STK1000: Løsningsforslag Uke 35

H2023

Check in (oppgave i tekst s.6) 1.4

Individene (cases) i dette datasettet er leiligheter (apartments) og vi har 6 variabler:

- Husleie: Kvantitativ
- Inkludert kabel-TV : Kategorisk
- Lov med Kjæledyr : Kategorisk
- Antall soverom: Kvantitativ
- Antall bad: Kvantitativ
- Avstand til campus: Kvantitativ

Kap 1.1 Oppgave 1.8

a) Her er statene individene (cases).

c) Her er både "antall studenter fra staten som går på college" og "antall studenter som går på college i hjemstaten" kvantitative.

Check in (oppgave i tekst s.13) 1.8

```r
x = c(92, 82, 75, 98, 94, 57, 80, 90, 92, 80, 87, 91, 65, 73, 70, 85, 83,
      61, 70, 90, 75, 75, 59, 68, 85, 78, 80, 94)
stem(x, width = 100, scale = 0.8)
```

```
  The decimal point is 1 digit(s) to the right of the |

  5 | 79
  6 | 158
  7 | 0035558
  8 | 00023557
  9 | 00122448
```

Vi ser at fordelingen ikke er symmetrisk, men i stedet venstreskjev.

Oppgave 1.47

En slik forskjell kan forekomme hvis vi har mange husstander med formue under 97300, og noen få med veldig mye større formue. Dette er et godt eksempel på farene med å kun oppgi et gjennomsnitt.

Oppgave 1.87 fra 9.utgave av læreboka

a) Hvis vi velger samme verdi for alle fire tallene får vi et standardavvik på 0.

b) Hvis vi velger 10, 10, 20, 20, får vi det størst mulige standard avvik.

c) I a) finnes det 11 mulige svar: (10, 10, 10, 10), (11, 11, 11, 11), ..., (20, 20, 20, 20). I b) finnes det bare en løsning, med mindre vi er interessert i forskjellige permutasjoner av løsningen: (10, 20, 10, 20), (20, 10, 10, 20), etc.

Oppgave 1.120 fra 8. utgave av læreboka

a) Vi standardiserer normalfordelte variabler x med formelen

$$z = \frac{x - \mu}{\sigma},$$

hvor μ og σ er forventning og standardavvik for fordelingen. Vi får da

```
x = c(62, 93, 54, 76, 73, 98, 64, 55, 80, 71)
z = (x - 72)/10
z
```

```
[1] -1.0  2.1 -1.8  0.4  0.1  2.6 -0.8 -1.7  0.8 -0.1
```

b) $P(Z \geq z) = 0.15$ gir at $P(Z < z) = 1 - 0.15 = 0.85$. Fra Tabell A i boken finner vi at dette gir $z = 1.04$. Så grensen for A i form av en standardisert poengsum er 1.04. Vi kan også regne i R:

```
z_a = qnorm(0.85)
z_a
```

```
[1] 1.036433
```

c) Vi må nå reversere standardiseringen får å få det tilbake på poeng-skalaen. Dette kan vi gjøre med å snu om på uttrykket i a)

$$x = z \cdot \sigma + \mu = 82.4.$$

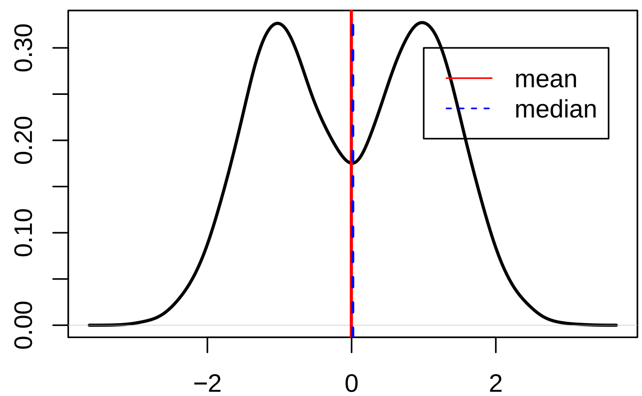
Så grensen for A er 82.4 poeng. Derfor er det to av studentene (93 og 98) som fikk A. Regnet med R:

```
z_a = qnorm(0.85)
z_a * 10 + 72
```

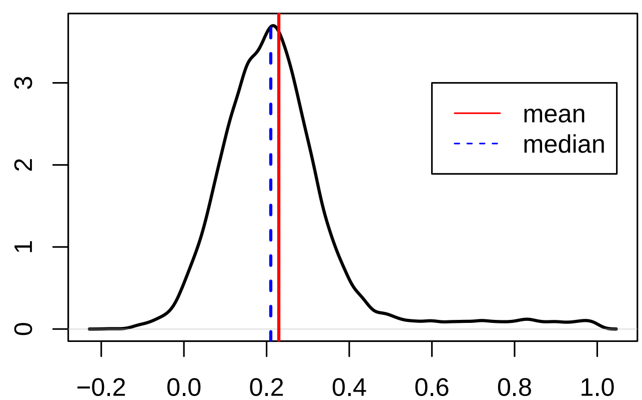
```
[1] 82.36433
```

Oppgave 1.62

a)

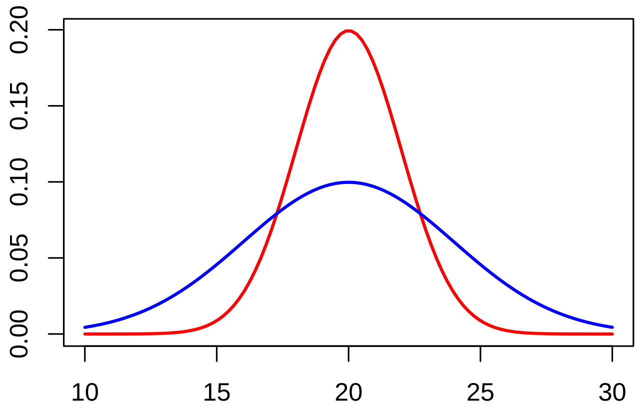


b)



Oppgave 1.63

a) og b)

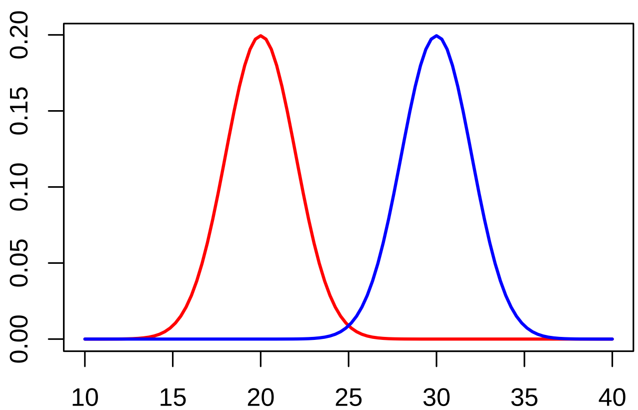


c)

Når vi forandrer standardavviket men holder forventningsverdien konstant ser vi at kurven blir bredere og smalere, men beholder samme midtpunkt.

Oppgave 1.64

a) og b)



c)

Vi ser nå at hvis vi forandrer forventningsverdien mens vi holder standardavviket konstant, forflytter vi kurven langs x-aksen, mens kurven beholder samme form.

Oppgave 1.71

a) 68-95-99.7 regelen tilsvarer avvik på 1, 2 og 3 standardavvik fra gjennomsnittet. Så for gjennomsnitt μ og standardavvik σ kan vi regne ut $\mu \pm \sigma \cdot \{1, 2, 3\}$. Vi får da at

   a) 68 % av kvinner sier mellom 7856 og 20738 ord per dag.

   b) 95 % av kvinner sier mellom 1415 og 27179 ord per dag.

   c) 99.7 % av kvinner sier mellom -5026 og 33620 ord per dag (merk at nedre grense er negativ).

b) Det er selvfølgelig ikke mulig å si et negativt antall ord, så fordelingen er ikke virkelig en normalfordeling.

c) Tilsvarende får vi

   a) 68 % av menn sier mellom 5004 og 23116 ord per dag.

   b) 95 % av menn sier mellom -4052 og 32172 ord per dag.

   c) 99.7 % av menn sier mellom -13108 og 41228 ord per dag.

   Vi får her også et negativt antall ord, som tilsier at fordelingen ikke er helt normal.

d) Vi ser at den gjennomsnittlige kvinne sier litt flere ord en den gjennomsnittlige mann i løpet av en dag, men forskjellene er små. Siden menn har større standardavvik enn kvinner, tilsier det at det er større variasjon mellom menn i antall ord de sier på en dag. En større andel av menn sier mer enn 33000 ord per dag, sammenlignet med kvinner. Vi kan ikke konkludere noe videre.

Oppgave 1.73

a) Figuren viser ikke noen y-akse, men vi vet at alle tetthetsfunksjoner skal ha areal lik 1. Derfor må høyden på funksjonen være 1. $1 * 0.75 = 0.75$, altså ligger 75 % av observasjonene under 0.75 i verdi.

b) 50 % av observasjonene under 0.5 i verdi.

c) $0.75 - 0.5 = 0.25$, altså ligger 25 % av observasjonene mellom 0.5 og 0.75.

d) Alle tetthetsfunksjoner skal ha areal lik 1

Oppgave 1.75

For en uniform fordeling mellom 0 og 1 er forventning og median begge 0.5. Kvartilene er $Q_1 = 0.25$ og $Q_2 = 0.75$.

Oppgave 1.79

Vi kan enten bruke Tabell A fra boken eller kommandoen `pnorm` i R for å regne ut $P(Z < z)$.

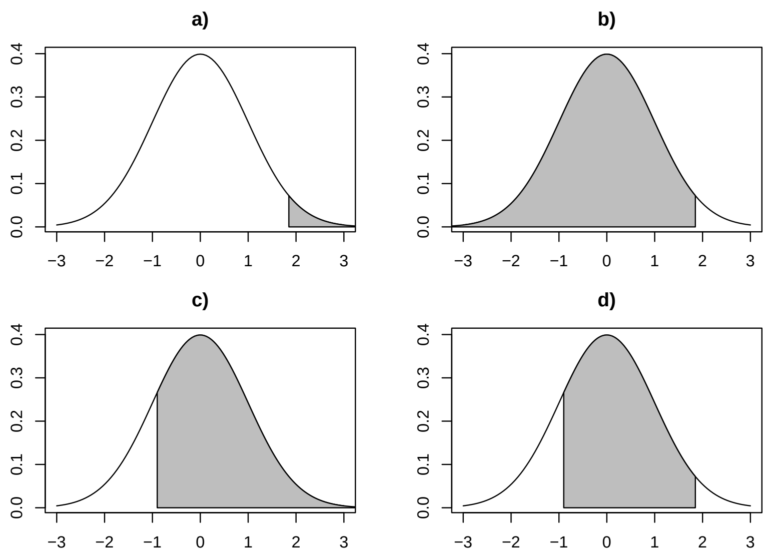
a) $P(Z > 1.85) = 1 - P(Z < 1.85) = 0.0322$.

b) $P(Z < 1.85) = 0.9678$.

c) $P(Z > -.90) = 1 - P(Z < 0.90) = 0.8159$.

d) $P(-0.90 < Z < 1.85) = P(Z < 1.85) - P(Z < -0.90) = 0.7838$.

Vi har skissert arealene i figuren under.



Oppgave 1.82

a) Vi bruker Tabell A i boken og ser at $z = 0.468$ passer.

b) $P(Z > z) = 0.122$ gir oss at $P(Z < z) = 1 - 0.122 = 0.878$, så vi får her $z = 1.165$.

Oppgave 1.95

a) Vi må først standardisere verdien $z = \frac{40-55}{15.5} = -0.97$. Fra tabell A får vi da at $P(Z < -0.97) = 0.1660$, så 16.6 % av kvinner har lave HDL-verdier.

b) $z = \frac{60-55}{15.5} = 0.32$. Fra tabell A har vi da at $P(Z \geq 0.32) = 1 - P(Z < 0.32) = 0.3745$. Altså har 37.45 % av kvinner har beskyttende verdier av HDL.

c) Her bruker vi resultatene fra a) og b). $P(-0.97 < Z < 0.32) = P(Z < 0.32) - P(Z < -0.97) = (1 - 0.3745) - 0.1660 = 0.4595$. Altså har 45.95 % av kvinner har HDL-verdier som ligger mellom 40 og 60 mg/dl.