

Løsning TALM1005 (statistikkdel) juni 2017

Oppgave 1

a) Har oppgitt at sannsynligheten for at én harddisk svikter er $p = 0,037$. Innfører hendelsene

A : harddisk 1 svikter

B : harddisk 2 svikter

C : harddisk 3 svikter

Datatap inntreffer dersom alle 3 harddiskene svikter samtidig, og når disse svikter uavhengig av hverandre, blir denne lik:

$$\begin{aligned}P(\text{datatap}) &= P(A \cap B \cap C) \\&= P(A) \cdot P(B) \cdot P(C) \\&= p^3 \\&= 0,037^3 \\&= \underline{\underline{5,07 \cdot 10^{-5}}}\end{aligned}$$

b) Vi skal bestemme sannsynligheten for at minst én av 150 datamaskin med en slik harddisk-konfigurasjon opplever datatap. Hvis X angir antallet datamaskiner som opplever datatap, er X binomisk fordelt med $X \sim \text{bin}(n, p) \sim \text{bin}(150, 5,07 \cdot 10^{-5})$, slik at

$$P(X = x) = \binom{150}{x} \cdot (5,07 \cdot 10^{-5})^x (1 - 5,07 \cdot 10^{-5})^{150-x}.$$

Det gir:

$$\begin{aligned}P(X \geq 1) &= 1 - P(X = 0) \\&= 1 - \binom{150}{0} \cdot (5,07 \cdot 10^{-5})^0 (1 - 5,07 \cdot 10^{-5})^{150-0} \\&= \underline{\underline{7,58 \cdot 10^{-3}}}\end{aligned}$$

c)

i) Vi skal bestemme antallet måter å sette sammen 3 harddisker fra to produsenter A og B, slik at rekkefølgen spiller en rolle. Vi har da 2 valgmuligheter for hver harddisk, og multiplikasjonsprinsippet gir at det er i alt

$$2^3 = \underline{\underline{8}}$$

forskjellige måter å gjøre dette på. Vi kan raskt sjekke at dette stemmer: vi får konfigurasjonene

AAA, AAB, ABA, ABB, BBB, BBA, BAB, BAA

ii) Dersom rekkefølgen ikke spiller noen rolle, skiller vi ikke mellom f.eks. AAB og ABA. Med enkel opplisting ser vi at vi får følgende muligheter:

AAA, BBB, AAB (to fra A; én fra B), BBA (én fra A, to fra B),

dvs. antallet forskjellige måter er 4.

Oppgave 2

a) La R angi andelen enheter i en dagsproduksjon som må repareres, og K andelen enheter som må kasseres. Hvis X angir bedriftens tap på en tilfeldig enhet, så er tapet gitt ved

$$X = 200R + 1000K,$$

slik at forventningsverdien (i kroner) blir

$$\begin{aligned}\mu = E(X) &= E(200R + 1000K) \\&= 200E(R) + 1000E(K) \\&= 200 \cdot 0,07 + 1000 \cdot 0,03 \\&= \underline{\underline{44}}\end{aligned}$$

Variansen σ^2 (i kroner²) kan vi beregne som

$$\begin{aligned}\sigma^2 &= E\left((X - \mu)^2\right) \\ &= E(X^2) - \mu^2 \\ &= \sum_{\text{alle } x} x^2 \cdot P(X = x) - \mu^2 \\ &= (0^2 \cdot P(X = 0) + 200^2 \cdot P(X = 200) + 1000^2 \cdot P(X = 1000)) - \mu^2 \\ &= 0^2 \cdot 0,90 + 200^2 \cdot 0,07 + 1000^2 \cdot 0,03 - 44^2 \\ &= \underline{\underline{30864}}\end{aligned}$$

b) La S angi bedriftens tap på en hel dagsproduksjon på 300 enheter. Dersom $X_1, X_2, X_3, \dots, X_{300}$ angir tapet på én enhet, så er det totale tapet gitt ved

$$S = X_1 + X_2 + \dots + X_{300}.$$

Ettersom alle enhetene antas å være uavhengige av hverandre, er S altså en sum av 300 uavhengige variable med samme fordeling. Da kan S antas normalfordelt med forventningsverdi $\mu_S = 300\mu = 300 \cdot 44 = 13200$ og varians $\sigma_S^2 = 300\sigma^2 = 300 \cdot 30864 = 9259200$.

c) Vi skal beregne sannsynligheten for at S blir større enn 10 000 kr, dvs.

$$\begin{aligned}P(S > 10000) &= 1 - P(S \leq 10000) \\ &= 1 - P\left(\frac{S - \mu_S}{\sigma} \leq \frac{10000 - 13200}{\sqrt{9259200}}\right) \\ &= 1 - P(Z \leq -1,05) \\ &= 1 - (1 - P(Z \leq 1,05)) \\ &= P(Z \leq 1,05) \\ &= \underline{\underline{0,8531}}\end{aligned}$$

Her er det brukt tabellverdier for standard normalfordeling, men vi kunne også brukt den kumulative normalfordelingen på kalkulator.

Oppgave 3

a) La X angi antall innkomne telefonsamtaler i løpet av en time, med et oppgitt gjennomsnitt på 10 samtaler per time. Betingelsene for at X skal være en Poisson-fordelt variabel:

i) X har kun diskrete verdier (åpenbart sant, ettersom X angir et antall)

ii) Hvorvidt hendelsen inntreffer, dvs. at noen ringer, er uavhengig av tidligere hendelser (også en grei antakelse her)

iii) Hendelsene er jevnt spredd over tidsrommet (kanskje den mest tvilsomme betingelsen, ettersom telefoner kan ha en tendens til å “hope seg opp” dersom det oppstår f.eks. dataproblemer e.l. som mange ringer inn om samtidig)

iv) To hendelser kan ikke inntreffe samtidig, dvs. man kan ikke ha flere samtidige innringere. Selv om Orakeltjenesten har flere inngående linjer, er det rimelig å anta at sannsynligheten for flere ringer inn samtidig, er liten (med mindre det er “telefonstorm” - se over).

Alt i alt er Poisson-fordeling en ganske rimelig antakelse.

b) Med X antatt Poisson-fordelt med $\lambda = 10$, skal vi beregne

$$\begin{aligned}P(X > 10) &= 1 - P(X \leq 10) \\ &= 1 - 0,5830 \\ &= \underline{\underline{0,4170}}\end{aligned}$$

Her har vi brukt tabellverdi fra den kumulative Poisson-fordelingen på formelarket, men vi kan også beregne dette på kalkulator.

c) Vi får oppgitt at i sommerferien er X Poisson-fordelt med en intensitet $\alpha = 2$ samtaler per time, dvs $\alpha = \frac{2}{60}$ samtaler per minutt. La Y angi antallet innkomne telefonsamtaler i løpet en periode (lunchpause) med varighet $t = 20$ minutter. Denne vil være Poisson-fordelt med

$$\begin{aligned}\lambda &= \alpha \cdot t \\ &= \frac{2}{60} \cdot 20 \\ &= \frac{2}{3}\end{aligned}$$

Sannsynligheten for at ingen ringer i løpet av dette tidsrommet blir altså

$$\begin{aligned}P(Y = 0) &= \frac{\lambda^0 e^{-\lambda}}{0!} \\ &= e^{-\lambda} \\ &= e^{-\frac{2}{3}} \\ &= \underline{\underline{0,513}}\end{aligned}$$

Oppgave 4

a) Vi får oppgitt at $X_1, X_2, \dots, X_{200}$ er kaffemengden som en tilfeldig gjest fyller i koppen sin. Den samlede kaffemengden som 200 gjester fyller, er da summen av disse:

$$S = X_1 + X_2 + \dots + X_{200},$$

og ettersom $X_1, X_2, \dots, X_{200}$ er normalfordelt med $\mu = 0,25$ og $\sigma = 0,10$, blir S via sentralgrenseteoremet også normalfordelt med forventningsverdi

$$\mu_S = 200\mu = 200 \cdot 0,25 = \underline{\underline{50}}$$

og varians

$$\sigma_S^2 = 200\sigma^2 = 200 \cdot 0,10^2 = \underline{\underline{2,0}}$$

b) Vi skal finne sannsynligheten for at den 50-liters kaffebeholderen er tom etter at 200 gjester har forsynt seg med én kopp hver, dvs. vi skal finne

$$\begin{aligned}P(S > 50) &= 1 - P(S \leq 50) \\ &= 1 - P\left(\frac{S - \mu_S}{\sigma_S} \leq \frac{50 - 50}{\sqrt{2,0}}\right) \\ &= 1 - P(Z \leq 0) \\ &= 1 - \frac{1}{2} \\ &= \underline{\underline{\frac{1}{2}}}\end{aligned}$$

Her har vi brukt tabellen for kumulativ fordeling av standard normalfordeling, men vi kan også beregne dette fra kalkulator.

c) Vi skal bestemme antall gjester n som må forsyne seg for at det skal være 99 % sannsynlig at beholderen er tom, dvs. vi skal bestemme en n som er slik at

$$P(S > 50) = 0,99$$

Ved å omgjøre til standard normalfordeling, får vi følgende likning:

$$\begin{aligned}P(S > 50) &= 0,99 \\ 1 - P(S \leq 50) &= 0,99 \\ P(S \leq 50) &= 0,01 \\ P\left(\frac{S - \mu_S}{\sigma_S} \leq \frac{50 - \mu_S}{\sigma_S}\right) &= 0,01\end{aligned}$$

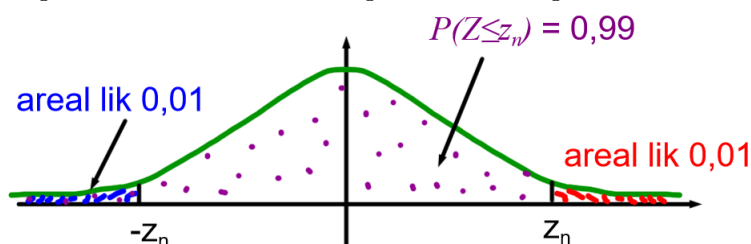
Her få vi bruk for å bestemme forventningsverdien μ_S og standardavviket σ_S for den samlede kaffemengden til de n gjestene. I fra sentralgrenseteoremet får vi

$$\begin{aligned}\mu_S &= n\mu = n \cdot 0,25 = 0,25n \\ \sigma_S^2 &= n\sigma^2 = n \cdot 0,10^2 \Rightarrow \sigma_S = \sqrt{n \cdot 0,10^2} = 0,10\sqrt{n}\end{aligned}$$

Likningen for å bestemme n blir da

$$\begin{aligned}P\left(\frac{S - \mu_S}{\sigma_S} \leq \frac{50 - 0,25n}{0,10\sqrt{n}}\right) &= 0,01 \\ P\left(Z \leq \frac{50 - 0,25n}{0,10\sqrt{n}}\right) &= 0,01\end{aligned}$$

Vi skal nå bestemme verdien av størrelsen $\frac{50-0,25n}{0,10\sqrt{n}}$, slik at den kumulative sannsynligheten blir lik 0,01. Her kan vi så klart prøve oss frem på kalkulator ved å finne på en verdi for n (f.eks. 250 gjester), beregne størrelsen $\frac{50-0,25n}{0,10\sqrt{n}}$, og så sjekke om den kumulative sannsynligheten blir lik 0,01. Hvis vi istedet ønsker å regne ut svaret, kan det være greit å minne seg på hvordan standard normalfordeling ser ut:



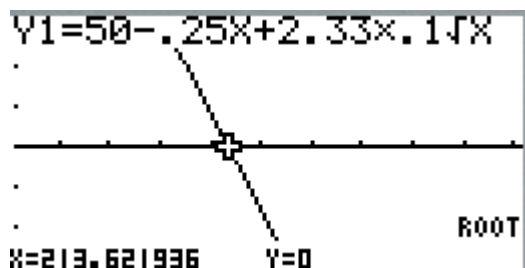
Uheldigvis står ikke $P(Z \leq z) = 0,01$ tabulert direkte, så vi må benytte oss symmetrien til standard normalfordeling for å kunne bruke de tabulerte verdiene. Størrelsen $\frac{50-0,25n}{0,10\sqrt{n}}$ er åpenbart negativ - la oss kalle den $-z_n$. Grafen viser at problemet vårt tilsvarer å finne en z_n som er slik at $P(Z \leq z_n) = 0,99$, og denne finner vi i tabellen:

$$P(Z \leq z_n) = 0,99 \Rightarrow z_n = 2,33$$

Likningen for å bestemme n blir altså

$$\begin{aligned}\frac{50 - 0,25n}{0,10\sqrt{n}} &= -z_n = -2,33 \\ 50 - 0,25n &= -2,33 \cdot 0,10\sqrt{n} \\ 50 - 0,25n + 2,33 \cdot 0,10\sqrt{n} &= 0\end{aligned}$$

Dette er en brysom likning å løse ved regning (må kvadrere for å fjerne rota osv.), derfor plotter vi venstresiden på en grafisk kalkulator, og leser av nullpunktene:



Vi leser av løsningen

$$\underline{n = 214}$$

Oppgave 5

a) Vi skal gjennomføre en hypotesetest med signifikansnivå $\alpha = 0,05$ på 3 forskjellige måter for å teste nullhypotesen H_0 mot mothypotesen H_1 , der

$$\begin{aligned}H_0 : \mu &= 1,50 \\ H_1 : \mu &< 1,50\end{aligned}$$

i) Beregne kritisk verdi: vi skal bestemme en kritisk verdi k som er slik at sannsynligheten for å forkaste nullhypotesen selv om H_0 er riktig, er lik signifikansnivået α , dvs.

$$P(\bar{X} \leq k | H_0) = 0,05$$

Dersom nullhypotesen er riktig, er $\mu = \mu_0 = 1,50$. Når vi skal omforme til standard normalfordeling, må vi huske at standardavviket til de $n = 15$ målingene er gitt ved $\frac{\sigma}{\sqrt{n}}$. Vi får:

$$P\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq \frac{k - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = 0,05$$

$$P\left(Z \leq \frac{k - 1,50}{\frac{0,05}{\sqrt{15}}}\right) = 0,05$$

Via symmetrien til standard normalfordeling, får vi her at

$$\frac{k - 1,50}{\frac{0,05}{\sqrt{15}}} = -U_\alpha = -1,645$$

$$k = -1,645 \cdot \frac{0,05}{\sqrt{15}} + 1,50$$

$$k = \underline{1,479}$$

Ettersom det målte gjennomsnittet er $\bar{x} = 1,467 < k$, vil vi altså forkaste nullhypotesen, og konkludere med at kaffeautomaten gir ut **for lite** kaffe.

ii) Vi beregner testobservatoren

$$U_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{1,467 - 1,50}{\frac{0,05}{\sqrt{15}}}$$

$$= \underline{-2,56}$$

Ettersom testobservatoren har en verdi

$$|U_0| = 2,56 > U_\alpha = 1,645,$$

vil vi også her forkaste nullhypotesen.

iii) Vi beregner signifikanssannsynligheten p , som er sannsynligheten for å få en minst like “ekstrem” verdi for kaffemengden som den faktisk observerte ($\bar{x} = 1,467$), gitt at nullhypotesen er riktig. Altså:

$$p = P(\bar{X} \leq \bar{x} | H_0)$$

$$= P\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P\left(Z \leq \frac{1,467 - 1,50}{\frac{0,05}{\sqrt{15}}}\right)$$

$$= P(Z \leq -2,56)$$

Dersom vi skal benytte oss av normalfordelingstabellen, må vi her bruke at

$$P(Z \leq -2,56) = 1 - P(Z \leq 2,56)$$

$$= 1 - 0,9948$$

$$= \underline{0,0052}$$

Ettersom signifikanssannsynligheten $p = 0,0052$ er mindre en signifikansnivået $\alpha = 0,05$, konkluderer vi med at det er svært “usannsynlig” å få så lite kaffe som vi faktisk fikk, dersom produsentens påstand om at “gjennomsnittlig kaffemengde er 1,50 dl” er riktig. Vi forkaster derfor også her nullhypotesen.

b) Vi skal beregne styrkefunksjonen

$$\beta(\mu = 1,40) = P(H_1 | \mu = 1,40)$$

Styrkefunksjonen angir altså sannsynligheten for at vi forkaster nullhypotesen, dvs. påstår mothypotesen H_1 , dersom det faktiske gjennomsnittet av kaffemengden er 1,40 dl. Ettersom produsenten hardnakket hevder at snittet er 1,50 dl, er vi i dette scenariet interessert i at hypotesetesten forkaster nullhypotesen (for den er beviselig feil i dette tilfellet) - styrkefunksjonen angir derfor “styrken” eller følsomheten til hypotesetesten for den gitte verdien av μ . Vi kan beregne styrkefunksjonen ved å støtte oss til den tidligere beregnede kritiske verdien: det å forkaste nullhypotesen og påstå H_1 tilsvarer at vi får et snitt på de 15 målingene mindre enn $k = 1,479$. Denne sannsynligheten er

$$\begin{aligned} P(\bar{X} \leq k | \mu = 1,40) &= P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{k - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \\ &= P\left(Z \leq \frac{1,479 - 1,40}{\frac{0,05}{\sqrt{15}}}\right) \\ &= P(Z \leq 6,1) \\ &\approx \underline{\underline{1}} \end{aligned}$$

Det er altså essensielt helt garantert at vi kommer til å forkaste H_0 dersom den faktiske kaffemengden er 1,40 dl (produsenten på sin side har hele tiden hevdet et snitt på 1,50 dl). I så måte har hypotesen “god styrke”, dvs. den er tilstrekkelig følsom.