

Supplerende opgave til kapitel 3 – Chauvenets kriterium

I kapitel 3 i *Statistik* (s. 56) angav vi Tukeys regel:

Tukeys regel

Et udfald er exceptionelt, hvis det er mere end 1,5 gange kvartilafstanden fra nærmeste kvartil.

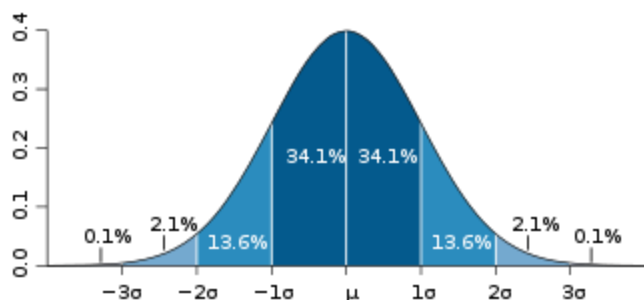
I denne opgave vil vi undersøge to andre mål for hvornår en måling er en exceptionel værdi og sammenligne de tre.

I kapitel 7, opgave 8 giver vi følgende definition:

Exceptionelt udfald: 3 σ -reglen

En værdi, der er mere end tre spredninger fra middelværdien er et exceptionelt udfald.

For en normalfordeling (se figuren nedenfor) betyder det, at vi er langt fra middelværdien, og sandsynligheden for at få sådan et udfald er 0,27 %, hvilket svarer til det yderste interval i hver side. Se mere i kapitel 7 om normalfordelingen.



En tredje, mere kompliceret, metode er:

Chauvenets kriterium for et exceptionelt udfald¹

Antag, at man laver et eksperiment og får n talværdier. Benyt normalfordelingen til at bestemme sandsynligheden p for at få en observation med samme værdi som den mistænkelige værdi (eller større). Multipliser dette tal med n . Hvis facit er mindre end 0,5, er observationen et exceptionelt udfald.

Denne metode er ikke pålidelig for små datasæt.

Eksempel

Vi har målt 9, 10, 10, 10, 11, 50. Dvs. $n = 6$. Værdien 50 ser mere end almindeligt tvivlsom ud. Kan man bortkaste den med 3 σ -reglen, Tukeys regel eller Chauvenets kriterium?

Middelværdien er 16,7. Spredningen er 16,34. 50 er lidt over to spredninger fra middelværdien ($16,7 + 2 \cdot 16,34 = 49,38$). Dvs. at vi ikke kan bortkaste værdien efter 3 σ -reglen.

¹ Kilde: William Chauvenet: *A Manual of Spherical and Practical Astronomy V. II*, 1863.

Kvartilsættet er 10, 10, 11, så kvartilbredden er 1. Tukey vil altså bortkaste alle værdier over 12,5 ($11 + 1,5 \cdot 1$).

Chauvenets metode giver $p = 0,05$, så $n \cdot p$ giver 0,3. $0,3 < 0,5$, så vi bortkaster værdien og får en ny fordeling med $n = 5$, middelværdi 10 og spredning 0,7.

I dette tilfælde gav de tre kriterier ikke samme resultat. Da vi logisk set skal bortkaste 50, konkluderer vi, at 3σ -reglen ikke er pålidelig, når vi kun har få værdier.

Efter bortkast af 50 er kvartilsættet 10, 10, 10. Det bekræfter vores påstand om, at medianen er mere robust overfor exceptionelle værdier end middelværdien, som ændrede sig fra 16,7 til 10.

Øvelse: Exceptionelle værdier

I et fysikforsøg har vi målt følgende værdier for trykket i en beholder i kPa: 101.2, 90.0, 99.0, 102.0, 103.0, 100.2, 89.0, 98.1, 101.5, 102.0.

- Bestem middelværdi og spredning for datasættet.
 - Undersøg, hvilke værdier der forkastes med hhv. Chauvenets kriterium, Tukeys regel og 3σ -reglen.
 - Når du har gennemført analysen og bortkastet de suspekterede data, skal du gentage analysen og se, om der i det nye, mindre datasæt er data, der skal bortkastes, selvom de ikke blev bortkastet i første omgang.
-