

Statistiek: Stam-bladdiagram en boxplot

6/12/2013

dr. Brenda Casteleyn



**Keu6**  
Coaching & Onderzoek

## 1. Theorie

Stam-bladdiagram en boxplot zijn methoden om visueel een verdeling voor te stellen.

1) Stam-bladdiagram: Wanneer we bijvoorbeeld de leeftijden van 20 studenten willen voorstellen:

Leeftijden: 22 23 24 22 23 25 18 19 20 23 18 19 18 22 23 25 22 23 22 30

We tellen hier 3 18-jarigen; 2 19-jarigen; 1 20-jarige; 5 22-jarigen; 5 23-jarigen; 1 24-jarige, 2 25-jarigen. en 1 30-jarige. Neem nu voor de stam het eerste getal dus het tiental en voor de bladeren de eenheden, dan krijgen we volgende figuur:

tientallen (stam) | eenheden (bladen)

↓      ↓

1	88899	--> drie van 18 jaar en twee van 19 jaar
2	1222233333455	--> de mensen van 21, 22, 23, 24 en 25 jaar
3	0	--> één dertigjarige

We zien hier in een oogopslag dat de meeste mensen twintigjarigen zijn.

Wanneer we de mannen en vrouwen afzonderlijk willen voorstellen kunnen we de mannen bv. links en de vrouwen rechts zetten:

988	1	89
5432	2	1222233335
	3	0

We zien dat de dertigjarige een vrouw was. Bij de mannen zijn er twee 18-jarigen; 1 19-jarige; en telkens één van 25; 24; 23 en 22 jaar. We stellen ook vast dat er links minder cijfers staan dan rechts. Er zijn dus minder mannen dan vrouwen in deze groep, nl. 7 mannen en 13 vrouwen.

Diepte: om de diepte te berekenen tel je het aantal waarnemingen van boven naar beneden en van beneden naar boven tot je ergens middenin bij een maximum komt. Een voorbeeld maakt dit duidelijk: Stel dat we volgend bladdiagram hebben voor de lengtes van respondenten in cm:

Diepte

15	6899	4	(= 4 waarden)
16	44	6	(= 6 waarden met vorige 4 erbij)
16	5568889	13	(= 13 waarden met vorige 6 erbij)
17	11123334	21	(= 21 waarden met vorige 13 erbij)
17	5556688889999	<b>33</b>	(= referentiewaarde mediaan)
18	00000001112234	31	(= 31 waarden met volgende 17 erbij)
18	556777888	17	(= 17 waarden met volgende 8 erbij)
19	11111223	8	(= 8 waarden)

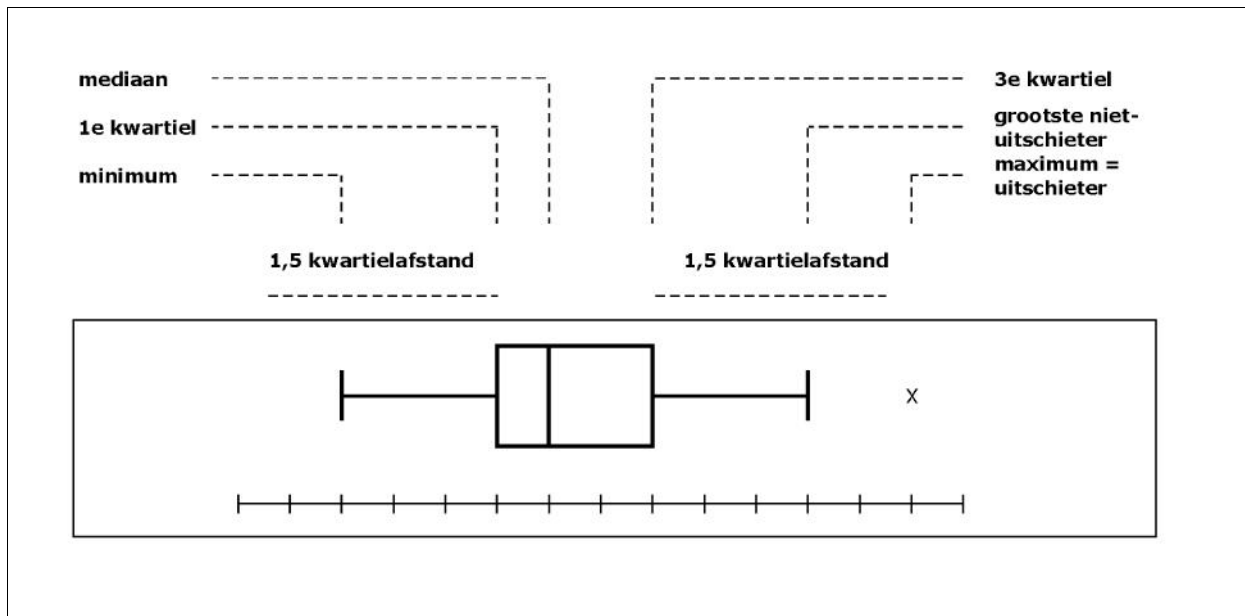
Wanneer weten we wanneer we van beneden naar boven of van boven naar beneden moeten optellen? Voor waarden kleiner dan de mediaan tellen we cumulatief van boven naar beneden en voor waarden groter dan mediaan tellen we van beneden naar cumulatief. In dit geval is de mediaan:  $(65+1)/2 =$  de 33ste waarde. 65 = het totaal aantal metingen, dus we tellen het aantal

nummertje rechts van de |, dat zijn er 65. De formule voor mediaan  $= (N+1)/2$ , dus de 33ste waarde. Deze bevindt zich bij het blad van 17. De mediaan is dus 17. De maximale diepte vinden we met de formule:  $\text{INT}[(N+1)/2]$ . In dit geval is de maximale diepte = de referentiewaarde van de mediaan:  $\text{INT}[(65+1)/2] = 33$  (INT betekent dat je enkel de cijfers voor de komma als uitkomst neemt. Zo is  $\text{INT}(2,6) = 2$  en  $\text{INT}(2,3) = 2$ ).

## 2) Boxplot of doosdiagram

Een boxplot is een grafiek waarin 5 getallen worden voorgesteld:

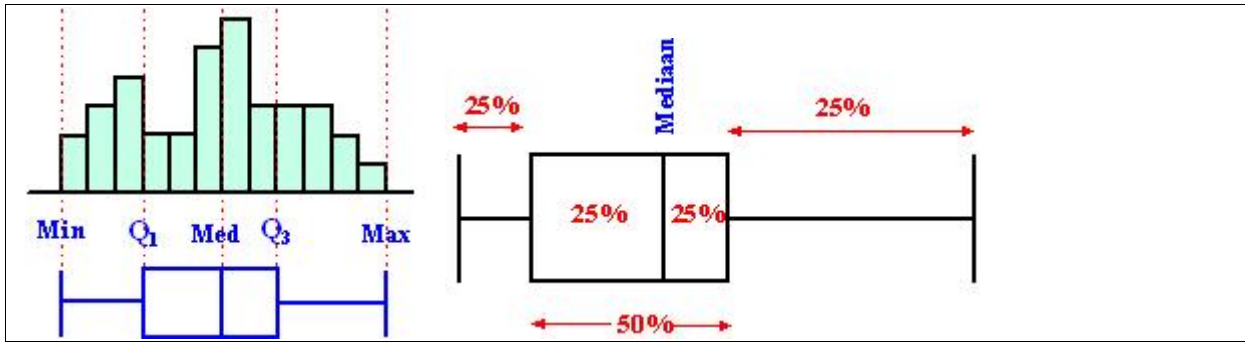
- Een centrale rechthoek die zich uitstrekt van het eerste tot het derde kwartiel
- De mediaan die wordt voorgesteld door een lijn binnen de rechthoek
- De lijnstukken die uit de rechthoek komen lopen van de kleinste tot de grootste waarneming
- Uitschieters die apart als punt of sterretje worden voorgesteld, dit zijn waarden die extreem groot of klein zijn (dus nog groter dan de kleinste of grootste waarde). Om te bepalen of een waarde een uitschieter is, gebruiken we de IKA-regel: waarden die  $1,5 \times \text{IKA}$  boven het derde kwartiel of onder het eerste kwartiel ligt is een uitschieter (zie oefening voor voorbeeld).



Boxplot: bron<sup>1</sup>

Hieronder zie je het verband tussen een boxplot en een histogram (links) en de verdeling van de waarnemingen in percentages (rechts)

<sup>1</sup> <http://upload.wikimedia.org/wikipedia/commons/e/ec/MadynoBoxplot01.png>



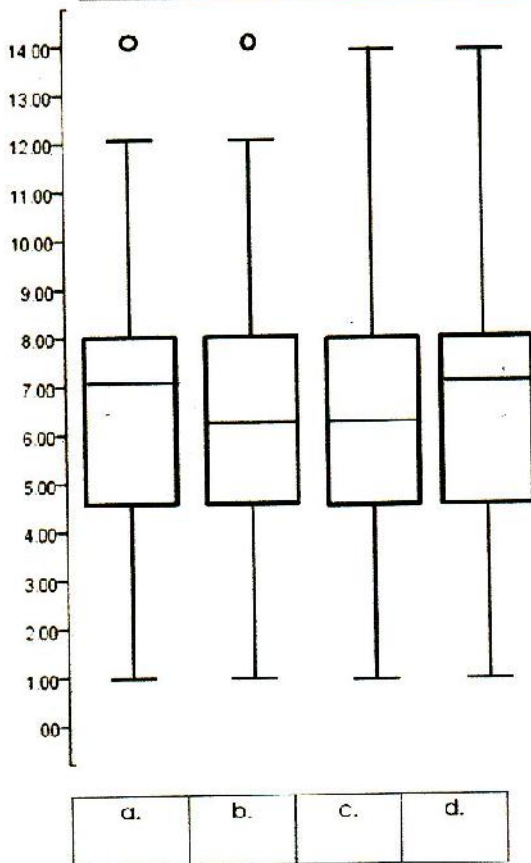
Boxplots: bron<sup>2</sup>

## 2. Oefeningen

### Proefexamen Thijssen 2012 Vraag 1

1) **Welke Box-Whiskerplot hoort bij de onderstaande verdeling:**

X	1	2	3	4	5	6	7	8	9	10	11	12	14
f	1	2	2	5	1	5	9	8	3	1	1	2	1



<sup>2</sup> <http://www.hhofstede.nl/modules/boxplot.htm> (hier zijn ook nog oefeningen te vinden)

## Proefexamen Thijssen 2012 Vraag 6

Hieronder staat een stam-blad diagram

3	3
4	39
5	156
6	1356899
7	022558
8	13
9	26

De maximale diepte van deze verdeling bedraagt

- A. 12
- B. 61
- C. 69
- D. 6

### 3. Oplossingen

#### Proefexamen Thijssen 2012 Vraag 1

Gegeven: Verdeling:

x	1	2	3	4	5	6	7	8	9	10	11	12	14
f	1	2	2	5	1	5	9	8	3	1	1	2	1

Gevraagd: Box-Wiskerplot

Oplossing: bepaalde waarden van mediaan.

Mediaan:  $(N+1)/2 = ?$  Bepaling N: totaal frequenties =  $1+2+2+5+1+5+9+8+3+1+1+2+1 = 41$

$Me = (41+1)/2 = 21$ ste waarneming --> bevindt zich bij de waarde voor  $x = 7$ . Dat betekent dat enkel mogelijk A of D nog in aanmerking komt. Bij A is 14 een uitschieter, bij D niet.

Om te weten of 14 een uitschieter is bepalen we de IKA.

$$IKA = Q_3 - Q_1$$

$Q_1 = (N+1)/4 = 42/4 = 10.5$ de waarneming --> bevindt halverwege tussen 4 en 5, dus 4,5

$Q_3 = [3(N+1)]/4 = 31.5$ de waarneming --> bevindt zich bij waarde 8

(we konden ook gewoon aflezen op de boxplots vermits alle antwoordopties dezelfde  $Q_1$  en  $Q_3$  hebben)

$$IKA = 8 - 4.5 = 3.5$$

$1.5 \times 3.5 = 5,25$ , dat wil zeggen dat alles wat meer 5,25 boven  $Q_3$  of onder  $Q_1$  ligt is een uitschieter. M.a.w alles groter dan  $= 13.25$  (nl.  $8 + 5,25$ ) of alles kleiner dan  $-0.75$  (nl.  $4.5 - 5.25$ ) zijn uitschieters. 14 is dus een uitschieter

→ Antwoord A

### Proefexamen Thijsen 2012 Vraag 6

Gegeven: Blad-stamdiagram

Gevraagd: bepaal de maximale diepte

Oplossing:

Formule voor maximale diepte:  $\text{INT}[(N+1)/2]$

Tel het aantal waarnemingen:  $N = 23$

Diepte =  $\text{INT}[(23+1)/2] = 12$

		diepte
3	3	1
4	39	3
5	156	6
6	1356899	12 = mediaan is $(23+1)/2 = 12$ de waarneming
7	022558	10
8	13	4
9	26	2

→ Antwoord A

### **Extra oefeningen**

Hier staan nog oefeningen met oplossingen:

[http://wiskundetrainer.nl/Kans\\_en\\_Statistiek\\_BB\\_files/boxplot.pdf](http://wiskundetrainer.nl/Kans_en_Statistiek_BB_files/boxplot.pdf)