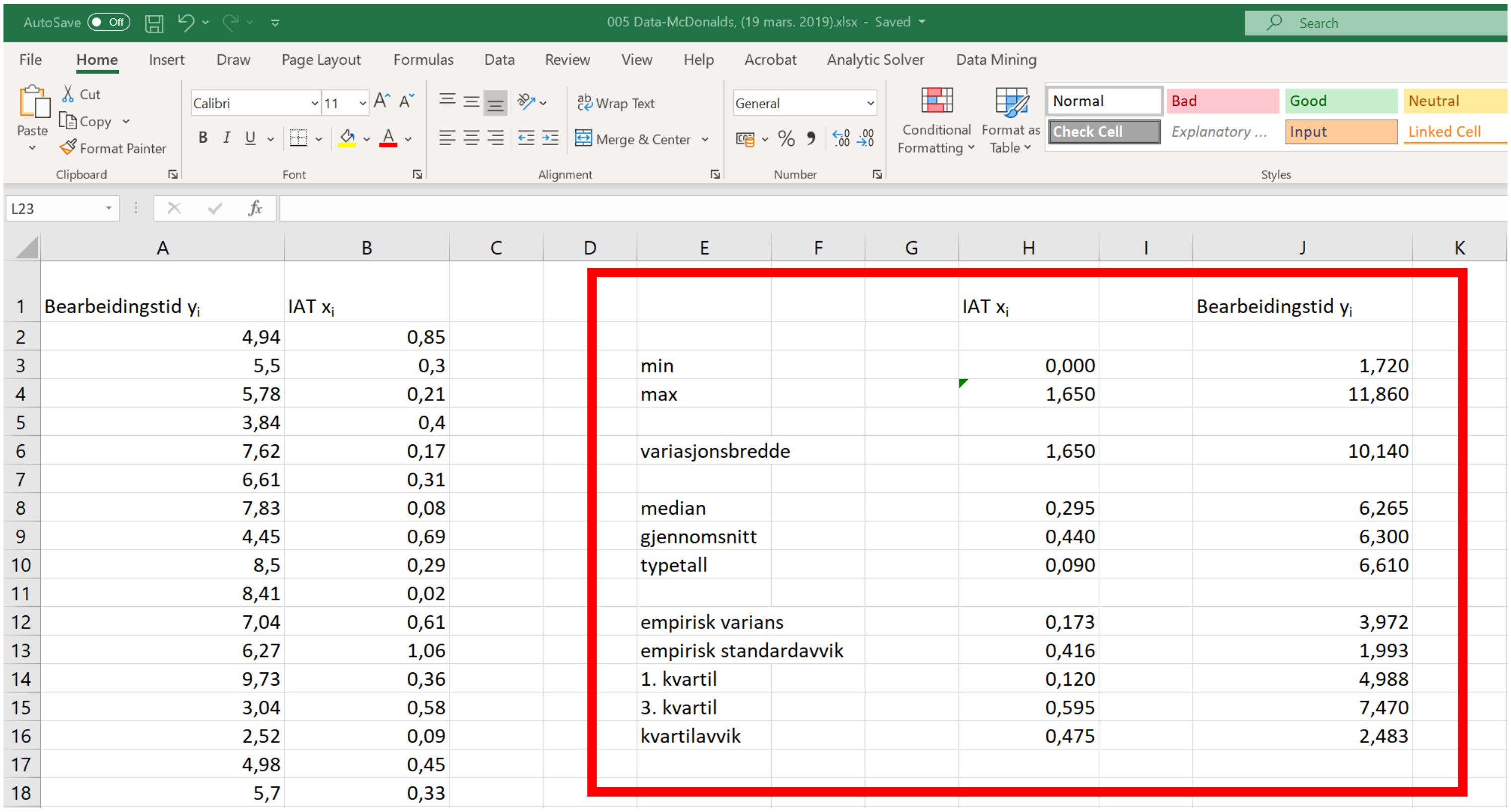
Tabell 2: f_r for y_i-ene (bearbeidingstid).

[1] At det er 34 kunder som har AIT-tid i intervallet $[0.0\,,\,0.2)$ finner man ved å telle. Men bruk gjerne Excel-filen som ligger på Canvas. Sorter, og deretter tell. Eller få Excel til å gjøre tellingen for dere.



Figur 1: f_r-stolpediagram for IAT-tidene (x_i) og bearbeidingstidene (y_i).

d) Bruker Excel-filen som ligger på Canvas. Da kan man regne ut følgende størrelser: [2]



	A	B	E	H	J
1	Bearbeidingstid y_i	IAT x_i		IAT x_i	Bearbeidingstid y_i
2	4,94	0,85			
3	5,5	0,3	min	0,000	1,720
4	5,78	0,21	max	1,650	11,860
5	3,84	0,4			
6	7,62	0,17	variasjonsbredde	1,650	10,140
7	6,61	0,31			
8	7,83	0,08	median	0,295	6,265
9	4,45	0,69	gjennomsnitt	0,440	6,300
10	8,5	0,29	typetall	0,090	6,610
11	8,41	0,02			
12	7,04	0,61	empirisk varians	0,173	3,972
13	6,27	1,06	empirisk standardavvik	0,416	1,993
14	9,73	0,36	1. kvartil	0,120	4,988
15	3,04	0,58	3. kvartil	0,595	7,470
16	2,52	0,09	kvartilavvik	0,475	2,483
17	4,98	0,45			
18	5,7	0,33			

Figur 2: Størrelser som beskriver nøkkeltall.

[2] I dette utklippet fra Excel så er det konsekvent valgt 3 desimaler. Dette kan kan justeres i Excel.

e) Rimelig?:

$$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} X \tag{2}$$

$$Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} Y \tag{3}$$

Identiske fordelte variabler:

At X_i-ene er identisk fordelt lik populasjonsvariabelen X er en greit antagelse å ta siden vi har luket ut alle dagene fra trekningen som *ikke* er representative for de dagene vi ønsker å analysere.

Tilsvarende for Y_i-ene og populasjonsvariabelen Y.

Uavhengige forsøksvariabler X_i og Y_i:

At X_i-ene er gjensidig *uavhengige* betyr at to påfølgende kunder *ikke* er koblet mot hverandre. Et brudd på denne antagelsen fås f.eks. dersom det kommer en vennegjeng. Siden vennene følger hverandre så er inter-arrival tidene den samme, dvs. inter-arrival tidene deres være avhengig av hverandre. Men personer som kommer samlet som *en* kunde. Dermed er uavhengigheten til X_i-ene rimelig å anta.

Tilsvarende for bearbeidingstidene Y_i.

f) X_i-ene er gjensidig uavhengige av Y_i:

Kunden endrer ikke bestillingen etter *når* forrige kunde ankom.
Antagelsen om uavhengighet er dermed rimelig.

g) Statistisk modell for X_i-ene: (inter-arrivel tidene)

Stolpediagrammet for x_i-ene i figur 1 indikerer at tetthetsfunksjonen synker raskt mot null for voksende inter-arrival tider. Vi kan spekulere i at den synker *eksponentielt*.

Vi søker på nettet på ordene "exponential' og "inter arrival time" Her er en link til Wikipedia om eksponentialfordelingen. Vi innser fra nettsiden at eksponential fordelingen er en ofte brukt fordeling for å modellere IAT, hvor kundene ankommer uavhengig av hverandre. Vi foreslår derfor følgende modell: [3]

Den statistiske modellen for de stokastiske forsøksvariablene $X_1, X_2, \ldots, X_{100}$ og populasjonsvariabelen X er gitt ved:

$$\boxed{X_1,\, X_2,\, X_3,\, \ldots\, ,X_n \overset{\text{i.i.d.}}{\sim} X \sim \text{Exp}[\lambda]_{\lambda>0} \qquad (4)}$$

hvor

$$(\,0, \infty\,) \quad = \quad \text{verdimengden } V \text{ til de stokastiske variablene} \tag{5}$$

$$\text{Exp}[\lambda]_{\lambda>0} \quad = \quad \text{familien av eksponential-fordelinger med parameter } \theta = \lambda \tag{6}$$

$$(\,0, \infty\,) \quad = \quad \text{parametermengden } \Theta, \tag{7}$$
$$\text{dvs. de mulige verdiene for parameteren } \theta = \lambda \text{ som er } \lambda > 0$$

[3]Dette med statistisk modell er litt teknisk og litt vanskelig. I kompenidet er det utarbeidet eksempler som er analoge til oppgavene i denne øvingen. Derfor er det både viktig og lurt å gå gjennom teorien først, og deretter gjøre øvingen.

Statistisk modell for Y_i-ene: (bearbeidingstidene)

Stolpediagrammet for y_i-ene i figur 1 indikerer at tetthetsfunksjonen følger en *normalfordeling*. Vi foreslår derfor følgende modell:

Den statistiske modellen for de stokastiske forsøksvariablene $Y_1, Y_2, \ldots, Y_{100}$ og populasjonsvariabelen Y er gitt ved:

$$\boxed{Y_1,\, Y_2,\, Y_3,\, \ldots\, , Y_n \overset{\text{i.i.d.}}{\sim} Y \;\sim\; N[\mu, \sigma]_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_+} \qquad (8)}$$

hvor [4]

$$\overbrace{\mathbb{R}}^{\text{reelle tall}} = \text{verdimengden } V \text{ til de stokastiske variablene} \qquad (9)$$

$$N[\mu, \sigma]_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_+} = \text{familien av normalfordelinger med parametre } \mu \text{ og } \sigma \qquad (10)$$

$$\mathbb{R}\times\mathbb{R}_+ = \text{parametermengden } \Theta, \qquad (11)$$
$$\text{dvs. de mulige verdiene for parameteren } \mu \text{ og } \sigma$$

■

[4] $\mathbb{R}_+$ = alle positive reelle tall.
Elementene i $\mathbb{R}\times\mathbb{R}_+$ er alle par (μ, σ), hvor $-\infty < \mu < \infty$ og $\sigma > 0$.

Oppgave 2: (estimering - logistikk , McDonald's)

a) Forventningen av estimatoren $\hat{\lambda}$:

$$\underline{\underline{E[\,\hat{\lambda}\,]}} \;=\; E[\,\overline{X}\,] \;=\; E\Big[\,\frac{1}{n}\sum_{i=1}^{n} X_i\,\Big] \;=\; \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{n}{n}\,\lambda \;=\; \underline{\underline{\lambda}} \tag{12}$$

altså $\hat{\lambda}$ er forventningsrett.

b) Punktestimat for λ: (realisering - AIT-tid)

$$\underline{\underline{\hat{\lambda}(\,x_1, x_2, \ldots, x_n\,)}} \;=\; \overline{x} \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i \;=\; \underline{\underline{0.440 \text{ minutter}}} \tag{13}$$

hvor vi bruk er at $\overline{x} = 0.440$ minutter fra oppgave **1d**.

Punktestimat for μ: (realisering - forventet bearbeidingstid)

$$\underline{\underline{\hat{\mu}(\, y_1, y_2, \ldots, y_n \,)}} \;=\; \overline{y} \;=\; \frac{1}{n}\sum_{i=1}^{n} y_i \;=\; \underline{\underline{6.300 \text{ minutter}}} \tag{14}$$

hvor vi bruk er at $\overline{y} = 6.300$ minutter fra oppgave **1d**.

Punktestimat for σ: (realisering - standardavvik av bearbeidingstid)

$$\underline{\underline{\hat{\sigma}(\, y_1, y_2, \ldots, y_n \,)}} \;=\; s_y \;=\; \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} \big(\, y_i - \overline{y} \,\big)} \;=\; \underline{\underline{1.993 \text{ minutter}}} \tag{15}$$

hvor vi bruk er at $s_y = 1.993$ minutter fra oppgave **1d**.

c) Asymptotisk $(1-\alpha)\,100\,\%$ konfidensintervall for

$$\hat{\mu} \;=\; \overline{Y} \;=\; \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{16}$$

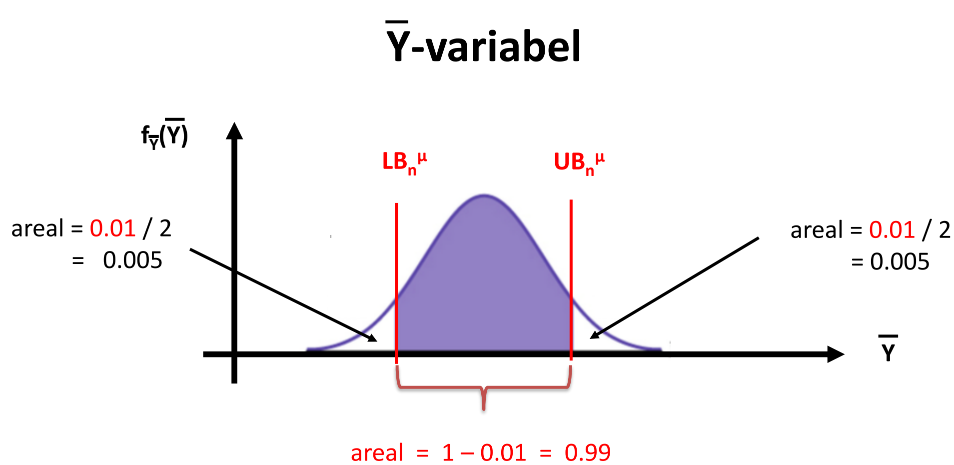
Fra oppgaven vet vi at:

$$Y_1,\, Y_2,\, Y_3,\, \ldots\,, Y_n \;\overset{\text{i.i.d.}}{\sim}\; Y \;\sim\; N[\,\mu\,,\,\sigma\,] \tag{17}$$

som blant annet betyr at $E[Y] = E[Y_i] = \mu$ og at $\sigma^2[Y] = \sigma^2[Y_i] = \sigma^2$.

Definisjonen av et asymptotisk konfidensintervall med $\theta \;=\; \mu$:

$$\boxed{\lim_{n\to\infty} P_\mu(\,LB_n^\mu \leq \mu \leq UB_n^\mu\,) \;=\; 1-\alpha \qquad (18)}$$



Figur 3: 99 %-konfidensintervall for μ.

Alle Y_i i lign.(17) oppfyller i.i.d. (med $Y_i \sim \mathrm{N}[\mu, \sigma]$).
Sum av normalfordelinger er forsatt normalfordelt:

$$\overline{Y} \quad \sim \quad N\big[\, E[\overline{Y}]\,,\, \sigma[\overline{Y}]\,\big] \tag{19}$$

Standardiserer: [5]

$$Z \;=\; \frac{\overline{Y} - E[\overline{Y}]}{\sigma[\overline{Y}]} \;=\; \frac{\overline{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \overset{n\to\infty}{\sim} \quad N[0,1] \tag{23}$$

[5] Estimatoren $\hat{\mu}$ er forventningsrett, dvs.:

$$E[\,\hat{\mu}\,] \;=\; E[\,\overline{Y}\,] \;=\; \mu \tag{20}$$

altså estimatoren $\hat{\mu}$ er forvetningsrett. Variansen til $\hat{\mu}$ er:

$$Var[\,\hat{\mu}\,] \;=\; Var[\,\overline{Y}\,] \;=\; \frac{\sigma^2}{n} \tag{21}$$

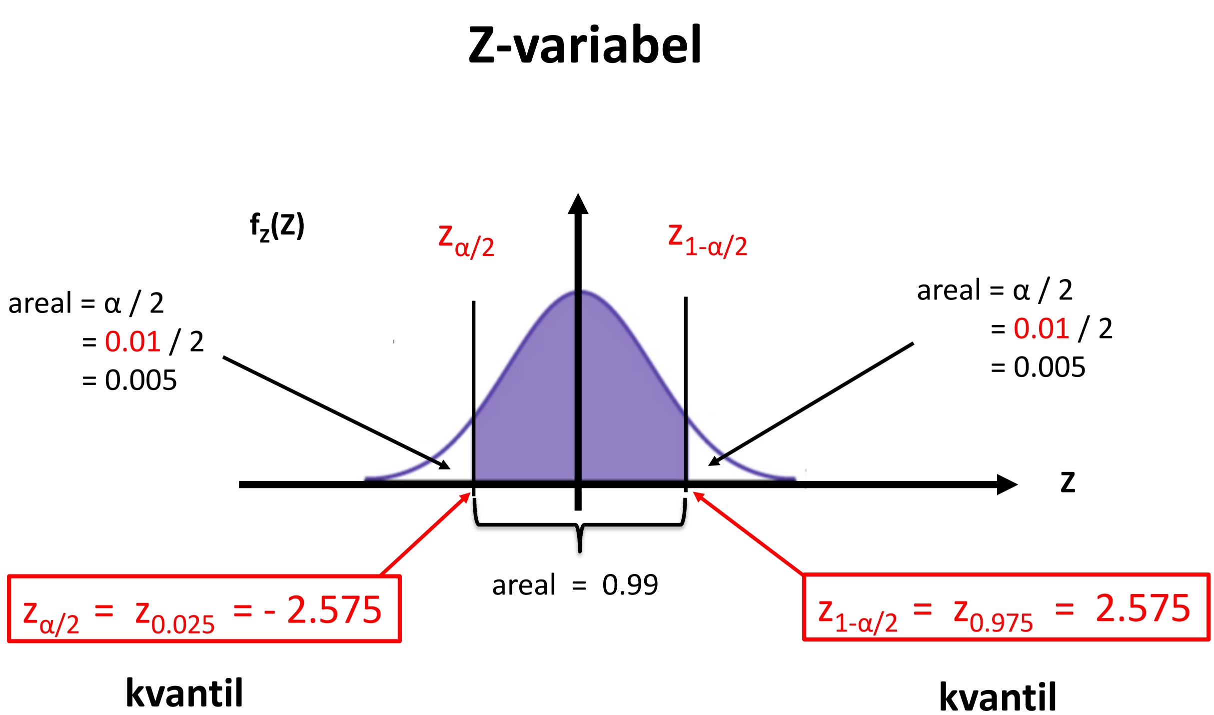
Generelt er $\sigma[Y] = \sqrt{Var[Y]}$, dermed:

$$\sigma[\,\hat{\mu}\,] \;=\; \sigma[\overline{Y}] \;=\; \frac{\sigma}{\sqrt{n}} \tag{22}$$

Kvantilene når $\alpha = 0.001$: (se figur 4) [6]

$$z_{\alpha/2} \;=\; z_{0.005} \;=\; -2.575 \quad , \quad z_{1-\alpha/2} \;=\; z_{0.995} \;=\; 2.575 \tag{24}$$

via omvendt tabelloppslag for normalfordelingen fra kompendiet.



Figur 4: Kvantil.

[6] Absoluttverdiene til $z_{\alpha/2}$ og $z_{1-\alpha/2}$ er like fordi normalfordelingen er symmetrisk.

Fra definisjonen i lign.(18):

$$\lim_{n\to\infty} P_{\mu}(LB_n^{\mu} \leq \mu \leq UB_n^{\mu}) \quad = \quad 1 - \alpha \tag{25}$$

Siden estimatoren $\hat{\mu}$ er normalfordelt, se lign.(19), så standardiserer vi:

$$\lim_{n\to\infty} P\Big(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2} \Big) \quad = \quad 1 - \alpha \tag{26}$$

hvor Z er gitt ved lign.(90):

$$\lim_{n\to\infty} P\left(z_{\alpha/2} \leq \frac{\overline{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2} \right) = 1 - \alpha \tag{27}$$

$$\Updownarrow \qquad \text{(algebra)}$$

$$\lim_{n\to\infty} P\left(\underbrace{\overline{Y} - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sigma}_{=\, LB_n^{\mu}} \leq \mu \leq \underbrace{\overline{Y} - \frac{z_{\alpha/2}}{\sqrt{n}}\sigma}_{=\, UB_n^{\mu}} \right) = 1 - \alpha \tag{28}$$

og hvor vi definerer nedre og øvre grenser:

$$LB_n^{\mu} \quad = \quad \overline{Y} \; - \; \frac{z_{1-\alpha/2}}{\sqrt{n}}\,\sigma \tag{29}$$

$$UB_n^{\mu} \quad = \quad \overline{Y} \; - \; \frac{z_{\alpha/2}}{\sqrt{n}}\,\sigma \tag{30}$$

Med $\hat{\sigma} \;=\; S_y$ innsatt for σ i lign.(29) og (30):

$$LB_n^{\mu} \quad = \quad \overline{Y} \; - \; \frac{z_{1-\alpha/2}}{\sqrt{n}}\,S_y \tag{31}$$

$$UB_n^{\mu} \quad = \quad \overline{Y} \; - \; \frac{z_{\alpha/2}}{\sqrt{n}}\,S_y \tag{32}$$

Intervallet

$$\underline{\underline{[\, LB_n^{\mu}\, ,\, UB_n^{\mu}\,] \;=\; \left[\, \overline{Y} \;-\; \frac{z_{1-\alpha/2}}{\sqrt{n}}\, S_y \;,\; \overline{Y} \;-\; \frac{z_{\alpha/2}}{\sqrt{n}}\, S_y \,\right]}} \tag{33}$$

er dermed et *asymptotisk* $(1-\alpha)\,100\,\%$-konfidensintervall for μ.

Fra løsningen av oppgave **1d**:

$$\overline{y} \quad = \quad 6.300 \text{ minutter} \tag{34}$$

$$s_y \quad = \quad 1.993 \text{ minutter} \tag{35}$$

Realisering: (tall)

$$\underline{LB_n^{\mu}(y_1, y_2, \ldots, y_n)} \quad = \quad \overline{y} \; - \; \frac{z_{1-\alpha/2}}{\sqrt{n}} \, s_y \tag{36}$$

$$= \quad \left(6.300 \; - \; \frac{2.575}{\sqrt{100}} \cdot 1.993 \right) \text{minutter} \tag{37}$$

$$= \quad \underline{5.7868 \text{ minutter}} \tag{38}$$

$$\underline{UB_n^{\mu}(y_1, y_2, \ldots, y_n)} \quad = \quad \overline{y} \; - \; \frac{z_{\alpha/2}}{\sqrt{n}} \, s_y \tag{39}$$

$$= \quad \left(6.300 - \frac{(-2.575)}{\sqrt{100}} \cdot 1.9393 \right) \text{minutter} \tag{40}$$

$$= \quad \underline{6.8132 \text{ minutter}} \tag{41}$$

som gir realiseringen

$$\underline{\underline{[\; LB_n^{\mu} \;,\; UB_n^{\mu} \;] \quad = \quad [\; 5.7868 \;,\; 6.8132 \;] \text{ minutter}}} \tag{42}$$

av det asymptotiske 99 %-konfidensintervallet for μ.

d) Konfidensintervallet finnes på samme måte som i oppgave **1c**, se lign.(33): [7]

$$\underline{\underline{[\,LB_n^{\mu}\,,\,UB_n^{\mu}\,] \;=\; \left[\;\overline{Y} \;-\; \frac{q_{1-\alpha/2}}{\sqrt{n}}\,S_y \;,\; \overline{Y} \;-\; \frac{q_{\alpha/2}}{\sqrt{n}}\,S_y\;\right]}} \tag{43}$$

hvor $q_{\alpha/2}$ og $q_{1-\alpha/2}$ er kvantilene til til Students t-fordelingen.

Kvantilene når $\alpha = 0.001$: (jfr. figur 4) [8]

$$q_{\alpha/2} \;=\; q_{0.005} \;=\; -2.626 \quad , \quad q_{1-\alpha/2} \;=\; z_{0.995} \;=\; 2.626 \tag{44}$$

siden $n = 100$.

Fra løsningen av øving 5 har vi at:

$$\overline{y} \;=\; 6.300 \text{ minutter} \tag{45}$$

$$s_y \;=\; 1.993 \text{ minutter} \tag{46}$$

[7] Finner lign.(43) i kompendiet.

[8] Absoluttverdiene til $q_{\alpha/2}$ og $q_{1-\alpha/2}$ er like fordi Student's t-fordelingen, i likhet med normalfordelingen, er symmetrisk.

Realisering: (tall)

$$\underline{LB_n^{\mu}(y_1, y_2, \ldots, y_n)} \;=\; \overline{y} \;-\; \frac{q_{1-\alpha/2}}{\sqrt{n}}\, s_y \tag{47}$$

$$= \left(6.300 \;-\; \frac{2.626}{\sqrt{100}} \cdot 1.993 \right) \text{ minutter} \tag{48}$$

$$= \underline{5.7766 \text{ minutter}} \tag{49}$$

$$\underline{UB_n^{\mu}(y_1, y_2, \ldots, y_n)} \;=\; \overline{y} \;-\; \frac{q_{\alpha/2}}{\sqrt{n}}\, s_y \tag{50}$$

$$= \left(6.300 - \frac{(-2.626)}{\sqrt{100}} \cdot 1.9393 \right) \text{ minutter} \tag{51}$$

$$= \underline{6.8234 \text{ minutter}} \tag{52}$$

som gir realiseringen

$$\underline{\underline{[\; LB_n^{\mu} \;,\; UB_n^{\mu} \;] \;=\; [\; 5.7766 \;,\; 6.8234 \;] \text{ minutter}}} \tag{53}$$

av det asymptotiske 99 %-konfidensintervallet for μ.

Siden Student's t-fordelingen har tykkere hale enn normalfordelingen
så er kvantilene også større.
Dermed blir også konfidensintervallet større. [9]

[9] Dette med at Student's t-fordelingen har tykkere hale enn normalfordelingen, står i kompendiet. Derfor er det viktig at dere kjenner til teorien i kompendiet før dere starter å løse øvingen.

e) Definisjonen av et asymptotisk konfidensintervall er: (se kompendium)

$$\lim_{n\to\infty} P_{\mu}(LB_n^{\sigma} \leq \sigma \leq UB_n^{\sigma}) \;=\; 1-\alpha \tag{54}$$

Siden Y_i-variablene er $i.i.d.$ så er gjennomsnittet $\overline{Y}$ normalfordelt.
Ved å standardisere og gjøre tilsvarende som i forrige eksempel så får man:

$$\underline{\underline{[\; LB_n^{\sigma} \;,\; UB_n^{\sigma}, \;] \;=\; \left[\; \sqrt{\frac{n-1}{(n-1)+z_{1-\alpha/2}\sqrt{2(n-1)}}}\, S_y \;,\; \sqrt{\frac{n-1}{(n-1)+z_{\alpha/2}\sqrt{2(n-1)}}}\, S_y \;\right]}} \tag{55}$$

Kvantilene når $\alpha = 0.001$: (se figur 4) [10]

$$z_{\alpha/2} \;=\; z_{0.005} \;=\; -2.575 \quad , \quad z_{1-\alpha/2} \;=\; z_{0.995} \;=\; 2.575 \tag{56}$$

via omvendt tabelloppslag fra kompendiet.
Fra oppgave **1d**:

$$s_y \;=\; 1.993 \text{ minutter} \tag{57}$$

[10] Absoluttverdiene til $z_{\alpha/2}$ og $z_{1-\alpha/2}$ er like fordi N-fordelingen er symmetrisk.

Realisering: (tall)

$$\underline{LB_n^{\sigma}(y_1, y_2, \ldots, y_n)} = \sqrt{\frac{n-1}{(n-1) + z_{1-\alpha/2}\sqrt{2(n-1)}}}\, s_y \tag{58}$$

$$= \sqrt{\frac{100-1}{(100-1) + 2.575\sqrt{2(100-1)}}} \cdot 1.993 \tag{59}$$

$$= \underline{1.705 \text{ minutter}} \tag{60}$$

$$\underline{UB_n^{\sigma}(y_1, y_2 \ldots, y_n)} = \sqrt{\frac{n-1}{(n-1) + z_{\alpha/2}\sqrt{2(n-1)}}}\, s_y \tag{61}$$

$$= \sqrt{\frac{100-1}{(100-1) - 2.575\sqrt{2(100-1)}}} \cdot 1.993 \tag{62}$$

$$= \underline{2.503 \text{ minutter}} \tag{63}$$

som gir realiseringen

$$\underline{\underline{[\ LB_n^{\sigma}\ ,\ UB_n^{\sigma}\ ] \ = \ [\ 1.705\ ,\ 2.503\ ] \text{ minutter}}} \tag{64}$$

av det <u>asymptotiske</u> 99 %-konfidensintervallet for standardavviket σ til bearbeidsingtiden Y.

f) Siden $Y_1, Y_2, \ldots, Y_n$ er *i.i.d.* normalfordelte stokastiske variabler, dvs. $Y_i \sim N[\mu, \sigma]$ så er den stokastiske variabelen

$$\frac{(n-1)S_y^2}{\sigma^2} \;\sim\; \chi^2_{n-1} \tag{65}$$

χ^2_{n-1}-fordelt med $n-1$ frihetsgrader.

Vi bruker χ^2_{n-1}-fordelingens α kvantiler, $\chi_{\alpha/2}$ og $\chi_{1-\alpha/2}$, for å konstruere et eksakt 99 %-konfidensintervall for σ:

$$P\left(\chi_{\alpha/2} \;\leq\; \frac{(n-1)S_y^2}{\sigma^2} \;\leq\; \chi_{1-\alpha/2} \right) \;=\; 1-\alpha \tag{66}$$

$$\Updownarrow \qquad \text{(algebra)}$$

$$P\left(\underbrace{\sqrt{\frac{n-1}{\chi_{1-\alpha/2}}}\, S_y}_{=\, LB_n^\sigma} \;\leq\; \sigma \;\leq\; \underbrace{\sqrt{\frac{n-1}{\chi_{\alpha/2}}}\, S_y}_{=\, UB_n^\sigma} \right) \;=\; 1-\alpha \tag{67}$$

hvor vi definerer dermed nedre og øvre grenser:

$$LB_n^\sigma \;=\; \sqrt{\frac{n-1}{\chi_{1-\alpha/2}}}\, S_y \tag{68}$$

$$UB_n^\sigma \;=\; \sqrt{\frac{n-1}{\chi_{\alpha/2}}}\, S_y \tag{69}$$

med

$$\chi_{\alpha/2} \;=\; \text{nedre kvantil til } \chi^2_{n-1}\text{-fordelingen} \tag{70}$$

$$\chi_{1-\alpha/2} \;=\; \text{øvre kvantil til } \chi^2_{n-1}\text{-fordelingen} \tag{71}$$

Intervallet

$$\underline{\underline{[\ LB_n^{\sigma}\,,\,UB_n^{\sigma}\ ]\ =\ \left[\ \sqrt{\frac{n-1}{\chi_{1-\alpha/2}}}\,S_y\ ,\ \sqrt{\frac{n-1}{\chi_{\alpha/2}}}\,S_y\ \right]}} \tag{72}$$

er dermed et <u>eksakt</u> $(1-\alpha)\,100\,\%$-konfidensintervall for σ for effekten Y.

Med $\alpha = 0.01$ og $n = 100$ for bearbeidsingstiden Y så finner vi f.eks. ved å google på nettet: [11] [12]

$$\chi_{\alpha/2} = \chi_{0.005} = 66.5 \tag{73}$$

$$\chi_{1-\alpha/2} = \chi_{0.995} = 139.0 \tag{74}$$

Fra løsningen av øving 5:

$$s_y = 1.993 \text{ minutter} \tag{75}$$

Realisering: (tall)

$$\underline{LB_n^\sigma(y_1, y_2, \ldots, y_n)} = \sqrt{\frac{n-1}{\chi_{1-\alpha/2}}}\, s_y \tag{76}$$

$$= \sqrt{\frac{100-1}{139.0}} \cdot 1.993 \tag{77}$$

$$= \underline{1.6820} \tag{78}$$

$$\underline{UB_n^\sigma(y_1, y_2, \ldots, y_n)} = \sqrt{\frac{n-1}{\chi_{\alpha/2}}}\, s_y \tag{79}$$

$$= \sqrt{\frac{100-1}{66.5}} \cdot 1.993 \tag{80}$$

$$= \underline{2.4317} \tag{81}$$

[11] Kvantilene $\chi_{\alpha/2}$ og $\chi_{1-\alpha/2}$ er forskjellige fordi χ-fordelingen ikke er symmetrisk, se f.eks. figur i kompendiet eller google på internett.

[12] Ved googling på internett så finner man mange nettsteder hvor man finner de aktuelle kvantilene, f.eks. https://keisan.casio.com/exec/system/1180573197. Husk at $n = 100$.

som gir realiseringen

$$\underline{\underline{[\ LB_n^{\sigma}\ ,\ UB_n^{\sigma}\ ] \quad = \quad [\ 1.6820\ ,\ 2.4317\ ]}} \tag{82}$$

av det <u>eksakte</u> 99 %-konfidensintervallet for standardavviket σ til bearbeidingstiden Y.

g) Asymptotisk $(1-\alpha)\,100\,\%$ konfidensintervall for

$$\hat{\lambda} \;=\; \overline{X} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \tag{83}$$

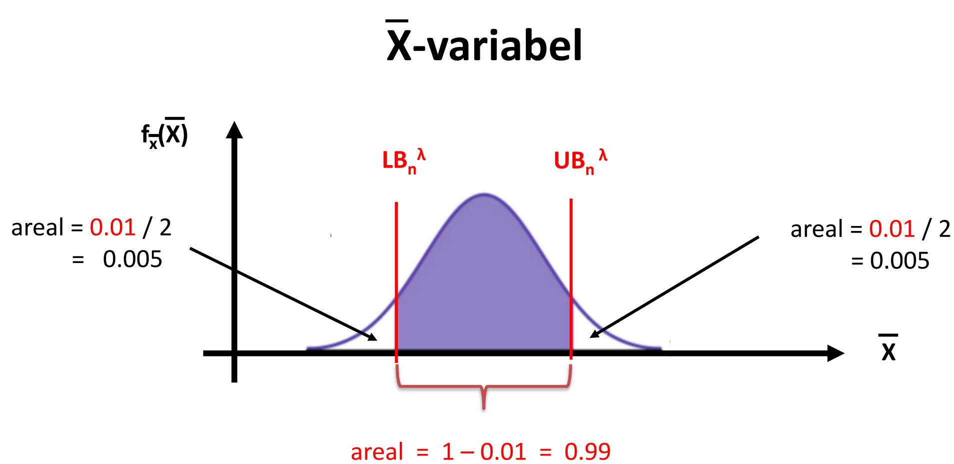
Fra oppgaven vet vi at:

$$X_1,\, X_2,\, X_3,\, \ldots\,,X_n \;\overset{\text{i.i.d.}}{\sim}\; X \;\sim\; \text{Exp}[\,\lambda\,] \tag{84}$$

I fotnoten i oppgaveteksten var det oppgitt att $E[X] = E[X_i] = \lambda$ og at $\sigma^2[X] = \sigma^2[X_i] = \lambda^2$.

Definisjonen av et asymptotisk konfidensintervall med $\theta \;=\; \lambda$:

$$\boxed{\lim_{n\to\infty} P_{\lambda}(\,LB_n^{\lambda} \leq \lambda \leq UB_n^{\lambda}\,) \;=\; 1-\alpha \qquad (85)}$$



Figur 5: Asymptotisk 99 %-konfidensintervall for λ.

Siden alle X_i i lign.(84) oppfyller i.i.d. (med $X_i \sim \text{Exp}[\lambda]$), får vi fra sentralgrenseteoremet at

$$\overline{X} \quad \overset{n\to\infty}{\sim} \quad N\left[\, E[\overline{X}]\, ,\, \sigma[\overline{X}]\,\right] \tag{86}$$

Standardiserer: [13]

$$\underline{Z} \;=\; \frac{\overline{X} - E[\overline{X}]}{\sigma[\overline{X}]} \;=\; \underline{\frac{\overline{X} - \lambda}{\frac{\lambda}{\sqrt{n}}}} \quad \overset{n\to\infty}{\sim} \quad N[0,1] \tag{90}$$

[13]Estimatoren $\hat{\lambda}$ er forventningsrett, dvs.:

$$E[\,\hat{\lambda}\,] \;=\; E[\,\overline{X}\,] \;=\; \lambda \tag{87}$$

Variansen til $\hat{\lambda}$ er:

$$Var[\,\hat{\lambda}\,] \;=\; Var[\,\overline{X}\,] \;=\; \frac{\lambda^2}{n} \tag{88}$$

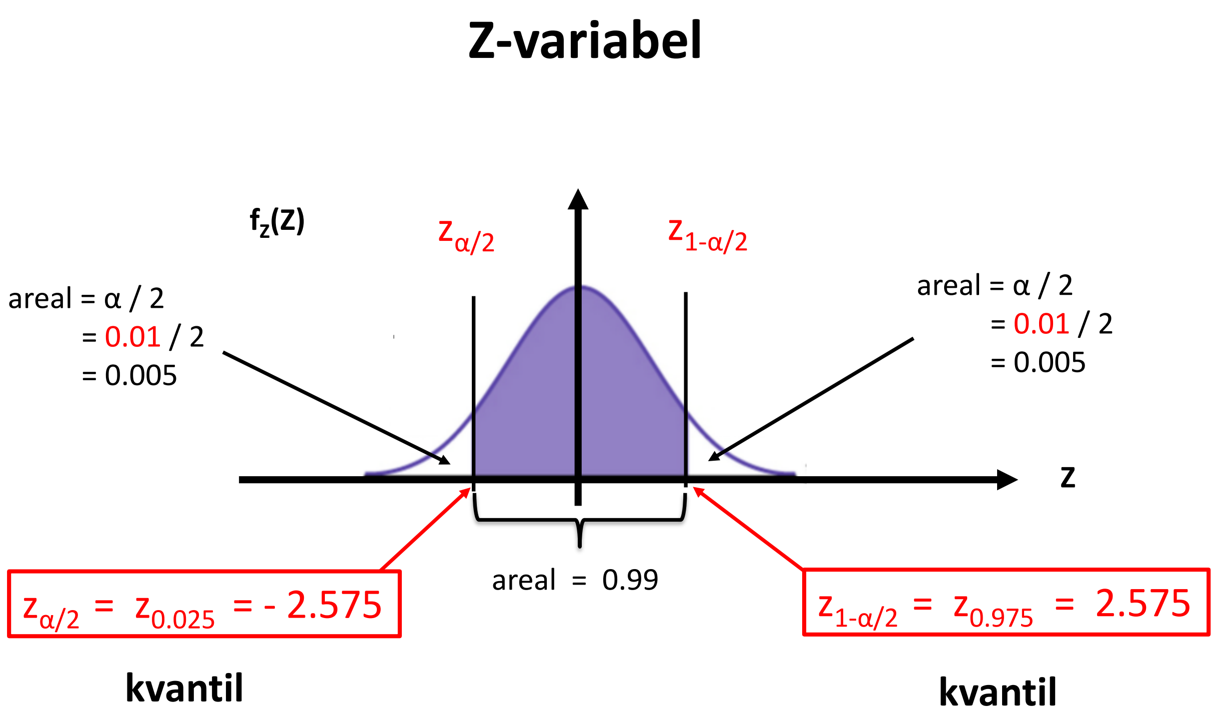
Generelt er $\sigma[X] = \sqrt{Var[X]}$, dermed:

$$\sigma[\,\hat{\lambda}\,] \;=\; \sigma[\overline{X}] \;=\; \frac{\lambda}{\sqrt{n}} \tag{89}$$

Kvantilene når $\alpha = 0.001$: (se figur 6) [14]

$$z_{\alpha/2} \;=\; z_{0.005} \;=\; -2.575 \quad , \quad z_{1-\alpha/2} \;=\; z_{0.995} \;=\; 2.575 \tag{91}$$

via omvendt tabelloppslag for normalfordelingen fra kompendiet.



Figur 6: Kvantil.

[14]Absoluttverdiene til $z_{\alpha/2}$ og $z_{1-\alpha/2}$ er like fordi normalfordelingen er symmetrisk.

Fra definisjonen i lign.(85):

$$\lim_{n\to\infty} P_\lambda(\, LB_n^\lambda \;\leq\; \lambda \leq UB_n^\lambda\,) \;\;=\;\; 1-\alpha \tag{92}$$

Siden estimatoren $\hat{\lambda}$ er normalfordelt, se lign.(86), så standardiserer vi:

$$\lim_{n\to\infty} P\big(\, z_{\alpha/2} \;\leq\; Z \leq z_{1-\alpha/2}\,\big) \;\;=\;\; 1-\alpha \tag{93}$$

hvor Z er gitt ved lign.(90):

$$\lim_{n\to\infty} P\left(\; z_{\alpha/2} \;\leq\; \frac{\overline{X}-\lambda}{\frac{\lambda}{\sqrt{n}}} \;\leq\; z_{1-\alpha/2} \right) \;=\; 1-\alpha \tag{94}$$

$$\Updownarrow \qquad \text{(algebra)}$$

$$\lim_{n\to\infty} P\Bigg(\; \underbrace{\overline{X} - \frac{z_{1-\alpha/2}}{\sqrt{n}}\,\lambda}_{=\;LB_n^\lambda} \;\leq\; \lambda \;\leq\; \underbrace{\overline{X} - \frac{z_{\alpha/2}}{\sqrt{n}}\,\lambda}_{=\;UB_n^\lambda} \Bigg) \;=\; 1-\alpha \tag{95}$$

og hvor vi definerer nedre og øvre grenser:

$$LB_n^\lambda \;\;=\;\; \overline{X} \;-\; \frac{z_{1-\alpha/2}}{\sqrt{n}}\,\lambda \tag{96}$$

$$UB_n^\lambda \;\;=\;\; \overline{X} \;-\; \frac{z_{\alpha/2}}{\sqrt{n}}\,\lambda \tag{97}$$

Med $\hat{\lambda} \;=\; \overline{X}$ innsatt for λ i lign.(96) og (97):

$$LB_n^\lambda \;\;=\;\; \overline{X} \;-\; \frac{z_{1-\alpha/2}}{\sqrt{n}}\,\overline{X} \tag{98}$$

$$UB_n^\lambda \;\;=\;\; \overline{X} \;-\; \frac{z_{\alpha/2}}{\sqrt{n}}\,\overline{X} \tag{99}$$

Intervallet

$$\underline{\underline{[\,LB_n^{\lambda}\,,\,UB_n^{\lambda}\,] \;=\; \left[\,\overline{X} \;-\; \frac{z_{1-\alpha/2}}{\sqrt{n}}\,\overline{X}\;,\;\overline{X} \;-\; \frac{z_{\alpha/2}}{\sqrt{n}}\,\overline{X}\,\right]}} \tag{100}$$

er dermed et *asymptotisk* $(1-\alpha)\,100\,\%$-konfidensintervall for λ.

Fra oppgave **1d**:

$$\overline{x} \quad = \quad 0.440 \text{ minutter} \tag{101}$$

Realisering: (tall)

$$\underline{LB_n^\lambda(x_1, x_2, \ldots, x_n)} \quad = \quad \overline{x} \; - \; \frac{z_{1-\alpha/2}}{\sqrt{n}}\,\overline{x} \tag{102}$$

$$= \quad \left(0.440 \; - \; \frac{2.575}{\sqrt{100}} \cdot 0.440 \right) \text{ minutter} \tag{103}$$

$$= \quad \underline{0.326 \text{ minutter}} \tag{104}$$

$$\underline{UB_n^\lambda(x_1, x_2, \ldots, x_n)} \quad = \quad \overline{x} \; - \; \frac{z_{\alpha/2}}{\sqrt{n}}\,\overline{x} \tag{105}$$

$$= \quad \left(0.440 - \frac{(-2.575)}{\sqrt{100}} \cdot 0.440 \right) \text{ minutter} \tag{106}$$

$$= \quad \underline{0.552 \text{ minutter}} \tag{107}$$

som gir realiseringen

$$\underline{\underline{[\; LB_n^\lambda \;,\; UB_n^\lambda \;] \quad = \quad [\; 0.327 \;,\; 0.553 \;] \text{ minutter}}} \tag{108}$$

av det asymptotiske 99 %-konfidensintervallet for λ.

$\blacksquare$