# QQ-plot

Vi har i dokumentet ’Kan et datasæt beskrives med normalfordelingen’ set på hvordan man estimér middelværdi og spredning for et datasæt og derefter sammenligner et histogram af datasættet med en normalfordeling med den estimerede middelværdi og spredning. Formålet var at undersøge om datasættet kunne beskrives vha. en normalfordeling. Vi har en smartere må at lave denne undersøgelse på vha. det man kalder et QQ-plot. Til det skal vi bruge den sætning vi beviste i ’Bevis for fordelingsfunktionen’:

### Sætning

Hvis er fordelingsfunktionen for en normalfordeling med middelværdi og spredning , så er

Sætningen kobler Fordelingsfunktionen sammen med fordelingsfunktionen for standardnormalfordelingen . F.eks. har vi at , og hvis vi aflæser på standardnormalfordelingen nedenfor, så får vi at . Hvilket giver at .

**Standardnormalfordelingen (**

Tæthedsfunktion:

Fordelingsfunktion:



**Normalfordelingen**

Tæthedsfunktion:

Fordelingsfunktion:

****

Når vi sammenligner et histogram af datasættet med en normalfordeling, så sammenligner vi frekvenserne i datasættet med en tæthedsfunktion . I stedet kunne vi også sammenligne de kumulerede frekvenser med en fordelingsfunktion . Men hvis vi bruger den inverse funktion af på

så får vi at

Dermed har vi at hvis datasættet tilnærmelsesvis følger en normalfordeling, så får man tilnærmelsesvis en ret linje hvis man tager på de kumulerede frekvenser. Nedenfor er illustreret hvordan det kommer til at se ud, og dette kaldes et QQ-plot.



### Opgave 1 (opgave 2 i Ugemat)

Vi har et datasæt hvor vi gerne vil undersøge om det kan beskrives vha. en normalfordeling. Vi kender en række kumulerede frekvenser for datasættet, f.eks. ved vi at af observationerne har en værdi på eller mindre. Se tabellen nedenfor.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| Kumuleret frekvens til  |  |  |  |  |  |  |

Ud fra datasættet har vi estimeret middelværdien til og spredning til .

Nedenfor er fordelingsfunktionen tegnet for en normalfordeling med denne middelværdi og spredning sammen med de kumulerede frekvenser. Vi kan se at det umiddelbart ser lovende ud, men det er svært at vurdere hvor godt datasættet følger normalfordelingen.



Vi har tidligere vist at hvor er standardnormalfordelingens fordelingsfunktion. Dermed har vi at er en lineær funktion.

1. Tegn grafen for i koordinatsystemet nedenfor.



1. Forklar hvordan man ud fra grafen for kan aflæse og .

For at kunne tegne de kumulerede frekvenser for datasættet sammen med skal vi ligeledes bruge på dem. Vi har f.eks. at den kumulerede frekvens til er . Vha. standardnormalfordelings fordelingsfunktion , får vi at , dvs. . Dette er gjort for alle de kumulerede frekvenser i tabellen nedenfor.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| Kumuleret frekvens til  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

1. Tegn punkterne for i koordinatsystemet ovenfor.
2. Vurdér om datasættet kan beskrives ved en normalfordeling ved at undersøge om den lineære funktion giver en god beskrivelse af punkterne.

### Kommando til QQ-plot

Kommandoen til at tegne et QQ-plot i Maple fungerer således: *QQplot(datasæt).*

### Opgave 2

I ’Kan et datasæt beskrives med normalfordelingen’ undersøgte vi om datasættene Normaldata1 og Normaldata2 kunne beskrives vha. normalfordelinger. Her vil vi gøre det samme vha. QQplot.

1. Gør rede for at datasættet Normaldata1 kan beskrives vha. en normalfordeling.
2. Gør rede for at datasættet Normaldata2 ikke kan beskrives vha. en normalfordeling.