

Onderzoek 5.

Uitvoeren van onderzoek Statistiek

Programma

1. Volgende week: presentaties
actualiteitsopdracht
2. Centrummaten
3. Spreidingsmaten
4. De normale verdeling
5. Grafieken

1. Presentaties volgende week

10 minuten per groep

**theorie hoofdstuk 1, 2 en 3 uit Baarda in
presentatie verwerken**

Power-Point twee dagen van tevoren inleveren

2. Centrummaten

Meetniveau

Kwalitatief

nominaal: opsomming

ordinaal: volgorde in de antwoorden

Kwantitatief

interval en ratio: antwoorden zijn getallen (geen codes)

Het meetniveau bepaalt de soort analyse die we kunnen uitvoeren

Centrumsmaten

Modus (elk meetniveau)

Mediaan (niet voor nominaal meetniveau)

(Rekenkundig) gemiddelde (alleen kwantitatief)

Modus

Modus: meest voorkomende getal

bv. a, a, b, b, b, c, c, d, d, d, d, f, h

de modus is d

Mediaan

**De mediaan is het middelste waarnemingsgetal
Antwoorden moeten wel op volgorde staan**

bv: 2, 2, 3, 3, 3, 4, 5, 6, 6, 6, 6, 7, 9

13 getallen

de mediaan is het 7^e getal

mediaan = 5

mediaan

bij een even aantal waarnemingen:

bv 2, 2, 3, 3, 3, 4, 5, 6, 6, 6, 6, 7, 9, 9

14 getallen

mediaan is het 7,5^e getal

mediaan = $(5+6)/2=5,5$

Rekenkundig gemiddelde:

Cijfer X	Frequentie f	
4	2	
5	5	
6	8	
7	7	
8	3	
9	1	

bel
ers

Rekenkundig gemiddelde:

X	f	Label
4	2	ers
5	5	
6	8	
7	7	
8	3	
9	1	
totaal	n = 26	

Rekenkundig gemiddelde:

X	f	X*f
4	2	8
5	5	25
6	8	48
7	7	49
8	3	24
9	1	9
totaal	n = 26	163

bel
ers

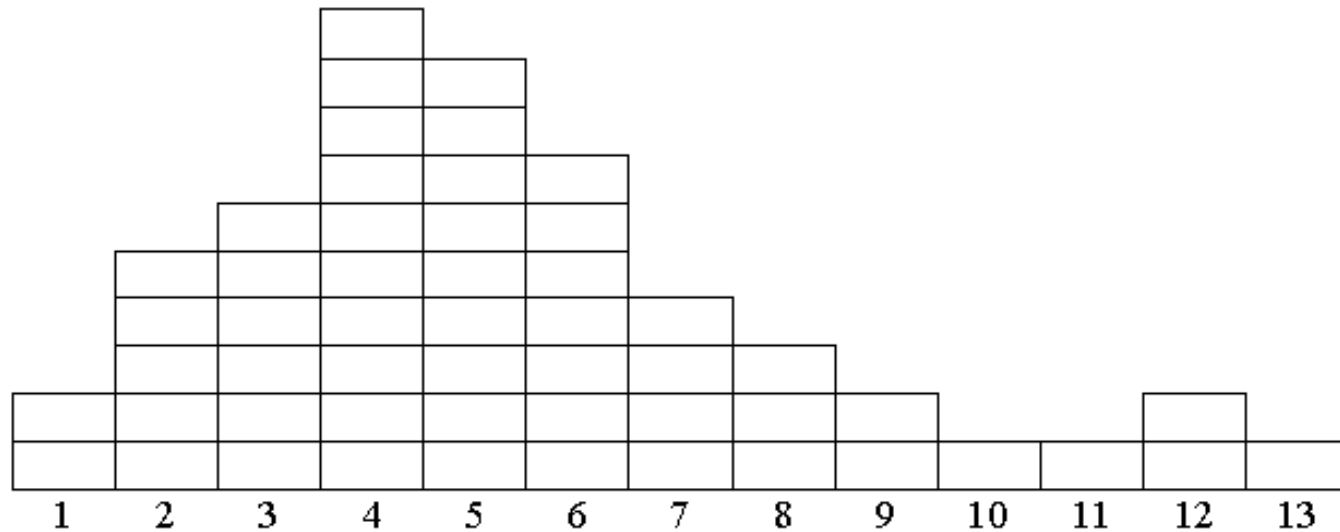
Rekenkundig gemiddelde

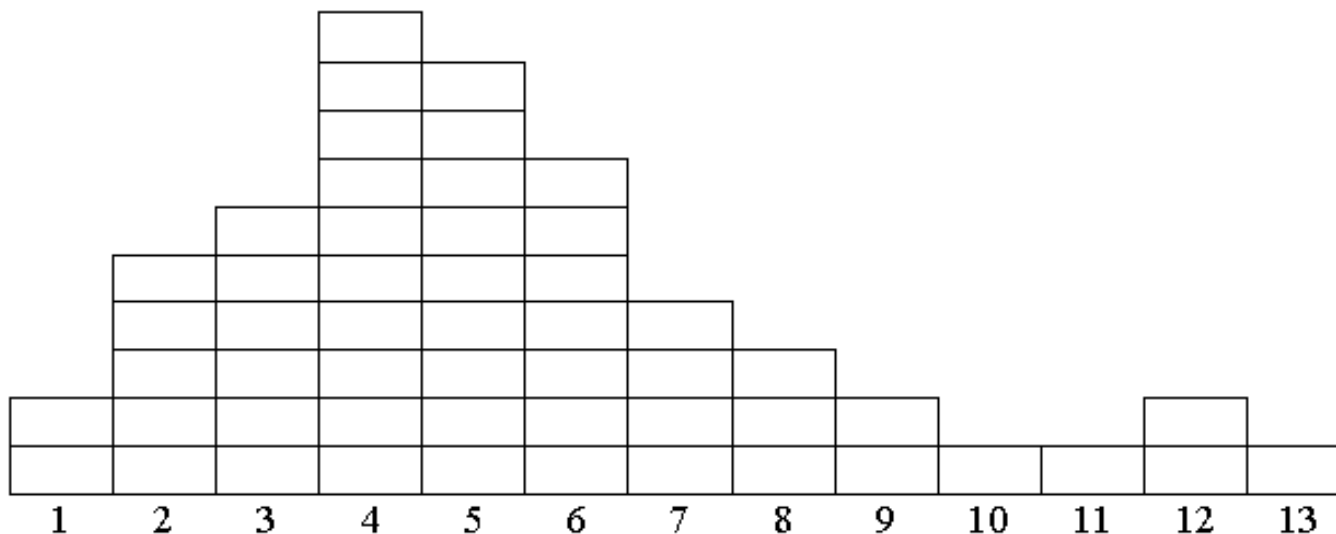
Het gemiddelde is nu $163/26=6,3$

Mediaan versus gemiddelde

**In welke situatie is de mediaan een betere
centrummaat dan het gemiddelde?**

Opgave: Bereken gemiddelde, modus en mediaan





De mediaan is gelijk aan de $(53+1)/2=27$ e waarneming,
dus de mediaan is gelijk aan 5.

Het rekenkundig gemiddelde is gelijk aan $285/53=5,38$

De modus = 4

Bij deze blokjesverdeling blijkt dat $\text{modus} < \text{mediaan} < \text{rekenkundig gemiddelde}$. In zo'n situatie spreekt men van een verdeling die scheef is naar rechts (de staart zit aan de rechterkant) of anders gezegd een positief scheve verdeling.

3. Spreidingsmaten

Spreiding en en frequentietabel

Kwartielafstand (kwantitatieve gegevens)

Standaarddeviatie (kwantitatieve gegevens)

Frequentietabel

X	f			
4	2			
5	5			
6	8			
7	7			
8	3			
9	1			

relatieve frequenties

X	f	Relatieve f		
4	2	8%		
5	5	19%		
6	8	30%		
7	7	27%		
8	3	12%		
9	1	4%		
	n = 26	100%		

cumulative frequenties

X	f	Cumulative f		
4	2	2		
5	5	7		
6	8	15		
7	7	22		
8	3	25		
9	1	26		
	n = 26			

cumulative percentages

X	f	Cumulative f	Relatieve f	Cumulatief %
4	2	2	8%	8%
5	5	7	19%	27%
6	8	15	30%	57%
7	7	22	27%	84%
8	3	25	12%	96%
9	1	26	4%	100%
	n = 26		100%	

conclusie

hoe trager het cumulatieve percentage oploopt
des te groter is de spreiding

Kwartielen

bv 2, 2, 3, 3, 3, 4, 5, 6, 6, 6, 6, 7, 9, 9

14 getallen, mediaan is het 7,5^e getal

mediaan = $(5+6)/2=5,5$

Bij de kwartielen verdeel je het linker- en het rechterdeel in tweeën

dan krijg je 4 delen van elk 25%.

Kwartielen

bv 2, 2, 3, 3, 4, 5, 6, 6, 6, 6, 7, 9, 9

$Q_2 = \text{mediaan} = 5,5$

$Q_1 = 3$

$Q_3 = 6$

(Inter)kwartielafstand = $Q_3 - Q_1$

$IQR = 6 - 3 = 3$

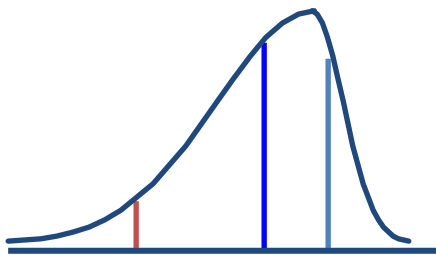
Boxplot

**Schematische weergave van:
de kwartielen
het minimum en maximum**

**hierin is tevens een scheve verdeling in te
herkennen**

Vorm van de verdeling en boxplot

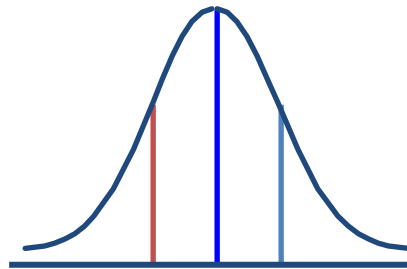
Links Scheef



Q1 Q2 Q3



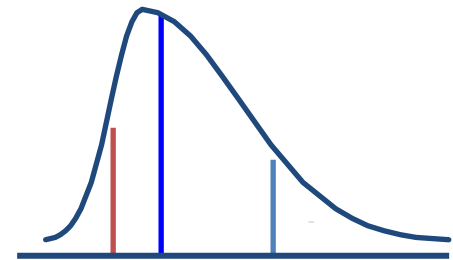
Symmetrisch



Q1 Q2 Q3



Rechts-Scheef



Q1 Q2 Q3



Standaarddeviatie

geeft de spreiding in de antwoorden aan
wordt uitgedrukt in een getal

berekening

betekenis

Handmatige berekening

X	f				
1	2				
2	3				
3	4				
4	3				
5	2				
totaal	N=14				

x	f	x*f			
1	2	2			
2	3	6			
3	4	12			
4	3	12			
5	2	10			
totaal	N=14	42			

1: Bereken eerst het gemiddelde

$$\text{Gemiddelde} = \mu = 42/14=3$$

X	f	X*f	d		
1	2	2	-2		
2	3	6	-1		
3	4	12	0		
4	3	12	1		
5	2	10	2		
totaal	N=14	42			

2: Bereken de deviatie d

$$d = X - \text{gemiddelde}$$

X	f	X*f	d	d ²	f*d ²
1	2	2	-2	4	8
2	3	6	-1	1	3
3	4	12	0	0	0
4	3	12	1	1	3
5	2	10	2	4	8
totaal	N=14	42			22

3: Bereken de som van $f \cdot d^2$

som = 22

X	f	X*f	d	d ²	f*d ²
1	2	2	-2	4	8
2	3	6	-1	1	3
3	4	12	0	0	0
4	3	12	1	1	3
5	2	10	2	4	8
totaal	N=14	42			22

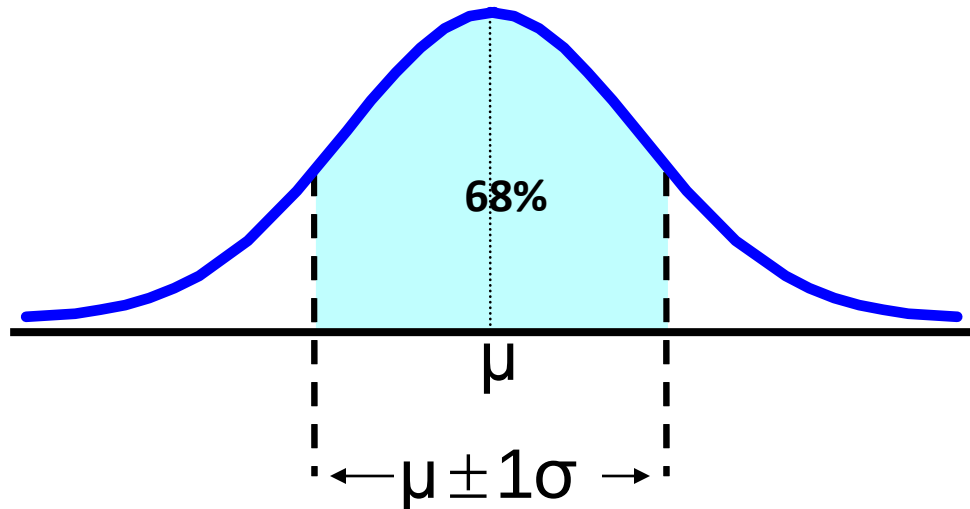
4: Standaarddeviatie = $\sigma = \sqrt{((\text{som van } f*d^2)/n)}$

Standaarddeviatie = $\sqrt{(22/14)}=1,25$

Betekenis standaarddeviatie

Als een variabele een klokvormige (normale) verdeling heeft, dan bevat het interval:

$\mu \pm 1\sigma$ ongeveer 68% van de waarden in de populatie



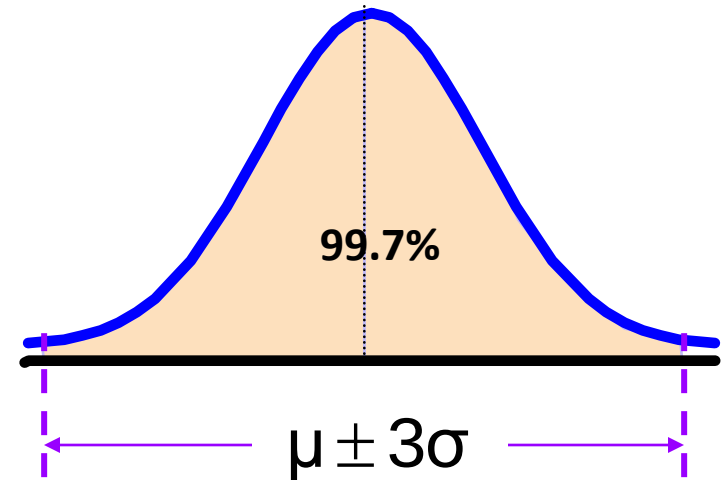
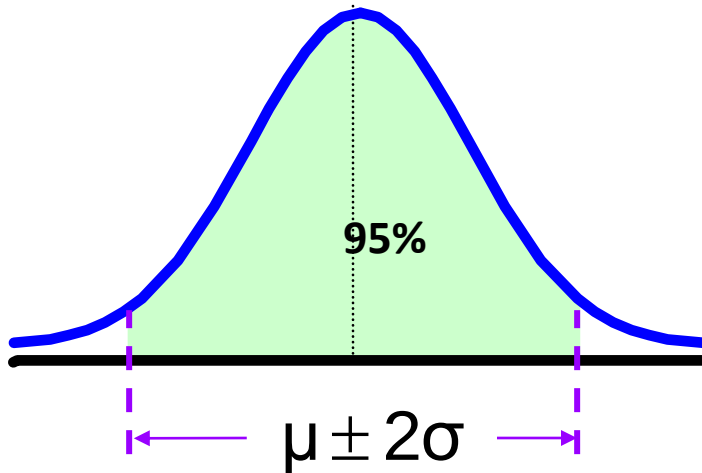
Vuistregels (vervolg)

$\mu \pm 2\sigma$ bevat ongeveer **95%** van de
populatie

$\mu \pm 3\sigma$ bevat ongeveer **99.7%** van de
populatie

waarden in de

waarden in de



Spreidingsmaten samengevat

Nominaal: frequentieverdeling

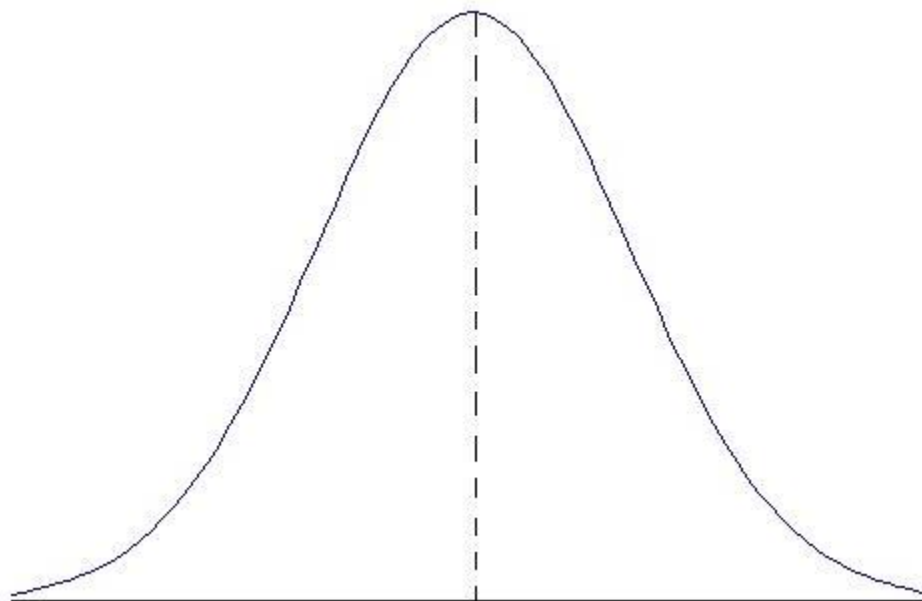
Ordinaal: cumulatieve frequentieverdeling

Interval of ratio: standaarddeviatie, boxplot

4. Normale verdeling

zuiver theoretische verdeling
voor verdeling van grote aantallen
symmetrische verdeling

Figuur



**de top van de grafiek valt samen met het
gemiddelde**

**de standaarddeviatie geeft aan hoe “breed” de
grafiek is**

Vuistregels normale verdeling

68% van de waarden heeft een afwijking van maximaal één keer de standaarddeviatie t.o.v. het gemiddelde.

95% van de waarden heeft een afwijking van maximaal twee keer de standaarddeviatie t.o.v. het gemiddelde.

99,7% van de waarden heeft een afwijking van maximaal drie keer de standaarddeviatie t.o.v. het gemiddelde.

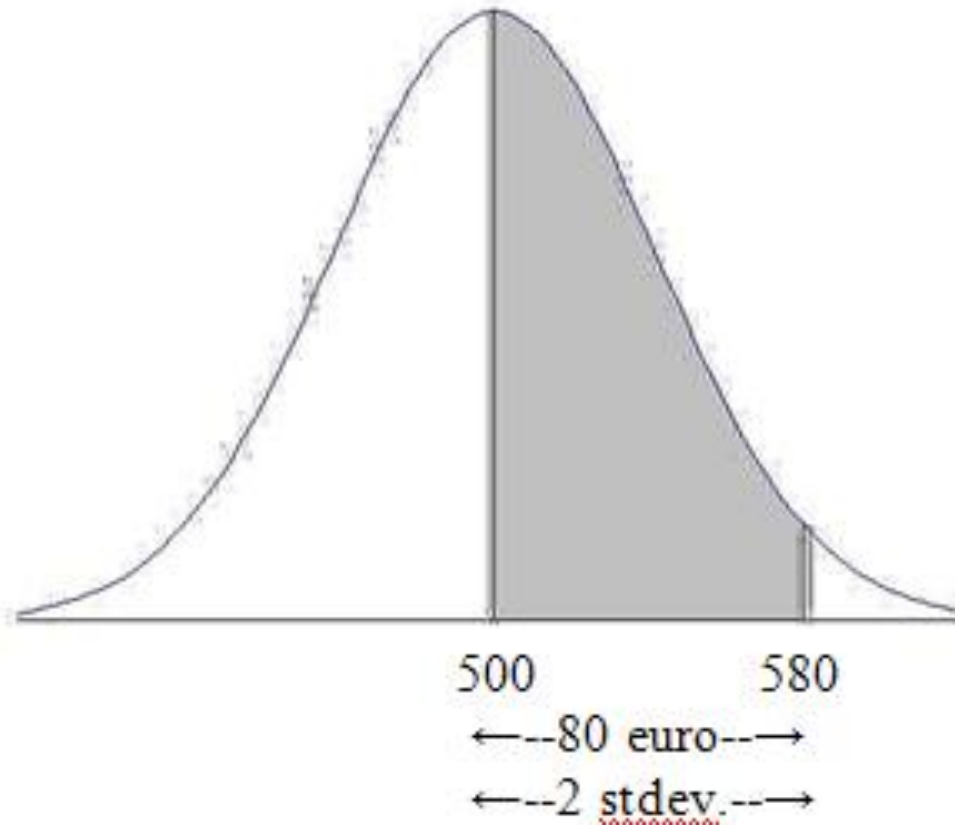
Voorbeeld

**Aan een product wordt gemiddeld € 500
uitgegeven**

de uitgaven zijn normaal verdeeld

de standaarddeviatie bedraagt € 40

**Hoeveel procent geeft tussen € 500 en € 580 aan
dit product uit?**



Gevraagd: gearceerde deel

uitwerking

**€ 80 komt overeen met twee keer de
standaarddeviatie
dus de helft van 95 %
is 47,5 %**

Maar..

deze methode werkt alleen als we met een geheel aantal malen de standaarddeviatie werken.

Hoe nu als dit niet zo is?

Zie tabel in de syllabus blz. 29

standaardnormale verdeling

In deze tabel staat de standaardnormale verdeling afgedrukt,
dan geldt: gemiddelde = 0 en standaarddeviatie = 1.

De waarde langs de x-as wordt de z-waarde genoemd,
die geeft aan hoeveel keer de standaarddeviatie je neemt.

Voorbeeld

zie het vorige voorbeeld

**bepaal hoeveel procent tussen € 500 en € 570
uitgeeft**

Uitwerking

**70 is $70/40=1,75$ keer de standaarddeviatie
kijk in de tabel bij $k=1,7$ en ga naar rechts tot de
kolom met 0,05
dat geeft 45,99
dus 45,99 % geeft tussen € 500 en € 570 uit.**

Symbolen voor Populaties:

We gebruiken de Griekse letters om de *populatie parameters* weer te geven.

Populatie gemiddelde = μ

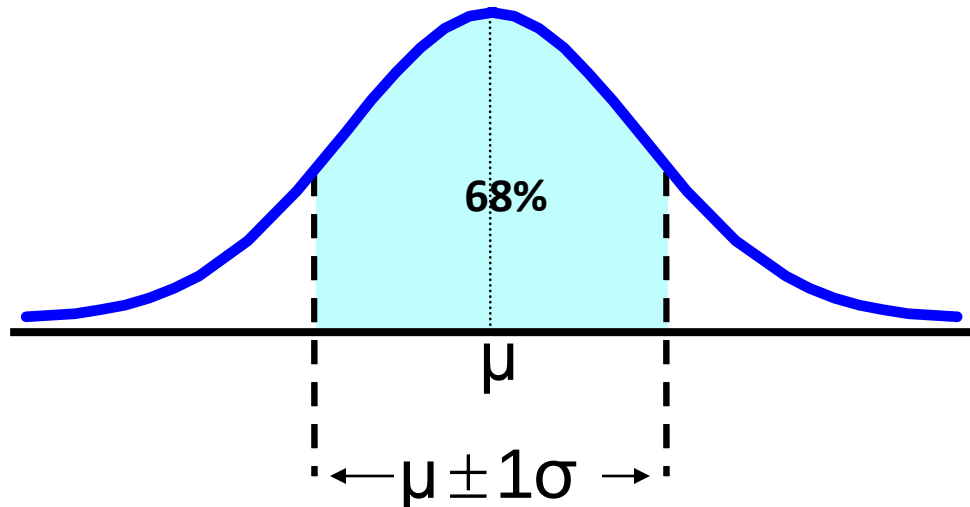
Populatie Standaarddeviatie = σ

Population Variantie = σ^2

Vuistregels

Als een variabele een **klokvormige of normale verdeling** heeft, dan bevat het interval:

$\mu \pm 1\sigma$ voor **68%** van de waarden in de populatie



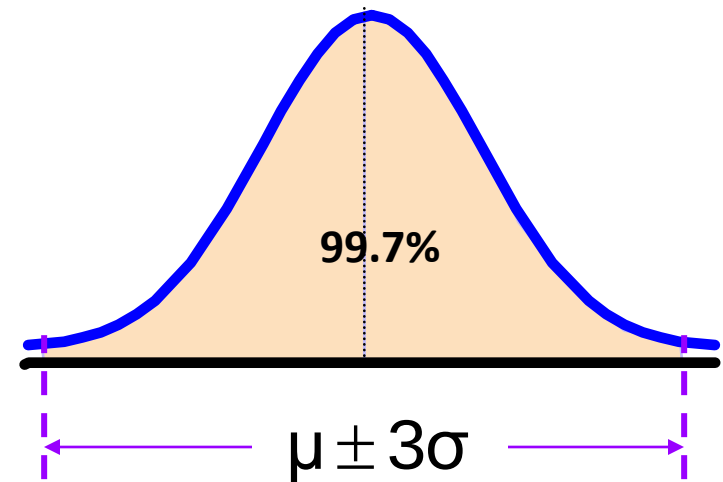
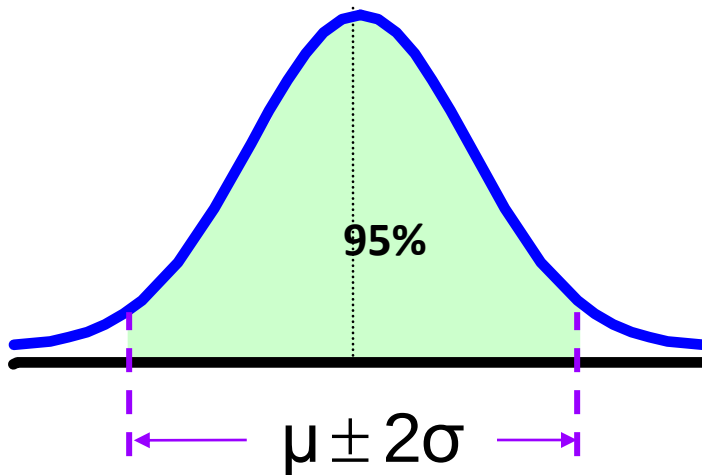
Vuistregels (vervolg)

$\mu \pm 2\sigma$ bevat ongeveer **95%** van de
populatie

$\mu \pm 3\sigma$ bevat ongeveer **99.7%** van de
populatie

waarden in de

waarden in de



z-tabel

e

Op een flesje bier staat: inhoud 33 cl.

Via Wikipedia vinden we dat dit symbool betekent dat bij 0 tot 50 cl de procentuele fout 9% mag zijn. In ons geval mag de fout dus maximaal 3 cl zijn (afgerond). We gaan er van uit dat 99,7% van de flesjes aan deze eis voldoet.

- a) Hoe groot is dan de standaarddeviatie?
- b) Tussen welke grenzen ligt de “middelste 90%” van de flesjes?
(Gebruik de z-tabel)
- b) Hoeveel procent van de flesjes heeft een inhoud kleiner dan 31 cl?

5. Grafieken

Het meetniveau van de vraag bepaalt het soort te gebruiken grafiek

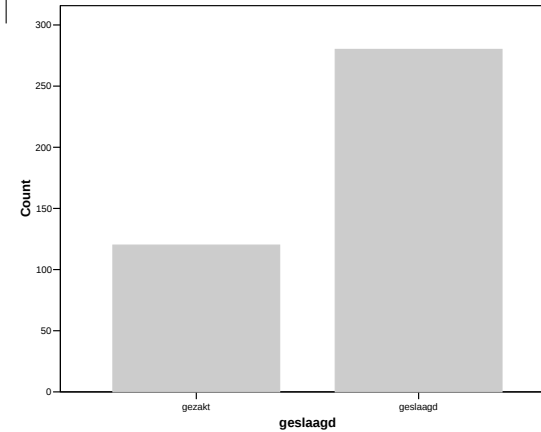
Dat geeft het volgende overzicht:

Overzicht

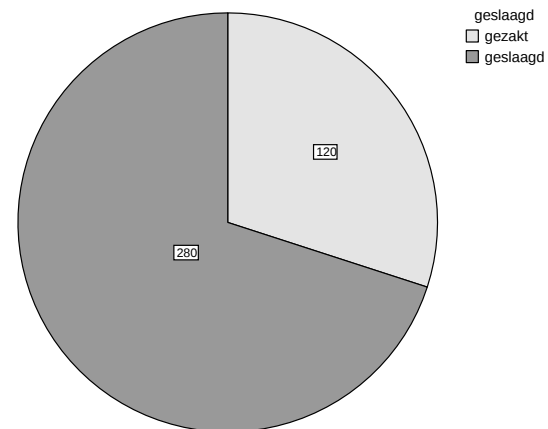
Nominaal	Ordinaal	Interval/ratio
• Frequentietabel • Tabel met percentages • Staafdiagram	• Frequentietabel • Tabel met percentages • Staafdiagram • Cumulatieve frequenties • Cumulatieve percentages En soms histogram, als de klassen even breed zijn en aansluiten	Eerst klassenindeling maken dan: • Frequentietabel • Tabel met percentages • Histogram • Cumulatieve frequenties • Cumulatieve percentages • Boxplot

Grafiek nominaal

○Staafdiagram

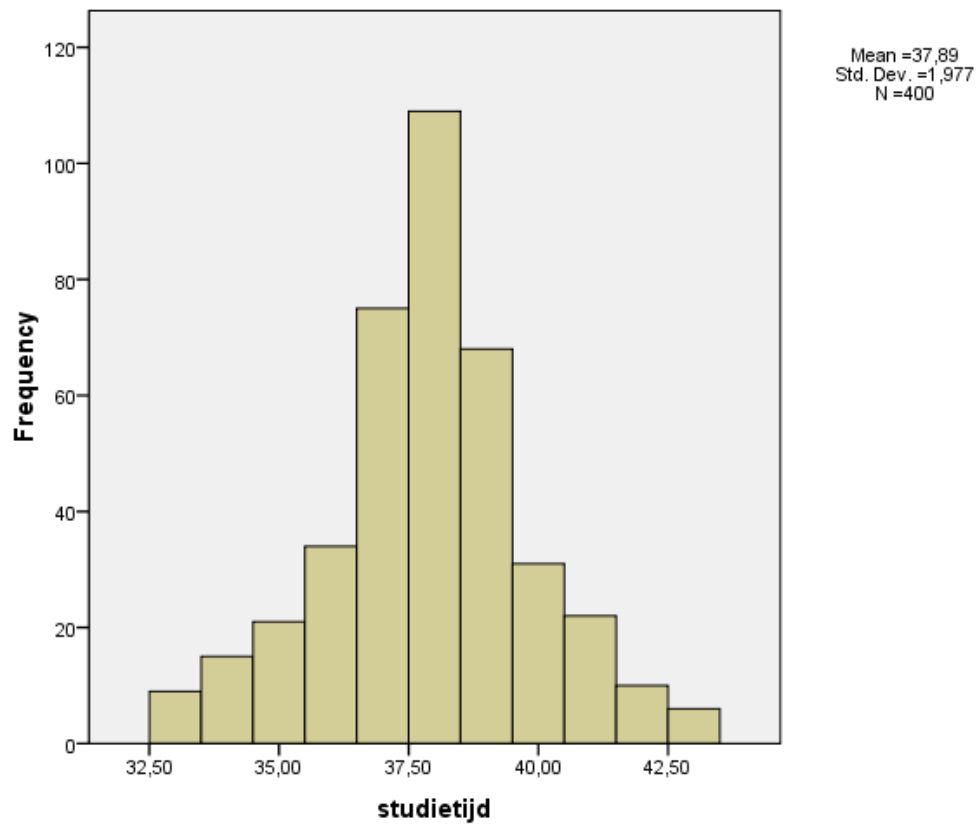


○Cirkeldiagram



Grafiek Ratio/Interval meetniveau

Histogram
normaalverdeling



Let op het verschil tussen staafdiagram en histogram

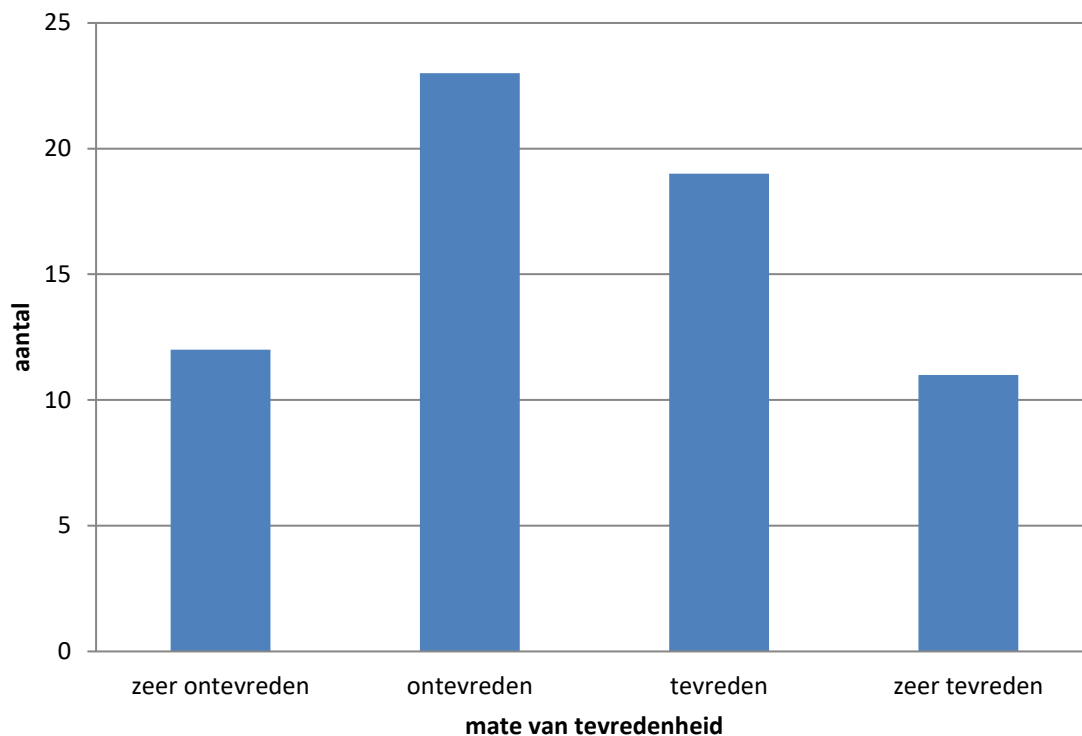
staafdiagram

- bij kwalitatieve vragen
- verschillende categorieën op de x-as
- staven los van elkaar

histogram

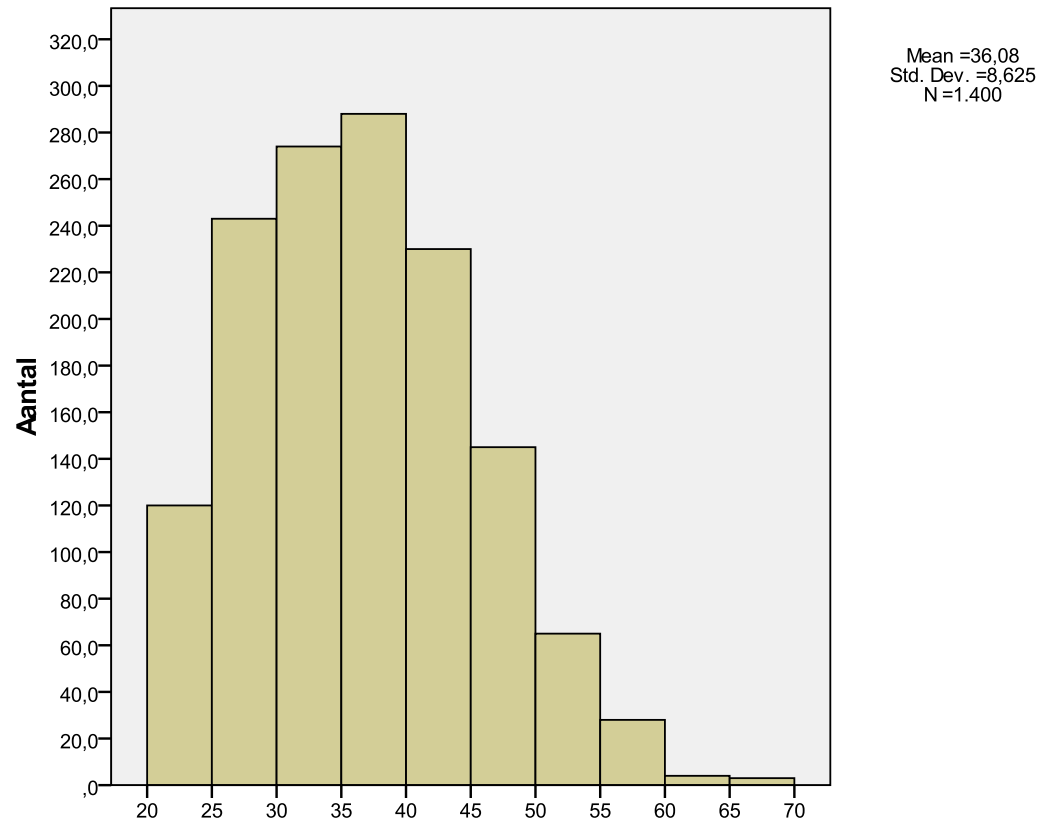
- bij kwantitatieve vragen
- doorlopende schaalverdeling op de x-as
- staven tegen elkaar aan

Oordeel over de dienstverlening



Staafdiagram

Leertijdsopbouw van respondenten Auto Online enquête



Histogram

Randvoorwaarden bij grafieken en tabellen

Titel (volledig en toch beknopt)

Bron

**Assen (schaalverdeling aangeven, geen grote
getallen)**

Eenheden

Legenda (bij grafieken)

Opdrachten

maken opdrachten hoofdstuk 3, 4 en 5