Ugrupperede fordelinger

Datasæt, der består af optællinger, kan repræsenteres grafisk ved såkaldte *ugrupperede fordelinger* hvor hver forskellig dataværdi betragtes som sin egen kategori. Betragtes f.eks. et datasæt, der er fremkommet ved at man for ti husstande har optalt hvor mange fjernsyn hver husstand har, kunne det se således ud:

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

Det tilsvarende sorterede datasæt er således:

$$X_s = \big(1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 5\big)$$

Datasættet indeholder som sagt optællinger hvor dataværdierne i dette tilfælde alle tilhører mængden $\mathbb{X} = \big\{1, 2, 3, 4, 5\big\}$. Hyppigheden af hver dataværdi kan nu optælles i en tabel hvorefter frekvensen samt den kumulerede frekvens kan beregnes:

Værdi	Hyppighed	Frekvens	Kumuleret frekvens
1	2	16.7%	16.7%
2	3	25.0%	41.7%
3	4	33.3%	75.0%
4	2	16.7%	91.7%
5	1	8.3%	100.0%
Sum	$n = 12$	100%	

Hyppigheden h er hvor mange gange den givne dataværdi forekommer i datasættet, *frekvensen* f er hvor stor en andel som dataværdien udgør ift. hele datasættet og beregnes ved:

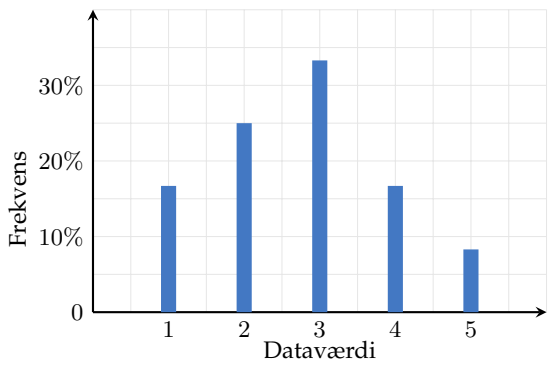
$$f_1 = \frac{h_1}{n} = \frac{2}{12} \approx 0.167 = 16.7\%$$
$$f_2 = \frac{h_2}{n} = \frac{3}{12} = 0.25 = 25\%$$
$$\vdots$$

Den *kumuleret frekvens* k er den løbende sum af frekvenserne op til den givne værdi og beregnes ved:

$$k_1 = f_1 \approx 16.7\%$$
$$k_2 = f_1 + f_2 \approx 16.7\% + 25\% = 41.7\%$$
$$\vdots$$

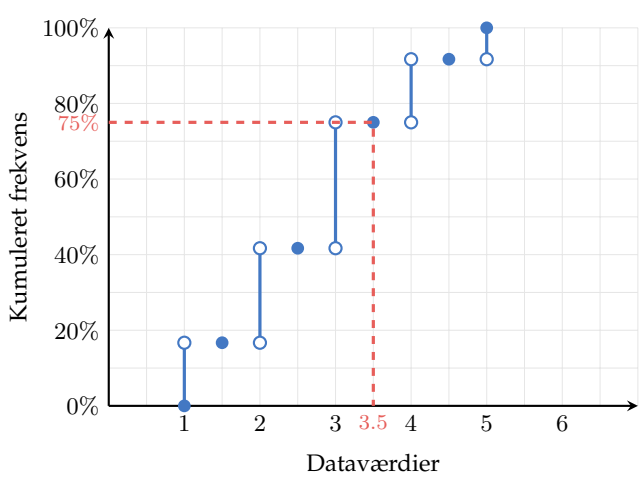
Bemærk at man ud fra tabellen også kan genskabe det sorterede datasæt men altså ikke det oprindelige datasæt idet informationen omkring værdiernes placering er gået tabt i optælling af hyppigheden.

Ud fra tabellen kan man nu opstille yderligere diagrammer som f.eks. søjlediagrammer over dataværdiernes hyppighed eller frekvens:



Af søjlediagrammet der viser datasættets fordeling ses, at størstedelen af datasættets værdier ligger omkrig værdierne 3 og 2 hvorimod værdien 5 er en del mindre repræsenteret.

Den kumulerede frekvens afbilledes ved en såkaldt *sumdiagram*:



Af sumkurven kan således anvendes til at aflæses hvor stor en del af datasættet der ligger under en bestemt dataværdi. I figuren kan f.eks. aflæses ved den stiplede røde linje, at mindst 75% af datasættet må forventes at ligge under en dataværdi på 3.5.

Geogebra: `BarChart(...)`
Først indtastes data i Geogebras regneark (spreadsheet).
Hvis man har selve datasættet indtastet i felterne A1 til A100 kan man plotte datasættets søjlediagram for frekvenstætheden ved følgende kommando:

$$\texttt{BarChart}\big(\texttt{A1} : \texttt{A100}, 0.2, \tfrac{1}{100}\big)$$

Her er talværdien 0.2 tykkelsen på søjlerne og talværdien $\frac{1}{100}$ er den skaleringsfaktoren, der for et datasæt på 100 talværdier omregner hyppigheden til frekvensen.
Hvis man i stedet har en fordelingstabel med grupperne i felterne A1 til A10 og frekvensen i B1 til B10 anvendes kommandoen:

$$\texttt{BarChart}\big(\texttt{A1} : \texttt{A10}, \texttt{B1} : \texttt{B10}, 0.2\big)$$