

howest



hogeschool

Basis van Onderzoek

Academiejaar 2019-2020

Miet Craeynest (OLOD-verantwoordelijke)

Helga Peeters

Annemie Schepens

Inhoudstafel van dit OLOD

0. Inleiding



- 1 hoorcollege
- Hfst uit boek: Craeynest et al. (2017): Algemene psychologie: **enkel p. 15-28**
- Hfst uit Blokwijzer deel 1: **paragraaf 5.3**

1. Dit is onderzoek

- 5 hoorcolleges
- Boek: Baarda, B. (2014). *'Dit is onderzoek'*



2. Kwantitatieve data-analyse

- 4 hoorcolleges, 2 oefencolleges
- Cursus: Craeynest, M., & De Bruycker, E. (2019). *Blokwijzer deel 1*.



3. Informatievaardigheden

- 1 hoorcollege, 1 oefencollege
- Digitale cursus op Leho: Craeynest, M., & De Bruycker, E. (2019). *Syllabus Informatievaardigheden*



Situering: waar zitten we?

Wat is onderzoek en welke **soorten** onderzoek zijn er?

- **Deel 1. Dit is onderzoek**
Vragenlijsten, interview,... survey versus experiment...

Analyseren van bevindingen

> Statistiek: mensen, toeval → kans-berekening

- **Deel 2. Kwantitatieve data- analyse**

Waar vind ik nu wetenschappelijke informatie?

- **Deel 3. Informatievaardigheden**

Lesmateriaal: Blokwijzer

Het lesmateriaal bij dit deel is de **Blokwijzer 'Kwantitatieve Data-analyse'**. De opbouw ervan kan je helpen bij het instuderen van de leerstof:

- Elk hoofdstuk begint met een **overzichtelijke inhoudstafel**. Daaronder staat ook welk online materiaal er op Sofia te vinden is.
 - Studietip: naast de slides en de oplossingen van de oefeningen vind je er ook de hele tekst van de Blokwijzer als pdf terug. Die biedt de mogelijkheid om via de zoekfunctie begrippen snel terug te vinden. Je kunt er ook zaken in markeren, er online notities in aanbrengen en die notities vervolgens ook downloaden.
- Daarna staan per hoofdstuk ook de **specifieke studiedoelstellingen** opgesteld. Die verwoorden wat je van de leerstof moet kennen en kunnen.
 - Studietip: overloop die na het studeren van dat hoofdstuk als zelftest
- De **linkerpagina is wit gelaten**, zodat er ruimte staat voor eigen notities.
 - Studietip: gebruik die witruimte voor eigen notities bij de tekst; dat kan je helpen om de leerstof verder te verwerken.
- Op het einde van elk hoofdstuk staat een **conceptmap**: daarin wordt de leerstof in de vorm van een mindmap samengevat.
 - Studietip: vul de ontbrekende concepten in (de oplossingen ervan vind je terug op Sofia)
- Elk hoofdstuk eindigt met ruimte om je **eigen reflecties en eventuele vragen** te noteren.
 - Studietip: neem die vragen mee naar de oefencolleges!
- Tot slot nog dit: alle **toepassingen in Excel** staan in de Blokwijzer in een groen kader gemarkeerd. Er horen ook instructie filmpjes bij. *Die hoef je voor dit OLOD nog niet te kunnen!* Dat komt pas aan bod in het OLOD Kwantitatief Onderzoek (Blok 1C).
 - Tip: Dat betekent dat we deze Blokwijzer er dan opnieuw bij zullen nemen; bewaar hem dus goed :-)

Inhoudstafel van deze les

0. Soorten statistiek
1. Voorbereiding op de kwantitatieve data-analyse
2. Beschrijvende analyses

0. Soorten statistiek

A. Beschrijvende statistiek

B. Toetsende statistiek

A. Beschrijvende statistiek

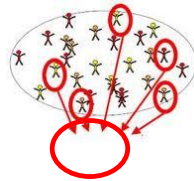
- **Doel:** data- of gegevensreductie op STEEKPROEFNIVEAU
- **Hoe:** presenteren van gegevens op een overzichtelijke, inzichtelijke manier.
 - tabellen, figuren en grafieken
 - percentages of gemiddelden
 - het weergeven van het verband tussen twee (of meer) kenmerken → correlatie
 - Bijv. *lengte en gewicht*
 - Bijv. *seksuele activiteit en partijvoorkeur*

B. Toetsende statistiek

- **Doel:** Generaliseren van de verzamelde informatie naar een groter geheel (Steekproef → **POPULATIE**)
- **Hoe:** Kansberekening
 - Hoe groot is de **kans** is dat onze conclusies over een bepaalde populatie op grond van een steekproef **onjuist** zijn?

- Voorbeeld: *Hoe zou de Vlaamse bevolking reageren op de aanstelling van een vrouw als minister-president van de Vlaamse regering?*

Steekproef:



→ Louter uitspraken over de steekproef =

beschrijvende statistiek

Vb: 74% van de ondervraagde personen vindt een vrouw niet even bekwaam als een man om minister-president te zijn, waarvan x aantal mannen, en y aantal vrouwen, woonachtig te...

Steekproef → Populatie



→ Analyses van de steekproef om uitspraken over de populatie te doen =
toetsende statistiek

Dwz **kansberekening**: kans bepalen dat het toeval speelde in die steekproef en dat we ons vergissen als we besluiten dat 74% *van de hele populatie* het niet eens is met de voorgelegde stelling (= kans op **toeval** bepalen).

1. Voorbereiding op de kwantitatieve analyse

1.1 Eenheden en variabelen

1.2 Soorten variabelen naargelang de waarde erop

1.3 Soorten variabelen naargelang het meetniveau

1.4 Datamatrix

1.1 Eenheden en variabelen

- **Eenheden**: de basisobjecten (personen, bedrijven, auto's,...) waarvan we bepaalde **kenmerken** willen weten en dus **meten** (intelligentie, aantal werknemers, remafstand)



Zie hoger: Thema *Dit is onderzoek!*

- Dat **gemeten kenmerk** = **Variabele** (letterlijk: 'veranderlijke')
- Bij een onderzoek **meten** we voor elke eenheid de waarde van elke variabele die belangrijk is voor onze onderzoeksvraag.
 - Onderzoekseenheid → variabele → waarde op die variabele
 - Vb: *Pieter* → *Heeft hij facebookvrienden?* → ja/nee
 - Vb: *Pieter* → *Hoeveel facebookvrienden heeft hij?* → 183

Meerdere
soorten

1.2 Soorten variabelen naargelang de waarde erop

A. Categorische of kwalitatieve variabele

- Kan niet in een (*zinvol) getal worden uitgedrukt, betreft **categorieën**
 - Bijv: *geslacht, bloedgroep,...*

B. Numerieke of kwantitatieve variabelen

- Drukken een **hoeveelheid** uit en worden verkregen door tellen, meten, wegen,...
 - » Bijv: *leeftijd, gewicht, aantal personen dat dagelijks een kruispunt oversteekt,...*
- Verder Onderscheid:
 - **Continue variabelen**: alle mogelijke waarden tussen bepaalde grenzen (dus ook komma-getallen en breuken)
 - **Discrete variabelen**: alleen hele waarden

1.3 Soorten variabelen naargelang het meetniveau

- Het meetniveau is richtinggevend voor het soort analyses
- Dwz: het geeft aan welke analyses je met die variabelen kunt doen
 - Bijv.: *Geslacht uitdrukken als Man = 1, Vrouw = 2.*
 - *Kun je dan van deze warden het gemiddelde nemen?*

A. Nominaal niveau

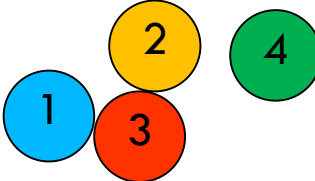
B. Ordinaal niveau

C. Intervalniveau

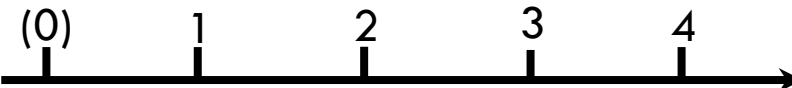
D. Rationiveau

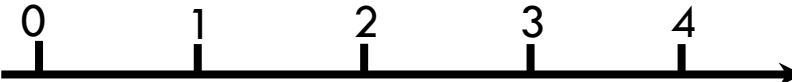
Typen variabelen – meetniveau

- Variabelen verschillen in **meetniveau**
- Het meetniveau bepaalt hoe we gegevens kunnen **verwerken, analyseren en interpreteren**

- **Nominaal** 

- **Ordinaal** 

- **Interval** 

- **Ratio** 

OVERZICHT

MEETNIVEAU	CLASSIFICATIE	ORDENING	VASTE MEETEENHEID	ABSOLUUT NULPUNT
NOMINAAL	JA	NEEN	NEEN	NEEN
ORDINAAL	JA	JA	NEEN	NEEN
INTERVAL	JA	JA	JA	NEEN
RATIO	JA	JA	JA	JA

! Likertschalen !

- Om **moeilijk kwantificeerbare** gegevens toch te kunnen meten en behandelen op een voldoende hoog meetniveau
 - Vb. de liefde voor statistiek
- Uitspraken die **gradueel van elkaar verschillen** en waaraan een **getalwaarde** wordt gekoppeld

	Helemaal mee oneens	Eerder mee oneens	Eerder mee eens	Helemaal mee eens
1. Ik ben dol op statistiek.	1	2	3	4
2. Ik hou van statistiek.	1	2	3	4

- Meetniveau: ordinaal → interval/ratio

1.4 Datamatrix

STAP 1: DATA-invoer

- De **ruwe gegevens** worden in een **datamatrix (datafile)** geplaatst om ermee aan de slag te kunnen (**Excel**).
 - Rijen: eenheden
 - Kolommen: kenmerken of variabelen

	Variabele1	Variabele2	Variabele3	Variabele4	Variabele5	...
Eenheid1						
Eenheid2						
Eenheid3						
Eenheid4						
Eenheid5						
Eenheid6						
...						

- Eenheden: identificatienummer geven (vb: 'id' of 'nr')

BESTAND

START

INVOEGEN

PAGINA-INDELING

FORMULES

GEGEVENS

CONTROLLEREN

BEELD

Van Access

Van web

Van tekst

Van andere bronnen

Bestaande verbindingen

Externe gegevens ophalen

Verbindingen

Eigenschappen

Koppelingen bewerken

Verbindingen

Sorteren

Filter

Geavanceerd

Sorteren en filteren

Wissen

Opnieuw toep.

Geavanceerd

Hulpmiddelen voor gegevens

Flash Fill

Duplicaten verwijderen

Gegevensvalidatie

Samenvoegen

Wat-als-analyse

Relaties

AA9

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Nr	leeftijd	geslacht	onderwijs	contact1	contact2	contact3	contact 4	contact5	contact6	contact7	subtielracisme1	subtielr
2	1	19	1	6	3	2	2	1	1	4	4	3	
3	2	21	2	5	4	4	5	4	3	3	3	5	
4	3	20	2	5	2	2	1	1	1	2	2	4	
5	4	20	2	4	1	1	1	1	1	1	1	3	
6	5	18	2	5	1	1	2	1	1	1	2	4	
7	6	19	2	4	1	1	1	1	1	1	1	4	
8	7	18	2	3	1	1	1	1	1	3	2	4	
9	8	51	1	2	1	1	1	1	1	1	1	7	
10	9	31	2	3	1	1	1	1	1	1	1	4	
11	10	51	1	5	2	2	1	1	1	2	2	4	
12	11	52	1	2	2	2	1	1	2	1	1	5	
13	12	21	2	5	2	1	1	1	1	1	1	2	
14	13	30	2	4	7	7	4	2	2	6	6	6	
15	14	21	2	3	1	1	1	1	1	1	1	3	
16	15	21	2	4	1	1	1	1	1	1	1	6	
17	16	21	1	4	2	2	2	2	2	2	2	5	
18	17	20	2	3	2	2	2	1	2	2	2	3	
19	18	46	2	4	2	2	2	2	2	2	2	2	
20	19	67	2	4	2	2	1	2	2	2	2	7	
21	20	40	2	6	7	7	3	7	7	7	7	1	
22	21	52	1	6	2	2	1	1	1	2	2	1	

- Geef (*liefst**) numerieke codes aan de antwoordmogelijkheden per variabele
 - Bijv. *man = 1, vrouw = 2 (of omgekeerd)*

(Hoeft niet per se voor variabelen met een nominaal meetniveau)*

→ Het overzicht van al de codes vormt je **codeboek**

VARIABLE	VRAAG	ANTWOORDCODE	MEETNIVEAU
DEMOGRAFISCHE VARIABLEN			
Leeftijd	Leeftijd in jaren: ... jaar	leeftijd in jaren	<i>Ratio</i>
Geslacht	O Man O Vrouw	1 = man 2 = vrouw	<i>Nominaal</i>
Onderwijstype	Onderwijstype in jaren: ... jaar	1 = lager onderwijs 2 = lager secundair 3 = hoger secundair 4 = hoger onderwijs korte type 5 = hoger onderwijs lange type 6 = universitair onderwijs	<i>Ordinaal</i>
CONTACT MET ALLOCHTONEN			
Contact_1	1. Hoeveel contact heb jij met allochtonen?	1 (nooit) - 2 - 3 - 4 - 5 - 6 - 7 (heel veel)	<i>Ordinaal (Likertschaal) → Interval/Ratio</i>
Contact_2	2. Hoe vaak heb je een gesprek met allochtonen?	1 (nooit) - 2 - 3 - 4 - 5 - 6 - 7 (heel veel)	
Contact_3	3. Hoe vaak heb je constructief contact met allochtonen?	1 (nooit) - 2 - 3 - 4 - 5 - 6 - 7 (heel veel)	
Contact_4	4. Hoe vaak had je reeds positieve ervaringen met allochtonen?	1 (nooit) - 2 - 3 - 4 - 5 - 6 - 7 (heel veel)	
Contact_5	5. Hoe vaak zie je allochtonen in de buurt waar je woont?	1 (nooit) - 2 - 3 - 4 - 5 - 6 - 7 (heel veel)	
Contact_6	6. Hoe vaak zie je allochtonen in de buurt waar je werkt of school loopt?	1 (nooit) - 2 - 3 - 4 - 5 - 6 - 7 (heel veel)	
Contact_7	7. Hoe vaak heb je aangenaam contact met allochtonen?	1 (nooit) - 2 - 3 - 4 - 5 - 6 - 7 (heel veel)	
RACISME			
SUBTIEL RACISME			
Subtielracisme_1	1. Allochtonen zouden het verstand moeten hebben zichzelf niet op te dringen op plaatsen waar ze bij voorbaat van weten dat ze gediscrimineerd zullen worden.	1 (helemaal oneens) - 2 - 3 - 4 - 5 - 6 - 7 (helemaal eens)	<i>Ordinaal (Likertschaal) → Interval/Ratio</i>
Subtielracisme_2	2. In het verleden bleek dat Italiaanse migranten zich vlot aanpasten aan onze cultuur. Turken en Marokkanen zouden dit ook moeten doen, zonder dat ze hiervoor speciaal beloond worden.	1 (helemaal oneens) - 2 - 3 - 4 - 5 - 6 - 7 (helemaal eens)	
Subtielracisme_3	3. Alle Marokkanen lijken op elkaar.	1 (helemaal oneens) - 2 - 3 - 4 - 5 - 6 - 7 (helemaal eens)	
Subtielracisme_4	4. Allochtonen leren aan hun kinderen waarden en vaardigheden die hun kansen op succes in onze maatschappij beperken.	1 (helemaal oneens) - 2 - 3 - 4 - 5 - 6 - 7 (helemaal eens)	
GROF RACISME			
Grofracisme_1	1. Allochtonen zijn een gevaar voor de tewerkstelling van de Belgen.	1 (helemaal oneens) - 2 - 3 - 4 - 5 - 6 - 7 (helemaal eens)	<i>Ordinaal (Likertschaal) → Interval/Ratio</i>
Grofracisme_2	2. Alle Marokkanen zijn dieven.	1 (helemaal oneens) - 2 - 3 - 4 - 5 - 6 - 7 (helemaal eens)	
Grofracisme_3	3. De Islam bedreigt onze Westerse cultuur.	1 (helemaal oneens) - 2 - 3 - 4 - 5 - 6 - 7 (helemaal eens)	
Grofracisme_4	4. Het blanke ras is superieur aan andere rassen.	1 (helemaal oneens) - 2 - 3 - 4 - 5 - 6 - 7 (helemaal eens)	
Grofracisme_5	5. Ik hou van Marokkanen**	1 (helemaal oneens) - 2 - 3 - 4 - 5 - 6 - 7 (helemaal eens)	
**Reversed item			

– Ontbrekende gegevens ('missing values'): cel leeglaten

– **Multiple response item:**

- Dat is een multiple choice-vraag waarop meerdere antwoorden mogen gegeven worden
- Invoer in Excel: 1variabele per antwoordmogelijkheid
- Voorbeeld: *Kruis aan welke huisdieren je hebt*

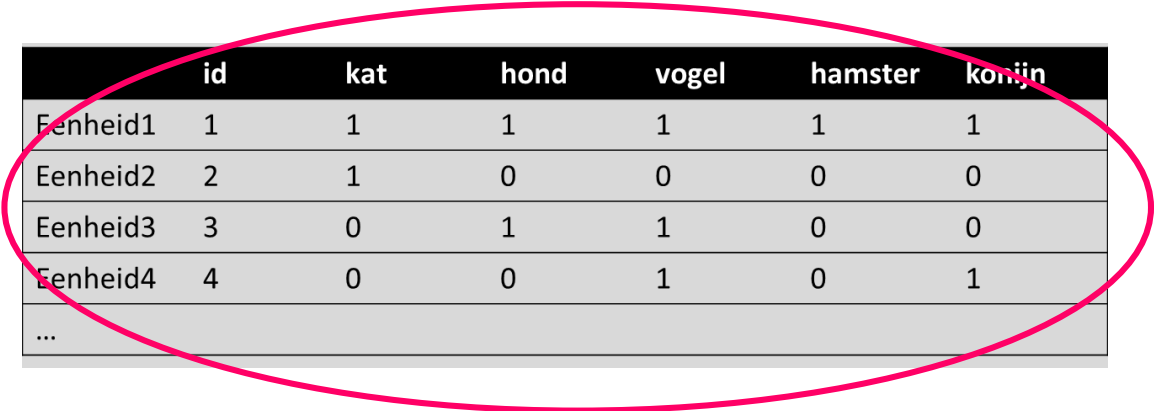
☐ *Kat* Ja = 1 / nee= 0

☐ *Hond* Ja = 1 / nee= 0

☐ *Vogel* Ja = 1 / nee= 0

☐ *Hamster* Ja = 1 / nee= 0

☐ *Konijn* Ja = 1 / nee= 0



	id	kat	hond	vogel	hamster	konijn
Eenheid1	1	1	1	1	1	1
Eenheid2	2	1	0	0	0	0
Eenheid3	3	0	1	1	0	0
Eenheid4	4	0	0	1	0	1
...						

STAP 2: DATA-PREPARATIE

- Controleer je datafile op **fout ingetikte gegevens** (onmogelijke waarden)
- Keer eventuele '**reversed items**' om
 - Score 1 → 4
 - Score 2 → 3
 - Score 3 → 2
 - Score 4 → 1
- Maak **Schaalscores** aan
 - GEM CONTACT: gemiddelde van contact _1 t.e.m. contact_4
 - SOM CONTACT: som van contact _1 t.e.m. contact_4

2. Beschrijvende analyses

- **Orderingstechnieken: tabellen en grafieken (2.1)**
- **Reductietechnieken: parameters van ligging, spreiding en vorm (2.2)**
- **Relaties tussen variabelen**
 - **op interval- en rationiveau (2.3)**
 - **op nominaal en ordinaal niveau (2.4)**

2.1 Ordeningstechnieken: tabellen en grafieken

- Doel: **datareductie**
 - Data **overzichtelijk** presenteren
 - En er **algemene uitspraken** over doen (dwz: deel van de info gaat wel verloren).
- Ordeningstechnieken:
 - 2.1.1 Frequentietabel
 - 2.2.2 Staaf- en lijndiagram
 - 2.2.3 Cirkeldiagram
 - 2.2.4 Indelen in klassen
 - 2.2.5 Histogram

2.1.1 Frequentietabel

- Geeft het aantal keer weer dat een bepaalde waarde van een variabele voorkomt

Voorbeeld.

Tabel 1. Frequentietabel van de bloedgroepen van een steekproef van 50 studenten

Bloedgroep	Absolute frequentie	Percentage
O	18	36
A	23	46
B	7	14
AB	2	4
	50	100

Voorbeeld.

Tabel 2. Frequentietabel van de mate van verstandelijke beperking bij een steekproef van 70 gebruikers van een voorziening voor personen met een beperking

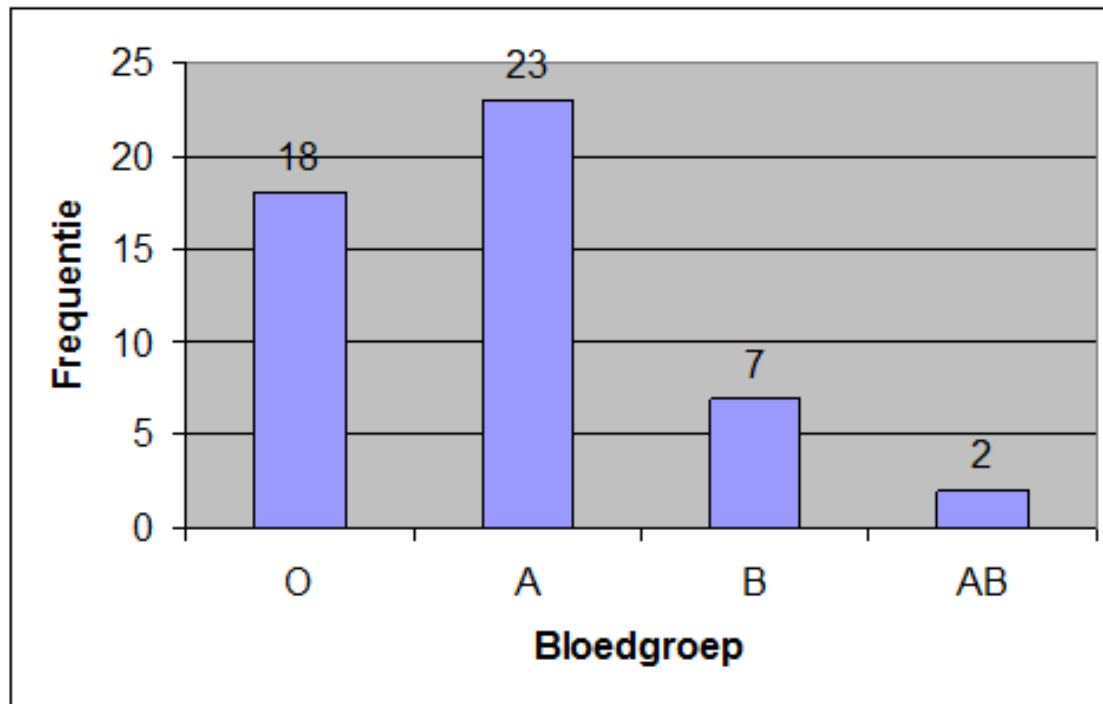


Verstandelijke beperking (VB)	Absolute frequentie	Percentage	Cumulatieve frequentie	Cumulatief percentage
Geen VB	3	4,3	3	4,3
Zwakbegaafd	7	10,0	10	14,3
Lichte VB	19	27,1	29	41,4
Matige VB	25	35,7	54	77,1
Ernstige VB	14	20,0	68	97,1
Diepe VB	2	2,9	70	100,0
	70	100	70	100,0



2.1.2 Staaf- en lijndiagram

- **Nominale en ordinale** gegevens zoals bovenstaande kunnen we grafisch voorstellen door een staafdiagram.
- Op de horizontale **X-as** plaatsen we de verschillende categorieën, op de verticale **Y-as** de frequentie.



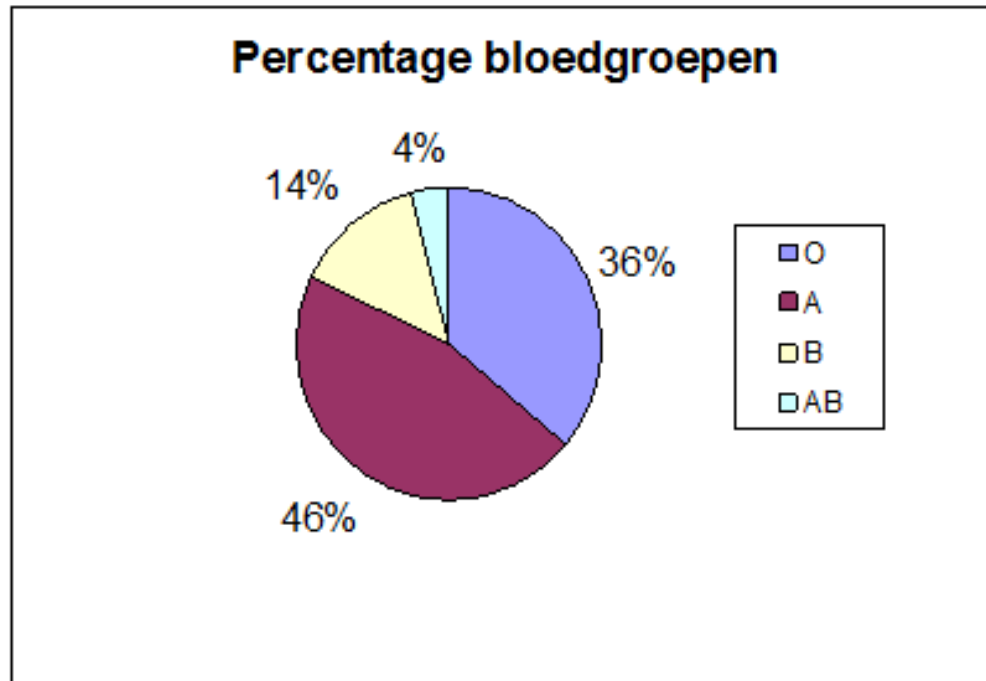
Figuur 1. Staafdiagram bloedgroepen.

- Lijndiagram: punten met elkaar verbinden

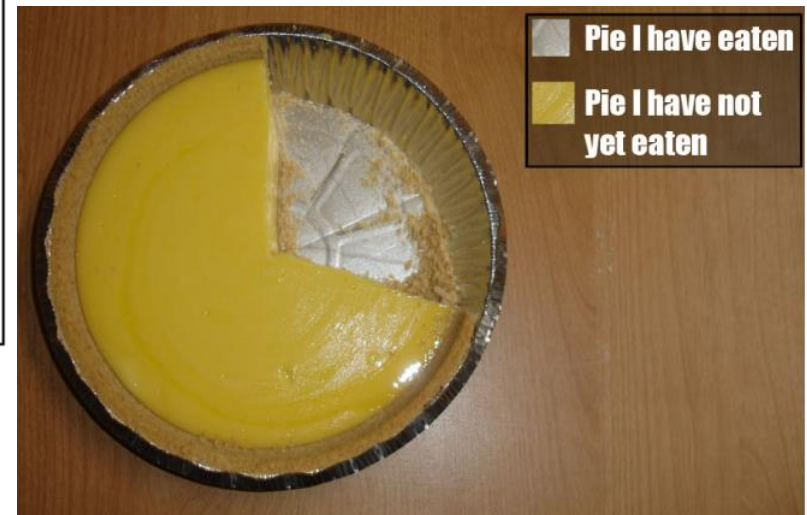


Figuur 2. Lijndiagram bloedgroepen

2.1.3 Taart- of Cirkeldiagram



Figuur 2. Cirkeldiagram bloedgroepen.



2.1.4 Indeling in klassen

- Voor variabele met heel uitgebreide waaier van waarden
- Meetwaarden groeperen of categoriseren in categorieën of klassen
- Vaak voor variabelen van **interval- en rationiveau**
- Variabele worden gereduceerd tot een categorische variabele

- *Gegeven:* frequentieverdeling van de leeftijd van de deelnemers aan een onderzoek.
- *Gevraagd:* Maak op basis van deze tabel een nieuwe frequentietabel, waarbij je de data groepeert in 5 leeftijdsklassen.

Leeftijd	Frequentie	Leeftijd	Frequentie
20	5	44	7
21	7	47	4
22	2	50	7
23	4	51	1
28	5	53	5
29	3	55	3
30	6	56	6
31	2	57	2
35	7	58	7
36	4	59	3
37	7	60	7
40	3	62	2
42	1	65	5
43	5		

Totaal	31
--------	----

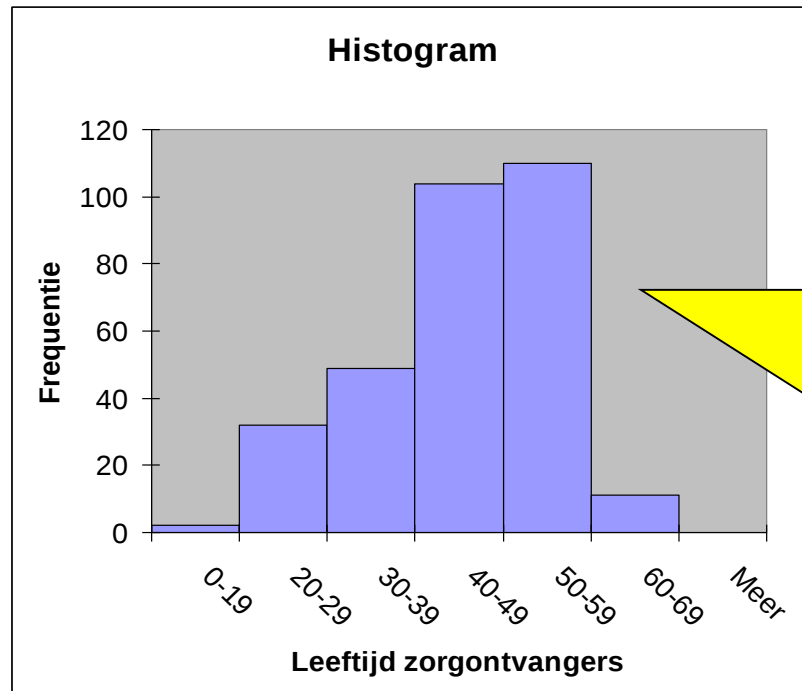


‘Leeftijdsklasse’ is nu een categorische/kwalitatieve variabele geworden (en was aanvankelijk een numerieke/kwantitatieve variabele, namelijk ‘leeftijd’)

Leeftijd	Absolute frequentie	Percentage	Cumulatieve frequentie	Cumulatief percentage
20-29	26	21,7	26	21,7
30-39	26	21,7	52	43,4
40-49	20	16,7	72	60,1
50-59	34	28,3	106	88,4
60+	14	11,7	120	100
Totaal	120	100		

2.1.5 Histogram

- Voor variabelen van hoger meetniveau (**interval-en rationiveau**) kan je een histogram gebruiken.
- Een histogram ziet eruit als een kolommendiagram, maar bij een histogram **sluiten de kolommen aan** en stellen ze **klassen van waarden** voor.



**Staafdiagram
≠
Histogram!!**

2. Beschrijvende analyses

- Ordeningstechnieken: tabellen en grafieken (2.1)
- **Reductietechnieken: parameters van ligging, spreiding en vorm (2.2)**
- Relaties tussen variabelen
 - op interval- en rationiveau (2.3)
 - op nominaal en ordinaal niveau (2.4)

2. Beschrijvende analyses

2.2 Reductietechnieken: parameters van ligging, spreiding en vorm

- 2.2.1 Parameters van positie of ligging

- Zeggen iets over **waar** de waarnemingen van een bepaalde variabele zich bevinden.

- 2.2.2 Parameters van spreiding

- Zeggen iets over hoe de waarnemingen zich **rond het gemiddelde** verdelen.

- 2.2.3 Parameters van vorm

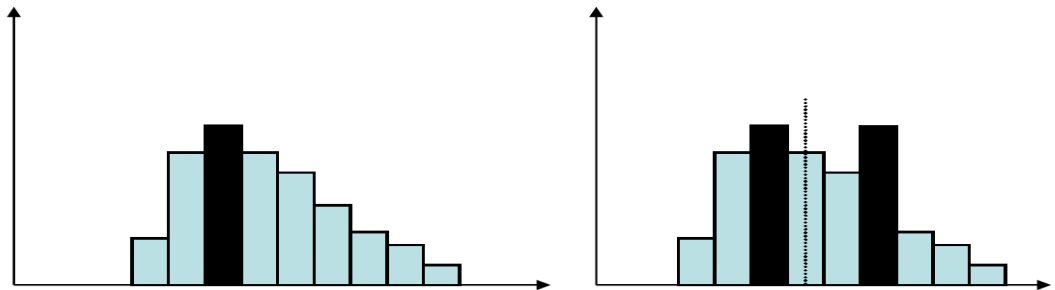
- **Symmetrie** (of scheefheid) en **spitsheid** (of platheid) van de verdeling.

2.2.1 Parameters van positie of ligging

- Modus
- Mediaan
- Gemiddelde
- Percentielen

Modus

- = de categorie of waarde in de frequentieverdeling die **het vaakst** voorkomt onder de waarnemingen.
- **Multimodale variabele**: meerdere modi
- **Unimodale variabele**: één modus



- **Modale klasse**

- **Modale klasse** = de klasse met de hoogste frequentie bij klassen van gelijke breedte.

Leeftijd	Absolute frequentie	Percentage	Cumulatieve frequentie	Cumulatief percentage
20-29	26	21,7	26	21,7
30-39	26	21,7	52	43,4
40-49	20	16,7	72	60,1
50-59	34	28,3	106	88,4
60+	14	11,7	120	100
Totaal	120	100		

Mediaan

- Waarde voor de variabele waarop de respondenten **in twee gelijke groepen** kunnen worden verdeeld qua aantallen.
 - Het is dus de waarde waarvoor evenveel respondenten worden gevonden die lager (of gelijk) scoren dan deze waarde, als dat er respondenten hoger of gelijk scoren.
 - = pc 50 (zie verder bij 'percentielen')

– Vb. 3 4 6 8 9 10 15 17 20

– Vb. 3 4 6 8 9 10 15 17



Gemiddelde

- Het **rekenkundige gemiddelde** is de waarde die we verkrijgen door alle waargenomen meetwaarden voor een variabele op te tellen en te delen door het aantal waarnemingen.

Opdracht 2.6

Bereken het rekenkundig gemiddelde bij de volgende frequentieverdeling

Aantal vragen juist	Aantal personen (FR)
1	1
3	1
4	1
5	3
9	3
10	1

Gemiddelde

$$= [1 + 3 + 4 + (3*5) + (3*9) + 10] / 10$$
$$= 60 / 10 = 6$$

- **Getrimd gemiddelde**: gemiddelde zonder extreme waarden

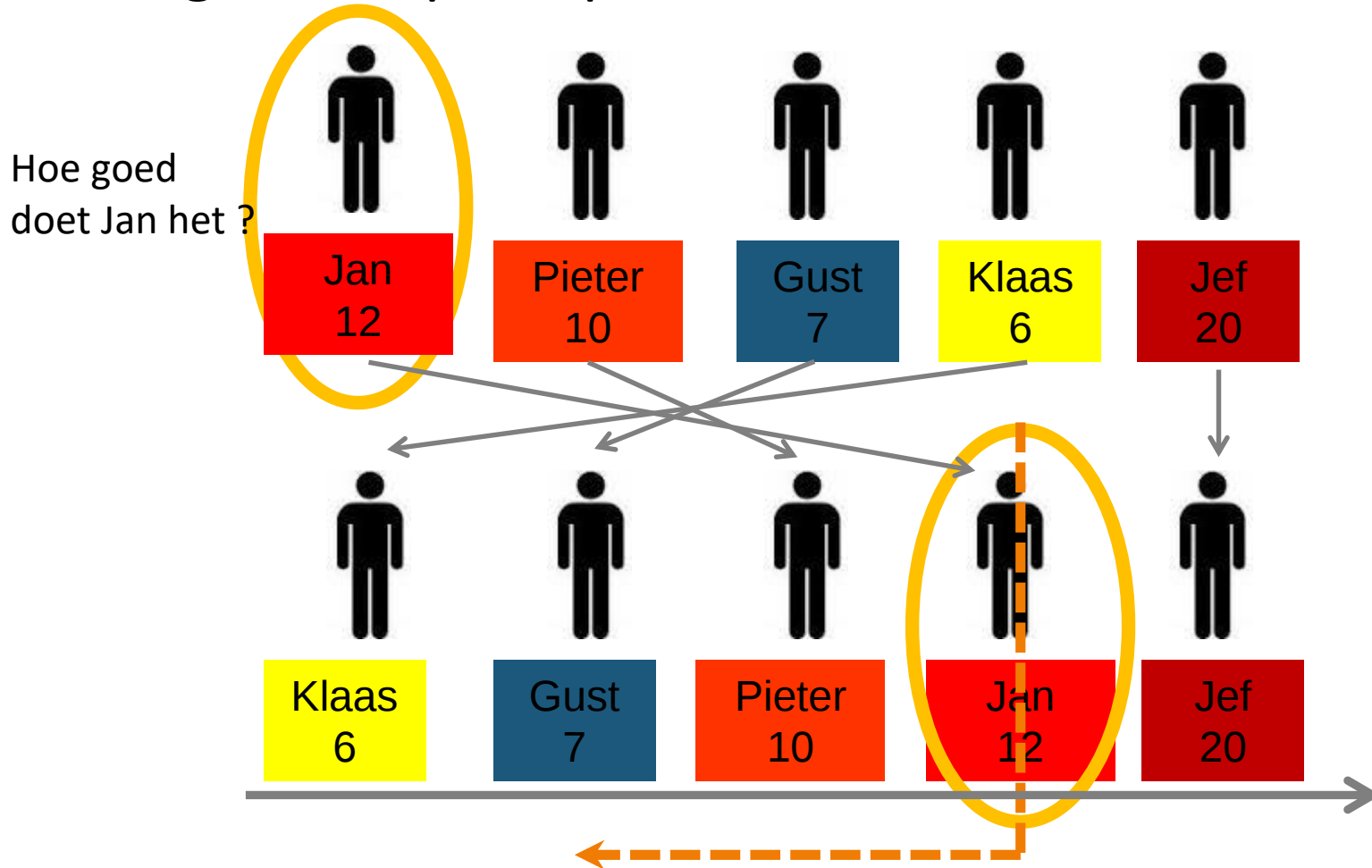


Percentielen

- Verdelen de waarnemingen (scores) in 100 groepen van gelijke grootte
- Elk groepje omvat 1% van de waarnemingen
 - Vb. pc 24: dat punt op de frequentieverdeling van een variabele waaronder 24% van de waarnemingen is gelegen
 - Vb. pc 95: dat punt op de frequentieverdeling van een variabele waaronder 95% van de waarnemingen is gelegen.
- Vb. *Iemand haalt een score 183 op een bepaalde test.*
 - Is dit goed? Slecht?
 - Bij de score 183 hoort een percentielscore 95. Weten we nu meer?

Percentielen = rang-ordening

- Algemeen principe:



X % doet het **minder of even** goed dan Jan; **DUS: Jan zit op Pc X**

– Percentielen

- Splitsen de waarnemingen in **100 klassen die elk 1%** van de waarnemingen bevatten.

– Decielen

- Splitsen de waarnemingen in **10 klassen die elk 10%** van de waarnemingen bevatten.
- Deciel 1 = percentiel 10,..., deciel 8 = percentiel 80

– Kwartielen

- Uitgedrukt in Q1, Q2, en Q3
- Splitsen de waarnemingen in **4 klassen die elk 25%** van de waarnemingen bevatten.
- Q1: grenswaarde waaronder 25% van de verdeling ligt en waarboven 75% van de verdeling ligt.
- Q1 = percentiel 25
- Q2 = percentiel 50
- Q3 = percentiel 75

Overzicht: wanneer kan je welke parameter van positie gebruiken?

Nominaal	Ordinaal	Interval	Ratio
Modus	Modus Mediaan Percentielen	Modus Mediaan Percentielen Gemiddelde	Modus Mediaan Percentielen Gemiddelde

**Belangrijk
schema!!**

2. Beschrijvende analyses

2.2 Reductietechnieken: parameters van ligging, spreiding en vorm

- 2.2.1 Parameters van positie of ligging

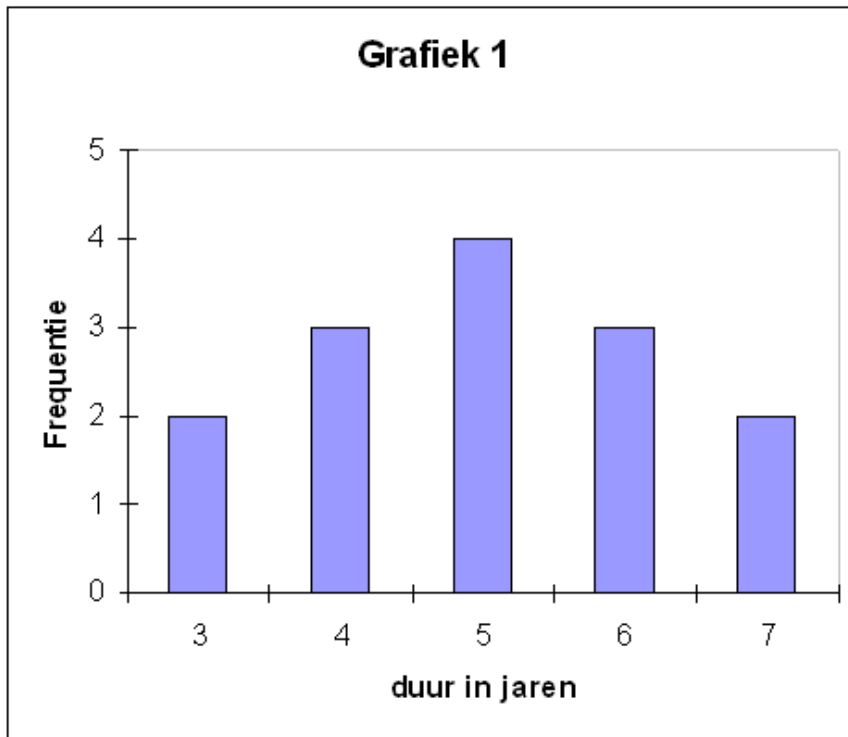
- Zeggen iets over *waar* de waarnemingen van een bepaalde variabele zich bevinden.

- 2.2.2 Parameters van spreiding

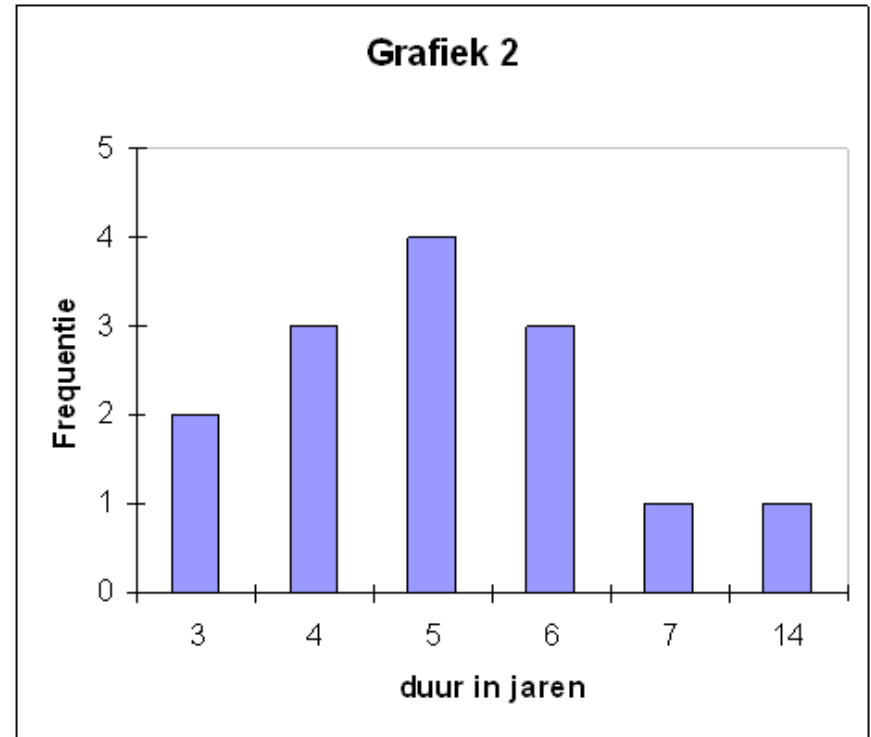
- Zeggen iets over hoe de waarnemingen zich *rond het gemiddelde* verdelen.

- 2.2.3 Parameters van vorm

- *Symmetrie* (of scheefheid) en *spitsheid* (of platheid) van de verdeling.



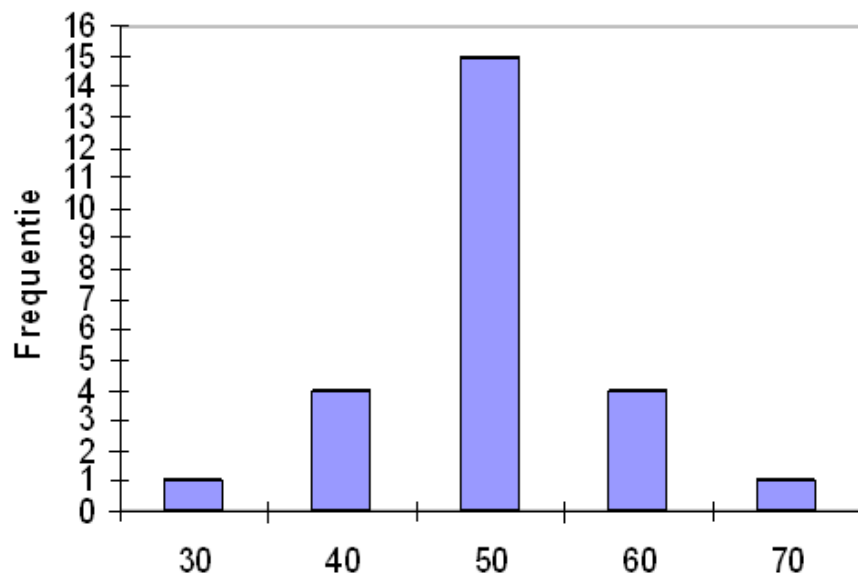
Modus: 5
 Mediaan: 5
 Gemiddelde: 5



Modus: 5
 Mediaan: 5
 Gemiddelde: 5,5

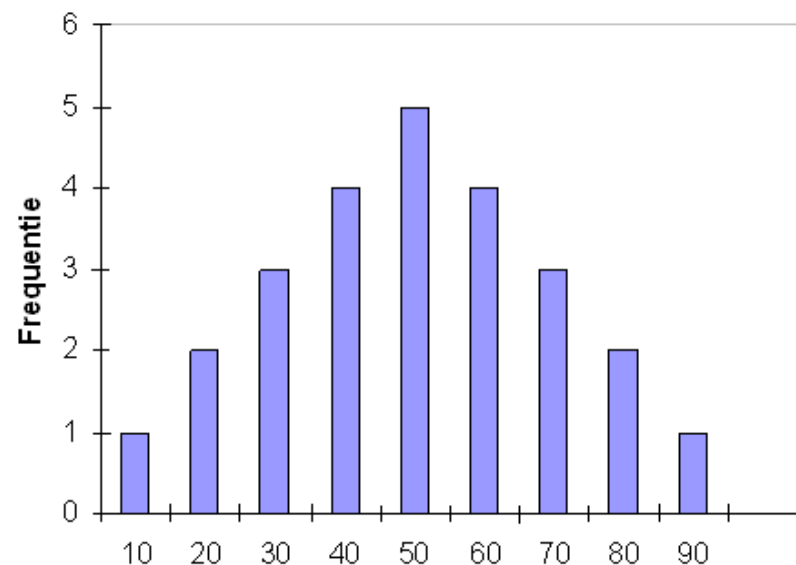
- Het gemiddelde is gevoelig aan extreme waarden (*daarom neemt men soms het getrimd gemiddelde*).
- Modus en mediaan niet.

Grafiek 3



Modus: 50
Mediaan: 50
Gemiddelde: 50

Grafiek 4



Modus: 50
Mediaan: 50
Gemiddelde: 50

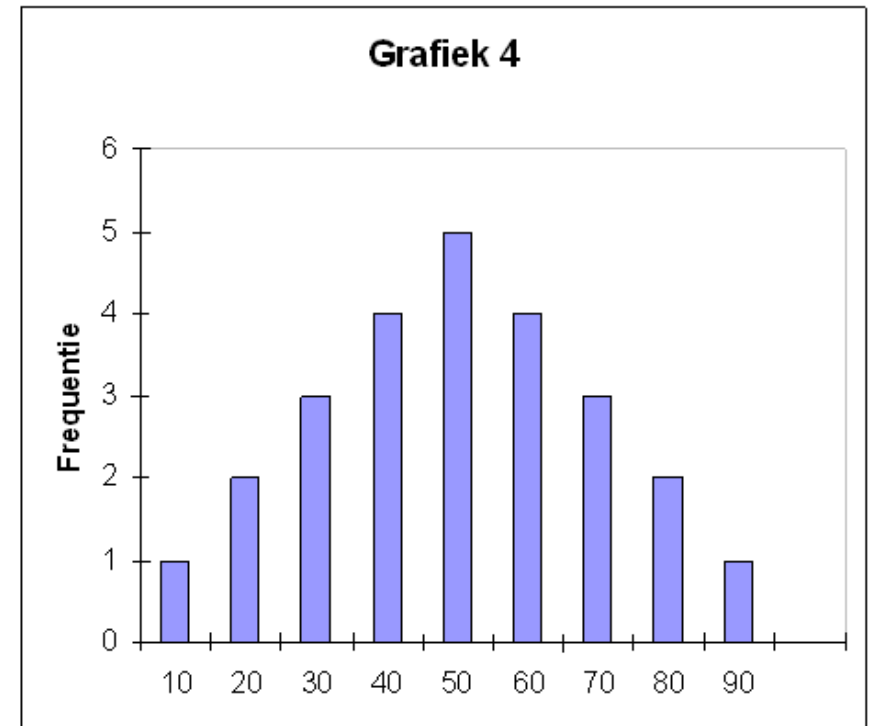
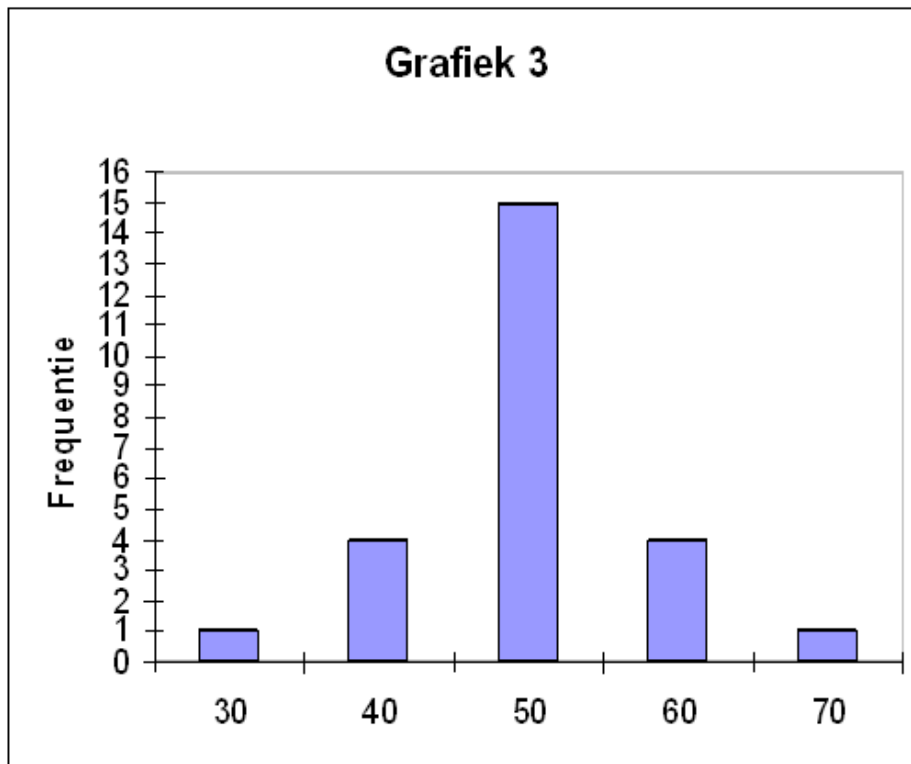
- De twee frequentieverdelingen hebben dezelfde modus, mediaan en gemiddelde. Toch zijn ze verschillend: grafiek 4 heeft een grotere **spreiding** dan grafiek 3.

2.2.2 Parameters voor spreiding

- = spreidingsmaten
- Geeft een beschrijving van de spreiding van de gegevens
- Geeft aan in hoeverre de gegevens juist **niet** gelijk zijn aan het gemiddelde.
- Maat voor de onderlinge verschillen tussen scores.
 - Variatiebreedte, range en bereik
 - Variantie en standaardafwijking

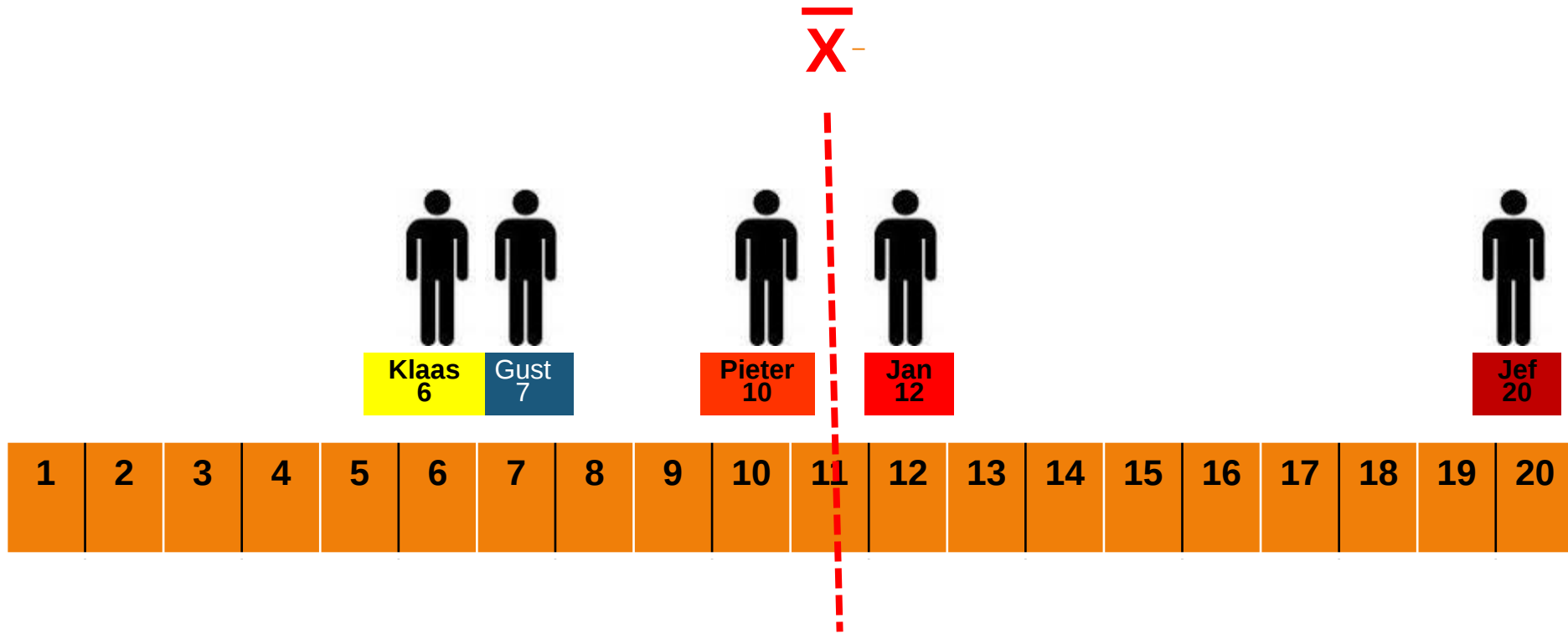
Variatiebreedte = range = bereik

- Verschil tussen de grootste en de kleinste waargenomen waarde (dus: maximumscore min minimumscore)



Variantie en standaarddeviatie

- De variantie geeft aan in hoeverre de getallen afwijken van het gemiddelde.
 - Meer bepaald is dat een uitdrukking van de **gemiddelde afwijking van het gemiddelde**
- De **variantie** is het kwadraat van de standaarddeviatie
- De **standaarddeviatie (SD)** is de vierkantswortel van de variantie
- Hoe wordt de variantie berekend?

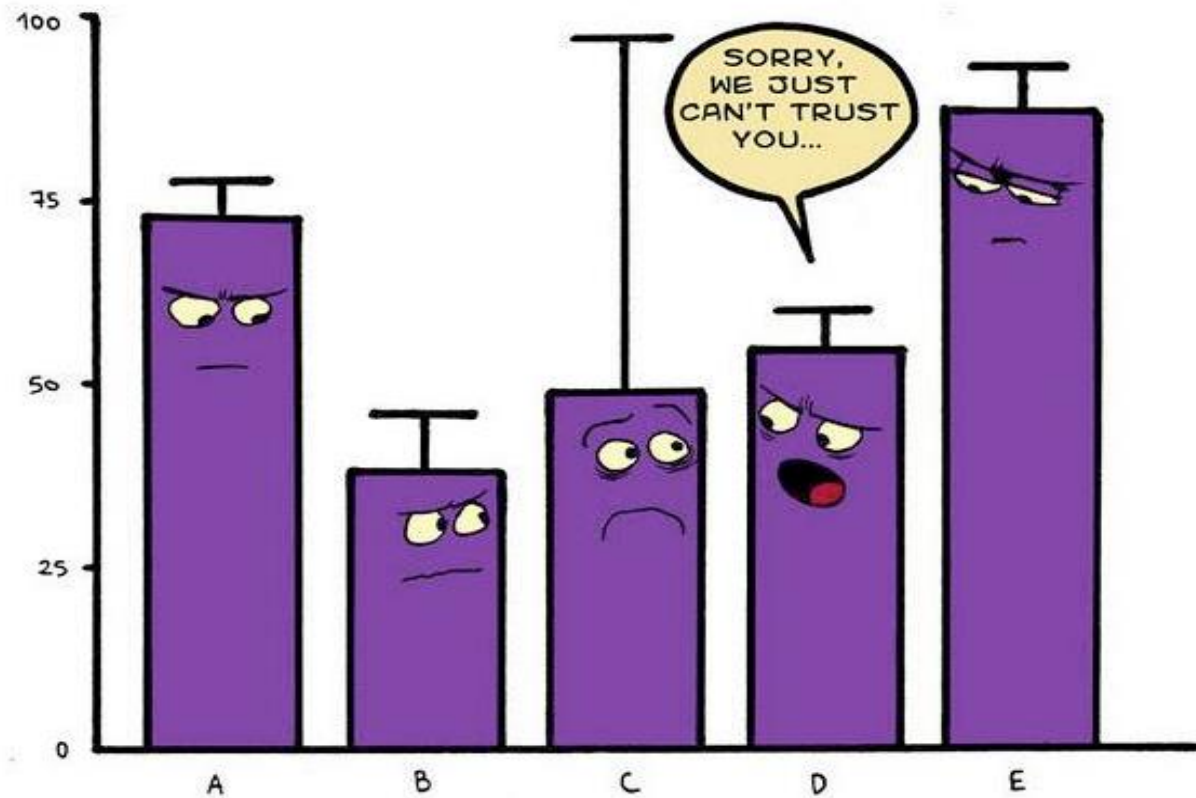


Klaas: $6 - 11 = -5$
 Gust: $7 - 11 = -4$
 Pieter: $10 - 11 = -1$
 Jan: $12 - 11 = +1$
 Jef: $20 - 11 = +9$

SD = de gemiddelde afwijking
van de gemiddelden = $\frac{\sum 0}{N}$

$$\rightarrow \sqrt{\sum (X - \bar{X})^2 / N}$$

- De standaarddeviaties worden in een staafdiagram vaak weergegeven aan de hand van **foutenbalken**



Notatie:

- Gemiddelde : M
- Standaarddeviatie: SD
- Steekproefgrootte: N



2. Beschrijvende analyses

2.2 Reductietechnieken: parameters van ligging, spreiding en vorm

- 2.2.1 Parameters van positie of ligging
 - Zeggen iets over *waar* de waarnemingen van een bepaalde variabele zich bevinden.
- 2.2.2 Parameters van spreiding
 - Zeggen iets over hoe de waarnemingen zich *rond het gemiddelde* verdelen.
- 2.2.3 Parameters van vorm
 - *Symmetrie* (of scheefheid) en *spitsheid* (of platheid) van de verdeling.

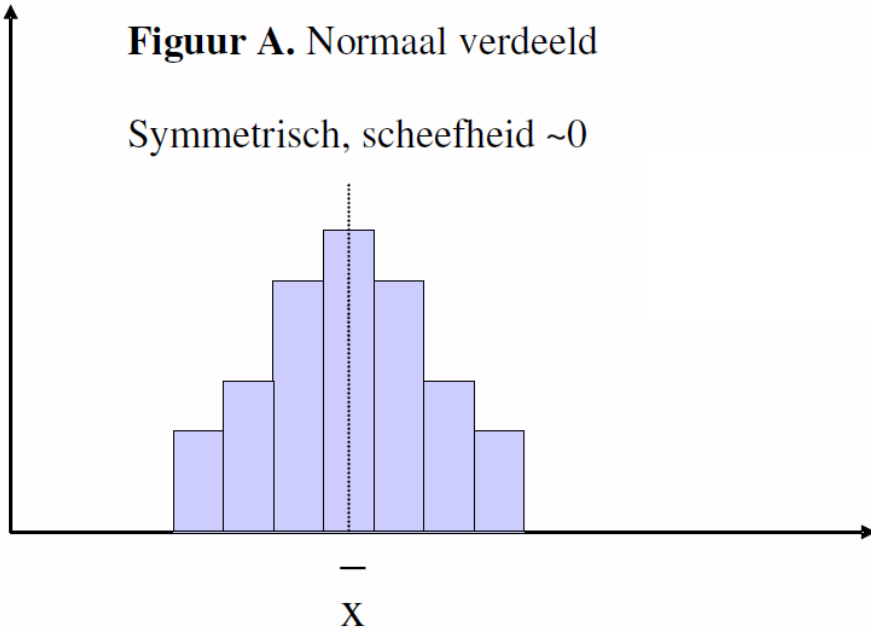
2.2.3 Parameters voor vorm

- Scheefheid van de curve
- Kurtosis ('platheid') van de curve

Scheefheid

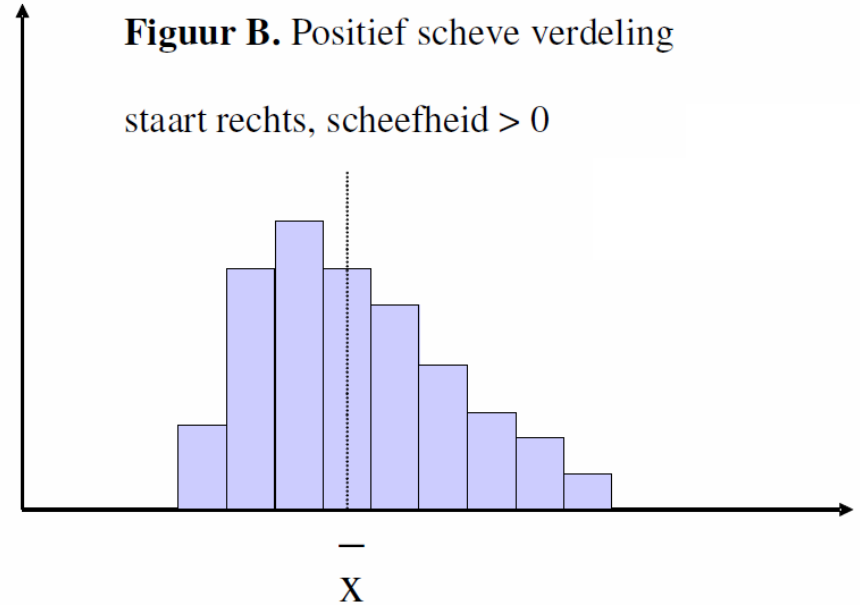
Figuur A. Normaal verdeeld

Symmetrisch, scheefheid ~ 0



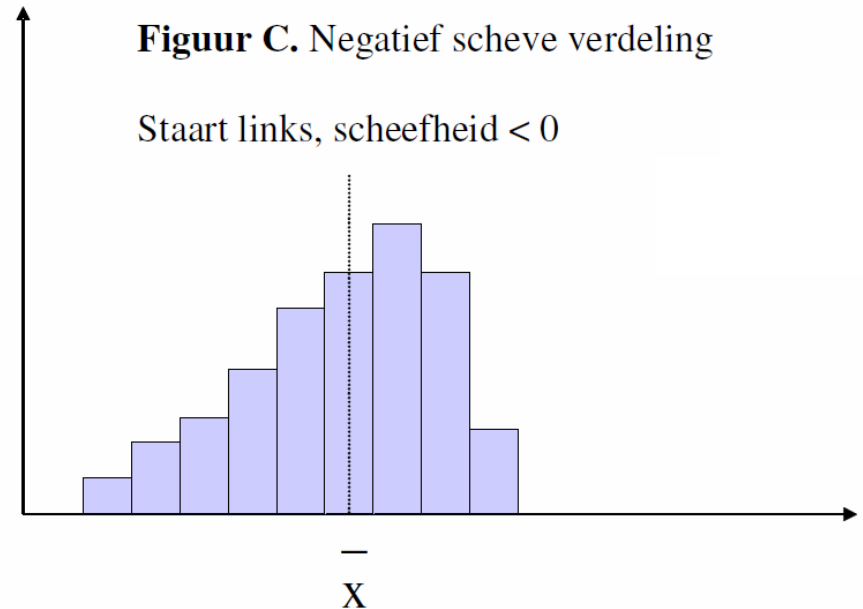
Figuur B. Positief scheve verdeling

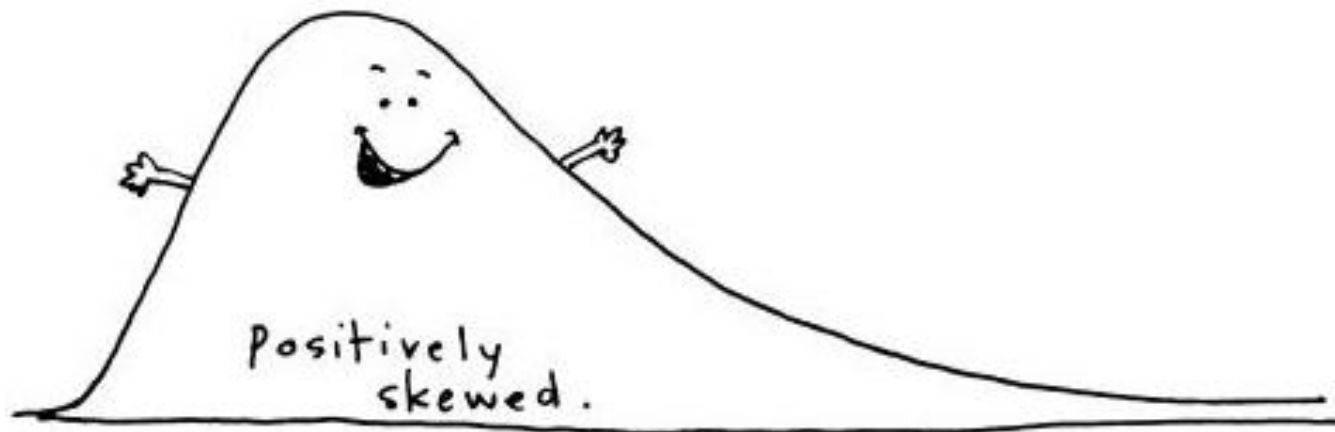
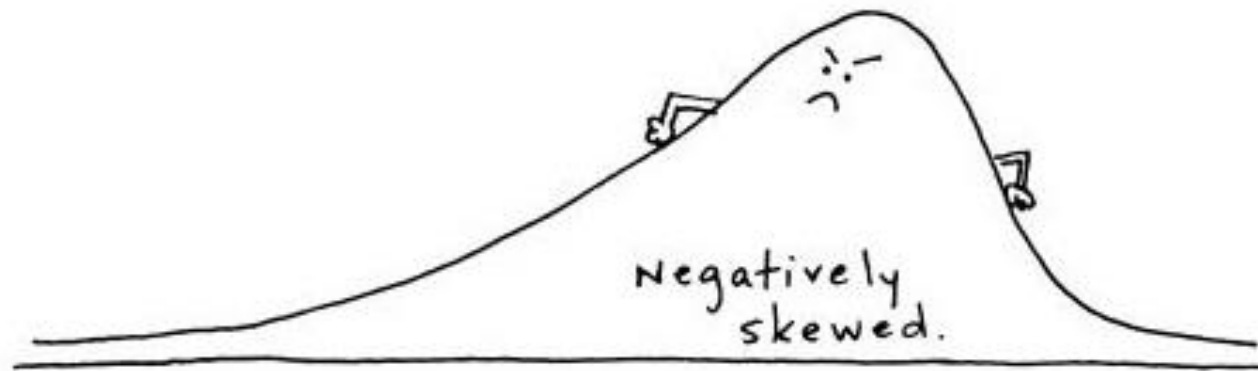
staart rechts, scheefheid > 0



Figuur C. Negatief scheve verdeling

Staat links, scheefheid < 0

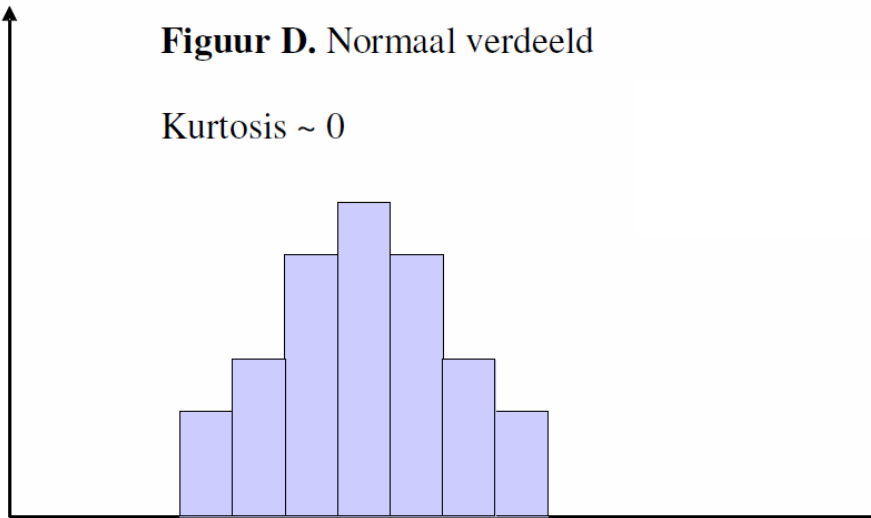




Kurtosis

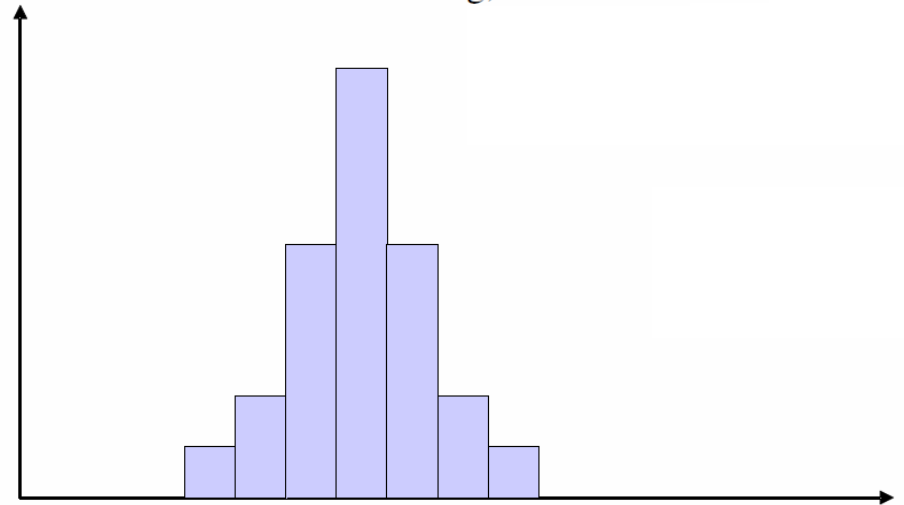
Figuur D. Normaal verdeeld

Kurtosis ~ 0



Figuur E. Spitsers in vergelijking

met normaal verdeling, kurtosis > 0



Figuur F. Platter in vergelijking

met normaal verdeling, kurtosis < 0

