Middelværdi

Den mest anvendte statistiske deskriptor er *middelværdien* som i daglig tale ofte kaldes *gennemsnittet*. Hvis man f.eks. betragter følgende datasæt:

$$X = (1, 1, 2, 3, 3, 3, 4, 4, 4, 5)$$

Idet alle datasættets værdier er kendte kan middelværdien μ af datasættet direkte beregnes ved ud fra gennemsnittet ved:

$$\mu = \frac{1+1+2+3+3+3+4+4+4+5}{10} = \frac{30}{10} = 3$$

Beregningen af middelværdien ud fra datasættet kan generaliseres ved brug af en generaliseret notation for et datasæt med n dataværdier:

$$X = (x_1, x_2, \dots, x_n)$$

Den tilsvarende generaliserede formel for middelværdien er således:

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Det bemærkes, at der summeres over alle elementerne i datasættet og middelværdien kan derfor kun beregnes direkte ved gennemsnittet hvis man kender hele datasættet.

Gennemsnitsberegningen kan dog omskrives til et såkaldt *vægtet gennemsnit* som kan anvendes hvis man kun har kendskab til datasættets hyppighedsfordeling. For ovenstående datasæt er den tilsvarende hyppighedsfordeling:

Dataværdi	1	2	3	4	5
Hyppighed	2	1	3	3	1

Summen af alle dataværdierne kan beregnes ved at gange hyppighederne med dataværdierne for alle m dataværdier og lægge disse sammen, hvilket giver følgende beregning:

$$\begin{aligned}\mu &= \frac{2\cdot 1 + 1\cdot 2 + 3\cdot 3 + 3\cdot 4 + 1\cdot 5}{10} \\ &= \frac{2+2+9+12+5}{10} = \frac{30}{10} = 3\end{aligned}$$

Den tilsvarende generaliserede formel for middelværdien beregnet ud fra hyppighederne $h_1, h_2, \dots, h_m$ er således:

$$\mu = \frac{h_1 \cdot x_1 + h_2 \cdot x_2 + \dots + h_m \cdot x_m}{n}$$

Bemærk her, at der anvendes n kan beregnes som summen af hyppighederne $n = h_1 + h_2 + \dots + h_m$.

Middelværdien kan også beregnes hvis man i stedet for hyppigheden har kendskab til datasættets frekvensfordeling. For at se dette omskrives det vægtede gennemsnit af hyppighederne ved:

$$\begin{aligned}\mu &= \frac{h_1 \cdot x_1 + h_2 \cdot x_2 + \dots + h_m \cdot x_m}{n} \\ &= \frac{h_1 \cdot x_1}{n} + \frac{h_2 \cdot x_2}{n} + \dots + \frac{h_m \cdot x_m}{n} \\ &= \frac{h_1}{n} \cdot x_1 + \frac{h_2}{n} \cdot x_2 + \dots + \frac{h_m}{n} \cdot x_m \\ &= f_1 \cdot x_1 + f_2 \cdot x_2 + \dots + f_m \cdot x_m\end{aligned}$$

I sidste omskrivning anvendes, at frekvensen beregnes ved $f_i = \frac{h_i}{n}$.

For ovenstående datasæt er den tilsvarende frekvensfordeling:

Dataværdi	1	2	3	4	5
Frekvens	0.2	0.1	0.3	0.3	0.1

Middelværdien beregnes nu ved et vægtet gennemsnit hvor hver dataværdi er vægtet med den tilhørende frekvens:

$$\mu = 0.2 \cdot 1 + 0.1 \cdot 2 + 0.3 \cdot 3 + 0.3 \cdot 4 + 0.1 \cdot 5 = 3$$

For en generel formel anvendes dataværdierne $x_1, x_2, \dots, x_m$ samt den tilhørende observationsfrekvens $f_1, f_2, \dots, f_m$, hvorved middelværdien μ beregnes ved:

$$\mu = f_1 \cdot x_1 + f_2 \cdot x_2 + \dots + f_m \cdot x_m$$

Det bemærkes her, at hvis alle dataværdier behandles som observationsværdier vil antallet af observationværdier være lig med størrelsen af datasættet $m = n$. Hyppigheden af hver dataværdi vil i så fald blive 1, hvilket vil resultere i, at samtlige dataværdier vil have frekvensen af $\frac{1}{n}$. Formlen for middelværdien vil i så fald omskrives til formlen for gennemsnittet.

For en grupperet fordeling hvor dataværdierne er grupperet i intervaller kan middelværdien beregnes ved at anvende midterværdien i intervallerne som $x_1, x_2, \dots, x_m$. Til eksempel betragtes følgende grupperede fordeling hvor midterværdien for værdiintervallet også er angivet:

Værdiinterval	$[0,1[$	$[1,2[$	$[2,3[$	$[3,4[$	$[4,5[$
Midterværdi	0.5	1.5	2.5	3.5	4.5
Frekvens	0.2	0.1	0.3	0.3	0.1

Middelværdien i dette tilfælde beregnes ud fra intervallernes midterværdi, der vægtes med den tilsvarende frekvens for gruppen:

$$\begin{aligned}\mu &= 0.2 \cdot 0.5 + 0.1 \cdot 1.5 + 0.3 \cdot 2.5 + 0.3 \cdot 3.5 + 0.1 \cdot 4.5 \\ &= 2.5\end{aligned}$$

Middelværdi

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

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n}$$

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

$$\mu = \frac{h_1 \cdot x_1 + h_2 \cdot x_2 + \dots + h_m \cdot x_m}{n}$$

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

$$\mu = f_1 \cdot x_1 + f_2 \cdot x_2 + \dots + f_m \cdot x_m$$

For en grupperet fordeling hvor dataværdierne er grupperet i intervaller beregnes middelværdien ved at anvende midterværdien i intervallerne som $x_1, x_2, \dots, x_m$.

Geogebra: `mean(...)`

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

`mean(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:

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