# Konfidensinterval for middelværdi

Lad os nu antage at vi har som er uafhængige stokastiske variable der er normalfordelt med middelværdi og spredning . Fra dokumentet ’Regneregler for middelværdi og spredning’ har vi at gennemsnittet af de stokastiske variable,

er normalfordelt med middelværdi og spredning . Det medfører at vi har

Hvor det kan ses at jo større er, jo tættere er fordelingen af gennemsnittet koncentreret omkring middelværdien.

### Test for middelværdi

Antag at vi nu har som er uafhængige stikprøver fra en normalfordeling med ukendt middelværdi og ukendt spredning . Vi vil undersøge om middelværdien har en bestemt værdi.

Hvis vi skal undersøge om nulhypotesen: kan forkastes på et signifikansniveau, så starter vi med at estimere middelværdien og spredningen ud fra stikprøverne:

Ud fra teorien ovenfor har vi at acceptområdet for er givet ved

Dvs. vi forkaster nulhypotesen hvis eller .

### Opgave 1 (på klassen)

Vi har et datasæt bestående af 100 stikprøver fra en normalfordeling med ukendt middelværdi og spredning hvor vi har estimeret middelværdien og spredning til

1. Undersøg om nulhypotesen: kan forkastes på signifikansniveau.

### Opgave 2



Vi antager nu at de 200 attenårige mænd udgør en repræsentativ stikprøve for alle attenårige mænd i Hongkong og at deres vægt er normalfordelt.

1. Estimér middelværdien og spredningen af attenårige mænd i Hongkong.
*Tip: Find dokumentet ’Kan et datasæt beskrives med normalfordelingen’.*
2. Undersøg om nulhypotesen nedenfor kan forkastes på et signifikansniveau:

*Middelværdien af vægten af attenårige mænd i Hongkong er .*

### Konfidensinterval for middelværdi

Hvis vi går tilbage til gennemsnittet af de stokastiske variable, , så kan vi i stedet for at undersøge om middelværdien har en bestemt værdi, undersøge hvilke værdier vi ikke kan forkaste. Det vil være de tilfælde hvor er et af de normale udfald:

Hvis vi ser på den første ulighed, så kan den omskrives således:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Ligeledes kan den anden ulighed omskrives således:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Hvis vi kombinerer (1) og (2), så får vi at

Dermed har vi nu et interval som indeholder alle de værdier vi ikke kan forkaste. Men vi har også at sandsynligheden er ca. for at intervallet indeholder middelværdien, så kalder vi intervallet for et  *konfidensinterval*.

Med uafhængige stikprøver som vi brugt til at estimere middelværdien og spredningen, kan vi estimere et konfidensinterval således:

Grunden til at vi ikke gør så meget ud af at sandsynligheden ovenfor hedder og ikke er at vi er nødt til at estimere spredningen og det gør at sandsynligheden bliver ca. . Derfor taler vi også om et estimat af et konfidensinterval.

Bemærk at middelværdien ligger fast og det i stedet er intervallet som forandrer sig hvis vi tager nye stikprøver, og teorien ovenfor viser at ca. af gangene så indeholder intervallet middelværdien.

### Opgave 3

Vi fortsætter her med opgave 2.

1. Estimér et konfidensinterval for middelværdien af vægten af attenårige mænd i Hongkong.
2. Hvis vi antager at estimaterne af middelværdi og spredning er uændret, hvor mange attenårige mænd skal vi så have for at en vægt på ikke er indeholdt i konfidensintervallet?

### Opgave 4

1. Forklar hvorfor bredden på vores estimat af et konfidensinterval er givet ved det nedenstående (som i samfundsfag kaldes den statistiske sikkerhed).

1. Bestem hvor meget bredden ændrer sig hvis vi har 4 gange så mange stikprøver og er uændret.
2. Hvad med 10 gange så mange stikprøver?