Deskriptiv statistik (ugrupperede 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) :
```

Hyppighed

Ud fra en liste med observationer kan hyppighederne optælles:

```
> obs := [4, 5, 3, 4, 6, 5, 2, 7, 7, 4, 5, 6, 6, 3, 3, 5]
```

$$obs := [4, 5, 3, 4, 6, 5, 2, 7, 7, 4, 5, 6, 6, 3, 3, 5] \quad \textbf{(1.1)}$$

```
> HYP := hyppighed(obs)
```

$$HYP := \begin{bmatrix} 2 & 1 \\ 3 & 3 \\ 4 & 3 \\ 5 & 4 \\ 6 & 3 \\ 7 & 2 \end{bmatrix} \quad \textbf{(1.2)}$$

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

Frekvens og kumuleret frekvens

```
> frekvens(obs)
```

(2.1)

$$\begin{bmatrix} 2 & 0.0625 \\ 3 & 0.188 \\ 4 & 0.188 \\ 5 & 0.250 \\ 6 & 0.188 \\ 7 & 0.125 \end{bmatrix} \quad \textbf{(2.1)}$$

eller

> *frekvens*(*HYP*)

$$\begin{bmatrix} 2 & 0.0625 \\ 3 & 0.187 \\ 4 & 0.187 \\ 5 & 0.250 \\ 6 & 0.187 \\ 7 & 0.125 \end{bmatrix} \quad \textbf{(2.2)}$$

Svaret er en matrix, hvor 1. søjle er observationerne, og 2. søjle er antal af hver observation omregnet til %, dvs. **frekvensen**.

> *kumuleretFrekvens*(*obs*)

$$\begin{bmatrix} 2 & 0.0625 \\ 3 & 0.250 \\ 4 & 0.437 \\ 5 & 0.687 \\ 6 & 0.875 \\ 7 & 1. \end{bmatrix} \quad \textbf{(2.3)}$$

eller

> *kumuleretFrekvens*(*HYP*)

$$\begin{bmatrix} 2 & 0.0625 \\ 3 & 0.250 \\ 4 & 0.437 \\ 5 & 0.687 \\ 6 & 0.875 \\ 7 & 1. \end{bmatrix} \quad \textbf{(2.4)}$$

> *frekvensTabel*(*obs*)

observation	hyppighed	frekvens	kumuleret
2	1	6.25	6.25
3	3	18.75	25
4	3	18.75	43.7
5	4	25	68.7
6	3	18.75	87.5
7	2	12.5	100

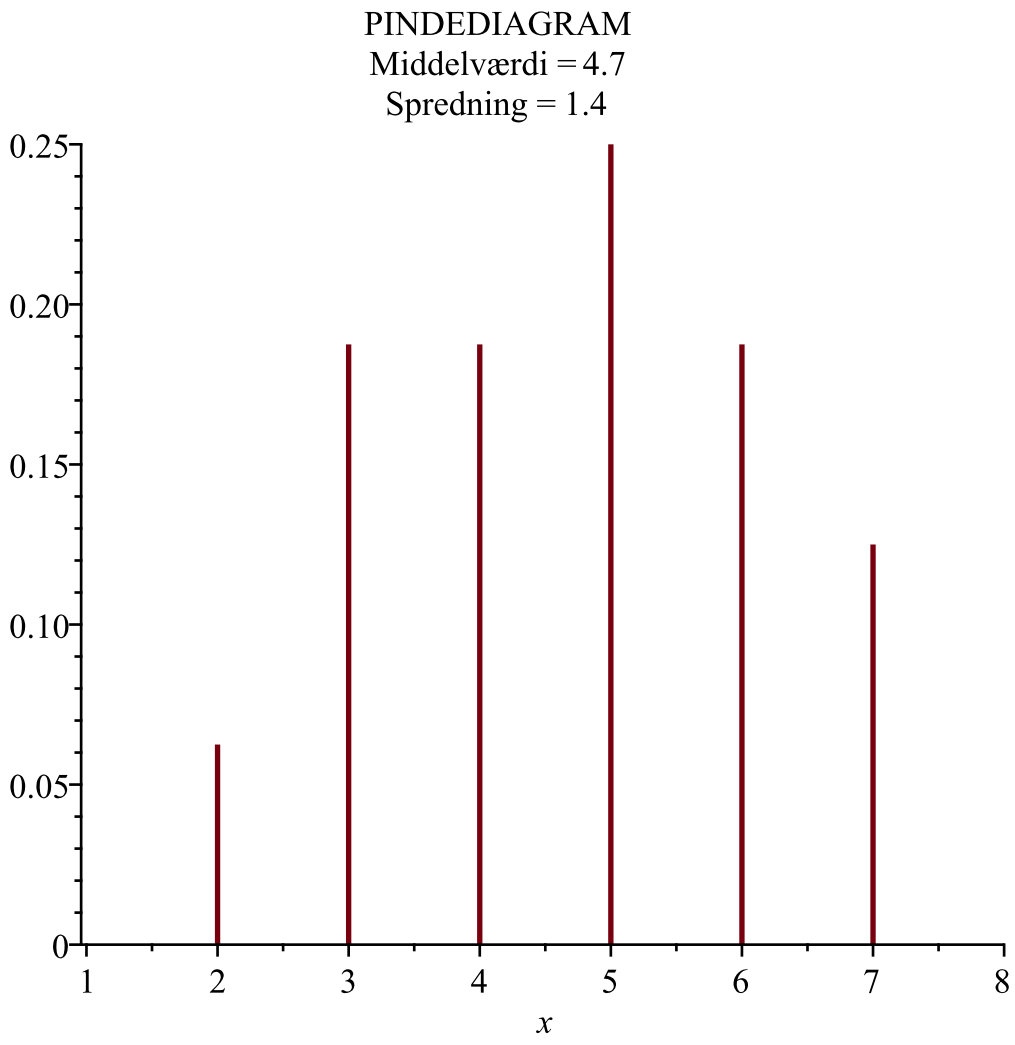
eller

```
> frekvensTabel(HYP)
```

observation	hyppighed	frekvens	kumuleret
2	1	6.25	6.25
3	3	18.75	25
4	3	18.75	43.7
5	4	25	68.7
6	3	18.75	87.5
7	2	12.5	100

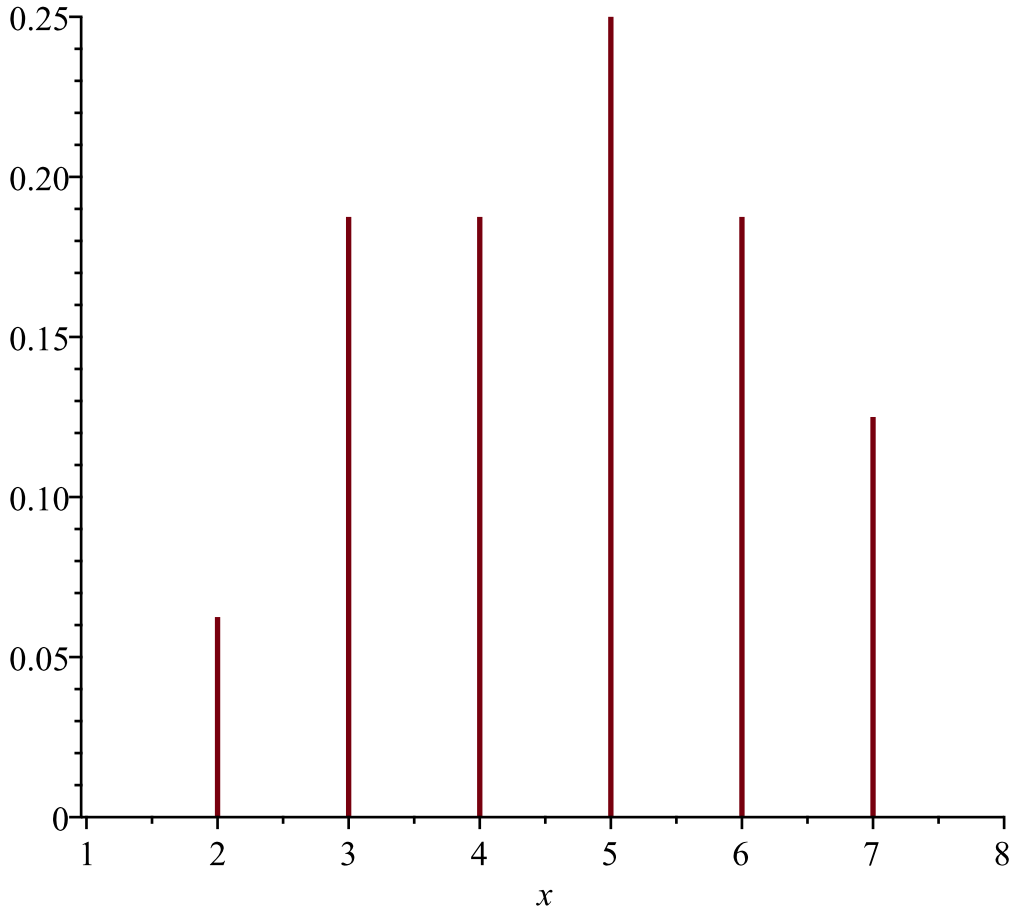
Stolpediagram = Pindediagram

```
> plotPindediagram(obs)
```



eller

```
> plotPindediagram(HYP)
```



Typetal

> *typetal*(*obs*)

[5] **(4.1)**

eller

> *typetal*(*HYP*)

[5] **(4.2)**

Dvs. **typetallet** er 5 (det er den mest hyppige observation).

Median

metode=1
Her returneres 50%-fraktilen. Denne metode benyttes som standard, hvis der er foretaget en optælling eller en gruppering af de rå data - dog bortset fra sortering.

metode=2
Her returneres den midterste observation i det sorterede observationssæt, hvis denne findes, ellers gennemsnittet af de to midterste observationer.
Denne metode benyttes som standard, hvis de rå data benyttes som input (med eller uden sortering).

> *median*(*obs*)

5. **(5.1)**

eller

> *median*(*HYP*)

5. **(5.2)**

Kvartiler

Kvartilsættet for et observationssæt. Her er to metoder tilgængelige:

metode=1
p-kvartiler bestemmes som den mindste observation, hvis kumulerede frekvens er p eller derover. Denne metode benyttes som standard, hvis der er foretaget en optælling eller en gruppering af de rå data - dog bortset fra sortering.

metode=2
I denne metode sorteres observationssættet. Medianen er den midterste observation, hvis denne findes, eller gennemsnittet af de to midterste observationer.
Herefter deles op i to halvdele. Hvis der er et ulige antal observationer, fjernes den midterste, så de to halvdele altid har det samme antal observationer. Medianen bestemmes tilsvarende for hver af det to halvdele, hvilket giver hhv. nedre kvartil og øvre kvartil.
Denne metode benyttes som standard, hvis de rå data benyttes som input (med eller uden sortering).

> *kvartiler*(*obs*)

[3.5000, 5., 6.] **(6.1)**

eller

> *kvartiler*(*obs*, *metode* = 2)

[3.5000, 5., 6.] **(6.2)**

alternativ

> *kvartiler*(*HYP*)

[3., 5., 6.] **(6.3)**

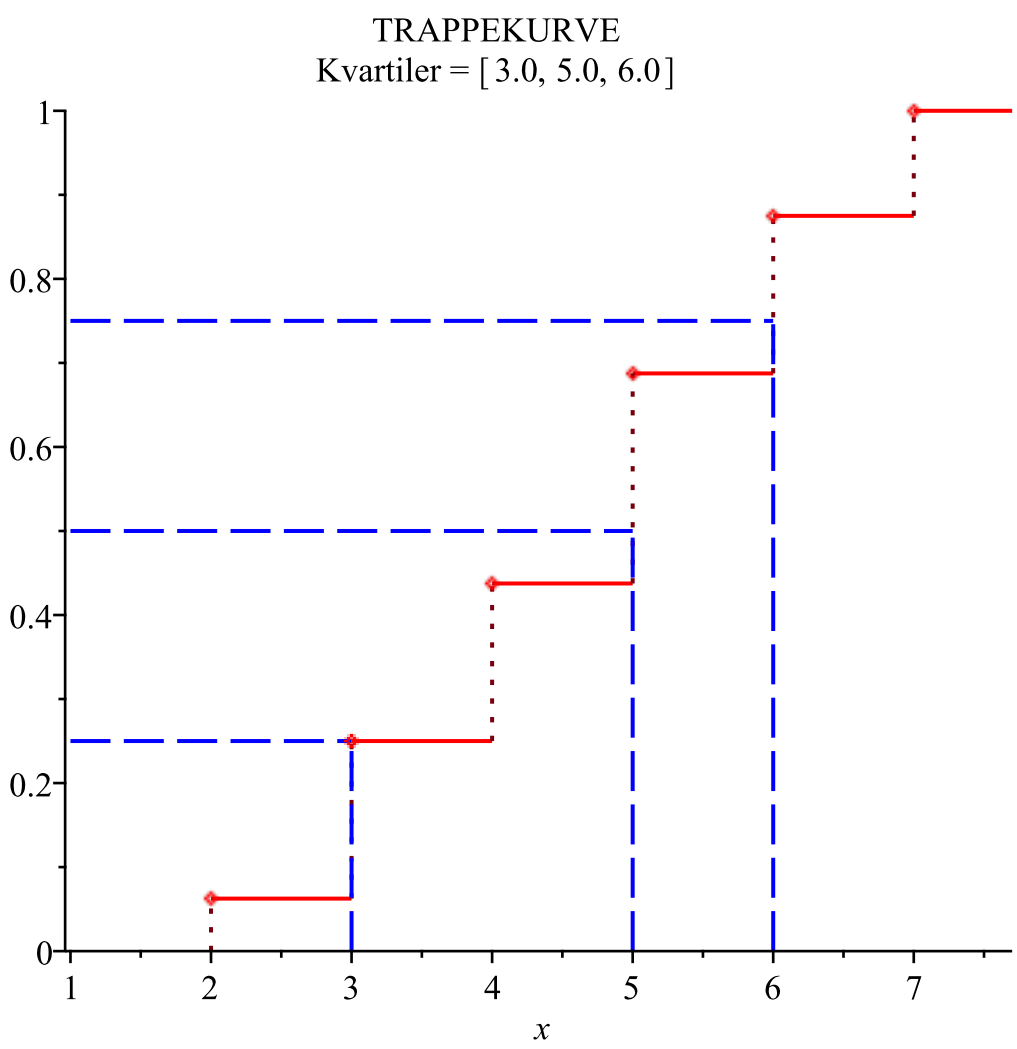
eller

> *kvartiler*(*HYP*, *metode* = 1)

[3., 5., 6.] **(6.4)**

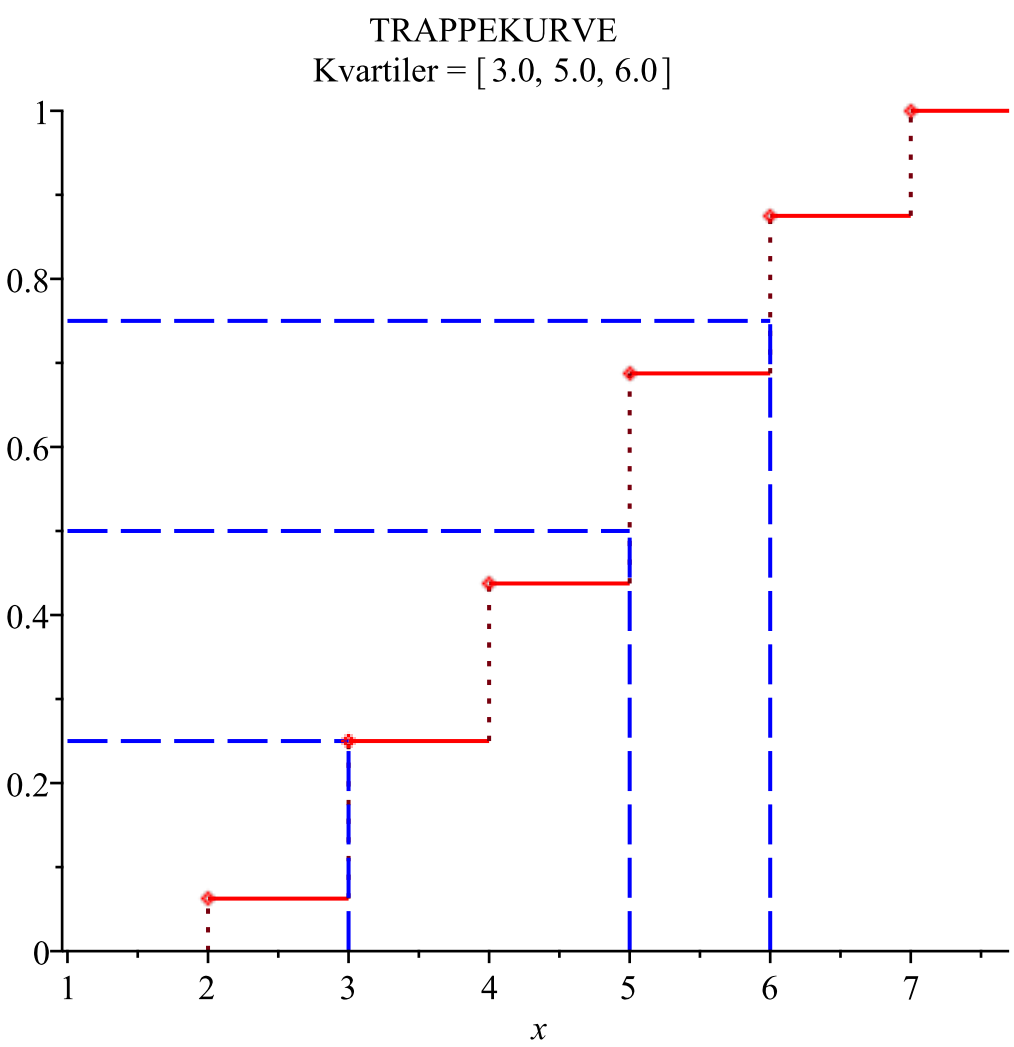
Trappediagram = trappekurve

> *plotTrappekurve*(*obs*)



eller

```
> plotTrappekurve(HYP)
```



Funktionen, som beskriver trappekurven, kan bestemmes:

```
> f := unapply(trappekurve(HYP, x), x) :
> 'f(x)'=f(x)
```

$$f(x) = \begin{cases} 0 & x < 2 \\ 0.0625000000000000 & 2 \le x \text{ and } x < 3 \\ 0.250000000000000 & 3 \le x \text{ and } x < 4 \\ 0.437500000000000 & 4 \le x \text{ and } x < 5 \\ 0.687500000000000 & 5 \le x \text{ and } x < 6 \\ 0.875000000000000 & 6 \le x \text{ and } x < 7 \\ 1 & 7 \le x \end{cases} \quad \textbf{(7.1)}$$

Så kan man f.eks. udregne værdien i 4.2 (som bliver 0.4375):

```
> f(4.2)
```

$$0.437500000000000 \quad \textbf{(7.2)}$$

Gennemsnit = middelværdi

```
> gennemsnit(obs)
```

$$4.687500000 \quad \textbf{(8.1)}$$

eller

> *middel*(*obs*)

4.687500000 **(8.2)**

eller

> *gennemsnit*(*HYP*)

4.68750000000000 **(8.3)**

eller

> *middel*(*HYP*)

4.68750000000000 **(8.4)**

Dvs. **gennemsnittet** eller **middelværdien** er ca. 4.69

Varians & spredning

> *varians*(*obs*)

2.08984375000000 **(9.1)**

eller

> *varians*(*HYP*)

2.08984375000000 **(9.2)**

Dvs. **variansen** er ca. 2.09

> *spredning*(*obs*)

1.44562918827755 **(9.3)**

eller

> *spredning*(*HYP*)

1.44562918827755 **(9.4)**

Dvs. **spredningen** er ca. 1.45