Løsningsforslag eksamen TALM1005 november 2018

Oppgave 1

a) Innfører hendelsene

$$
\begin{aligned}
&\text{B: trekker byggstudent, } P(B) = 0,20 \\
&\text{M: trekker maskinstudent, } P(M) = 0,10 \\
&\text{S: trekker student fra storgruppe, } P(S) = 0,70 \\
&\text{T: student får toppkarakter}
\end{aligned}
$$

Får i tillegg oppgitt at $P(T|B) = 0,13$, $P(T|M) = 0,09$, $P(T|S) = 0,08$.

i) Sannsynligheten for at en tilfeldig student får toppkarakter finner vi fra loven om total sannsynlighet:

$$
\begin{aligned}
P(T) &= P(T|B) \cdot P(B) + P(T|M) \cdot P(M) + P(T|S) \cdot P(S) \\
&= 0,13 \cdot 0,20 + 0,09 \cdot 0,10 + 0,08 \cdot 0,70 \\
&= \underline{\underline{0,091}}
\end{aligned}
$$

ii) Sannsynligheten for at en tilfeldig student er en maskinstudent som fikk toppkarakter ut i fra definisjonen av betinget sannsynlighet:

$$
\begin{aligned}
P(T|M) &= \frac{P(T \cap M)}{P(M)} \Rightarrow P(T \cap M) = P(T|M) \cdot P(M) \\
P(T \cap M) &= 0,09 \cdot 0,10 \\
&= \underline{0,0090}
\end{aligned}
$$

iii) Sannsynligheten for at studenten er en byggstudent, gitt at vi kun ser på studentene som fikk toppkarakter, finner vi fra Bayes' formel:

$$
\begin{aligned}
P(B|T) &= \frac{P(T|B) \cdot P(B)}{P(T)} \\
&= \frac{0,13 \cdot 0,20}{0,091} \\
&= \underline{\underline{0,286}}
\end{aligned}
$$

b) Skal sette sammen en referansegruppe med 6 studenter på to forskjellige måter:

i) To "likeverdige" representanter fra hver klasse

Vi skal nå foreta et uordnet utvalg av 2 studenter fra totalt antall studenter i hver klasse - dvs. vi skal trekke 2 studenter uordnet, uten tilbakelegging (vi kan tenke oss vi trekker to navnelapper for hver klasse fra en eske med navn på alle i klassen). For eksempel: vi skal trekke 2 fra de 100 byggstudentene, og dette kan gjøres på $\binom{100}{2}$ forskjellige måter. Figuren under viser totalt antall valgmuligheter:

$\binom{100}{2}$	$\binom{50}{2}$	$\binom{350}{2}$
Bygg	Maskin	Storgruppe

Det totale antall muligheter er da

$$
\binom{100}{2} \cdot \binom{50}{2} \cdot \binom{350}{2} = \underline{\underline{3,70 \cdot 10^{11}}}
$$

ii) To representanter med én som hovedrepresentant og én som vara

Vi må nå foreta et ordnet utvalg av de to studentene fra hver klasse, ettersom det nå spiller en rolle hvem som har hvilken funksjon. Figuren under viser antall valgmuligheter for hver posisjon:

<table>
<tr><td>100</td><td>99</td><td>50</td><td>49</td><td>350</td><td>349</td></tr>
<tr><td>Hoved</td><td>Vara</td><td>Hoved</td><td>Vara</td><td>Hoved</td><td>Vara</td></tr>
<tr><td colspan="2">Bygg</td><td colspan="2">Maskin</td><td colspan="2">Storgruppe</td></tr>
</table>

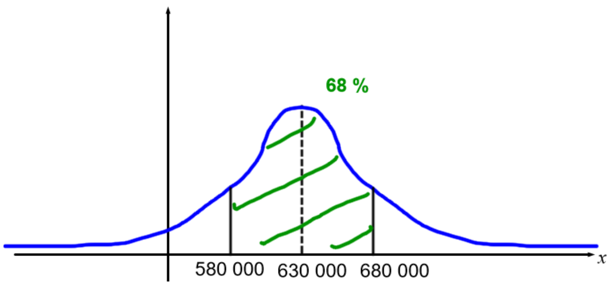
Det totale antall muligheter er da

$$100 \cdot 99 \cdot 50 \cdot 49 \cdot 350 \cdot 349 = \underline{\underline{2,96 \cdot 10^{12}}}$$

Oppgave 2

a)

i) Skal skissere sannsynlighetsfordelingen for årslønna $X \sim N\left(\mu, \sigma^2\right) \sim N\left(630\,000, 50\,000^2\right)$. Kurven er symmetrisk om toppunktet i $x = \mu$, og 68 % av arealet under grafen ligger innenfor intervallet $[\mu - \sigma, \mu + \sigma]$. Skisse:



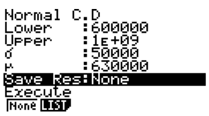
ii) Sannsynligheten for at en tilfeldig ingeniør tjener mer enn 600 000 kr per år, er lik

$$P\left(X > 600\,000\right) = 1 - P\left(X \leq 600\,000\right)$$

Denne kan vi finne enten på kalkulator, eller ved å omforme til standard normalfordeling og bruke tabeller:

$$\begin{aligned} P\left(X > 600\,000\right) &= 1 - P\left(X \leq 600\,000\right) \\ &= 1 - P\left(\frac{X-\mu}{\sigma} < \frac{600\,000-\mu}{\sigma}\right) \\ &= 1 - P\left(Z \leq \frac{600\,000 - 630\,000}{50\,000}\right) \\ &= 1 - P\left(Z \leq -0,60\right) \\ &= 1 - 0,2743 \\ &= \underline{\underline{0,7257}} \end{aligned}$$

Kan finne verdien direkte på kalkulator:



b) Skal finne sannsynligheten for at minst 5 av 10 ingeniører har en årslønn på mer en 600 000. La X angi antall ingeniører med årslønn over 600 000 kr. Dersom vi antar at ingeniørenes lønninger er uavhengige variable, er X binomisk fordelt med $n = 10$ og p lik verdien vi fant i oppg. a) ii), dvs. $p = 0,7257$. Vi skal altså finne

$$P\left(X \geq 5\right) = 1 - P\left(X \leq 4\right),$$

der vi kan bruke binomisk kumulativ fordeling på kalkulator eller i tabell. Vi får at

$$\begin{aligned} P\left(X \geq 5\right) &= 1 - P\left(X \leq 4\right) \\ &= 1 - 0,0310 \\ &= \underline{\underline{0,969}} \end{aligned}$$

c) Skal finne sannsynligheten for at to ingeniører til sammen har en inntekt på minst 1 000 000 kr. Hvis hver ingeniørs inntekt er X_1og X_2, så skal vi altså bestemme

$$P\left(X_1 + X_2 > 1\,000\,000\right)$$

Lineærkombinasjonen $Y = X_1 + X_2$ er også normalfordelt med forventningsverdi

$$E\left(Y\right) = E\left(X_1 + X_2\right) = E\left(X_1\right) + E\left(X_2\right) = \underline{2\mu}$$

Variansen er (antar at lønnsvariablene er uavhengige)

$$Var\left(Y\right) = Var\left(X_1 + X_2\right) = Var\left(X_1\right) + Var\left(X_2\right) = \underline{2\sigma^2}$$

Det gir (bruker standard metode med omforming til standard normalfordeling):

$$\begin{aligned}
P\left(Y > 1\,000\,000\right) &= 1 - P\left(Y \leq 1\,000\,000\right) \\
&= 1 - P\left(\frac{Y - 2\mu}{\sqrt{2}\sigma} \leq \frac{1\,000\,000 - 2\mu}{\sqrt{2}\sigma}\right) \\
&= 1 - P\left(Z \leq \frac{1\,000\,000 - 2 \cdot 630000}{\sqrt{2} \cdot 50\,000}\right) \\
&= 1 - P\left(Z \leq -3,68\right) \\
&= 1 - \left(1 - P\left(Z \leq 3,68\right)\right) \\
&= \underline{\underline{0,9999}}
\end{aligned}$$

Det er essensielt 100 % sannsynlig at to slike ingeniører har en samlet lønn på mer enn én million.

d) Skal finne sannsynligheten for at gjennomsnittslønna $\overline{X}$ blant 50 ingeniører er mer enn 640 000 kr. Gjennomsnittet er normalfordelt som

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{50}\right) \sim N\left(630\,000, \frac{50\,000^2}{50}\right)$$

Vi får at

$$\begin{aligned}
P(\overline{X} > 640\,000) &= 1 - P(\overline{X} \leq 640\,000) \\
&= 1 - P(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{640\,000 - \mu}{\frac{\sigma}{\sqrt{n}}}) \\
&= 1 - P\left(Z \leq \frac{640\,000 - 630\,000}{\frac{50\,000}{\sqrt{50}}}\right) \\
&= 1 - P(Z \leq 1,41) \\
&= 1 - 0,9207 \\
&= \underline{\underline{0,0793}}
\end{aligned}$$

Oppgave 3

a) La X angi antall besøkende til serveren i løpet av ett minutt. Kriteriene for at X skal være Poisson-fordelt er:

i. **Uavhengighet**: Antall besøkende på ett tidspunkt påvirker ikke antall besøkende på et senere tidspunkt

ii: **Konstant sannsynlighet** for å få en besøkende innenfor tidsintervallet (dvs. uavhengig av tid på dagen)

iii. **Ikke overlapp**: Kan ikke ha to besøkende eksakt samtidig

b) Med antakelsen om at X er Poisson-fordelt med $\lambda = 15$:

i) Sannsynligheten for 20 eller flere besøkende i løpet av et minutt (kan bruke kalkulator eller tabell for kumulativ Poisson-sannsynlighet):

$$\begin{aligned} P(X \geq 20) &= 1 - P(X \leq 19) \\ &= 1 - 0,8752 \\ &= \underline{\underline{0,1248}} \end{aligned}$$

ii) Sannsynligheten for minst 150 potensielle besøkende i løpet av en "nedetid" på 10 minutter: ettersom vi har en intensitet $\alpha = 15\,\text{besøkende/min}$, er forventet antall besøkende i dette tidsintervallet lik

$$\lambda = 10 \cdot 15 = 150.$$

La Y angi antall besøkende i løpet av 10 minutter. Vi skal bestemme

$$P(Y \geq 150) = 1 - P(X \leq 149)$$

Her kan vi enten bruke kalkulator, eller vi kan bruke normaltilnærmingen, ettersom $\lambda = 150$ og altså mye høyere enn "grenseverdien" $\lambda = 10$ på formelarket. Poisson-fordelingen har den egenskapen (beskrevet på formelarket) at

$$\begin{aligned} \mu &= E(Y) = \lambda = 150 \\ \sigma^2 &= Var(Y) = \lambda = 150 \end{aligned}$$

Dersom vi bruke normaltilnærming, blir altså forventningsverdi og varians hhv.

$$\begin{aligned} \mu &= 150 \\ \sigma^2 &= 150 \end{aligned}$$

Med Poisson-fordeling på kalkulator, får vi

$$\begin{aligned} P(Y \geq 150) &= 1 - P(X \leq 149) \\ &= 1 - 0,4891 \\ &= \underline{\underline{0,5119}} \end{aligned}$$

c) La $X_1, X_2, ..., X_{50}$ angi antall besøkende til hver av de 50 uavhengige serverne. Skal finne sannsynligheten for at det totale antallet besøkende $X = X_1 + X_2 + ... + X_{50}$ i løpet av ett minutt er mindre enn 750.

Vi har altså 50 uavhengige, identisk fordelte variable, der hver variabel har hhv. forventning og varians lik (se forrige oppgave - dette framkommer fra egenskapene til Poisson-fordelingen, som beskrevet på formelarket)

$$\begin{aligned} \mu &= E(X) = 15 \\ \sigma^2 &= Var(X) = 15 \end{aligned}$$

Fra sentralgrenseteoremet er da summen X av disse variablene tilnærmet normalfordelt som

$$X \sim N\left(n\mu, n\sigma^2\right) \sim N(50 \cdot 15, 50 \cdot 15) \sim N(750, 750)$$

slik at

$$\begin{aligned} P(X < 750) &\approx P\left(\frac{X - n\mu}{\sqrt{n}\sigma} < \frac{750 - 750}{\sqrt{50} \cdot \sqrt{15}}\right) \\ &= P(Z < 0) \\ &= \underline{\underline{\frac{1}{2}}} \end{aligned}$$

Oppgave 4

a) Levetiden T til driftsreima er Weibullfordelt som $T \sim W(\eta = 15\,000, \beta = 4)$. Skal finne

$$\begin{aligned}\mu = E(T) &= \eta\Gamma\left(1 + \frac{1}{\beta}\right)\\ &= 15\,000 \cdot \Gamma\left(1 + \frac{1}{4}\right)\\ &= 15\,000 \cdot \Gamma(1,25)\\ &= 15\,000 \cdot 0,9064\\ &= \underline{\underline{13\,596}}\end{aligned}$$

Forventet levetid er 13 596 h.

b) Sannsynligheten for at reima ryker innen 10 000 timer er

$$\begin{aligned}P(T < 10\,000) &= 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}}\\ &= 1 - e^{-\left(\frac{10\,000}{15\,000}\right)^{4}}\\ &= \underline{\underline{0,1792}}\end{aligned}$$

Oppgave 5

Innledende kommentar: det er en trykkfeil i oppgaveteksten - tabellen inneholder 11 målinger, mens teksten sier at det er 10 målinger. I dette løsningsforslaget tas det utgangspunkt i 10 målinger, og at oppgitt gjennomsnitt og standardavvik i oppgaven gjelder for 10 målinger i (dvs. man bruker $n = 10$ og $\overline{X} = 118,6$), men det har *ingen betydning for vurderingen om du bruker 10 eller 11 målinger.*

a)

i) Et 95 % konfidensintervall for utslippet X, som antas normalfordelt med $\mu = 117\,\text{g/km}$ og kjent $\sigma = 5\,\text{g/km}$, blir lik

$$\begin{aligned}\overline{X} \pm u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} &= 118,6 \pm u_{\frac{0,05}{2}} \frac{5}{\sqrt{10}}\\ &= 118,6 \pm 1,96 \cdot \frac{5}{\sqrt{10}}\\ &= 118,6 \pm 3,099\\ &= \underline{\underline{[115,5,\ 121,7]}}\end{aligned}$$

ii) Et 95 % konfidensintervall har den egenskapen at dersom "forsøket" gjentas mange ganger (dvs. man foretar en stikkprøve med 10 biler fra en langt større produksjonslinje med biler), og vi konstruerer et konfidensintervall for hver av de "mange" stikkprøvene, så vil intervallet inneholde den "sanne" verdien av forventningsverdien μ i 95 % av tilfellene.

b) Hypotesetesten som skal gjennomføres, er her

$$\begin{aligned}H_0 &: \mu = \mu_0 = 117\,\text{g/km}\\ H_1 &: \mu > 117\,\text{g/km}\end{aligned}$$

Gjennomfører hypotesetesten på 3 ulike måter (det er kun én som trengs for å få full uttelling på oppgaven):

i) Viser på 3 forskjellige måter (kun én kreves for full uttelling):

Kritisk verdi

Den kritiske verdien bestemmes ut i fra definisjonen av signifikansnivået:

$$P\left(\overline{X} \geq k | H_0\right) = \alpha$$

Lager standard normalfordeling, dvs. introduserer variabelen $Z \sim N(0,1)$:

$$\begin{aligned}
P\left(\frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \geq \frac{k-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right) &= \alpha \\
P\left(Z \geq \frac{k-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right) &= \alpha \\
P\left(Z \geq \frac{k-117}{\frac{5}{\sqrt{10}}}\right) &= 0,05
\end{aligned}$$

Vi skal altså bestemme den verdien av $\frac{k-117}{\frac{5}{\sqrt{10}}}$ som er slik at 5 % av "grafen" ligger **over** denne verdien. Men dette tilsvarer pr. def. 5 %-kvantilen til standard normalfordeling, dvs.

$$\frac{k-117}{\frac{5}{\sqrt{10}}} = u_{0,05} = 1,645 \Rightarrow k = 117 + 1,645 \cdot \frac{5}{\sqrt{10}} = \underline{119,6}$$

Ettersom det faktiske snittet er $\overline{X} = 118,6$ er **lavere** enn den kritiske verdien, vil vi beholde nullhypotesen - dvs. vi kan ikke påstå at $\mu > 117\,\mathrm{g/km}$.

Signifikanssannsynlighet (p-verdi)

p-verdien er sannsynligheten for å få en minst like "ekstrem" stikkprøve som den faktiske verdien $x = 118,6\,\mathrm{g/km}$, *gitt at nullhypotesen er riktig*. Altså:

$$\begin{aligned}
p &= P\left(\overline{X} \geq x | H_0\right) \\
&= P\left(\frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \geq \frac{x-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right) \\
&= P\left(Z \geq \frac{118,6-117}{\frac{5}{\sqrt{10}}}\right) \\
&= P(Z \geq 1,012) \\
&= 1 - P(Z \leq 1,012) \\
&= 1 - 0,8461 \\
&= \underline{0,1639}
\end{aligned}$$

Ettersom denne p-verdien er (vesentlig) større enn signifikansnivået $\alpha = 0,05$, er det "ganske sannsynlig" å få det faktiske snittet i stikkprøven dersom nullhypotesen er riktig - dvs. H_0 er mest sannsynlig riktig. Vi beholder altså H_0.

Testobservator

Bruker her observatoren

$$\begin{aligned}
U_0 &= \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \\
&= \frac{118,6-117}{\frac{5}{\sqrt{10}}} \\
&= \underline{1,012}
\end{aligned}$$

Beslutningsgrunnlaget her er at vi forkaster H_0 dersom $|U_0| > u_\alpha = u_{0,05} = 1,645$. Ser her at dette **ikke** er tilfelle, dvs. vi beholder nullhypotesen.

Konklusjon ved alle tre metoder: vi kan **ikke** på grunnlag av stikkprøven forkaste nullhypotesen, og kan **ikke** hevde at den annonserte utslippsverdien på 117 g/km er overskredet.

ii) Signifikansnivået α er sannsynligheten for å forkaste H_0, selv om denne er riktig - i dette tilfellet: sannsynligheten for å påstå at utslippet overstiger den påståtte verdien, selv om dette i realiteten ikke er tilfellet. Med $\alpha = 5$ %, er dette altså sannsynligheten for å feilaktig forkaste H_0.

c) Skal bestemme styrkefunksjonen $\beta(\mu = 122)$. Ut i fra definisjonen er

$$\beta(\mu = 122) = P(\text{akseptere } H_1 \text{—} \mu = 122)$$

At vi "aksepterer H_1" tilsvarer her til at stikkprøven ligger "utenfor" området med kritisk verdi, dvs. $\overline{X} \geq k$, der vi tidligere bestemte kritisk verdi til $k = 119,6$. Vi skal altså beregne

$$\begin{aligned}
\beta(\mu = 122) &= P\left(\overline{X} \geq k | \mu = 122\right) \\
&= P\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{k - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \\
&= P\left(Z \geq \frac{119,6 - 122}{\frac{5}{\sqrt{10}}}\right) \\
&= P(Z \geq -1,52) \\
&= 1 - P(Z \leq -1,52) \\
&= 1 - (1 - P(Z \leq 1,52)) \\
&= P(Z \leq 1,52) \\
&= \underline{\underline{0,9357}}
\end{aligned}$$

Det er altså ca. 93, 6 % sannsynlig at vi, på grunnlag av en stikkprøve, forkaster nullhypotesen og aksepterer mothypotesen, dersom det faktiske, reelle utslippstallet er $\mu = 122$. Ettersom H_1 faktisk er sann i dette tilfellet (vi ligger godt over den kritiske verdien på $k = 119,6$), er det "veldig sannsynlig" at vi kommer til å trekke riktig konklusjon i dette tilfellet.

d) Dersom standardavviket hadde vært ukjent, hadde vi brukt t-test til å avgjøre hypotesetesten. Testobservatoren i dette tilfellet er

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

der det empiriske standardavviker er oppgitt i oppgaven

$$S = 2,67$$

Får her at

$$\begin{aligned}
T_0 &= \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} \\
&= \frac{118,6 - 117}{\frac{2,67}{\sqrt{10}}} \\
&= \underline{1,89}
\end{aligned}$$

Sammenlikner $|T_0|$ med kvantilen $t_{\alpha,n-1} = t_{0,05,9} = 1,833$. Ser at i dette tilfellet er $|T_0| > t_{0,05,9}$, slik at vi i dette tilfellet får motsatt konklusjon av forrige oppgave: vi vil forkaste H_0 og påstå H_1, dvs. at dataene tyder på at utslippstallet er høyere enn de annonserte 117 g/km.