7 Kontinuerlige fordelinger

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 tabellen over standard normalfordeling i bilag A, og tegn en skisse av kurven til hver oppgave.

Formelark

KONTINUERLIGE FORDELINGER

Sannsynlighet som areal $P(a \leqslant X \leqslant b)$ er arealet under tetthetsfunksjonen.

Én verdi

$$P(X = a) = 0$$

Uniform fordeling

$$f(x) = \frac{1}{b-a}$$

Uniform: forventning og standardavvik

$$\mu = \frac{a+b}{2} \qquad \sigma = \frac{b-a}{\sqrt{12}}$$

- For kontinuerlige variabler er $P(X < a) = P(X \leqslant a)$.

NORMALFORDELINGEN

68–95–99,7-regelen $\mu \pm \sigma$: 68 %. $\mu \pm 2\sigma$: 95 %. $\mu \pm 3\sigma$: 99,7 %.

Standardisering

$$Z = \frac{X-\mu}{\sigma}$$

Symmetri

$$P(Z \leqslant -z) = 1 - P(Z \leqslant z)$$

Grenseverdi

$$x = \mu + z \cdot \sigma$$

Python `norm.cdf(x, mu, sigma)` og `norm.ppf(p, mu, sigma)`

- Tabellen i bilag A gir arealet til venstre for z.
- Sjekk alltid om normalfordelingen passer, særlig i halene.

7.1.1 Fra histogram til tetthetsfunksjon

7.1 Er variabelen diskret eller kontinuerlig?

a) Vekten til et eple.

b) Antall epler i en pose.

c) Tiden det tar å løpe 3000 meter.

7.2 I et histogram over medianinntekt i kommunene har klassen 650–700 (tusen kr) 108 av 357 kommuner.

a) Finn andelen.

b) Finn tettheten (høyden på søylen) når arealet skal være andelen.

7.3 $f(x) = 0{,}25$ for $0 \leqslant x \leqslant 4$. Vis at f er en tetthetsfunksjon, og finn $P(1 \leqslant X \leqslant 3)$.

7.4 $f(x) = \dfrac{x}{8}$ for $0 \leqslant x \leqslant 4$.

a) Vis at det totale arealet er 1.

b) Finn $P(X \leqslant 2)$.

c) Finn $P(X > 3)$.

7.5 X er kontinuerlig. Hva er $P(X = 2)$? Forklar.

7.6 Kan $f(x) = 2$ for $0 \leqslant x \leqslant 0{,}5$ være en tetthetsfunksjon?

7.7 Kan $f(x) = 0{,}3$ for $0 \leqslant x \leqslant 3$ være en tetthetsfunksjon?

7.8 Forklar hvorfor $P(X < 5) = P(X \leqslant 5)$ når X er kontinuerlig, men ikke når X er antall kron på ti myntkast.

7.9 Hva blir utskriften?

Python

```
tider = [12.4, 9.8, 15.1, 11.0,
         10.5, 13.9, 8.7, 12.0]
antall = 0
for t in tider:
    if 10 <= t < 13:
        antall = antall + 1
print(antall / len(tider))
```

7.10 Et histogram har klassene 0–10 med andel 0,4 og 10–30 med andel 0,6. Finn tettheten i hver klasse. Hvorfor er ikke søylen for 10–30 høyest?

7.11 Finn medianen når $f(x) = \dfrac{x}{8}$ for $0 \leqslant x \leqslant 4$.

7.1.2 Uniform kontinuerlig fordeling

7.12 X er uniformt fordelt fra 0 til 20. Finn

a) $f(x)$,

b) $P(X \leqslant 5)$,

c) $P(X > 15)$,

d) μ.

7.13 X er uniformt fordelt fra 2 til 8. Finn $P(3 \leqslant X \leqslant 5)$, μ og σ.

7.14 U er tallet fra `rng.random()`. Finn $P(U > 0{,}7)$.

7.15 En buss går hvert 12. minutt, og du kommer på et tilfeldig tidspunkt. Finn sannsynligheten for å vente mer enn 8 minutter.

7.16 Finn standardavviket til ventetiden i oppgaven over.

7.17 X er uniformt fordelt fra 0 til b, og $P(X \leqslant 3) = 0{,}2$. Finn b.

7.18 En målestokk viser lengder rundet av til nærmeste millimeter. Avrundingsfeilen er uniform fra $-0{,}5$ til 0,5 mm. Finn sannsynligheten for at feilen er mindre enn 0,1 mm i absoluttverdi.

7.19 Hvilket uttrykk gir et tall uniformt fordelt fra 5 til 9?

7.20 Hva blir utskriften?

Python

```
u = [0.12, 0.55, 0.81, 0.33, 0.97]
antall = 0
for x in u:
    if x < 0.5:
        antall = antall + 1
print(antall)
```

7.21 Gi et eksempel på en situasjon der en uniform kontinuerlig modell passer, og en der den ikke passer.

7.22 To uavhengige tall er uniformt fordelt fra 0 til 1. Finn sannsynligheten for at begge er under 0,5.

7.2.1 Normalfordelingen

EKSEMPEL *IQ*

IQ-tester er laget slik at resultatene i befolkningen er normalfordelte med $\mu = 100$ og $\sigma = 15$.

a) Mellom hvilke verdier ligger de midterste 95 %?

b) Hvor mange prosent har IQ over 115?

c) Hvor mange prosent har IQ under 55?

Løsning

a) $100 - 2 \cdot 15 = 70$ til $100 + 2 \cdot 15 = 130$.

b) 68 % ligger mellom 85 og 115, så 32 % ligger utenfor, halvparten på hver side. Over 115: 16 %.

c) $55 = 100 - 3 \cdot 15$. Utenfor $\mu \pm 3\sigma$ ligger 0,3 %, halvparten på hver side: 0,15 %.

■

EKSEMPEL *Medianinntekt i kommunene*

Medianinntekten i de 357 kommunene har gjennomsnitt 663 og standardavvik 61 (tusen kr).

Bruk 68–95–99,7-regelen som om fordelingen var normal, og sammenlikn med dataene: 71,4 % av kommunene ligger innenfor ett standardavvik, 93,6 % innenfor to og 100 % innenfor tre.

Løsning

Ett standardavvik: 602 til 724. To: 541 til 785. Tre: 480 til 846.

Andelene i dataene, 71 %, 94 % og 100 %, ligger nær 68 %, 95 % og 99,7 %. En normalfordeling er en ganske god modell, men ikke perfekt. I 7.2.4 ser vi nærmere på det. ■

7.23 $X \sim N(60, 5)$. Mellom hvilke verdier ligger omtrent 68 %, 95 % og 99,7 % av verdiene?

7.24 $X \sim N(60, 5)$. Bruk 68–95–99,7-regelen og finn

a) $P(X > 65)$,

b) $P(X < 50)$,

c) $P(55 < X < 70)$.

7.25 IQ er $N(100, 15)$. Hvor mange prosent har IQ mellom 85 og 130?

7.26 Hvilken kurve er bredest og lavest: $N(20, 2)$ eller $N(20, 5)$? Begrunn.

7.27 For en normalfordeling: Hvordan ligger gjennomsnittet, medianen og typetallet i forhold til hverandre?

7.28 Forklar hvorfor summen av ti terninger er omtrent normalfordelt, mens én terning ikke er det.

7.29 Vekten av brødposer er $N(500, 4)$ gram. Hvor mange prosent veier under 492 gram?

7.30 Høyden til 120 elever er omtrent $N(172, 8)$. Omtrent hvor mange elever er mellom 164 og 180 cm?

7.31 Hva blir utskriften?

Python

```
mu = 60
sigma = 5
data = [58, 66, 61, 53, 64, 59]
inne = 0
for x in data:
    if abs(x - mu) <= sigma:
        inne = inne + 1
print(inne)
```

7.32 Eva skriver $X \sim N(10, 4)$ når variansen er 4. Hva er feil?

7.33 Hvor stor andel av verdiene i en normalfordeling ligger mer enn tre standardavvik fra μ?

7.2.2 Standardisering og tabeller

EKSEMPEL *IQ med tabell*

$X \sim N(100, 15)$. Finn

a) $P(X \leqslant 120)$,

b) $P(X > 130)$,

c) $P(90 \leqslant X \leqslant 110)$.

Løsning

a) $z = \dfrac{120 - 100}{15} \approx 1{,}33$. Tabellen gir $P(Z \leqslant 1{,}33) = 0{,}9082$.

b) $z = \dfrac{30}{15} = 2{,}00$. $P(Z > 2) = 1 - 0{,}9772 = 0{,}0228$.

c) $z = \pm\, 0{,}67$. $P(Z \leqslant 0{,}67) = 0{,}7486$ og $P(Z \leqslant -0{,}67) = 1 - 0{,}7486 = 0{,}2514$. Svaret er $0{,}7486 - 0{,}2514 = 0{,}4972$. ■

7.34 $X \sim N(100, 15)$. Finn z-verdien for $x = 130$ og $x = 91$.

7.35 Bruk tabellen. Finn

a) $P(Z \leqslant 1{,}5)$,

b) $P(Z > 0{,}84)$,

c) $P(Z \leqslant -1{,}2)$,

d) $P(-0{,}5 \leqslant Z \leqslant 1{,}5)$.

7.36 $X \sim N(170, 8)$. Finn

a) $P(X \leqslant 180)$,

b) $P(X > 186)$,

c) $P(160 \leqslant X \leqslant 175)$.

7.37 Ola fikk 64 på en prøve med $\mu = 52$ og $\sigma = 8$, og 81 på en annen med $\mu = 72$ og $\sigma = 5$. Hvilken prøve gikk best sammenliknet med de andre?

7.38 Mia regner ut $z = \dfrac{60 - 50}{25}$ for $X \sim N(50, 5)$. Hva er feil?

7.39 Bruk modellen $N(663, 61)$ for medianinntekten i kommunene. Finn sannsynligheten for over 750 000 kr. I dataene er andelen 8,7 %. Sammenlikn.

7.40 $X \sim N(3500, 450)$. Finn $P(X < 3000)$.

7.41 Hva blir utskriften?

Python

```
mu = 50
sigma = 5
for x in [45, 60, 52.5]:
    print((x - mu) / sigma)
```

7.42 Finn $P(Z < -1{,}96 \text{ eller } Z > 1{,}96)$.

7.43 Melkekartonger er $N(1010, 5)$ ml. Finn sannsynligheten for under 1000 ml.

7.44 $X \sim N(80, 10)$. Hvilken verdi har $z = -0{,}5$?

7.2.3 Normalfordeling med digitale verktøy

EKSEMPEL *Melkekartonger*

En maskin fyller kartonger med $\mu = 1010$ ml og $\sigma = 6$ ml.

a) Hvor stor andel av kartongene inneholder mindre enn 1000 ml?

b) Produsenten vil at høyst 1 % skal inneholde mindre enn 1000 ml. Hva må μ stilles inn på når σ er den samme?

Løsning

a) `norm.cdf(1000, 1010, 6)` gir 0,0478, altså nesten 5 %. Med tabell: $z \approx -1{,}67$, og $1 - 0{,}9525 = 0{,}0475$.

b) Arealet 0,01 til venstre gir $z = -2{,}33$ (tabellen: $P(Z \leqslant 2{,}33) = 0{,}9901$). Da er $1000 = \mu - 2{,}33 \cdot 6$, og $\mu = 1000 + 13{,}98 \approx 1014$ ml. Med Python: `1000 - norm.ppf(0.01) * 6` gir 1013,96. ∎

EKSEMPEL *De fattigste kommunene*

Bruk modellen $X \sim N(663, 61)$. Finn grensen for de 10 % kommunene med lavest medianinntekt, og sammenlikn med dataene, der den tiende persentilen er 593,6.

Løsning

`norm.ppf(0.10, 663, 61)` gir 584,8. Med tabell: $z = -1{,}28$, og $663 - 1{,}28 \cdot 61 \approx 584{,}9$.

Modellen gir en grense på omtrent 585 000 kr, mens dataene gir 594 000 kr. Igjen ligger modellen litt for lavt i venstre ende. ∎

7.45 Bruk tabellen og finn z slik at $P(Z \leqslant z) = 0{,}75$.

7.46 Finn z slik at de 10 % høyeste verdiene ligger over z.

7.47 IQ er $N(100, 15)$. Hvor høy IQ må du ha for å være blant de 5 % høyeste?

7.48 $X \sim N(170, 8)$. Finn grensen for de 20 % laveste.

7.49 Hva gir `norm.cdf(180, 170, 8)` og `norm.ppf(0.9, 170, 8)`? Svar omtrentlig med tabellen.

7.50 Levetiden til et batteri er $N(8, 1{,}5)$ år. Hvor lang garanti kan produsenten gi hvis høyst 2 % skal ryke innenfor garantitiden?

7.51 En maskin fyller med $\sigma = 5$ ml. Hva må μ være for at bare 2,5 % skal inneholde under 1000 ml?

7.52 $X \sim N(50, \sigma)$. Bare 5 % av verdiene er over 60. Finn σ.

7.53 Mellom hvilke IQ-verdier ligger de midterste 90 %?

7.54 Hva blir utskriften?

Python

```
from scipy.stats import norm
print(norm.ppf(0.5, 170, 8))
```

7.55 Hva gir `norm.ppf(0.975)` (standard normalfordeling)?

7.2.4 Passer modellen?

EKSEMPEL *Hva gjør vi da?*

Hvis normalfordelingen ikke passer, kan vi

- bruke dataene direkte, for eksempel finne andelen eller persentilen i dataene i stedet for i modellen,
- bruke en annen modell (ikke pensum),
- se på gjennomsnitt av mange observasjoner, som ofte er normalfordelt selv om dataene ikke er det (kapittel 8).

Hva er best for andelen kommuner med over 45 % høyere utdanning?

Løsning

Vi har data for alle 357 kommunene, så vi teller dem: 9 kommuner, altså 2,5 %. Modellen trengs ikke, og den ville gitt feil svar. ∎

7.56 For et datasett er $\bar{x} = 120$ og medianen 118. For et annet er $\bar{x} = 120$ og medianen 95. Hvilket datasett kan være normalfordelt?

7.57 Skriv sjekklisten for å vurdere om en normalfordeling passer.

7.58 Antall timer elevene bruker på mobilen per dag har $\bar{x} = 3$ og $s = 2$. Kan det være normalfordelt?

7.59 I et datasett ligger 70 % innenfor ett standardavvik, 96 % innenfor to og alle innenfor tre. Passer en normalfordeling?

7.60 Andelen med høyere utdanning i kommunene har en lang hale mot høyre. Hva betyr det for en normalfordelingsmodell?

7.61 Hvorfor er halene mest følsomme for valg av modell?

7.62 Hva blir utskriften?

Python

```
data = [2, 3, 3, 4, 4, 4, 5, 5, 6, 14]
data.sort()
n = len(data)
median = (data[4] + data[5]) / 2
snitt = sum(data) / n
print(median, snitt)
```

7.63 Hva kan du gjøre når en normalfordeling ikke passer?

7.64 Reaksjonstider har ofte noen svært lange verdier. Hvilken form får histogrammet, og hvordan påvirker det gjennomsnittet?

7.65 Hvorfor er det vanskelig å avgjøre om data er normalfordelte når det bare er 15 observasjoner?

7.3.1 Undersøkelse: Fødes det like mange barn hver måned?

7.66 I 2025 ble det født 55 401 barn, og 5206 av dem i juli. Bruk modellen der alle dager er like sannsynlige.

a) Finn forventet antall fødsler i juli.

b) Finn standardavviket.

c) Finn z-verdien.

7.67 I juni 2025 ble det født 4904 barn. Finn z-verdien. Juni har 30 dager.

7.68 Oktober 2025 hadde 4696 fødsler og $z = -0{,}1$. Betyr det at modellen passer for oktober?

7.69 Hvorfor er det lettere å se at fødslene ikke er jevnt fordelt med tall for hele landet enn med tall for en liten kommune?

Fasit

7.1.1 Fra histogram til tetthetsfunksjon

7.1
a) Kontinuerlig b) Diskret c) Kontinuerlig

7.2
a) $\frac{108}{357} \approx 0{,}303$ b) $\frac{0{,}303}{50} \approx 0{,}0061$

7.3
Arealet er $4 \cdot 0{,}25 = 1$. $P(1 \leqslant X \leqslant 3) = 2 \cdot 0{,}25 = 0{,}5$

7.4
a) Trekant med grunnlinje 4 og høyde $f(4) = 0{,}5$: $\frac{4 \cdot 0{,}5}{2} = 1$.

b) Trekant med grunnlinje 2 og høyde 0,25: 0,25.

c) $P(X \leqslant 3) = \frac{3 \cdot 0{,}375}{2} = 0{,}5625$, så $P(X > 3) = 0{,}4375$.

7.5
0. En enkelt verdi har ikke noe areal under kurven.

7.6
Ja. Arealet er $0{,}5 \cdot 2 = 1$. At $f(x)$ er større enn 1, er ikke noe problem, fordi $f(x)$ ikke er en sannsynlighet.

7.7
Nei. Arealet er 0,9, ikke 1.

7.8
For en kontinuerlig variabel er $P(X = 5) = 0$. For antall kron er $P(X = 5) > 0$, så $P(X \leqslant 5) = P(X < 5) + P(X = 5)$ er større.

7.9
`0.5`. Verdiene 12,4, 11,0, 10,5 og 12,0.

7.10
$\frac{0{,}4}{10} = 0{,}04$ og $\frac{0{,}6}{20} = 0{,}03$. Den andre klassen er dobbelt så bred, så andelen fordeles over et lengre intervall.

7.11
$P(X \leqslant m) = \frac{m \cdot \frac{m}{8}}{2} = \frac{m^2}{16} = 0{,}5$, så $m^2 = 8$ og $m = \sqrt{8} \approx 2{,}83$.

7.1.2 Uniform kontinuerlig fordeling

7.12
a) 0,05 b) 0,25 c) 0,25 d) 10

7.13
$\frac{2}{6} = \frac{1}{3}$, $\mu = 5$ og $\sigma = \frac{6}{\sqrt{12}} \approx 1{,}73$.

7.14
0,3

7.15
$\frac{4}{12} = \frac{1}{3}$

7.16
$\frac{12}{\sqrt{12}} = \sqrt{12} \approx 3{,}46$ minutter

7.17
$\frac{3}{b} = 0{,}2$, så $b = 15$.

7.18
Intervallet $-0{,}1$ til 0,1 har bredde 0,2: $\frac{0{,}2}{1} = 0{,}2$.

7.19
```
5 + 4 * rng.random()
```

7.20
2

7.21
Passer: ventetid på en buss med fast intervall når du kommer tilfeldig. Passer ikke: høyden til voksne, der verdier nær gjennomsnittet er mye vanligere.

7.22
$0{,}5 \cdot 0{,}5 = 0{,}25$

7.2.1 Normalfordelingen

7.23
55–65, 50–70 og 45–75.

7.24
a) 16 % b) 2,5 % c) 34 % + 47,5 % = 81,5 %

7.25
34 % + 47,5 % = 81,5 %

7.26
$N(20, 5)$. Større standardavvik gir en bredere kurve, og siden arealet er 1, blir den lavere.

7.27
De er like, fordi kurven er symmetrisk om μ.

7.28
Summen av mange uavhengige bidrag blir omtrent normalfordelt. Én terning er uniform med seks like sannsynlige verdier.

7.29
$492 = \mu - 2\sigma$, så 2,5 %.

7.30
68 % av 120 ≈ 82 elever.

7.31
4. 66 og 53 ligger mer enn 5 fra 60.

7.32
Andre parameter er standardavviket. Med varians 4 er $\sigma = 2$, så det skal være $N(10, 2)$.

7.33
0,3 %, altså 0,15 % på hver side.

7.2.2 Standardisering og tabeller

7.34
2 og $-0{,}6$

7.35
a) 0,9332 b) $1 - 0{,}7995 = 0{,}2005$ c) $1 - 0{,}8849 = 0{,}1151$ d) $0{,}9332 - 0{,}3085 = 0{,}6247$

7.36
a) $z = 1{,}25$: 0,8944 b) $z = 2$: $1 - 0{,}9772 = 0{,}0228$ c) $z = -1{,}25$ og 0,63: $0{,}7357 - 0{,}1056 = 0{,}6301$

7.37
$z = 1{,}5$ og $z = 1{,}8$. Den andre prøven.

7.38
Hun har delt på variansen. Riktig er $z = \frac{10}{5} = 2$.

7.39
$z = \frac{87}{61} \approx 1{,}43$, og $1 - 0{,}9236 = 0{,}0764$. Modellen gir 7,6 %, litt mindre enn dataene, fordi fordelingen er svakt skjev mot høyre.

7.40
$z \approx -1{,}11$, og $1 - 0{,}8665 = 0{,}1335$.

7.41
`-1.0`, `2.0` og `0.5`

7.42
$2 \cdot (1 - 0{,}9750) = 0{,}05$

7.43
$z = -2$: $1 - 0{,}9772 = 0{,}0228$

7.44
$80 - 0{,}5 \cdot 10 = 75$

7.2.3 Normalfordeling med digitale verktøy

7.45
$z \approx 0{,}67$ (0,7486)

7.46
Arealet til venstre er 0,90: $z \approx 1{,}28$.

7.47
$z = 1{,}645$: $100 + 1{,}645 \cdot 15 \approx 124{,}7$

7.48
$P(Z \leqslant 0{,}84) = 0{,}7995$, så $z \approx -0{,}84$: $170 - 0{,}84 \cdot 8 \approx 163{,}3$.

7.49
$P(X \leqslant 180) \approx 0{,}894$ og grensen med 90 % til venstre, $170 + 1{,}28 \cdot 8 \approx 180{,}2$.

7.50
$z \approx -2{,}05$: $8 - 2{,}05 \cdot 1{,}5 \approx 4{,}9$ år.

7.51
$z = -1{,}96$: $\mu = 1000 + 1{,}96 \cdot 5 =$ 1009,8 ml.

7.52
$60 = 50 + 1{,}645\ \sigma$, så $\sigma \approx 6{,}08$.

7.53
$100 \pm 1{,}645 \cdot 15$: fra 75,3 til 124,7.

7.54
170.0. Halvparten av arealet ligger til venstre for μ.

7.55
1,96

7.2.4 Passer modellen?

7.56
Det første. I det andre er gjennomsnittet mye høyere enn medianen, så fordelingen er trolig skjev mot høyre.

7.57
Histogram (symmetrisk, én topp, følger kurven), gjennomsnitt mot median, andelene innenfor 1, 2 og 3 standardavvik, uteliggere, og konteksten.

7.58
Neppe. $\bar{x} - 2s = -1$, og tid kan ikke være negativ. Fordelingen er trolig skjev mot høyre.

7.59
Ja, omtrent. Andelene ligger nær 68 %, 95 % og 99,7 %. Vi bør også se på histogrammet.

7.60
Modellen gir for få kommuner med høy andel. Sannsynligheter for høye verdier blir for små.

7.61
Der er det få observasjoner, og små endringer i formen gir store relative forskjeller i sannsynlighetene.

7.62
4.0 5.0. Gjennomsnittet er større enn medianen på grunn av 14.

7.63
Bruke dataene direkte (andeler og persentiler i dataene), bruke en annen modell, eller se på gjennomsnitt av mange observasjoner.

7.64
Skjev mot høyre. De lange verdiene trekker gjennomsnittet opp over medianen.

7.65
Med få observasjoner varierer histogrammet og andelene mye av tilfeldige grunner, så selv normalfordelte data kan se ujevne ut.

7.3.1 Undersøkelse: Fødes det like mange barn hver måned?

7.66
a) $55\,401 \cdot \dfrac{31}{365} \approx 4705$

b) $\sqrt{55\,401 \cdot \dfrac{31}{365} \cdot \dfrac{334}{365}} \approx 65{,}6$

c) $\dfrac{5206 - 4705}{65{,}6} \approx 7{,}6$

7.67
$\mu \approx 4554$ og $\sigma \approx 64{,}6$, så $z \approx 5{,}4$.

7.68
Oktober passer med modellen, men én måned er ikke nok. De fleste månedene ligger langt utenfor $\mu \pm 2\sigma$, så modellen passer ikke for året.

7.69
Med mange fødsler er den tilfeldige variasjonen liten sammenliknet med antallet (standardavviket vokser bare med $\sqrt{n}$). Med få fødsler drukner mønsteret i tilfeldig variasjon.