Varians og spredning

Det 2. statistiske moment er den såkaldte *varians* som ofte betegnes ved σ^2 hvor σ kaldes for datasættets *spredning*. Som navnene antyder angiver variansen (og spredningen) hvor meget et datasæt varierer omkring dets middelværdi.

Variansen kan beregnes ud fra datasættet ved:

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + ... + (x_n - \mu)^2}{n}$$

Hvis man i stedet for selve datasættet kender hyppighedsfordelingen af datasættet, dvs. dataværdierne $\mathbb{X} = \{x_1, x_2, ..., x_m\}$ samt den tilhørende hyppighed $h_1, h_2, ..., h_m$, kan man beregne variansen ved:

$$\sigma^2 = \frac{h_1 \cdot (x_1 - \mu)^2 + h_2 \cdot (x_2 - \mu)^2 + ... + h_m \cdot (x_m - \mu)^2}{n}$$

Hvis man i stedet for hyppigheden kender dataværdiernes frekvenser $f_1, f_2, ..., f_m$, kan man beregne variansen ved:

$$\sigma^2 = (x_1 - \mu)^2 \cdot f_1 + (x_2 - \mu)^2 \cdot f_2 + ... + (x_m - \mu)^2 \cdot f_m$$

Ud fra variansen bestemmes spredningen ved $\sigma = \sqrt{\sigma^2}$. Spredningen σ har samme enhed som selve dataværdierne og kan derfor tolkes som et mål for den afstand som dataværdierne i gennemsnit afviger fra middelværdien. Spredningen er derfor også typisk anvendt når man vil kvantificere måleusikkerheden for en målemetode.

Varians og spredning

For et datasæt, der er givet med de n talværdier $x_1, x_2, ..., x_n$, beregnes variansen af datasættet ved gennemsnittet:

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu) + ... + (x_n - \mu)^2}{n}$$

For et datasæt, der er givet ved en hyppighedsfordeling af m dataværdier $x_1, x_2, ..., x_m$ med de tilsvarende hyppigheder $h_1, h_2, ..., h_m$, beregnes variansen ved det vægtede gennemsnit:

$$\sigma^2 = \frac{h_1 \cdot (x_1 - \mu)^2 + h_2 \cdot (x_2 - \mu) + ... + h_m \cdot (x_m - \mu)^2}{n}$$

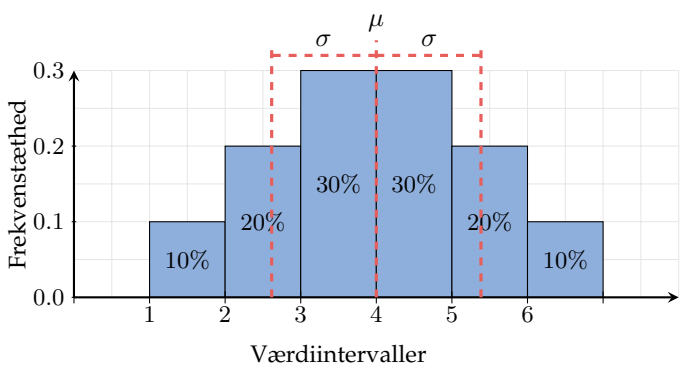
For et datasæt, der er givet ved en frekvensfordeling af m dataværdier $x_1, x_2, ..., x_m$ med de tilsvarende frekvenser $f_1, f_2, ..., f_m$, beregnes variansen ved det vægtede gennemsnit:

$$\sigma^2 = f_1 \cdot (x_1 - \mu)^2 + f_2 \cdot (x_2 - \mu)^2 + ... + f_m \cdot (x_m - \mu)^2$$

Spredningen σ beregnes ud fra variansen ved:

$$\sigma = \sqrt{\sigma^2}$$

Middelværdien repræsenterer det 1. statistiske moment. Begrebet *moment* er helt generelt en matematisk metode til at transformere en fordeling til et sæt af værdier, der beskriver fordelingens egenskaber ud fra en slags ligevægtsprincip. I dette perspektiv angiver det 1. moment det midtpunkt hvor differencen mellem alle dataværdier og midtpunktet summerer til nul. Af denne grund anvendes det 1. moment ligeledes til at bestemme massemidtpunktet (også kaldet tyngdepunktet) af et fysisk objekt hvilket er en god analogi til at skabe sig en intuitiv forståelse af hvad middelværdien repræsenterer.



Geogebra: `Variance(...)`

Først indtastes data i Geogebras regneark (spreadsheet).
Hvis man har selve datasættet indtastet i felterne A1 til A100 kan man beregne variansen ved følgende kommando i Algebra-vinduet:

`Variance(A1 : A100)`

Hvis man i stedet har en fordelingstabel med dataværdierne i felterne A1 til A10 og de tilhørende frekvenser i felterne B1 til B10 anvendes kommandoen:

`Variance(A1 : A10, B1 : B10)`