Deskriptiv statistik (grupperede observationer)

Gym-pakken har en del relevante kommandoer.

deskriptiv statistik

antalobs	boksplot	fraktil	frekvens
frekvensTabel	gennemsnit	grupperData	hyppighed
kvartiler	kumuleretFrekvens	median	middel
plotHistogram	plotPindediagram	plotSumkurve	plotTrappekurve
spredning	sumkurve	trappekurve	typeinterval
typetal	varians		

```
> restart
> with(Gym) :
```

Gruppering af observationerne (intervalhyppigheder)

```
> obs := [21.3, 13.7, 7.4, 13.4, 12.8, 9.2, 8.9, 4.2, 15.5, 11.9, 18.2, 14.1, 10.9, 21.7, 10.1,
    10.2, 4.2, 23.3, 21.7, 9.9, 16.1, 22.4, 8.5, 13.1, 15.3, 19.0, 14.4, 15.6, 18.2, 14.9, 10.8,
    13.7, 11.5, 24.8, 13.7, 14.6, 21.1, 10.1, 24.7, 15.6, 17.2, 12.4, 16.1, 12.9, 15.2, 24.9,
    26.1, 19.4, 19.4, 10.7] :
```

Observationerne grupperes.
Intervallet [0;30] inddeles i 5 lige store intervaller.

```
> gruppeDATA := grupperData(obs, [0..30], 5)
```

$$gruppeDATA := \begin{bmatrix} 0..6. & 2 \\ 6..12. & 13 \\ 12..18. & 20 \\ 18..24. & 11 \\ 24..30 & 4 \end{bmatrix} \quad \textbf{(1.1)}$$

Svaret er en matrix, hvor 1. søjle er intervallerne, og 2. søjle er antal af hver observation, dvs. **intervalhyppigheden.**

Intervalfrekvens og kumuleret intervalfrekvens

```
> frekvens(gruppeDATA)
```

(2.1)

$$\begin{bmatrix} 0..6. & 0.0400 \\ 6..12. & 0.260 \\ 12..18. & 0.400 \\ 18..24. & 0.220 \\ 24..30 & 0.0800 \end{bmatrix} \quad \textbf{(2.1)}$$

Svaret er en matrix, hvor 1. søjle er intervallerne, og 2. søjle er omregnet til %.

> *kumuleretFrekvens*(*gruppeDATA*)

$$\begin{bmatrix} 0..6. & 0.0400 \\ 6..12. & 0.300 \\ 12..18. & 0.700 \\ 18..24. & 0.920 \\ 24..30 & 1. \end{bmatrix} \quad \textbf{(2.2)}$$

Nu er dataene blevet kumuleret.

> *frekvensTabel*(*gruppeDATA*)

observation	hyppighed	frekvens	kumuleret
0 .. 6.	2	4	4
6. .. 12.	13	26	30
12. .. 18.	20	40	70
18. .. 24.	11	22	92
24. .. 30	4	8	100

Typeinterval

> *typeinterval*(*gruppeDATA*)

$$[12..18.] \quad \textbf{(3.1)}$$

Dvs. **typeintervallet** er [12;18], dvs. der er størst intervalhyppighed.

Median

> *median*(*gruppeDATA*)

$$15.000 \quad \textbf{(4.1)}$$

Dvs. **medianen** er 15.0

Kvartiler

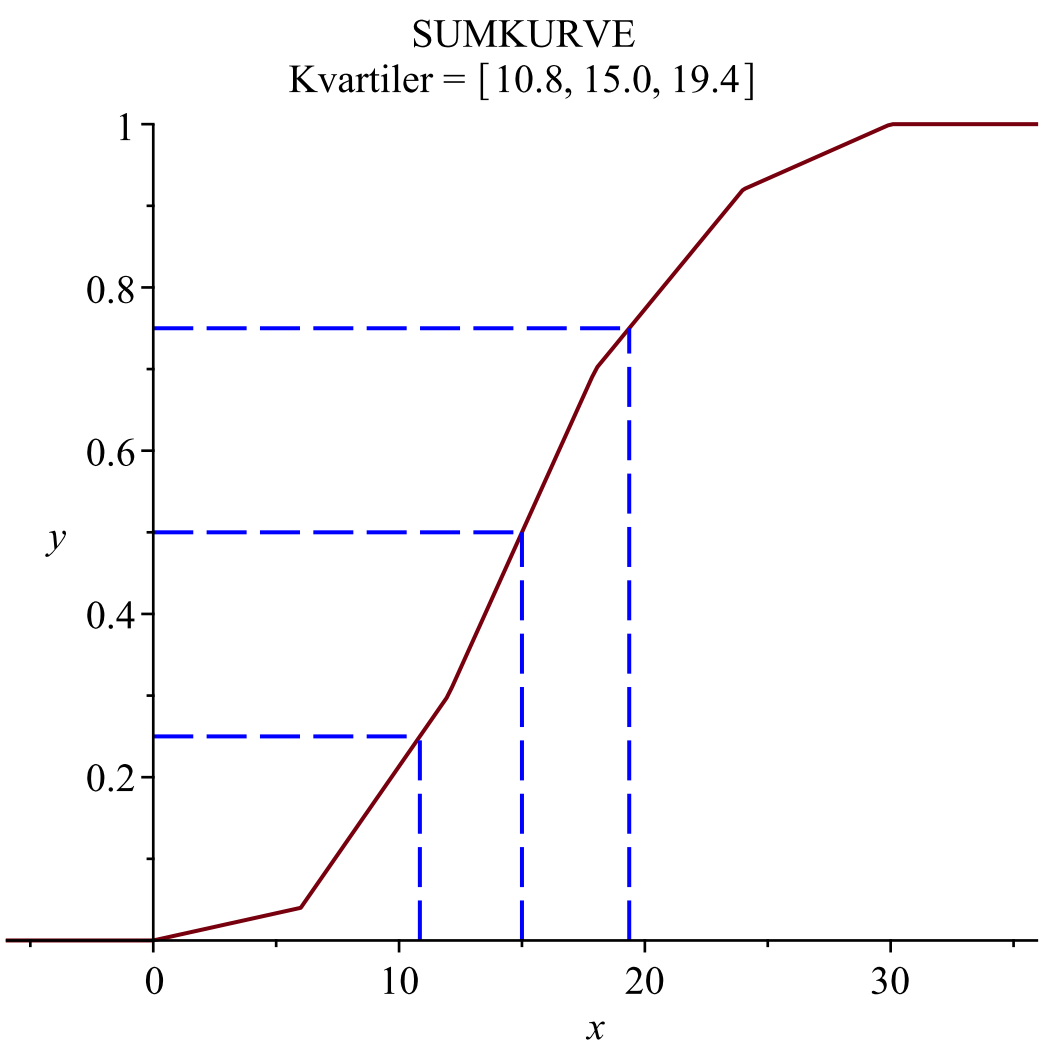
> *kvartiler*(*gruppeDATA*)

$$[10.846, 15.000, 19.364] \quad \textbf{(5.1)}$$

Dvs. **nedre kvartil** = 10.85, **medianen** = 15.00, **øvre kvartil** = 19.36

Sumkurve

> *plotSumkurve*(*gruppeDATA*)



Funktionen, som beskriver sumkurven, kan bestemmes:

```
> f := unapply(sumkurve(gruppeDATA, x), x) :
> 'f(x)'=f(x)
```

$$f(x) = \begin{cases} 0 & x < 0 \\ 0.00666666666666667\,x & 0 \le x \text{ and } x < 6. \\ 0.0433333333333333\,x - 0.220000000000000 & 6. \le x \text{ and } x < 12. \\ 0.0666666666666667\,x - 0.500000000000000 & 12. \le x \text{ and } x < 18. \\ 0.0366666666666667\,x + 0.0399999999999997 & 18. \le x \text{ and } x < 24. \\ 0.0133333333333333\,x + 0.600000000000000 & 24. \le x \text{ and } x < 30 \\ 1 & 30 \le x \end{cases} \quad \textbf{(6.1)}$$

Så kan man f.eks. udregne værdien i 12.6 (som bliver 0.34):

```
> f(12.6)
```

$$0.340000000000000 \quad \textbf{(6.2)}$$

Eller beregne, hvad x er, når y er 0.60:

```
> solve(f(x) = 0.60, x)
```

$$16.50000000 \quad \textbf{(6.3)}$$

Dvs. man har faktisk beregnet **60%-fraktilen** til 16.5

Spredning

> *spredning*(*gruppeDATA*)

5.87387436024980 **(7.1)**

Dvs. **spredningen** er ca. 5.87

Gennemsnit = middelværdi

> *gennemsnit*(*gruppeDATA*)

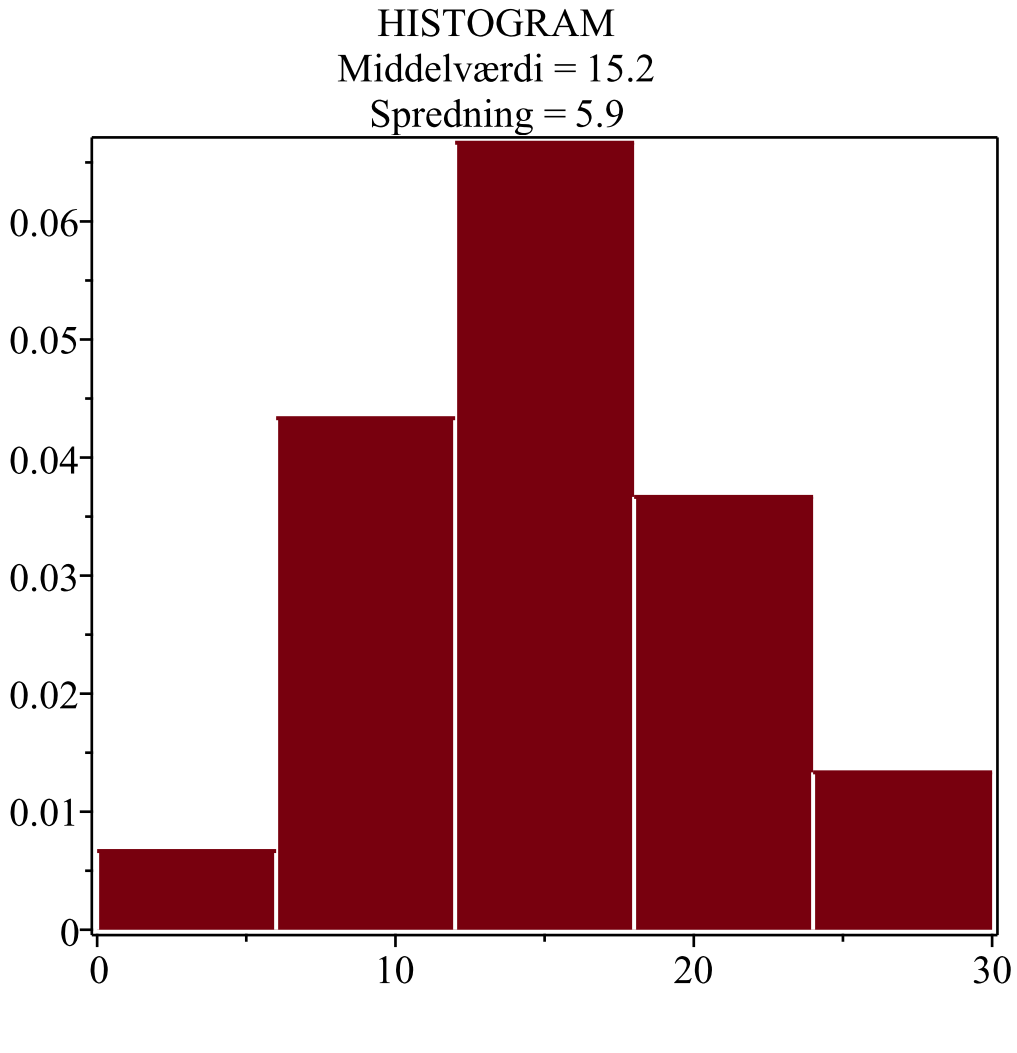
15.2400000000000 **(8.1)**

eller

> *middel*(*gruppeDATA*)

15.2400000000000 **(8.2)**

Dvs. **gennemsnittet** eller **middelværdien** er 15.24

Histogram

> *plotHistogram*(*gruppeDATA*)