8 Utvalgsfordelinger

OPPGAVEHEFTE · STATISTIKK

Overskriftene viser til delkapitlene i boka. Først kommer en oppsummering av hele kapittelet. Under hver overskrift står gjennomregnede eksempler med løsning, og deretter oppgaver du skal løse uten hjelpemidler. Fasiten står bakerst. Bruk tabellene i bilag A.

Formelark

UTVALG OG REGNEREGLER

Parameter og observator μ, σ, p er faste. $\bar{x}$, s, $\hat{p}$ varierer.

Sum (uavhengige) $Var(X+Y) = Var(X) + Var(Y)$

Sum av n

$$n\mu \qquad \sigma\sqrt{n}$$

Gjennomsnitt

$$\mu \qquad \frac{\sigma}{\sqrt{n}}$$

Andel

$$p \qquad \sqrt{\frac{p(1-p)}{n}}$$

> - Legg sammen variansene, ikke standardavvikene.

SENTRALGRENSESETNINGEN OG NORMALTILNÆRMING

Sentralgrensesetningen

$$\bar{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Standardisere et gjennomsnitt

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Normaltilnærming $X \approx N\left(np, \sqrt{np(1-p)}\right)$ når $np \geqslant 10$ og $n(1-p) \geqslant 10$

Halvkorreksjon $P(X \geqslant k) \approx P(Y \geqslant k - 0{,}5)$

> - Sentralgrensesetningen gjelder for store n (tommelfingerregel $n \geqslant 30$), uansett populasjonens form.

8.1.1 Variasjon fra utvalg til utvalg

EKSEMPEL *En liten populasjon*

En gruppe på fire elever sov 6, 7, 8 og 9 timer. Vi trekker to av dem tilfeldig (uten tilbakelegging) og regner ut gjennomsnittet $\bar{x}$.

a) Finn populasjonsgjennomsnittet μ.

b) Skriv opp alle mulige utvalg og gjennomsnittene deres.

c) Lag utvalgsfordelingen til $\bar{x}$, og finn forventningsverdien.

Løsning

a) $\mu = \dfrac{6+7+8+9}{4} = 7{,}5$

b) $\{6,7\}$: 6,5. $\{6,8\}$: 7. $\{6,9\}$: 7,5. $\{7,8\}$: 7,5. $\{7,9\}$: 8. $\{8,9\}$: 8,5.

c)

$\bar{x}$	6,5	7	7,5	8	8,5
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$E(\bar{x}) = \dfrac{6{,}5+7+2\cdot 7{,}5+8+8{,}5}{6} = \dfrac{45}{6} = 7{,}5$, det samme som μ. Gjennomsnittet treffer ikke alltid μ, men i gjennomsnitt treffer det. ∎

8.1 Er tallet en parameter eller en observator?

a) Gjennomsnittshøyden til alle 18-åringer i Norge.

b) Andelen i en spørreundersøkelse som sier de trener ukentlig.

c) Standardavviket i en stikkprøve på 50 batterier.

8.2 Skriv symbolene for gjennomsnitt, standardavvik og andel, både for populasjonen og for utvalget.

8.3 En populasjon består av tallene 2, 4 og 9. Vi trekker to uten tilbakelegging.

a) Finn μ.

b) Finn alle mulige gjennomsnitt.

c) Vis at gjennomsnittet av dem er μ.

8.4 Hva er en utvalgsfordeling?

8.5 To klasser regner ut gjennomsnittlig søvn for 20 tilfeldige elever hver på samme skole og får 7,4 og 7,9 timer. Betyr det at én av klassene har regnet feil?

8.6 Hvilken utvalgsfordeling er smalest: $\hat{p}$ med $n = 100$ eller med $n = 1000$?

8.7 En nettavis spør leserne om de er fornøyde med kommunen, og 8000 svarer. Gir det store antallet et godt anslag for hele befolkningen?

8.8 Hva blir utskriften?

Python

```
pop = [2, 4, 9]
snitt = []
for i in range(3):
    for j in range(i + 1, 3):
        snitt.append((pop[i] + pop[j]) / 2)
print(snitt)
```

8.9 I en simulering med $n = 100$ har $\hat{p}$ standardavvik 0,045. Omtrent hvor stort blir standardavviket med $n = 400$?

8.10 Hvorfor er det nyttig å simulere en utvalgsfordeling?

8.1.2 Summer og gjennomsnitt

EKSEMPEL *To terninger*

Én terning har $\mu = 3{,}5$ og $\text{Var} = \dfrac{35}{12}$ (6.1.3). Finn forventningsverdien og standardavviket til summen av to terninger.

Løsning

$E = 3{,}5 + 3{,}5 = 7$ og $\text{Var} = \dfrac{35}{12} + \dfrac{35}{12} = \dfrac{35}{6} \approx 5{,}83$, så $\sigma = \sqrt{5{,}83} \approx 2{,}42$.

Det stemmer med fordelingen til summen i 6.1.1. Legger vi sammen standardavvikene, får vi $2 \cdot 1{,}71 = 3{,}42$, som er for mye. ∎

EKSEMPEL *Heisen*

Ti personer med vekt $\mu = 80$ kg og $\sigma = 12$ kg går inn i heisen. Vektene er uavhengige.

a) Finn forventningsverdien og standardavviket til den samlede vekten.

b) Finn forventningsverdien og standardavviket til gjennomsnittsvekten.

Løsning

a) $E = 10 \cdot 80 = 800$ kg og $\sigma_{\text{sum}} = 12\sqrt{10} \approx 37{,}9$ kg.

b) $E(\bar{X}) = 80$ kg og $\sigma_{\bar{X}} = \dfrac{12}{\sqrt{10}} \approx 3{,}79$ kg.

I 8.1.3 finner vi sannsynligheten for at heisen blir overbelastet. ∎

EKSEMPEL *Gjennomsnittsalder*

Alderen i Norge har $\mu = 41{,}3$ år og $\sigma = 23{,}4$ år. Finn standardavviket til gjennomsnittsalderen i et tilfeldig utvalg med $n = 10$ og $n = 100$ personer.

Løsning

$n = 10$: $\dfrac{23{,}4}{\sqrt{10}} \approx 7{,}4$ år. $n = 100$: $\dfrac{23{,}4}{10} = 2{,}34$ år.

En simulering med 5000 utvalg gir 7,55 og 2,36, nesten det samme. ∎

EKSEMPEL *Meningsmåling*

Ap har $p = 0{,}28$. Finn standardavviket til $\hat{p}$ i en meningsmåling med 1000 spurte.

Løsning

$$\mathrm{SD}(\hat{p}) = \sqrt{\frac{0{,}28 \cdot 0{,}72}{1000}} \approx 0{,}0142$$

Omtrent 1,4 prosentpoeng. Det stemmer med simuleringen i 8.1.1. ∎

8.11 X og Y er uavhengige med $\mathrm{SD}(X) = 8$ og $\mathrm{SD}(Y) = 6$. Finn $\mathrm{SD}(X + Y)$.

8.12 $\mathrm{E}(X) = 12$ og $\mathrm{E}(Y) = 7$. Finn $\mathrm{E}(X + Y)$ og $\mathrm{E}(3X)$.

8.13 $\mathrm{SD}(X) = 4$. Finn $\mathrm{SD}(5X)$.

8.14 En populasjon har $\mu = 50$ og $\sigma = 20$. Finn forventning og standardavvik for

a) summen av 16 observasjoner,

b) gjennomsnittet av 16 observasjoner.

8.15 Finn standardavviket til summen av fire terningkast. Én terning har $\mathrm{Var} = \frac{35}{12}$.

8.16 $p = 0{,}3$ og $n = 400$. Finn $\mathrm{E}(\hat{p})$ og $\mathrm{SD}(\hat{p})$.

8.17 Hvorfor er standardavviket til summen av to uavhengige variabler mindre enn summen av standardavvikene?

8.18 Hvor mange observasjoner trengs for at standardfeilen skal bli en firedel av σ?

8.19 En pakke har 8 kjeks som hver veier $\mu = 12$ g med $\sigma = 1$ g, uavhengig av hverandre. Finn forventning og standardavvik for vekten av pakken.

8.20 Hva blir utskriften?

Python

```
sigma = 20
for n in [4, 16, 100]:
    print(sigma / n**0.5)
```

8.21 Hvor mange må spørres for at $\mathrm{SD}(\hat{p})$ skal være høyst 0,01 når $p = 0{,}2$?

8.22 Forklar hvorfor regelen for variansen til en sum krever uavhengighet. Gi et eksempel der den ikke gjelder.

8.1.3 Sentralgrensesetningen

EKSEMPEL *Overbelastet heis*

Ti personer går inn i en heis som er godkjent for 900 kg. Vekten til hver person har $\mu = 80$ kg og $\sigma = 12$ kg, og vektene er uavhengige og omtrent normalfordelte. Finn sannsynligheten for at den samlede vekten er over 900 kg.

Løsning

$z = \frac{900 - 800}{37{,}9} \approx 2{,}64$. Tabellen gir $\mathrm{P}(Z \leqslant 2{,}64) = 0{,}9959$, så

$$\mathrm{P}(\mathrm{sum} > 900) = 1 - 0{,}9959 = 0{,}0041$$

Omtrent 4 av 1000 turer med ti personer overbelaster heisen. Heiser har derfor en sikkerhetsmargin langt over 900 kg. ∎

EKSEMPEL *Gjennomsnittsalder*

Vi trekker 50 tilfeldige personer i Norge. Finn sannsynligheten for at gjennomsnittsalderen er over 45 år. Bruk $\mu = 41{,}3$ og $\sigma = 23{,}4$.

Løsning

Etter sentralgrensesetningen er $\bar{X} \approx \mathrm{N}(41{,}3;\, 3{,}31)$, fordi $\frac{23{,}4}{\sqrt{50}} \approx 3{,}31$.

$z = \frac{45 - 41{,}3}{3{,}31} \approx 1{,}12$, og $\mathrm{P}(\bar{X} > 45) = 1 - 0{,}8686 = 0{,}1314$.

Det skjer i omtrent 13 % av utvalgene. ∎

EKSEMPEL *Lykkehjulet på skoleballet*

I 6.1.4 ble lykkehjulet spilt 250 ganger, og det betalte ut gjennomsnittlig 18,4 kr per spill. For ett spill er $\mu = 16$ kr og $\sigma = 30{,}4$ kr. Hvor sannsynlig er det at gjennomsnittet blir minst 18,4 kr?

Løsning

$\bar{X} \approx \mathrm{N}(16;\, 1{,}92)$, fordi $\frac{30{,}4}{\sqrt{250}} \approx 1{,}92$.

$z = \frac{18{,}4 - 16}{1{,}92} = 1{,}25$, og $\mathrm{P}(\bar{X} \geqslant 18{,}4) = 1 - 0{,}8944 = 0{,}1056$.

Det skjer omtrent hver tiende gang, så resultatet på skoleballet var ikke spesielt uvanlig. Elevrådet hadde bare litt uflaks. ∎

8.23 Hva sier sentralgrensesetningen?

8.24 En populasjon har $\mu = 60$ og $\sigma = 15$, og $n = 36$. Finn $\mathrm{P}(\bar{X} > 64)$.

8.25 Ventetiden i en kundeservice har $\mu = 5$ min og $\sigma = 4$ min, og fordelingen er skjev. Finn sannsynligheten for at gjennomsnittet for 64 kunder er over 5,5 min.

8.26 Hvorfor kan vi ikke bruke sentralgrensesetningen til å finne sannsynligheten for at én kunde venter over 5,5 min?

8.27 Vekten av epler er $N(150, 20)$ gram. En pose har 9 epler. Finn sannsynligheten for at posen veier over 1400 gram.

8.28 En heis tåler 1000 kg. Tolv personer med $\mu = 80$ kg og $\sigma = 12$ kg går inn. Finn sannsynligheten for overbelastning.

8.29 Når gjelder $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ for alle n, også små?

8.30 Hva blir utskriften?

Python

```
mu = 60
sigma = 15
n = 36
xbar = 64
z = (xbar - mu) / (sigma / n**0.5)
print(z)
```

8.31 Et lodd gir 0 kr med sannsynlighet 0,99 og 1000 kr ellers. Er $n = 30$ nok til at gjennomsnittet er normalfordelt? Begrunn.

8.32 Poengsummen på en prøve har $\mu = 42$ og $\sigma = 12$. Finn sannsynligheten for at gjennomsnittet i en tilfeldig gruppe på 64 elever er under 40.

8.33 Mellom hvilke verdier ligger gjennomsnittet i oppgaven over med 95 % sannsynlighet?

8.1.4 Normaltilnærming

EKSEMPEL *Ap i en meningsmåling*

Vi spør 1000 tilfeldige velgere, og $p = 0{,}28$. Finn sannsynligheten for at minst 300 sier Ap.

Løsning

$np = 280$ og $n(1-p) = 720$, begge over 10. $\mu = 280$ og $\sigma = \sqrt{1000 \cdot 0{,}28 \cdot 0{,}72} \approx 14{,}2$.

$z = \dfrac{300 - 280}{14{,}2} \approx 1{,}41$, og $P(X \geqslant 300) \approx 1 - 0{,}9207 = 0{,}0793$.

Den eksakte binomiske sannsynligheten er 0,0855 (`1 - binom.cdf(299, 1000, 0.28)`).

■

EKSEMPEL *Gutter i Tromsø*

I Tromsø ble det født 854 barn i 2025, og 431 var gutter. Med $p = 0{,}514$ er $X \sim B(854; 0{,}514)$, $\mu \approx 439{,}0$ og $\sigma \approx 14{,}6$ (6.2.3). Finn sannsynligheten for høyst 431 gutter, med og uten halvkorreksjon.

Løsning

Uten: $z = \dfrac{431 - 439}{14{,}6} \approx -0{,}55$, og $1 - 0{,}7088 = 0{,}2912$. Med: $z = \dfrac{431{,}5 - 439}{14{,}6} \approx -0{,}51$, og $1 - 0{,}6950 = 0{,}3050$.

Det eksakte svaret er 0,3048. Halvkorreksjonen gir et svært godt svar.

■

EKSEMPEL *Lav oppslutning*

Hva er sannsynligheten for at en meningsmåling med 1000 spurte gir Ap 25 % eller mindre, når $p = 0{,}28$?

Løsning

$z = \dfrac{0{,}25 - 0{,}28}{0{,}0142} \approx -2{,}11$, og $P(\hat{p} \leqslant 0{,}25) = 1 - 0{,}9826 = 0{,}0174$.

Det skjer i under 2 % av målingene. Får en meningsmåling Ap ned i 25 %, er det trolig fordi oppslutningen faktisk har falt, ikke bare tilfeldigheter. Slike spørsmål er tema for kapittel 9 og 10.

■

8.34 Er betingelsen for normaltilnærming oppfylt?

a) $B(40; 0{,}5)$

b) $B(100; 0{,}05)$

c) $B(500; 0{,}03)$

8.35 $X \sim B(500; 0{,}4)$. Finn μ og σ.

8.36 Bruk oppgaven over og finn $P(X \geqslant 215)$ uten og med halvkorreksjon.

8.37 Du kaster en terning 60 ganger. Bruk normaltilnærming med halvkorreksjon og finn sannsynligheten for minst 15 seksere.

8.38 30 % av velgerne støtter et forslag. En meningsmåling spør 400. Finn sannsynligheten for at $\hat{p}$ blir over 0,33.

8.39 Hvorfor gir halvkorreksjon et bedre svar?

8.40 Skriv grensen med halvkorreksjon:

a) $P(X \leqslant 40)$

b) $P(X > 40)$

c) $P(X < 40)$

8.41 Hva blir utskriften?

Python

```
n = 500
p = 0.4
print(n * p >= 10 and n * (1 - p) >= 10)
print((n * p * (1 - p))**0.5 > 10)
```

8.42 Hvorfor passer normaltilnærming dårlig for $B(20; 0{,}05)$?

8.43 Det blir født 2000 barn i en by. Bruk $p = 0{,}514$ og finn sannsynligheten for at minst 1050 er gutter (med halvkorreksjon).

Fasit

8.1.1 Variasjon fra utvalg til utvalg

8.1
a) Parameter b) Observator c) Observator

8.2
Populasjon: μ, σ, p. Utvalg: $\bar{x}$, s, $\hat{p}$.

8.3
a) 5 b) 3, 5,5 og 6,5 c) $\frac{3+5{,}5+6{,}5}{3} = 5$

8.4
Fordelingen av verdiene en observator (for eksempel $\bar{x}$ eller $\hat{p}$) kan få når vi trekker alle mulige utvalg av samme størrelse fra populasjonen.

8.5
Nei. Gjennomsnittet varierer fra utvalg til utvalg av tilfeldige grunner.

8.6
Med $n = 1000$. Større utvalg gir mindre variasjon.

8.7
Nei, ikke nødvendigvis. Utvalget er selvselektert og trolig skjevt. Et stort utvalg gir lite tilfeldig variasjon, men retter ikke opp skjevhet.

8.8
`[3.0, 5.5, 6.5]`

8.9
Halvparten, omtrent 0,022.

8.10
For store populasjoner er det for mange mulige utvalg til å skrive opp. En simulering viser hvordan observatoren varierer, og hvor langt fra parameteren den typisk havner.

8.1.2 Summer og gjennomsnitt

8.11
$\sqrt{64+36} = 10$

8.12
19 og 36

8.13
20

8.14
a) 800 og $20\sqrt{16} = 80$ b) 50 og $\frac{20}{4} = 5$

8.15
$\sqrt{4 \cdot \frac{35}{12}} = \sqrt{11{,}67} \approx 3{,}42$

8.16
0,3 og $\sqrt{\frac{0{,}21}{400}} \approx 0{,}0229$

8.17
Avvikene utjevner hverandre delvis: Når den ene er høy, er den andre ofte ikke like høy. Variansene legges sammen, og $\sqrt{a^2+b^2} < a+b$.

8.18
$\sqrt{n} = 4$, så $n = 16$.

8.19
96 g og $\sqrt{8} \approx 2{,}83$ g

8.20
`10.0`, `5.0` og `2.0`

8.21
$\frac{0{,}16}{n} \leqslant 0{,}0001$ gir $n \geqslant 1600$.

8.22
Hvis variablene henger sammen, utjevner avvikene hverandre mindre (eller mer). Eksempel: høyden til to tvillinger, eller temperaturen to dager på rad.

8.1.3 Sentralgrensesetningen

8.23
Gjennomsnittet (og summen) av mange uavhengige observasjoner er omtrent normalfordelt, uansett hvilken fordeling populasjonen har.
$\bar{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$.

8.24
$\sigma_{\bar{X}} = 2{,}5$, $z = 1{,}6$: $1 - 0{,}9452 = 0{,}0548$

8.25
$\sigma_{\bar{X}} = \frac{4}{8} = 0{,}5$, $z = 1$: $1 - 0{,}8413 = 0{,}1587$

8.26
Setningen gjelder gjennomsnitt av mange observasjoner. Én ventetid har den skjeve fordelingen til populasjonen.

8.27
Summen er $N(1350; 60)$. $z = \frac{50}{60} \approx 0{,}83$: $1 - 0{,}7967 = 0{,}2033$

8.28
Summen har $\mu = 960$ og $\sigma = 12\sqrt{12} \approx 41{,}6$. $z \approx 0{,}96$: $1 - 0{,}8315 = 0{,}1685$

8.29
Når populasjonen selv er normalfordelt.

8.30
`1.6`

8.31
Nei. Populasjonen er svært skjev, og med 30 lodd er gjennomsnittet nesten alltid 0 eller 33,3 kr. Det trengs mye større n.

8.32
$\sigma_{\bar{X}} = 1{,}5$, $z \approx -1{,}33$: $1 - 0{,}9082 = 0{,}0918$

8.33
$42 \pm 1{,}96 \cdot 1{,}5$: fra 39,1 til 44,9.

8.1.4 Normaltilnærming

8.34
a) Ja (20 og 20) b) Nei ($np = 5$) c) Ja (15 og 485)

8.35
$\mu = 200$ og $\sigma = \sqrt{120} \approx 10{,}95$

8.36
Uten: $z \approx 1{,}37$, $1 - 0{,}9147 = 0{,}0853$.
Med: $z = \frac{214{,}5 - 200}{10{,}95} \approx 1{,}32$, $1 - 0{,}9066 = 0{,}0934$.

8.37
$\mu = 10$, $\sigma \approx 2{,}89$. $z = \frac{14{,}5 - 10}{2{,}89} \approx 1{,}56$: $1 - 0{,}9406 = 0{,}0594$

8.38
$SD(\hat{p}) \approx 0{,}0229$, $z \approx 1{,}31$: $1 - 0{,}9049 = 0{,}0951$

8.39
X er diskret, og sannsynligheten for hver verdi svarer til en søyle som går et halvt steg til hver side. Med halvkorreksjon kommer hele søylen med.

8.40
a) $P(Y \leqslant 40{,}5)$ b) $P(Y \geqslant 40{,}5)$ c) $P(Y \leqslant 39{,}5)$

8.41
`True` og `True` ($\sigma \approx 10{,}95$)

8.42
$np = 1$. Fordelingen er svært skjev og klemt inn mot 0, og en normalfordeling ville gitt sannsynlighet for negative verdier.

8.43
$\mu = 1028$, $\sigma = \sqrt{2000 \cdot 0{,}514 \cdot 0{,}486} \approx 22{,}35$. $z = \frac{1049{,}5 - 1028}{22{,}35} \approx 0{,}96$: $1 - 0{,}8315 = 0{,}1685$