STK1000: Løsningsforslag Uke 40

2022

Oppgave 5.15

a)

```
x = c(118, 24, 89, 85, 74, 135, 116, 107, 60, 99)
mu = mean(x)
mu
```

```
## [1] 90.7
```

b)

Vi trekker tre tilfeldige individer og regner ut gjennomsittet av de tre.

```
sub = sample(x, 3)
sub
```

```
## [1]  99 135  74
```

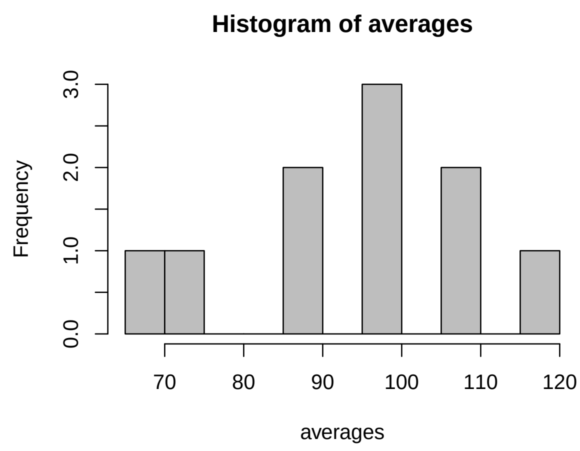
```
mean(sub)
```

```
## [1] 102.6667
```

c)

Vi repeterer b) 10 ganger med en for-løkke

```
n = 10
averages = numeric(n) # lager en liste med 10 nullere (som vi skal fylle inn senere)
for (i in 1:n) { # repeterer 10 ganger
  sub = sample(x, 3) # trekker 3 tilfeldige individer
  avg = mean(sub) # regner ut gjennomsnitt
  averages[i] = avg # fyller inn plass "i" med gjennosnittet
}
hist(averages, col = 'gray', breaks = 10)
```



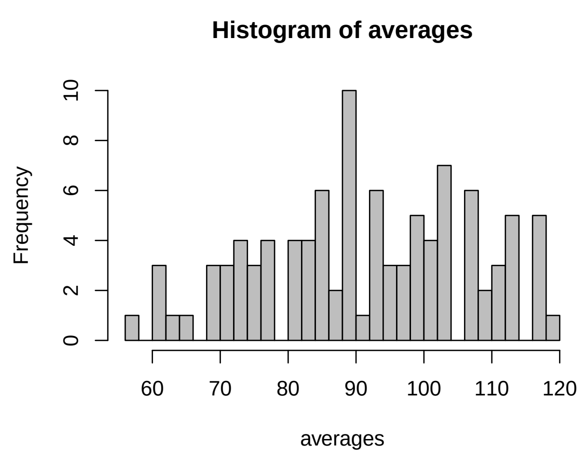
d)

Vi ser at midten av histogrammet ikke er så langt fra populasjonsgjennomsnittet på 90.7. Vi vet at gjennomsnittet av utvalget er en forventningsrett (unbiased) estimator av populasjonsgjennomsnittet. Det vil si at utvalgsgjennomsnittene våre skal fordele seg rundt populasjonsgjennomsnittet.

tilleggsoppgave 'e')

Som vi diskuterte i d) ser vi nå at histogrammet fordeler seg rundt populasjonsgjennomsnittet på 90.7.

```
n = 100
averages = numeric(n) # lager en liste med 100 nullere (som vi skal fylle inn senere)
for (i in 1:n) { # repeterer 100 ganger
  sub = sample(x, 3) # trekker 3 tilfeldige individer
  avg = mean(sub) # regner ut gjennomsnitt
  averages[i] = avg # fyller inn plass "i" med gjennosnittet
}
hist(averages, col = 'gray', breaks = 30)
```



Oppgave 5.16

a) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.67}{\sqrt{60}} = 0.2156.$

b) I 95 % av utvalgene vil vi ha et gjennomsnitt i intervallet $\mu \pm 2\sigma_{\bar{x}} = 7.13 \pm 2 \cdot 0.2156 = [6.70, 7.56]$.

c) $P(\bar{x} < 6.9) = P(Z < \frac{6.9-7.13}{0.2156}) = P(Z < -1.0668) = 0.14$.

Oppgave 5.17

a) Vi trenger et større utvalg enn 60, ettersom intervallet i Oppgave 5.16 b) er bredere enn det vi ønsker det skal være.

b) Ettersom vi ønsker $\mu + 2\sigma_{\bar{x}} \leq \mu + 0.17$ trenger vi $\sigma_{\bar{x}} \leq \frac{0.17}{2} = 0.085$.

c) Vi ønsker $\sigma_{\bar{x}} = \frac{1.67}{\sqrt{n}} \leq 0.085$. Hvis vi snur om på uttrykket får vi at $n \geq \left(\frac{1.67}{0.085}\right)^2 = 386.0069$. Vi trenger (minst) $n = 387$.

Oppgave 5.23

a) At $\bar{x}$ er forventningsrett (ikke er biased) betyr at $\bar{x}$ ikke er systematisk for høy eller for lav. Hvis man gjentar utvalget vil man få $\bar{x}$'er som er spredd rundt den sanne verdien μ.

b) Jo større utvalget vårt er, jo nærmere vil $\bar{x}$ typisk være den sanne verdien μ.

Oppgave 5.24

a) Vi har bruker definisjonene på side 248 og 256 og får

$$\mu = 4 \cdot 0.161 + 3 \cdot 0.343 + 2 \cdot 0.292 + 1 \cdot 0.096 + 0 \cdot 0.098 = 2.353.$$
$$\sigma^2 = (4 - 2.353)^2 \cdot 0.161 + (3 - 2.353)^2 \cdot 0.343 + (2 - 2.353)^2 \cdot 0.292 + (1 - 2.353)^2 \cdot 0.096 + (0 - 2.353)^2 \cdot 0.098 = 1.335$$
$$\sigma = \sqrt{1.335025} = 1.16.$$

b) Vi har at $\mu_{\bar{x}} = \mu = 2.353$ og $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{25}} = 0.231$.

c) $P(X \geq 3) = P(X = 4) + P(X = 3) = 0.161 + 0.341 = 0.502$.

d) $P(\bar{x} \geq 3) = P(Z \geq \frac{3-2.353}{0.231}) = P(Z \geq 2.80) = 1 - P(Z < 2.81) = 0.00256$.

e) utvalgsgjennomsnittet er mindre variabelt enn en enkeltobservasjon, og vi får lavere sannsynlighet for verdier over 3 (husk at 3 er større enn forventningsverdien 2.353)

Oppgave 5.30

a) Nei. Du kan ikke telle tid siden det er kontinuerlig.

b) Ja, her er det naturlig å anta at $X \sim B(20, p)$, hvor p er sannsynligheten for et defekt par.

c) Ja, her kan vi anta at $X \sim B(n, p)$ hvor n er størrelsen på utvalget og p er sannsynligheten for at en tilfeldig valgt student sier han spiser minst 5 om dagen.

d) Ja, til dels. Vi kan anta at $X \sim B(n, p)$ hvor n er antall dager i skoleåret og p er sannsynligheten for at du skulker. Her må vi i så fall anta at alle dagene er uavhengig av hverandre (om du skulker i dag påvirker ikke om du skulker i morgen).

Oppgave 5.43

a) Siden vi har et intensitet på 0.5 per cm^2 og en overflate på 100 cm^2 har vi en Poisson-fordeling med forventning på $0.5 \cdot 100 = 50$.

b) Vi har $\mu = 50$ og $\sigma = \sqrt{50}$. $P(X > 60) = P(Z > \frac{60-50}{\sqrt{50}}) = P(Z > 1.41) = 0.0793$.

Oppgave 5.47

a) $\mu_{\bar{x}} = \mu = 0.43$ og $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.95}{\sqrt{150}} = 0.0776$.

b) $P(\bar{x} > 0.55) = P(Z > \frac{0.55-0.43}{0.0776}) = P(Z > 1.547) = 1 - P(Z \leq 1.547) = 0.0609$. Her har vi brukt sentralgrenseteoremet da vi antok at $\bar{x}$ er (tilnærma) normalfordelt.

c) Ja, når n er så stor som 150 er det en rimelig antagelse at $\bar{x}$ er normalfordelt.

Oppgave 5.63

a) Man har $p = \frac{1}{4} = 0.25$ sannsynlighet for gjette riktig på et vilkårlig kort.

b) Her har vi en binomisk fordeling $B(20, 0.25)$ som gir

$$\begin{aligned} P(X \geq 10) &= P(X = 10) + P(X = 11) + \ldots + P(X = 20) \\ &= 0.0099 + 0.0030 + 0.0008 + 0.0002 + 0 + \ldots + 0 \\ &= 0.0139. \end{aligned}$$

Vi har brukt Tabell C i boken for å finne tallene.

c) Vi har forventning $\mu = np = 20 \cdot 0.25 = 5$, med standardavvik $\sigma = \sqrt{np(1-p)} = \sqrt{3.75} = 1.9365$.

d) Nei, vi kan ikke lenger bruke binomisk fordeling her. Det står at forsøket utføres med å gå gjennom **ALLE** kortene og vi vet at det er 5 av hver type, så forsøkene er ikke uavhengig av hverandre. Hvis man i stedet hadde utført forsøket med å trekke tilfeldig fra bunken med tilbakelegging, kunne vi brukt binomisk fordeling.

Oppgave 5.72

a) Hvis man bare selger 100 poliser vil sannsynligheten for å få noen krav være liten. Men på den andre siden vil et enkelt krav fort være større enn verdien av alle polisene (så selskapet går konkurs hvis det blir et krav). Hvis man selger mange tusen poliser sier sentralgrenseteoremet at gjennomsnittlig krav være cirka lik forventningsverdien, $\bar{x} \approx \mu$. Slik vil man være (neste helt) sikker på at verdien av alle polisene er større enn kravene.

b) Vi har $\mu_{\bar{x}} = \mu = 600$ og $\sigma_{\bar{x}} = \frac{12000}{\sqrt{50000}} = 53.66563$. Sannsynligheten for at gjennomsnittlig krav skal være større enn prisen for polisen er derfor $P(\bar{x} > 700) = P(Z > \frac{700-600}{53.66563}) = P(Z > 1.86) = 0.0312$. Altså vil dette bar inntreffe ca hvert 30. år.

Oppgave 5.74

a) Vi har $n = 1020$ og $p = 0.41$ som vil si at np og $n(1-p)$ er begge rundt 500. Vi kan derfor trygt bruke Normal-approksimasjonen her. Da har vi at $\mu_{\hat{p}} = 0.41$ og $\sigma_{\hat{p}} = \sqrt{\frac{0.41\cdot(1-0.41)}{1020}} = 0.0154$. Vi har ha sannsynligheten for at estimert $\hat{p}$ ligger innenfor oppgitt feilmargin (på ± 0.04)

$$\begin{aligned} P(0.44 < \hat{p} < 0.52) &= P\left(\frac{0.37 - 0.41}{0.0154} < Z < \frac{0.45 - 0.41}{0.0154}\right) \\ &= P(-2.597 < Z < 2.597) = 0.9953 - 0.0047 = 0.9906. \end{aligned}$$

Vi ser at det er veldig sannsynlig at estimatet ligger innen for en feilmargin på ± 0.04.

b) Herunder gjentar vi bare utregningene for $n = 300$ og $n = 5000$, som vil si at vi vi må regne ut $\sigma_{\hat{p}}$ på nytt.

For $n = 300$ får vi da $P(0.37 < \hat{p} < 0.45) = P(-1.408649 < Z < 1.408649) = 0.841$, og for $n = 5000$ får vi $P(0.44 < \hat{p} < 0.52) = P(-5.750784 < Z < 5.750784) \approx 1$.

c) Altså ser vi at større utvalg gir større sannsynlighet for å gi estimater nærmere den den sanne populasjonsparameteren.

Oppgave 5.76

Antar at den sanne populasjonsparameteren er $p = 0.41$. Vi har da $\sigma_{\hat{p}} = \sqrt{\frac{0.41\cdot(1-0.41)}{n}} = 0.005$. Hvis vi snur om på dette uttrykket får vi at $n = \frac{0.41\cdot(1-0.41)}{0.005^2} = 9676$.

Med en utvalgsstørrelse n=9676, vil omtrent 95% av repeterte uavhengige utvalg (rent hypotetisk) gi $\hat{p}$ innenfor 0.01 av den sanne populasjonsparameteren p.

Oppgave 5.62a fra 9.de utgave av boka

a) Individene er trolig uavhengige av hverandre, og uten mer informasjon kan vi anta at alle har samme sannsynlighet for å fortsette med nedlastingen på $p = 0.33$. Vi kan derfor anta at $X \sim B(50, 0.33)$.

Oppgave 5.64 fra 9.de utgave av boka

a) Forventningsverdien til X vil være $\mu_X = 15 \cdot 0.75 = 11.25$. Forventningsverdien til $\hat{p}$ vil være $\mu_{\hat{p}} = \frac{\mu_X}{n} = 0.75$.

b) For $n = 150$ har vi $\mu_X = 150 \cdot 0.75 = 112.5$ og for $n = 1500$ har vi $\mu_X = 150 \cdot 0.75 = 1125$. På den annen side vil alltid $\mu_{\hat{p}} = 0.75$. Når antallet i utvalget øker (n øker) vil forventet antall øke, mens forventet proporsjon vil alltid være konstant.

Midtveiseksamen Høsten 2009

1) $p_3 = 1 - 0.3 - 0.5 = 0.2$ som gir b).

2) $-1 \cdot \frac{1}{6} + 0 + z \cdot \frac{1}{6} = 1.5$. Det betyr at $z = 1.5 \cdot 6 + 1 = 10$ som gir d).

4) Median er $\frac{x+700}{2} = 650$ som betyr at $x = 600$ som gir d).

5) $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y = 3^2 + 1^2 + 2 \cdot 0.5 \cdot 3 \cdot 1 = 13$ som gir $\sigma_{X-Y} = 3.606$, som gir a).

6) $P(X > 37.5) = P(Z > \frac{37.5-37}{0.3}) = P(Z > \frac{5}{3}) = 0.0478$, som gir c).

7) $P(Z < z) = 0.75$ gir at $z = 0.647$. Vi får derfor temperaturen $0.647 \cdot 0.3 + 37.0 = 37.19$ som gir a).

8) c).

10) Vi har d) fordi

$$\begin{aligned} P(M) &= P(M \cap S) + P(M \cap S^C) \\ &= P(M \mid S)P(S) + P(M \mid S^C)P(S^C) \\ &= 0.95 \cdot 0.007 + 0.035 \cdot (1 - 0.007) \\ &= 0.041405. \end{aligned}$$