Løsning eksamen 8. september 2016

Oppgave 1

a)

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_x xP(X=x) = 0\cdot 0.13 + 1\cdot 0.55 + 2\cdot 0.31 + 3\cdot 0.01 = \underline{\underline{1.2}} \\
\mathrm{E}(X^2) &= \sum_x x^2 P(X=x) = 0^2\cdot 0.13 + 1^2\cdot 0.55 + 2^2\cdot 0.31 + 3^3\cdot 0.01 = 1.88 \\
\mathrm{Var}(X) &= \mathrm{E}(X^2) - \mathrm{E}(X)^2 = 1.88 - 1.2^2 = \underline{\underline{0.44}} \\
P(X_1+X_2=3) &= P(X_1=3\cap X_2=0) + P(X_1=2\cap X_2=1) \\
&\quad + P(X_1=1\cap X_2=2) + P(X_1=0\cap X_2=3) \\
&\overset{uavh}{=} 2\cdot 0.13\cdot 0.01 + 2\cdot 0.55\cdot 0.31 = \underline{\underline{0.34}}
\end{aligned}
$$

b)

$$
\begin{aligned}
\mathrm{E}(Y) &= \sum_{i=1}^{40}\mathrm{E}(X_i) = \sum_{i=1}^{40} 1.2 = 40\cdot 1.2 = \underline{\underline{48}} \\
\mathrm{Var}(Y) &= \mathrm{Var}(\sum_{i=1}^{40} X_i) \overset{uavh.}{=} \sum_{i=1}^{40}\mathrm{Var}(X_i) = \sum_{i=1}^{40} 0.44 = 40\cdot 0.44 = \underline{\underline{17.6}} \\
P(Y\le 50) &= P(\frac{Y-48}{\sqrt{17.6}} \le \frac{50+0.5-48}{\sqrt{17.6}}) \approx P(Z\le 0.60) = \underline{\underline{0.73}} \\
P(Y\le a) &= 0.90 \\
P(\frac{Y-48}{\sqrt{17.6}} \le \frac{a+0.5-48}{\sqrt{17.6}}) &= 0.90 \\
\Rightarrow \quad \frac{a+0.5-48}{\sqrt{17.6}} &\approx 1.28 \quad \Rightarrow \quad a = 48 - 0.5 + 1.28\cdot\sqrt{17.6} = 52.9
\end{aligned}
$$

Dvs minst $\underline{\underline{53}}$ parkeringsplasser. Eventuelt kan man finne dette ved prøving og feiling ved å regne ut $P(Y\le a)$ for a lik 51, 52, 53 og vil da se at første gang sannsynligheten er større enn 0.90 er når $a=53$. (Om man ikke bruker heltallskorreksjonen +0.5 blir $P(Y\le 50) = 0.68$ og $a = 53.4$, dvs 54 plasser.)

Oppgave 2

La T være levetiden. Med $\lambda = 0.08$ blir sannsynlighetstettheten $f(t) = 0.08e^{-0.08t}$ for $t>0$ (og $f(t)=0$ ellers).

a)

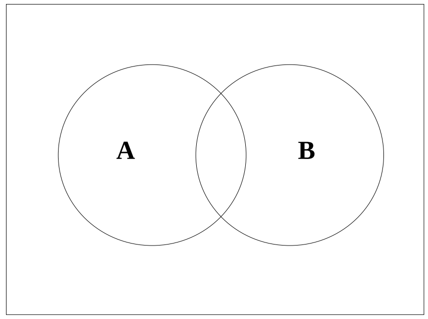
$$
\begin{aligned}
P(T<2) &= \int_0^2 0.08e^{-0.08t}dt = [-e^{-0.08t}]_0^2 = -e^{-0.08\cdot 2} - (-e^{-0.08\cdot 0}) = \underline{\underline{0.15}} \\
\mathrm{E}(T) &= 1/\lambda = 1/0.08 = \underline{\underline{12.5}} \\
P(10<T<20) &= \int_{10}^{20} 0.08e^{-0.08t}dt = [-e^{-0.08t}]_{10}^{20} = -e^{-0.08\cdot 20} - (-e^{-0.08\cdot 10}) = \underline{\underline{0.25}}
\end{aligned}
$$

Eventuelt kan man bruke den kumulative fordelingsfunksjonen til å beregne sannsynlighetene:

$F(t) = P(T<t) = 1 - e^{-\lambda t} = 1 - e^{-0.08t}$ som gir

$$
\begin{aligned}
P(T<2) &= F(2) = 1 - e^{-0.08\cdot 2} = \underline{\underline{0.15}} \\
P(10<T<20) &= P(T<20) - P(T<10) = F(20) - F(10) = 1 - e^{-0.08\cdot 20} - (1 - e^{-0.08\cdot 10}) = \underline{\underline{0.25}}
\end{aligned}
$$

b)



Pga uavhengighet er $P(A \cap B) = P(A) \cdot P(B) = 0.15 \cdot 0.15 = \underline{\underline{0.0225}}$. Dette er sannsynligheten for at begge instrumentene feiler i løpet av garantitiden.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.15 + 0.15 - 0.0225 = \underline{\underline{0.28}}$. Dette er sannsynligheten for at minst ett av instrumentene feiler i løpet av garantitiden.

La X være antallet av 10 instrumenter som feiler i løpet av garantitiden. Vi har en situasjon av typen:

- Gjentatte delforsøk som gir "suksess"/ikke "suksess" - flere instrument som feiler i løpet av to år eller ikke.
- Lik sannsynlighet p i alle delforsøk - samme sannsynlighet $p = 0.15$ for alle instrument for å feile.
- Uavhengige delforsøk - uavhengig fra instrument til instrument om det feiler eller ikke.
- Et bestemt antall, n, delforsøk - et bestemt antall $n = 10$ instrumenter.

Dvs $X \sim Bin(10, 0.15)$ og:

$$
\begin{aligned}
P(X \geq 2) &= 1 - P(X < 2) = 1 - (P(X = 0) + P(X = 1)) \\
&= 1 - (\binom{10}{0} 0.15^0 (1 - 0.15)^{10-0} + \binom{10}{1} 0.15^1 (1 - 0.15)^{10-1}) = 1 - (0.1969 + 0.3474) = \underline{\underline{0.46}}
\end{aligned}
$$

Oppgave 3

a) Modell: $Y_{ij} = \mu_i + e_{ij},$ der e_{ij} uavh. $\mathrm{N}(0, \sigma)$ og der μ_i er forventa bruddstyrke for solceller fra produsent i og e_{ij} er feilleddet (tilfeldig variasjon).

Dvs, vi antar normalfordeling og samme varians i hver gruppe, og uavhengighet mellom målingene.

(Alternativ formulering av modellen: $Y_{ij} = \mu + \alpha_i + e_{ij},$ der e_{ij} uavh. $\mathrm{N}(0, \sigma)$ der μ er gjennomsnittlig forventa levetid og α_i er effekten av produsent i.)

$$H_0: \ \mu_A = \mu_B = \mu_C = \mu_D \qquad \text{mot} \qquad H_1: \ \text{minst en } \mu_i \text{ ulik de andre}$$

Vi har fra pensum/formelarket at vi baserer denne testen på at under nullhypotesen er

$$F = \frac{SS_G/(k-1)}{SS_E/(N-k)} \sim F(\alpha, k-1, N-k)$$

og vi forkaster nullhypotesen dersom F blir stor. Fra den oppgitte variansanalysetabellen ser vi at p-verdien for testen er $0.0024 < 0.05$, dvs $\underline{\underline{\text{vi forkaster } H_0 \text{ på } 5\% \text{ nivå}}}$. Forventet bruddstyrke er ulik for solceller fra de ulike produsentene.

I en variansanalysemodell er residualet definert som: $\underline{\underline{\epsilon_{ij} = y_{ij} - \hat{\mu}_i = y_{ij} - \bar{y}_i}}$. Fra plottet av residualene mot gruppe (produsent) kan vi sjekke om antagelsen om like varians i hver gruppe ser ok ut - dette ser ut til å være en rimelig antagelse her. I tillegg burde vi hatt et histogram eller normalplott for å vurdere om antagelsen om normalfordeling ser ut til å være oppfylt. Dersom rekkefølgen målingene er utført i var kjent kunne vi også plotte residualene mot rekkefølge for å se om der er avhengighet mellom etterfølgende målinger.

b)

$$H_0: \ \mu_A = \mu_B \qquad \text{mot} \qquad H_1: \ \mu_A \neq \mu_B$$

Siden vi har en uparet sammenligning av to utvalg med ukjent varians baserer vi testen på (pensum/formelark)

$$T = \frac{\bar{Y}_A - \bar{Y}_B}{S_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \sim t(n_A + n_B - 2)$$

Med nivå 5%, dvs $\alpha = 0.05$, tosidig test og $n_A = n_B = 4$ forkaster vi H_0 dersom $T \leq -t_{0.025,6} = -2.447$ eller dersom $T \geq t_{0.025,6} = 2.447$.

For å regne ut testobservatoren må vi først regne ut s_p:

$$s_p^2 \quad = \quad \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} = \frac{3 \cdot 0.29 + 3 \cdot 0.42}{4 + 4 - 2} = 0.355 \quad \Rightarrow \quad s_p = \sqrt{0.355} = 0.60$$

Observert verdi på testobservatoren blir da: $t = \frac{6.41 - 4.19}{0.60\sqrt{\frac{1}{4}+\frac{1}{4}}} = 5.23$

Siden 5.23.63 > 2.447 blir konklusjonen at vi $\underline{\underline{\text{forkaster } H_0}}$. Dataene gir grunnlag for å konkludere at det er forskjell i forventet bruddstryke for solceller fra produsent A og B. (Og vi ser fra de gjennomsnittlige bruddstyrkene at det er solcellene fra produsent A som har høyest bruddstyrke.)

Vi antar $\underline{\underline{\text{normalfordeling,}}}$ $\underline{\underline{\text{like varians i begge gruppene}}}$ og $\underline{\underline{\text{uavhengige målinger.}}}$

Oppgave 4

a)

$$\begin{aligned}
P(X > 80) &= 1 - P(\frac{X - 77}{2.4} < \frac{80 - 77}{2.4}) = 1 - P(Z < 1.25) = 1 - 0.8944 = \underline{\underline{0.106}} \\
P(75 < X < 80) &= P(X < 80) - P(X < 75) = P(\frac{X - 77}{2.4} < \frac{80 - 77}{2.4}) - P(\frac{X - 77}{2.4} < \frac{75 - 77}{2.4}) \\
&= P(Z < 1.25) - P(Z < -0.83) = 0.8944 - 0.2033 = \underline{\underline{0.691}} \\
P(X > 80 | X > 75) &= \frac{P(X > 80 \cap X > 75)}{P(X > 75)} = \frac{P(X > 80)}{P(X > 75)} = \frac{0.1056}{1 - 0.2033} = \underline{\underline{0.132}} \\
P(\overline{X} > 80) &= 1 - P(Z < \frac{80 - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{Var}(\overline{X})}}) = 1 - P(Z < \frac{80 - \mu}{\sqrt{\sigma^2/n}}) = P(Z < \frac{80 - 77}{\sqrt{2.4^2/5}}) \\
&= P(Z < 2.80) = 1 - 0.9974 = \underline{\underline{0.0026}}
\end{aligned}$$

Gjennomsnittet av målinger har lavere varians enn enkeltmålinger, og det er derfor mindre sannsynlig at gjennomsnittet av fem målinger avviker så mye fra forventningsverdien enn at en enkeltmåling gjør det.

b)

$$\hat{\mu} = \overline{X}$$

$$Z = \frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{Var}(\overline{X})}} = \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1)$$

$$\begin{aligned}
P(-1.96 \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq 1.96) &= 0.95 \\
P(-1.96\frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu \leq 1.96\frac{\sigma}{\sqrt{n}}) &= 0.95 \\
P(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}) &= 0.95
\end{aligned}$$

Dvs et 95% konfidensintervall for μ er gitt ved:

$$\underline{\underline{[\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}]}}$$

Innsatt $\overline{x} = 67.2$, $n = 9$ og $\sigma = 2.4$ gir dette 95% konfidensintervall for μ:

$$[67.2 - 1.96\frac{2.4}{\sqrt{9}}, 67.2 + 1.96\frac{2.4}{\sqrt{9}}] = \underline{\underline{[65.6, 68.8]}}$$

Lengden til konfidensintervallet er:

$$\begin{aligned} L &= \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - (\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \\ \text{Dvs: } L \leq 2 &\Rightarrow 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq 2 \\ & \quad z_{\alpha/2}\sigma \leq \sqrt{n} \\ & \quad n \geq (z_{\alpha/2}\sigma)^2 = (1.96 \cdot 2.4)^2 = 22.1 \end{aligned}$$

Dvs det må gjøres minst $\underline{\underline{23}}$ målinger for å få et konfidensintervall med ønsket lengde.

c) Husk først at med $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ der $X_1, \ldots X_n$ er uavhengige med $\mathrm{E}(X) = \mu$ og $\mathrm{Var}(X) = \sigma^2$ så vil $\mathrm{E}(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$ og $\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \sigma^2/n$. (Dette har vi også brukt tidligere i oppgaven og det står bl.a. nederst på side 2 av formelarkene.) Vi får da:

$$\begin{aligned} \mathrm{E}(\hat{\mu}_1) &= \frac{1}{2}\mathrm{E}(\bar{X}) + \frac{1}{2}\mathrm{E}(Y) = \frac{1}{2}\mu + \frac{1}{2}\mu = \underline{\underline{\mu}} \\ \mathrm{E}(\hat{\mu}_2) &= 0.36 \cdot \mathrm{E}(\bar{X}) + 0.64\mathrm{E}(Y) = 0.36\mu + 0.64\mu = \underline{\underline{\mu}} \\ \mathrm{Var}(\hat{\mu}_1) &= \mathrm{Var}(\frac{1}{2}\bar{X} + \frac{1}{2}Y) \overset{uavh}{=} \frac{1}{2^2}\mathrm{Var}(\bar{X}) + \frac{1}{2^2}\mathrm{Var}(Y) = \frac{1}{2^2} \cdot \frac{2.4^2}{9} + \frac{1}{2^2} \cdot 0.6^2 = \underline{\underline{0.25}} \\ \mathrm{Var}(\hat{\mu}_2) &= \mathrm{Var}(0.36\bar{X} + 0.64\bar{Y}) \overset{uavh}{=} 0.36^2\mathrm{Var}(\bar{X}) + 0.64^2\mathrm{Var}(Y) = 0.36^2 \cdot \frac{2.4^2}{9} + 0.64^2 \cdot 0.6^2 = \underline{\underline{0.23}} \end{aligned}$$

(Utregningen av $\mathrm{Var}(\hat{\mu}_2)$ ble ikke spurt om, den er bare tatt med her for illustrasjonens skyld.)

d) Estimatoren $\underline{\underline{\hat{\mu}_2 \text{ er best}}}$ siden den har minst varians og begge estimatorene er forventningsrette.

Estimat for μ: $\quad \hat{\mu}_2 = 0.36 \cdot 67.2 + 0.64 \cdot 66.4 = \underline{\underline{66.7}}$.

Siden $\hat{\mu}_2 = 0.64 \cdot \bar{X} + 0.36 \cdot Y$ er en lineærkombinasjon av normalfordelte variable er $\hat{\mu}_2$ normalfordelt. Siden $\hat{\mu}_2$ også er forventningsrett har vi (gjør samme type utledning som i punkt b) eller se notatene til kap. 6 eller regel 6.8 i boka) at et $(1 - \alpha)100\%$ konfidensintervall for μ er gitt ved

$$[\hat{\mu}_2 - z_{\alpha/2}\mathrm{SD}(\hat{\mu}_2), \hat{\mu}_2 + z_{\alpha/2}\mathrm{SD}(\hat{\mu}_2)]$$

Med $\alpha = 0.05$ har vi $z_{\alpha/2} = 1.96$. Videre sette vi inn $\hat{\mu}_2 = 66.9$ og $\mathrm{SD}(\hat{\mu}_2) = \sqrt{\mathrm{Var}(\hat{\mu}_2)} = \sqrt{0.23} = 0.480$, og vi får da 95% konfidensintervallet for μ:

$$[66.7 - 1.96 \cdot 0.480, 66.7 + 1.96 \cdot 0.480] = \underline{\underline{[65.8, 67.6]}}$$

Merk at vi nå har fått et konfidensintervall som har lengde mindre enn 2 med bare en ekstra måling siden den ekstra målingen ble utført med et instrument med vesentlig bedre presisjon.