Løsning eksamen 13. mai 2015

Oppgave 1:

a)

$$\begin{aligned}
P(X<8) &= P\left(\frac{X-8.7}{0.35}<\frac{8-8.7}{0.35}\right)=P(Z<-2)=\underline{\underline{0.0228}}\\
P(8<X<9) &= P(X<9)-P(X<8)=P\left(\frac{X-8.7}{0.35}<\frac{9-8.7}{0.35}\right)-P\left(\frac{X-8.7}{0.35}<\frac{8-8.7}{0.35}\right)\\
&= P(Z<0.86)-P(Z<-2)=0.8051-0.0228=\underline{\underline{0.782}}\\
P(X>b) &= 0.99\\
P(X<b) &= P\left(Z<\frac{b-8.7}{0.35}\right)=0.01 \quad\Rightarrow\quad \frac{b-8.7}{0.35}=-2.33 \quad b=-2.33\cdot 0.35+8.7=\underline{\underline{7.9}}
\end{aligned}$$

Det siste svaret kan man sjekke ved å regne ut om det stemmer at $P(X>7.9)=0.99$.

b) Vi har uavhengige normalfordelte målinger med kjent standardavvik $\sigma=0.35$ og skal teste:

$$H_0: \ \mu \le 8.7 \qquad \text{mot} \qquad H_1: \ \mu > 8.7$$

Estimator: $\hat{\mu}=\bar{X}$. Siden σ kjent har vi dersom H_0 er korrekt følgende nullfordeling:

$$Z=\frac{\overline{X}-\mathrm{E}(\overline{X})}{\sqrt{\mathrm{Var}(\overline{X})}}=\frac{\overline{X}-\mu_0}{\sqrt{\sigma^2/n}}=\frac{\overline{X}-8.7}{0.35/\sqrt{n}}\sim \mathrm{N}(0,1)$$

Med signifikansnivå $\alpha=0.05$ forkaster vi H_0 dersom $Z\ge z_{0.05}=1.645$. Observerte data gir:

$$z_{obs}=\frac{8.91-8.7}{0.35/\sqrt{5}}=1.34<1.645.$$

Dvs. utfallet er ikke i forkastningsområdet. Konklusjon: Vi $\underline{\underline{\text{forkaster ikke } H_0}}$. Vi kan ikke konkludere at forventet bæreevne er hevet.

$$p-\text{verdi}=P(Z>z_{obs})=P(Z>1.34)=1-P(Z<1.34)=1-0.9099=\underline{\underline{0.09}}$$

c) Merk at vi har har en test av typen hvor $H_1: \ \mu>\mu_0$, og vi har da fra formelarket at vi skal bruke $\gamma(\mu)=1-P(Z\le z_\alpha+\frac{\mu_0-\mu}{\sigma/\sqrt{n}})$. Med $\sigma=0.35$, $n=5$, $z_\alpha=z_{0.05}=1.645$ og $\mu_0=8.7$ blir beregningen:

$$\gamma(9)=1-P\left(Z\le 1.645+\frac{8.7-9}{0.35/\sqrt{5}}\right)=1-P(Z\le -0.27)=1-0.3936=\underline{\underline{0.61}}$$

Dette betyr i praksis at dersom $\mu=9$ er det 61% sannsynlighet for at testen i punkt b) vil gi forkastning når $n=5$.

En styrke på 90% betyr at $1-\beta=0.90$ eller $\beta=0.10$ (der $\beta=$P(type II feil)). Da er $z_\beta=z_{0.10}=1.282$. Formelen på formelarket gir oss da at nødvendig utvalgsstørrelser blir

$$n=\frac{(z_\beta+z_\alpha)^2\sigma^2}{(\mu_0-\mu)^2}=\frac{(z_{0.1}+z_{0.05})^2\sigma^2}{(\mu_0-\mu)^2}=\frac{(1.282+1.645)^2 0.35^2}{(8.7-9)^2}=11.7$$

Dvs de må gjøres minst $\underline{\underline{12}}$ målinger for å få styrke på minst 0.90 (som er det samme som en sannsynlighet for type II feil på maks 0.10).

d) $\hat{\mu} = \bar{x} = 104.7/10 = \underline{\underline{10.47}}$ og $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{10-1} \cdot 0.901} = \underline{\underline{0.316}}$.

Vi er nå i situasjonen med normalfordeling med ukjent μ og ukjent σ og $(1-\alpha)100\%$ konfidensintervallet for μ er da gitt ved

$$[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \quad \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}].$$

Med $\alpha = 0.05$ blir $t_{\alpha/2,n-1} = t_{0.025,9} = -2.262$ og 95% konfidensintervalletet blir:

$$[10.47 - 2.262\frac{0.316}{\sqrt{10}}, 10.47 + 2.262\frac{0.316}{\sqrt{10}}] = \underline{\underline{[10.24\ , 10.70]}}$$

Antagelsene intervallet bygger på er at dataene er uavhengige og normalfordelte.

Oppgave 2:

a) $P(D) = \underline{\underline{0.09}}$, $P(T|D) = \underline{\underline{0.85}}$ og $P(\overline{T}|\overline{D}) = \underline{\underline{0.94}}$

$$\begin{aligned}
P(\overline{D}) &= 1 - P(D) = 1 - 0.09 = \underline{\underline{0.91}} \\
P(T|\overline{D}) &= 1 - P(\overline{T}|\overline{D}) = 1 - 0.94 = \underline{\underline{0.06}} \\
P(T) &= P(T \cap D) + P(T \cap \overline{D}) = P(T|D)P(D) + P(T|\overline{D})P(\overline{D}) \\
&= 0.85 \cdot 0.09 + 0.06 \cdot 0.91 = \underline{\underline{0.13}} \\
P(D|\overline{T}) &\overset{Bayes}{=} \frac{P(\overline{T}|D)P(D)}{P(\overline{T})} = \frac{(1 - P(T|D))P(D)}{1 - P(T)} = \frac{(1 - 0.85) \cdot 0.09}{1 - 0.13} = \underline{\underline{0.016}}
\end{aligned}$$

Oppgave 3:

a) La T være levetiden. Siden vi har for eksponentialfordeling at $\mathrm{E}(T) = 1/\lambda$ får vi her at $1/\lambda = 4.5$ dvs $\lambda = 1/4.5$.

Vi kan enten bruke sannsynlighetstettheten eller den kumulative fordelingsfunksjonen for å regne sannsynlighetene. Det er enklest å bruke den kumulative fordelingsfunksjonen, $P(T < t) = F(t) = 1 - e^{-\lambda t} = 1 - e^{-t/4.5}$ (merk at siden eksponentialfordelingen er en kontinuerlig fordeling er $P(T < t) = P(T \leq t)$) og vi får da:

$$\begin{aligned}
P(T < 1) &= F(1) = 1 - e^{-1/4.5} = \underline{\underline{0.20}} \\
P(T > 5) &= 1 - P(T \leq 5) = 1 - F(5) = 1 - (1 - e^{-5/4.5}) = \underline{\underline{0.33}} \\
P(2 < T < 4) &= P(T < 4) - P(T < 2) = F(4) - F(2) \\
&= 1 - e^{-4/4.5} - (1 - e^{-2/4.5}) = \underline{\underline{0.23}}
\end{aligned}$$

b) Vi har en situasjon karakterisert ved:

- Gjentatte delforsøk som gir "suksess"/ikke "suksess" - flere lysrør som fungerer i mindre enn ett år eller ikke.
- Lik sannsynlighet p i alle delforsøk - samme sannsynlighet $p = 0.20$ for alle lysrør for å fungere i mindre enn ett år.
- Uavhengige delforsøk - uavhengig fra lysrør til lysrør om det fungerer i mindre enn ett år eller ikke.
- Et bestemt antall, n, delforsøk - et bestemt antall n lysrør som undersøkes.

Dvs, alle betingelsene for binomisk fordeling er oppfylte og vi har dermed at X = antall lysrør som fungerer i mindre enn ett år er binomisk fordelt, $X \sim Bin(n, 0.20)$.

$$\begin{aligned}
P(X \geq 2) &= 1 - P(X < 2) = 1 - (P(X=0) + P(X=1)) \\
&= 1 - (\binom{5}{0} 0.20^0 (1-0.20)^{5-0} + \binom{5}{1} 0.20^1 (1-0.20)^{5-1}) = 1 - (.328 + 0.410) = \underline{\underline{0.26}}
\end{aligned}$$

Når $n = 120$ har vi at $np(1-p) = 120 \cdot 0.2 \cdot (1-0.2) = 19.2 > 5$ dvs vi kan bruke tilnærming til normalfordeling:

$$\begin{aligned}
P(X \geq 20) &= 1 - P(X \leq 19) \approx 1 - P(Z \leq \frac{19 + 0.5 - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}}) \\
&= 1 - P(Z \leq \frac{19 + 0.5 - 120 \cdot 0.2}{\sqrt{120 \cdot 0.2 \cdot (1-0.2)}}) = 1 - P(Z \leq -1.03) = 1 - 0.1515 = \underline{\underline{0.85}}
\end{aligned}$$

Om man utelater heltallskorreksjonen $+0.5$, får man enten svaret 0.87 dersom man starer ut som over, eller svaret 0.82 dersom man starter ut med $P(X \geq 20) = 1 - P(X < 20)$).

Oppgave 4:

a) Den generelle lineæare regresjonsmodellen er $Y = \alpha + \beta x + e$ der vi antar at $e \sim N(0, \sigma)$ og vi antar at feilleddene $e_1, \ldots, e_n$ for ulike målinger er uavhengige.

En praktisk tolkning av β-parameteren er at den sier oss hvor mye forventet porøsitet endrer seg når impedansen øker en enhet. Estimert regresjonslinje: $\underline{\underline{\hat{y} = 0.347 - 0.022 \cdot x}}$.

Det er sammenheng mellom x og Y dersom $\beta \neq 0$ så vi skal teste:

$$H_0: \quad \beta = 0 \qquad \text{mot} \qquad H_1: \quad \beta \neq 0$$

Vi leser p-verdien for testen rett ut fra datautskriften. Siden p-verdien=0.0043<0.05 ser vi at vi forkaster H_0 og kan konkludere med sammenheng.

Fra datautskriften har vi at andel forklart variasjon er $r^2 = \underline{\underline{0.41}}$.

b) Estimert forventet porøsitet ved $x = 5.3$ er: $\hat{y} = 0.347 - 0.022 \cdot 5.3 = \underline{\underline{0.23}}$.

Et $(1-\alpha)100\%$ prediksjonsintervall for regresjonslinja er gitt ved:

$$\hat{\alpha} + \hat{\beta} x \pm t_{\alpha/2, n-2} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{(s/SE(\hat{\beta}))^2}}$$

Vi har allerede regnet ut at for $x = 5.3$ er $\hat{\alpha} + \hat{\beta} x = 0.23$. Fra datautskriften finner vi videre at: $s = 0.047$, $SE(\hat{\beta}) = 0.0066$ og $n = 18$. Vi har også fra t-fordelingstabellen at $t_{\alpha/2, n-2} = t_{0.025,16} = 2.120$. Vi mangler da bare $\bar{x}$ som vi fra den oppgitte informasjonen under dataene får at er: $\bar{x} = 133.46/18 = 7.41$. Intervallet blir da:

$$0.23 \pm 2.120 \cdot 0.047 \sqrt{1 + \frac{1}{18} + \frac{(5.3 - 7.41)^2}{(0.047/0.0066)^2}} = 0.23 \pm 0.107 = \underline{\underline{[0.12,\ 0.34]}}$$

Prediksjonsintervallet forteller oss at det er ca 95% sannsynlighet at porøsiteten er i dette intervallet når impedansen er 5.3.

Oppgave 5:

a) Første skritt i testen er å regne ut forventet antall i hver celle i tabellen under antagelsen om lik fordeling i gruppene. Vi trenger da også rad- og kolonnesummene. Disse og de forventede verdiene er gitt i tabellen under (de forventede verdiene står i parantesene i hver celle). For eksempel finner vi forventet verdi for kombinasjonen "produsent A" og "feilet" som: $100 \cdot (12/200) = 6$ og for kombinasjonen "produsent B" og "feilet ikke" som: $100 \cdot (188/200) = 94$.

	feilet	feilet ikke	totalt
produsent A	8 (6)	92 (94)	100
produsent B	4 (6)	96 (94)	100
totalt	12	188	200

Testobservatoren blir da:

$$\begin{aligned} \mathcal{Q} &= \sum_{\text{alle celler}} \frac{(\text{observert-forventet})^2}{\text{forventet}} \\ &= \frac{(8-6)^2}{6} + \frac{(92-94)^2}{94} + \frac{(4-6)^2}{6} + \frac{(96-94)^2}{94} = 1.42 \end{aligned}$$

Denne verdien skal vi sammenligne med 5% kvantilen i kjikvadratfordelingen med parameter (frihetsgrader) $(r-1) \cdot (k-1) = (2-1) \cdot (2-1) = 1$. Fra kjikvadrat-tabellen finner vi at denne kvantilen har verdi 3.84. Siden $\mathcal{Q} = 1.42 < 3.84$ blir konklusjonen av vi ikke forkaster nullhypotesene om like andeler feilet/ikke feilet hos de to produsentene. Dvs, vi kan ikke konkludere at der er noen generell forskjell i andel pumper som feiler innen garantitiden hos de to produsentene.

Forutsetningen for testen er at forventet antall observasjoner i hver celle er minst 5. Dette ser vi er oppfylt her.