Statistiek voor MBW/KW 2023-2024

Dr. Hans Melissen
algo2@live.nl (jbm.melissen@mindef.nl)
(Breda A.232/ Den Helder E.2.044)

- Voorstellen
- Organisatie van het vak

- Wat is statistiek?
- Waarom statistiek?

- Voorkennis Statistiek

Voorstellen

Hans Melissen algo2@live.nl (jbm.melissen@mindef.nl)
Breda A2.032, Den Helder E2.044

Studie Wiskunde/Natuurkunde Universiteit Utrecht

Dienstplicht: Seedorf Bbt 41 Afdva
Philips Research (+ promotie in eigen tijd)
TU Delft
Hogescholen Fontys, Breda, Den Haag, Rotterdam
Sinds 2020 bij NLDA universitair docent Operations Research/Wiskunde/Statistiek MTW

Frank Heijnen assistent bij werkcolleges in Breda. Masterstudent Universiteit Tilburg

Cursusdoelen Statistiek

De student is in staat om:	Eindterm nummer:
1. de principes van discrete en continue kansmodellen en verdelingen toe te passen.	11
2. te rekenen met een aantal veelgebruikte discrete en continue verdelingen.	11
3. enkele grootheden (verwachting, variantie, covariantie, regressievergelijking en correlatiecoëfficient) uit te rekenen.	11
4. de centrale limietstelling te gebruiken.	11
5. met enkele schattingsmethoden te werken.	11
6. met enkele betrouwbaarheidsintervallen te werken.	11
7. met enkele toetsingsmethoden te werken.	11
8. de steekproefgrootte te bepalen.	11

Organisatie van het vak Statistiek voor MBW/KW

Uren per week

Statistiek deel 1

- Week 1-5: 1 blokuur (2 lesuur) hoorcollege theorie, 1 blokuur gastcollege Roy Lindelauf (week 1), 1 blokuur werkcollege. Slides en uitwerking van opgaven zijn beschikbaar op de elo.
- Week 6: Een blokuur hoorcollege oefententamen Statistiek deel 1. Een blokuur werkcollege.
- Week 7: Voor de cadetten: Een blokuur achterstallige opgaven en tentamenvoorbereiding.
  Voor allen: Tentamen Statistiek deel 1.

Statistiek deel 2

- Week 8-13: blokuur hoorcollege theorie. Een blokuur begeleid werkcollege.
- Week 14: Tentamen Statistiek deel 2.

- Beide deeltentamens tellen even zwaar mee. Per deeltentamen minimaal 5,0 halen. Gemiddeld over de twee minimaal 5,5
- De tentamens zijn “open boek”

Organisatie van het vak Statistiek voor MBW/KW

Materialen:

- Slides + Video's op de elo. De slides zijn leidend voor de tentamenstof.
- Uitwerking van de meeste opgaven als pdf/video op de elo.
- Reader van Drs. Van der Ven (met oefententamens) op de elo.
- Boek A. Buijs Statistiek om mee te werken Theorieboek (Als achtergrondinfo, tabellen niet gebruiken!)
- Boek A. Buijs Statistiek om mee te werken Opgavenboek (Opgaven, antwoorden + sommige uitwerkingen)
- Informatie op Moodle (mededelingen, forum, oefententamens)

Statistiek voor KW/MBW - Voorkennis

Dr. Hans Melissen
algo2@live.nl (jbm.melissen@mindef.nl)
(A.232/E.2.044)

Voorkennis voor Statistiek MBW/KW

Kansrekening

Er wordt uitgegaan van een basiskennis op het gebied van kansrekening/beschrijvende statistiek.
- Wiskunde A of D: Je achtergrondkennis zou ruim voldoende moeten zijn. Kijk deze slides door om jezelf te herinneren.
- Wiskunde B of HBO: Je mist waarschijnlijk deze achtergrond. Werk deze slides door/bekijk de video.

Analyse/Calculus

Voorkennis is het kunnen rekenen met variabelen, functies, afgeleiden en integralen (zie reader). Integralen doen we met de GR.

Excel

Statistische berekeningen kunnen ook worden gedaan met behulp van Excel. Om je hiermee vertrouwd te maken voor het vervolg van de studie zal er tijdens de cursus aandacht aan worden besteed. Als je nog nooit met Excel hebt gewerkt is het verstandig om je alvast wat basisfuncties van Excel eigen te maken zodat je niet met je oren gaat flapperen als er tijdens colleges voorbeelden worden behandeld. Voor de tentamens is kennis van Excel niet noodzakelijk.

Grafische rekenmachine

Om zelf berekeningen te kunnen doen en voor het tentamen is het essentieel dat je beschikt over een grafische rekenmachine. Voorbeelden op het college gaan uit van de TI-84. Als je een Casio, HP of andere grafische rekenmachine hebt gaan de berekeningen analoog aan de TI, maar je moet zelf zorgen dat je met je eigen rekenmachine kunt werken.

Voorkennis voor Statistiek MBW/KW

Combinatoriek (Tijdens Statistiek wordt dit nauwelijks gebruikt)

- Permutaties (3.1.1)
- Variaties (3.1.2)
- Combinaties (3.1.3)
- Met of zonder terugleggen (3.1.4)

Kansrekening

- Wat is een kans? (3.2.1)
- Rekenregels voor kansen (3.2.2)
- Voorwaardelijke kansen (3.2.3 + 3.3)

Beschrijvende statistiek

- Kansvariabelen (1.1.4)
- Meetniveau's (1.1.5)
- Frequentieverdelingen (1.2.1 + 1.2.2)
- Kruistabellen (1.2.3)
- Grafische representaties (1.3)

Voorkennis voor Statistiek MBW/KW

Beschrijvende statistiek (vervolg)

- Diagrammen (1.4.1 -3)
- Histogram (1.5.2)
- Cumulatieve frequentieverdeling (1.6.1)
- Maatstaven voor ligging (2.1 + 2.2)
- Maatstaven voor spreiding (2.4)
- Boxplot (2.5)
- Populatie of steekproef? (2.7)

Wat moet je doen om Statistiek te halen?

- Zorg dat je voorkennis op peil is

- Doe vanaf het begin mee en zorg dat je niet achter gaat lopen.

- Maak **alle** opgaven zelf. Van doorlezen van de uitwerking leer je niet hoe je zelf een opgave oplost. (Als je niet kunt zwemmen kun je 100 videos van Ranomi Kromowidjojo bekijken, maar als je in het kanaal valt verzuip je nog steeds)

- Werk met anderen samen (behalve op het tentamen), motiveer elkaar en leer van anderen

- Maak minstens twee oefententamens.

Wat is Statistiek?

Kansrekening (*Probability theory*)

Wiskunde voor het rekenen met kansen

- Kans, kansverdelingen, maatruimtes
- Stochastiek

Statistiek (*Statistics*)

Wiskundige hulpmiddelen om binnen situaties met onzekerheid gefundeerde keuzes te maken/conclusie te trekken (beslissingsondersteunend)

- **Beschrijvende statistiek**: Verzamelen, analyseren en representeren van data

- **Voorspellende statistiek**: Trekken van conclusies en verbanden, hypothesetesten

Let op: Statistiek heeft twee betekenissen!

What the f@ck is Statistics?

What the f@ck is Statistics?

Bron: https://en.wikiquote.org/wiki/Joseph_Stalin#/media/File:39_Zis_Stalin.jpg

Joseph Stalin (1878-1953)

Ioseb Vissarionovich Dzhugashvili

Beroemde quote:

When one man dies it's a tragedy.

When thousands die it's statistics.

Bron: A. Antonov-Ovseyenko (1981): Портрет тирана (Portrait of a Tyrant), p. 278

Waarom is Statistiek belangrijk in je academische opleiding?

- Belangrijk om op de hoogte te zijn van kwantitatieve methoden die beslissingen en conclusies kunnen onderbouwen.

- Te gebruiken bij je bachelor- en masteronderzoek (MTO).

- Belangrijk voor Methoden en Technieken in Onderzoek, Operations Management / OR, Logistiek, analyse van militaire tactieken, inzet van wapensystemen

- Belangrijk om kennis te hebben van nieuwe ontwikkelingen: Big Data, Machine Learning, datageletterdheid

- Defensie is een IGO (Informatie Gestuurde Organisatie). “datageletterdheid” heeft een prominente rol in het nieuwe profiel van een officier bij Defensie.

- Generaal Kersbergen (KMAR): “Mijn officieren hebben hun wapen niet op hun kont maar in hun kop!”

Combinatoriek: Permutaties

Voor bepaalde problemen in de kansrekening is het belangrijk om te kunnen rekenen met aantallen mogelijkheden van een eindig aantal dingen (combinatoriek). Vaak gaat het om vazen met ballen erin.

Voorbeeld: De Marechaussee krijgt van een buitenlandse veiligheidsdienst de namen van vier in Nederland verblijvende personen (Abdelhakim, Loes, Mourad, Samir), door waarvan vermoed wordt dat ze een terroristische aanval op een strategisch belangrijk bedrijf in Nederland aan het voorbereiden zijn. Men wil deze personen achtereenvolgens aan een nader onderzoek onderwerpen. Op hoeveel manieren kan dit onderzoek worden gedaan? (hoeveel verschillende volgordes zijn er waarin deze verdachten onderzocht kunnen worden?)

Antwoord: Dit kan op de volgende manieren: ALMS, ALSM, AMLS, AMSL, ASLM, ASML, LAMS, LASM, LMAS, LMSA, LSAM, LSMA, MALS, MASL, MLAS, MLSA, MSAL, MSLA, SALM,SAML, SLAM, SLMA, SMAL, SMLA. 24 in totaal.

Als vaasprobleem: Uit een vaas met n genummerde ballen worden alle ballen één voor één getrokken. Op hoeveel manieren kan dit als de volgorde van trekken belangrijk is?

Antwoord: Voor de eerste bal zijn n mogelijkheden, voor de tweede $n-1$, etc., dus $n \cdot (n-1) \cdot (n-2) \cdots 1 = n!$

Combinatoriek: Permutaties 3.1.1

ALMS, ALSM, AMLS, AMSL, ASLM, ASML,
LAMS, LASM, LMAS, LMSA, LSAM, LSMA,
MALS, MASL, MLAS, MLSA, MSAL, MSLA,
SALM,SAML, SLAM, SLMA, SMAL, SMLA.

LMS, LSM,
MLS, MSL,
SLM, SML,

MS,
SM

S

4 X 3 X 2 X 1 = 24

Met de Grafische Rekenmachine: 4 MATH PRB ! = 24

De beste volgorde is natuurlijk: ASML.

Combinatoriek: Permutaties 3.1.1

Permutaties: Kies uit een groep van *n* verschillende objecten *k* verschillende objecten. Volgorde is belangrijk!

Voorbeeld: In een project team van 15 leden benoem:

- Een projectleider
- Een voorzitter
- Een notulist
- Een kwaliteitsfunctionaris

Op hoeveel manieren kan dit?

$${}_{15}P_4 = 15 \cdot 14 \cdot 13 \cdot 12 = \frac{15!}{11!} = 32760 \quad (= 15 \text{ nPr } 4)$$

$$Formula: \quad {}_nP_k = \frac{n!}{(n-k)!}$$

Op een TI-84 Plus:
Voer in: 15
kies MATH
→→→ PRB
↓ nPr
enter
4
Enter
32760

Combinatoriek: Variaties 3.1.2

Een iets ander probleem krijg je als je als je uit 5 personen er 3 wilt onderzoeken. Hoeveel mogelijke volgorden zijn er dan?

$$5 \times 4 \times 3 = 60$$

Met de GR: 5 MATH PRB nPr 3 = 60.

Algemeen: Op hoeveel manieren kun je k dingen kiezen uit n dingen, als de volgorde waarin je getrokken hebt wel van belang is?

Antwoord: $n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}$

Combinatoriek: Combinaties 3.1.3

Nog iets moeilijker probleem: Je wilt uit 5 personen er 3 kiezen, maar de volgorde van die drie is niet van belang. Je selecteert eerst de 3 belangrijkste verdachten voor nader onderzoek, maar de volgorde waarin je die drie onderzoekt is (nog) niet van belang. Als je dit met variaties doet krijg je teveel, want ALM wordt dan zes keer geteld (ALM, AML, LAM, LMA, MAL, MLA i.p.v. één keer. Je moet dus het aantal variaties van 3 uit 5 nog delen door het aantal volgorden van drie dingen, dus

$$\frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10$$

Met de GR: 5 MATH PRB nCr 3 = 10.

Algemeen: Op hoeveel manieren kun je k dingen kiezen uit n dingen, als de volgorde waarin je getrokken hebt NIET van belang is?

Antwoord: $n \cdot (n-1) \cdot (n-2) \cdots \frac{(n-k+1)}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$ (de laatste notatie heet wel een binomiaalcoëfficiënt)

Wat is eigenlijk een kans?

1. Wat is de kans dat je met een dobbelsteen 6 gooit?

Wat is eigenlijk een kans?

1. Wat is de kans dat je met een dobbelsteen 6 gooit?

Antwoord: 1/6, want er zijn zes mogelijkheden

Wat is eigenlijk een kans?

1. Wat is de kans dat je met een dobbelsteen 6 gooit?

Antwoord: 1/6, want er zijn zes mogelijkheden

Antwoord 2: ≤ 1/6, want er zijn zeven mogelijkheden
(niet gooien kan ook)

Wat is eigenlijk een kans?

1. Wat is de kans dat je met een dobbelsteen 6 gooit?

Antwoord: 1/6, want er zijn zes mogelijkheden

Antwoord 2: ≤ 1/6, want er zijn zeven mogelijkheden
(niet gooien kan ook)

Antwoord 3: Kweenie, de dobbelsteen is misschien vals

Wat is eigenlijk een kans?

1. Wat is de kans dat je met een dobbelsteen 6 gooit?

Antwoord: 1/6, want er zijn zes mogelijkheden

Antwoord 2: ≤ 1/6, want er zijn zeven mogelijkheden
(niet gooien kan ook)

Antwoord 3: Kweenie, de dobbelsteen is misschien vals

2. Wat is de kans dat je een prijs wint bij de volgende trekking van de staatsloterij? (als je meedoet)

Wat is eigenlijk een kans?

1. Wat is de kans dat je met een dobbelsteen 6 gooit?

Antwoord: 1/6, want er zijn zes mogelijkheden

Antwoord 2: ≤ 1/6, want er zijn zeven mogelijkheden (niet gooien kan ook)

Antwoord 3: Kweenie, de dobbelsteen is misschien vals

2. Wat is de kans dat je een prijs wint bij de volgende trekking van de staatsloterij? (als je meedoet)

Je kans op een prijs is het hoogst van alle loterijen

De kans op een prijs bij de Staatsloterij wisselt per trekking en is in 2020 gemiddeld over alle 16 trekkingen 51,2%. Dat betekent dat er gemiddeld op 51,2% van de verkochte lotnummers een prijs van minimaal €5 valt (heel lot). De kans dat je een prijs wint, bij deelname aan 16 trekkingen, is dus maar liefst 1 op 2! Dat is de hoogste winkans van alle Nederlandse loterijen.

Bron: https://staatsloterij.nederlandseloterij.nl/spel/winkans

Wat is eigenlijk een kans?

1. Wat is de kans dat je met een dobbelsteen 6 gooit?

Antwoord: 1/6, want er zijn zes mogelijkheden

Antwoord 2: ≤ 1/6, want er zijn zeven mogelijkheden (niet gooien kan ook)

Antwoord 3: Kweenie, de dobbelsteen is misschien vals

2. Wat is de kans dat je een prijs wint bij de volgende trekking van de staatsloterij? (als je meedoet)

De staatsloterij vergeet voor het gemak te vertellen dat vrijwel alle prijzen kleiner zijn dan je inleg.

Je kans op een prijs is het hoogst van alle loterijen

De kans op een prijs bij de Staatsloterij wisselt per trekking en is in 2020 gemiddeld over alle 16 trekkingen 51,2%. Dat betekent dat er gemiddeld op 51,2% van de verkochte lotnummers een prijs van minimaal €5 valt (heel lot). De kans dat je een prijs wint, bij deelname aan 16 trekkingen, is dus maar liefst 1 op 2! Dat is de hoogste winkans van alle Nederlandse loterijen.

Bron: https://staatsloterij.nederlandseloterij.nl/spel/winkans

Zuivere en onzuivere dobbelstenen en munten

Bij een **zuivere** (eerlijke) dobbelsteen is de kans op elke uitkomst gelijk, bij een onzuivere niet.

Aantal ogen	Zuivere dobbelsteen	Verzwaarde dobbelsteen
1	$^1/_6$	0,01
2	$^1/_6$	0,05
3	$^1/_6$	0,05
4	$^1/_6$	0,05
5	$^1/_6$	0,05
6	$^1/_6$	0,79

Hetzelfde geldt voor munten
Wat is eigenlijk een onzuivere munt?
Recent onderzoek (2023 UvA https://doi.org/10.48550/arXiv.2310.04153):
De kans dat bij een tos de bovenkant van de munt boven blijft is 50,8%

Wat is een kans? Dagelijks gebruik

- Het begrip kans in de Nederlandse taal is ambigu
  - Kans is een mogelijkheid (een kans krijgen, waarnemen of mislopen)
  - Kans is een maat voor waarschijnlijkheid (Volkskrant: “Kans op economische crisis minimaal”)

Wat is een kans? Dagelijks gebruik

- Het begrip kans is in de Nederlandse taal ambigu
  - Kans is een mogelijkheid (een kans krijgen, waarnemen of mislopen)
  - Kans is een maat voor waarschijnlijkheid (Volkskrant: “Kans op economische crisis minimaal”)

- Kwalitatief, maar ook kwantitatief, meestal als percentage, of als 1 op de 200
  - Ik weet het 200% zeker
  - Ik weet het niet, het kan vriezen, het kan dooien, fifty-fifty, 50%

Wat is een kans? Dagelijks gebruik

- Het begrip kans is in de Nederlandse taal ambigu
  - Kans is een mogelijkheid (een kans krijgen, waarnemen of mislopen)
  - Kans is een maat voor waarschijnlijkheid (Volkskrant: “Kans op economische crisis minimaal”)

- Kwalitatief, maar ook kwantitatief, meestal als percentage, of als 1 op de 200
  - Ik weet het 200% zeker
  - Ik weet het niet, het kan vriezen, het kan dooien, fifty-fifty, 50%
  - De kans bestaat dat het college van volgende week niet doorgaat
    Dit is een contaminatie (taalfout) van de “de mogelijkheid bestaat” en “de kans is niet nul”

- Feit: Kansen zijn best lastig in te schatten, vooral heel kleine kansen (besmettingskansen bij Covid).
  Voor veel mensen is een kans van 0,01 en 0,0001 allebei “bijna nooit”

Wat is een kans? Dagelijks gebruik

- Het begrip kans is in de Nederlandse taal ambigu
  - Kans is een mogelijkheid (een kans krijgen, waarnemen of mislopen)
  - Kans is een maat voor waarschijnlijkheid (Volkskrant: “Kans op economische crisis minimaal”)

- Kwalitatief, maar ook kwantitatief, meestal als percentage, of als 1 op de 200
  - Ik weet het 200% zeker
  - Ik weet het niet, het kan vriezen, het kan dooien, fifty-fifty, 50%
  - De kans bestaat dat het college van volgende week niet doorgaat
    Dit is een contaminatie (taalfout) van de “de mogelijkheid bestaat” en “de kans is niet nul”

- Feit: Kansen zijn best lastig in te schatten, vooral heel kleine kansen (besmettingskansen bij Covid).
  Voor veel mensen is een kans van 0,01 en 0,0001 allebei “bijna nooit”

- Feit: Kinderen vanaf 5 jaar hebben al ongeveer hetzelfde gevoel voor kansen als volwassenen.

Hoe groot is de kans dat in de komende 50 jaar de Militaire Willems-Orde voor het eerst ooit aan een vrouw wordt uitgereikt?

a. Minstens 0,5
b. Tussen 0,1 en 0,5
c. Klein maar niet 0, hoogstens 0,1
d. 0

Twee vrouwen kregen tot dusver de Militaire Willems-Orde

Prinses Wilhelmina 1948

Twee vrouwen kregen tot dusver de Militaire Willems-Orde

Prinses Wilhelmina 1948

Jos Gemmeke 1955

Twee vrouwen kregen tot dusver de Militaire Willems-Orde

Prinses Wilhelmina 1948

Jos Gemmeke 1955

Kansrekening: Wat is een kans? (3.2.1)

Linda is 31 jaar oud, vrijgezel, uitgesproken en erg slim. Ze studeerde filosofie. Als student was ze erg betrokken bij kwesties als discriminatie en sociale rechtvaardigheid. Ze nam tijdens haar studie ook deel aan #MeToo en Black Lives Matter demonstraties.

Welke kans is groter:

1 Linda werkt als beleidsmedewerker op een ministerie.

2 Linda werkt als beleidsmedewerker op een ministerie en is actief in de feministische beweging.


This is not Linda

Kansrekening: Wat is een kans? (3.2.1)

Linda is 31 jaar oud, vrijgezel, uitgesproken en erg slim. Ze studeerde filosofie. Als student was ze erg betrokken bij kwesties als discriminatie en sociale rechtvaardigheid. Ze nam tijdens haar studie ook deel aan #MeToo en Black Lives Matter demonstraties.


This is not Linda

Welke kans is groter:

1 Linda werkt als beleidsmedewerker op een ministerie.

2 Linda werkt als beleidsmedewerker op een ministerie en is actief in de feministische beweging.

Daniel Kahneman (Nobelprijs Economie): 80% kiest voor 2 (*conjunction fallacy*)

Kansrekening: Wat is een kans? (3.2.1)

Kansrekening: Wat is een kans? (3.2.1)

Kansdefinitie van Pascal:

$$\text{Kans op een gebeurtenis} = \frac{\text{aantal gunstige gebeurtenissen}}{\text{totaal aantal gebeurtenissen}}$$

Beperkingen: - Het gaat om eindig veel gebeurtenissen
- Alle gebeurtenissen zijn even waarschijnlijk.

Voorbeeld 1: Je gooit met een zuivere dobbelsteen. Wat is de kans dat je zes ogen gooit?

Antwoord 1: De uitkomstenruimte is {1, 2, 3, 4, 5, 6}. De kans op het gooien van zes ogen is dus $\frac{1}{6}$.

Kansrekening: Wat is een kans? (3.2.1)

Kansdefinitie van Pascal:

$$\text{Kans op een gebeurtenis} = \frac{\text{aantal gunstige gebeurtenissen}}{\text{totaal aantal gebeurtenissen}}$$

Beperkingen: - Het gaat om eindig veel gebeurtenissen
- Alle gebeurtenissen zijn even waarschijnlijk.

Voorbeeld 1: Je gooit met een zuivere dobbelsteen. Wat is de kans dat je zes ogen gooit?

Antwoord 1: De uitkomstenruimte is {1, 2, 3, 4, 5, 6}. De kans op het gooien van zes ogen is dus $\frac{1}{6}$.

Voorbeeld 2: Je gooit met twee zuivere dobbelstenen. Wat is de kans dat je zes ogen gooit?

Antwoord 2: De uitkomstenruimte is {2, 3, 4, …, 11, 12}. De kans op het gooien van zes ogen is dus $\frac{1}{11}$.

Kansrekening: Wat is een kans? (3.2.1)

Kansdefinitie van Pascal:

$$\text{Kans op een gebeurtenis} = \frac{\text{aantal gunstige gebeurtenissen}}{\text{totaal aantal gebeurtenissen}}$$

Beperkingen: - Het gaat om eindig veel gebeurtenissen
- Alle gebeurtenissen zijn even waarschijnlijk.

Voorbeeld 1: Je gooit met een zuivere dobbelsteen. Wat is de kans dat je zes ogen gooit?

Antwoord 1: De uitkomstenruimte is {1, 2, 3, 4, 5, 6}. De kans op het gooien van zes ogen is dus $\frac{1}{6}$.

Voorbeeld 2: Je gooit met twee zuivere dobbelstenen. Wat is de kans dat je zes ogen gooit?

Antwoord 2: De uitkomstenruimte is {2, 3, 4, …, 11, 12}. De kans op het gooien van zes ogen is dus $\frac{1}{11}$.

Correct antwoord 2: Uitkomstenruimte: {1+1, 1+2, 1+3, …, 6+5, 6+6}. Kans op zes ogen is dus $\frac{5}{36}$.

Wiskundeacademie (NL 9:30): https://www.youtube.com/watch?v=ERvx0pngqqc

Kansrekening: Wat is een kans? (3.2.1)

Juiste oplossing: De uitkomstenruimte S bestaat niet uit totalen, maar uit paren:

S = {(n,m) | n, m = 1,2,3, ..., 6} =

{(1,1), (1,2), ..., (1,6), (2,1, (2,2) ,

(2,3), ..., (2,6), (3,1), ..., (6,5), (6,6)}.

6 x 6 = 36 mogelijke uitkomsten.

Gunstige uitkomsten:

1+5, 2+4, 3+3, 4+2, 5+1.

5 gunstige uitkomsten.

6 gooien met een paar dobbelstenen						
Dobbel 1 → Dobbel 2↓	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$$P(6) = \frac{5}{36} = 0.138\ldots$$

De totale kans van alle uitkomsten in de uitkomstenruimte is altijd 1: P(S) = 1

Vb. De kans om 2 or 3 or 4 ... or 12 te gooien met twee dobbelstenen is 1.

Wat is een kans? Dagelijks gebruik

Vraag: Op 24 februari 2022 startte president Poetin van Rusland een vredesmissie om buurland Oekraïne te bevrijden van fascistische invloeden, desnoods met inzet van nucleaire middelen. Wat is de kans dat deze actie binnen een jaar voorbij is?

Wat is een kans? Dagelijks gebruik

Vraag: Op 24 februari 2022 startte president Poetin van Rusland een vredesmissie om buurland Oekraïne te bevrijden van fascistische invloeden, desnoods met inzet van nucleaire middelen. Wat is de kans dat deze actie binnen een jaar voorbij is?

Kritische vragen:

- Op welk moment in tijd wordt deze vraag gesteld ?
- Is dit een “kansexperiment” dat je kunt herhalen?

Wat is een kans? Dagelijks gebruik

Vraag: Op 24 februari 2022 startte president Poetin van Rusland een vredesmissie om buurland Oekraïne te bevrijden van fascistische invloeden, desnoods met inzet van nucleaire middelen. Wat is de kans dat deze actie binnen een jaar voorbij is?

Kritische vragen:

- Op welk moment in tijd wordt deze vraag gesteld ?
- Is dit een “kansexperiment” dat je kunt herhalen?

“The probability that the war between Russia and Ukraine will be over within a year is 50%. The probability that it will last for another three years or more is twenty percent.”

Fokkink, R., & Lindelauf, R. When Will It End? Assessing the Duration of Putin’s War with Ukraine. https://library.oapen.org/bitstream/handle/20.500.12657/87676/9789400604742.pdf?sequence=1#page=538.

Kansrekening: Rekenregels voor kansen (3.2.2)

Twee gebeurtenissen zijn:

- Onafhankelijk: het wel of niet optreden van een gebeurtenis is niet afhankelijk van de tweede.

  bv. Het gooien van 6 met een dobbelsteen is onafhankelijk van het gooien van 6 in een eerdere worp.

  Alle trekkingen met terugleggen zijn onafhankelijk.

- Afhankelijk:

  bv. De kans dat iemand ouder dan 80 wordt is afhankelijk van rookgewoontes.

  Lootjes trekken hangt af van eerdere trekkingen.

  Een verkeerd product uit een productie kan afhankelijk zijn van eerdere foute producten uit dezelfde productielijn.

Kansrekening: Rekenregels voor kansen (3.2.2)

Voorbeeld: kies een kaart uit een normaal spel van 52 kaarten.

A = je pakt harten

B = je pakt een rode kaart

$P(A) = \frac{13}{52} = \frac{1}{4}$

$P(B) = \frac{26}{52} = \frac{1}{2}$

$P(A \cap B) = P(A) = \frac{1}{4}$ ($\cap$ betekent: doorsnede, “en”)

$P(A \cup B) = P(B) = \frac{26}{52} = \frac{1}{2}$ ($\cup$ betekent: vereniging, “of”)

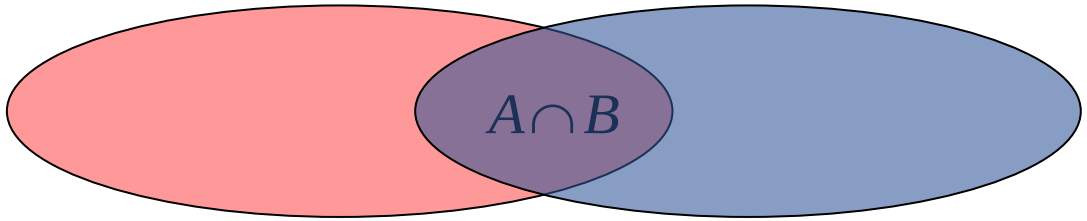
$P(A|B) = \frac{1}{26}$ (| betekent “onder voorwaarde dat”, conditionele kans)

$P(B|A) = \frac{1}{1} = 1$

In het algemeen is $P(A|B) \neq P(B|A) \neq P(A \cap B)$

Kansrekening: Rekenregels voor kansen (3.2.2)

Optelregel (voor overlappende verzamelingen):



A=meisjes B=brildragers

$$P(A \cup B) = P(A \text{ of } B) = P(A) + P(B) - P(A \cap B)$$

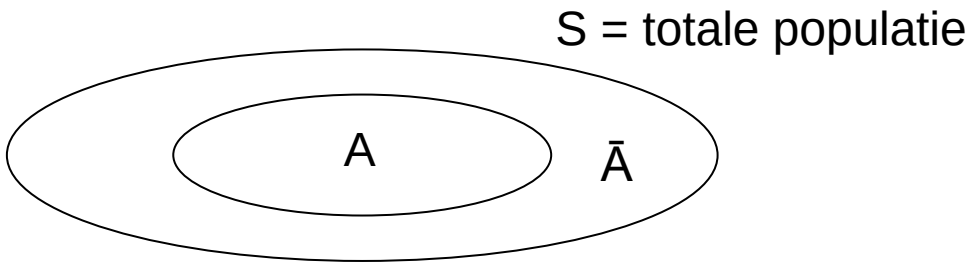
Speciaal geval for niet-overlappende verzamelingen:

$$P(A \cup B) = P(A \text{ or } B) = P(A) + P(B)$$

A=draagt contact lenzen B=draagt bril

Kansrekening: Rekenregels voor kansen (3.2.2)

Complement regel:



$$P(S) = 1$$

$$P(A) = 1 - P(\overline{A})$$

De complementregel is handig als P(A) moeilijkler te bepalen is dan P(Ā)

Kansrekening: Voorbeeld Verjaardagsparadox

Wat is de kans dat in een groep van n personen er twee op dezelfde dag jarig zijn?

Dit is lastig uit te rekenen, maar eenvoudig als complementaire kans:

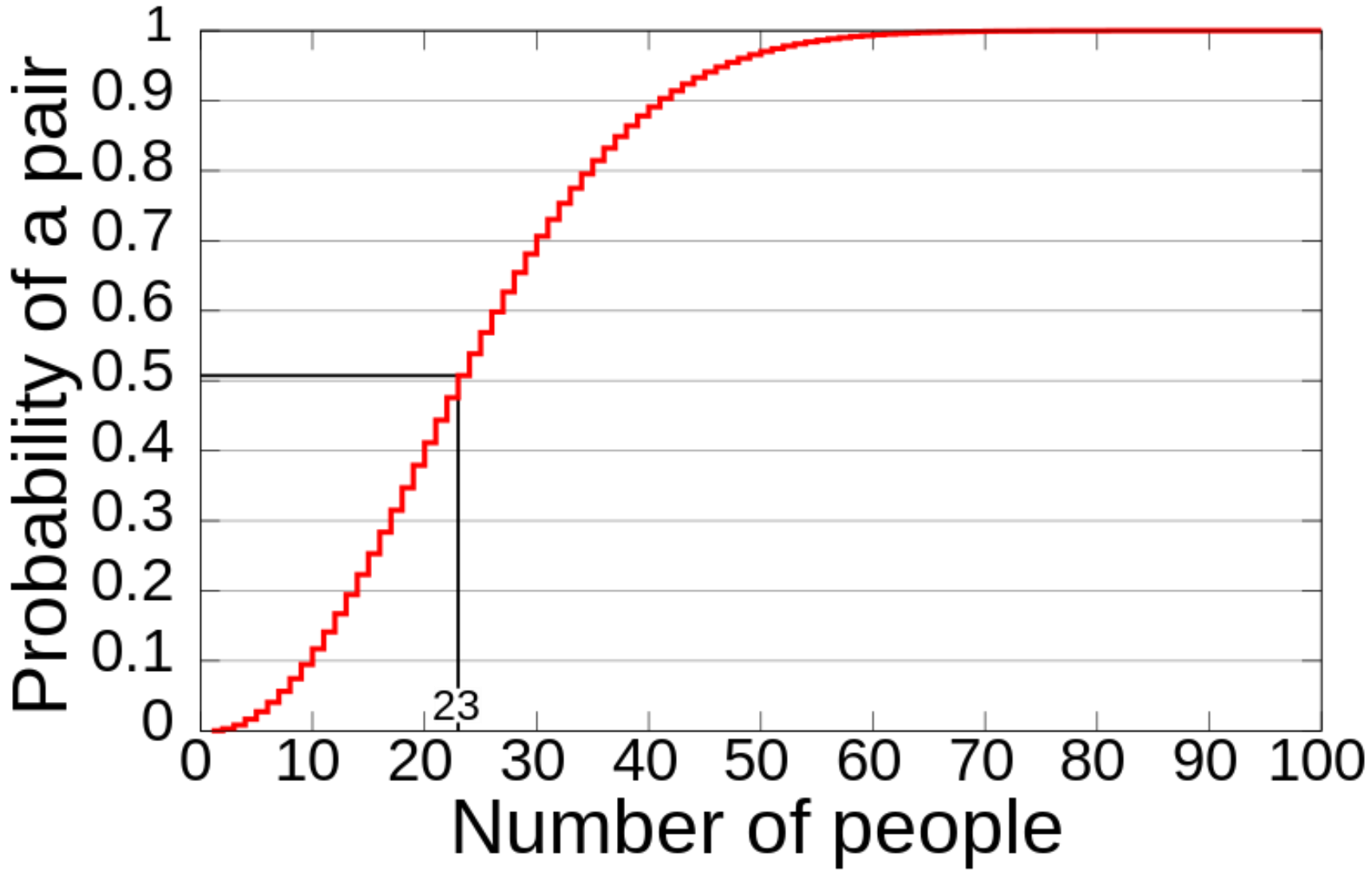
Wat is de kans dat alle verjaardagen op verschillende dagen vallen?

$$P(\bar{A}) = \frac{365 \cdot 364 \cdot 363 \cdots (365 - n + 1)}{365 \cdot 365 \cdot 365 \cdots 365}$$

De kans dat et minstens twee op dezelfde dag jarig zijn is:

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{365 \cdot 364 \cdot 363 \cdots (365 - n + 1)}{365 \cdot 365 \cdot 365 \cdots 365}$$

Kansrekening: Voorbeeld Verjaardagsparadox



In een groep van 23 personen is de kans dat er twee op dezelfde datum jarig zijn al meer dan 50%

Kansrekening: Voorwaardelijke kansen (3.2.3 + 3.3)

Het kiezen van de uitkomstenruimte S is belangrijk. De kans dat een persoon een vrouw is kan behoorlijk van de normale 50% afwijken als je bijvoorbeeld kijkt naar vrachtwagenchauffeurs of verplegend personeel.

Notatie:

P(A) = De kans dat “A” optreedt (A is een verzameling van gebeurtenissen)

bv. P(parkeerplaats is bezet op een willekeurig moment) = 0.42

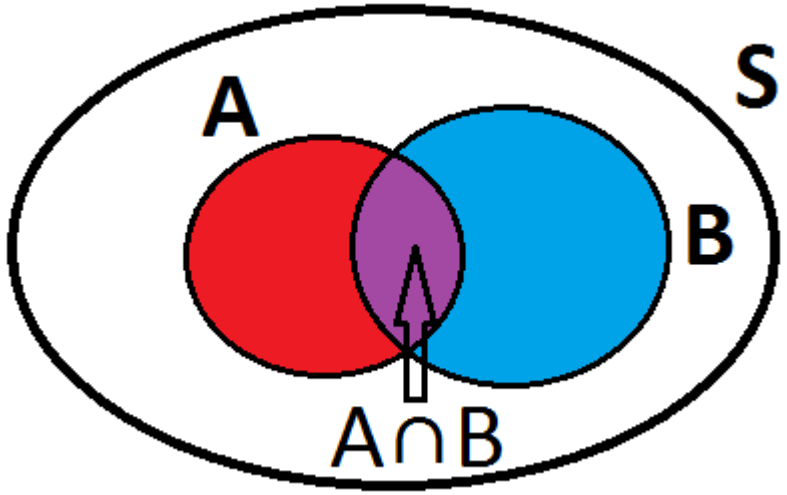
P(Ā) = De kans dat “A” **niet** optreedt

P(A|B) = de kans dat “A” optreedt terwijl je weet dat “B” ook optreedt

bv. P(kop | vorige worp was munt) = 50% = P(kop *(onderling onahankelijk)*

P(het regent morgen | vandaag schijnt de zon) = 10% *(afhankelijk)*

Kansrekening: Voorwaardelijke kansen (3.2.2)



Rekenregel voor conditionele kansen:

$$P(A|B) = \frac{n(A \cap B)}{n(B)} = \frac{n(A \cap B)}{n(S)} \cdot \frac{n(S)}{n(B)} = \frac{p(A \cap B)}{p(B)}$$

A en B zijn onafhankelijk betekent: P(A|B) = P(A),

ofwel P(A∩B) = P(A)·P(B)

Beschrijvende Statistiek: Kansvariabelen (1.1.4)

In de kansrekening wordt vaak gekeken naar een bepaalde eigenschap of karakteristiek. Zo'n eigenschap heet een **kansvariabele** als de waarde ervan van het toeval afhangt. Vb. De kleur van de volgende auto die passeert, de politieke partij waarop een geïnterviewde gaat stemmen. Het netto maandsalaris van een persoon. Of een verdachte een terrorist is.

Als de eigenschap in een getal is uit te drukken wordt er gesproken van een **kwantitatieve (meetbare) variabele**, anders is het een **kwalitatieve (categorische) variabele**.

Een variabele kan op voorhand bekend zijn (**deterministische variabele**), of niet op voorhand te bepalen, maar onderhevig aan toeval (**kansvariabele**).

Als een variabele eindig veel mogelijkheden heeft is het een **discrete variabele**, als tussenliggende reële waarden worden aangenomen heet het een **continue variabele**.

Beschrijvende Statistiek: Meetschalen (1.1.5)

Statistiek wordt gedaan aan de hand van data. Voor het analyseren van gegevens zijn veel verschillende statistische methoden en technieken beschikbaar. Voor toepassen van deze analysetechnieken is het vaak noodzakelijk dat de data een bepaalde structuur hebben.

Voorbeeld: Een gemiddelde kun je alleen uitrekenen als de gegevens uit getallen bestaan, waarbij zo'n gemiddelde iets betekent, dus wel bij jaarinkomens of verbruikte munititie, maar niet bij type drones of telefoonnummers.

Er wordt een onderscheid gemaakt tussen **kwantitatieve** (getallen) en **kwalitatieve** gegevens (geen getallen maar eigenschappen, kwalificaties, types, etc.)

Hierbinnen zijn weer twee onderverdelingen die samen de vier **meetschalen/ meetniveau's** van de statistiek vormen.

Belangrijk om te weten, maar niet in dit vak. Ze komen pas aan de orde bij MTO3 als je methodes moet selecteren die je gaat toepassen.

Beschrijvende Statistiek: Meetschalen (1.1.5)

- Nominaal (geen ordening):
  Vb: telefoonnummer, kleur, ja/nee, dienstvak

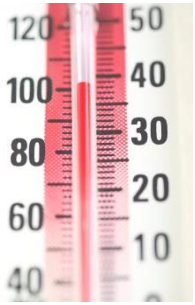
Type of products	code
Pharmaceutical	1
Food	2
Spareparts	3

- Ordinaal (wel ordening)
  Geen vast nulpunt, verschillen hebben geen betekenis
  Vb: Grootte-omschrijving, sterren-score, opleidingsniveau, intervallen/klassen: 2001-2010, 2011-2015, 2015-2024

Type of products	code
Large	1
Medium	2
Small	3

Kwalitatief

- Interval
  Ordinaal, dus geordend, maar verschillen hebben vaste betekenis, “7 graden warmer dan” is een zinvolle uitspraak. Er is geen vast nulpunt.
  Vb: tijd, temperatuur.



- Ratio
  Interval, met een vast referentiepunt. Verhoudingen hebben betekenis, “driemaal zo groot als” is een zinvolle uitspraak.
  Vb: lengte, gewicht

Kwantitatief

Deze schalen zijn opklimmend in structuur. Waar zitten militaire rangen?

Beschrijvende Statistiek: Frequentieverdelingen (1.2.1 + 1.2.2)

VOORBEELD 1.3A

Voor een bestand van 120 woningen is gegeven hoeveel kamers iedere woning heeft. We zouden dit kunnen weergeven door de frequentieverdeling in tabel 1.3.

TABEL 1.3 Klassenindeling naar aantal kamers per woning

Klasse	Aantal kamers
1	3
2	4
3	5 en 6
4	7 en 8
5	9 en 10
6	11 en 12

VOORBEELD 1.3B

Voor de verdeling van het aantal kamers levert het turven ons de frequentieverdeling uit tabel 1.4.

TABEL 1.4 Turftabel aantal kamers

Klasse	Aantal kamers	Turven	Frequentie
1	3	𝍸 𝍸 𝍸 ////	19
2	4	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	35
3	5 en 6	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //	47
4	7 en 8	𝍸 𝍸 𝍸	15
5	9 en 10	///	3
6	11 en 12	/	1
Totaal			120

De gemeten variabelen zijn hier aantal kamers per woning. Maar niet als getal maar in klassen, bv. “5 of 6”. Deze variabele is ordinaal, maar niet interval.

Beschrijvende Statistiek: Frequentieverdelingen (1.2.1 + 1.2.2)

VOORBEELD 1.4

Voor de 120 woningen uit het bestand willen we een klassenindeling naar bouwjaar maken. Allereerst moeten we vaststellen wat de hoogste en de laagste waarneming is. Dit blijken de bouwjaren 1910 en 1988 te zijn. Vervolgens kiezen we de klassengrenzen zodanig, dat alle waarnemingen onder te brengen zijn. Het lijkt logisch om hier klassen van 10 jaar breed te kiezen, want er moet altijd op worden gelet dat de grenzen een beetje 'mooi' uitkomen. Dat leidt tot de indeling in tabel 1.5.

TABEL 1.5 Klassenindeling naar bouwjaar

Klasse	Bouwjaar
1	1910 – < 1920
2	1920 – < 1930
3	1930 – < 1940
4	1940 – < 1950
5	1950 – < 1960
6	1960 – < 1970
7	1970 – < 1980
8	1980 – < 1990

VOORBEELD 1.5

Voor het woningenbestand maken we een frequentieverdeling van de variabele 'vraagprijs'. We kiezen de klassen 50 000 euro breed, te beginnen vanaf 100 000 euro. Het is duidelijk dat er veel meer huizen zijn met een relatief lage vraagprijs dan met een hoge. Om die reden is het wenselijk om bij de hoge prijzen bredere klassen te kiezen, want anders zouden we een aantal klassen krijgen met geen enkele waarneming. Vandaar dat vanaf een vraagprijs van 500 000 euro, de klassen 250 000 euro breed zijn. De klasseninde-ling die zo ontstaat is weergegeven in tabel 1.6.

TABEL 1.6 Frequentieverdeling van de vraagprijs

Klassengrens × € 1 000	Aantal
100 – < 150	11
150 – < 200	30
200 – < 250	13
250 – < 300	13
300 – < 400	17
400 – < 500	11
500 – < 750	19
750 – < 1 000	5
1 000 – < 1 250	1
Totaal	120

De klasse 1910 -< 1920 betekent dat 1910 er wel in zit, maar 1920 niet (die zit in de volgende klasse)

Beschrijvende Statistiek: Relatieve frequentieverdeling (1.2.2)

VOORBEELD 1.6

Voor de huizenprijzen (zie voorbeeld 1.5) berekenen we de relatieve frequenties. We doen dit door de absolute frequenties voor iedere klasse te delen door 120. De resultaten zijn weergegeven in tabel 1.7.

TABEL 1.7 Relatieve frequenties van de huizenprijzen

Klassengrenzen × € 1 000	Aantal	Relatieve frequentie
100 – < 150	11	0,092
150 – < 200	30	0,250
200 – < 250	13	0,108
250 – < 300	13	0,108
300 – < 400	17	0,142
400 – < 500	11	0,092
500 – < 750	19	0,158
750 – < 1 000	5	0,042
1 000 – < 1 250	1	0,008
Totaal	120	1,000

Beschrijvende Statistiek: Kruistabellen (1.2.3)

Kruistabellen zijn bedoeld om het verband tussen twee variabelen met twee of niet veel meer mogelijkheden te bestuderen.

en	A	Niet A	Totaal
B	$P(A \cap B)$	$P(\bar{A} \cap B)$	$P(B)$
Niet B	$P(A \cap \bar{B})$	$P(\bar{A} \cap \bar{B})$	$P(\bar{B})$
Totaal	$P(A)$	$P(\bar{A})$	1

TABEL 1.10 Griep en griepprik

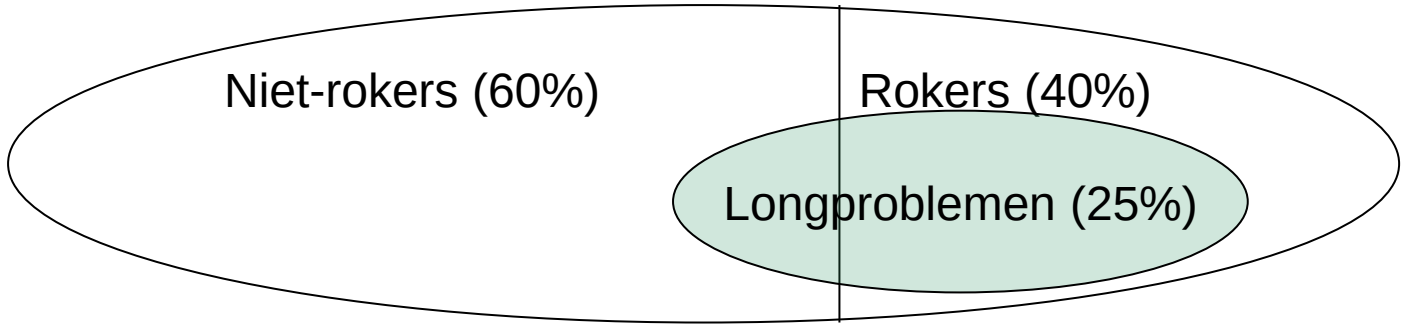
Behandeling	**Resultaat**		
	Griep	*Geen griep*	*Totaal*
Griepprik	15 (a)	135 (b)	150
Geen griepprik	24 (c)	60 (d)	84
Totaal	39	195	234

Bron: huisartsenpraktijk Van Kraaij, Gronsveld

∩ betekent: doorsnede, wat in beide verzamelingen zit.
$\bar{A}$ is het complement van A, alles wat niet in A zit.

Beschrijvende Statistiek: Kruistabellen (1.2.3)

	Rokers	Niet-rokers	
Longproblemen	200	50	250
Geen longproblemen	200	550	750
	400	600	1000



$P(L) = 0.25$

$$P(L|NR) = \frac{P(L \text{ and } NR)}{P(NR)} = 0.08$$

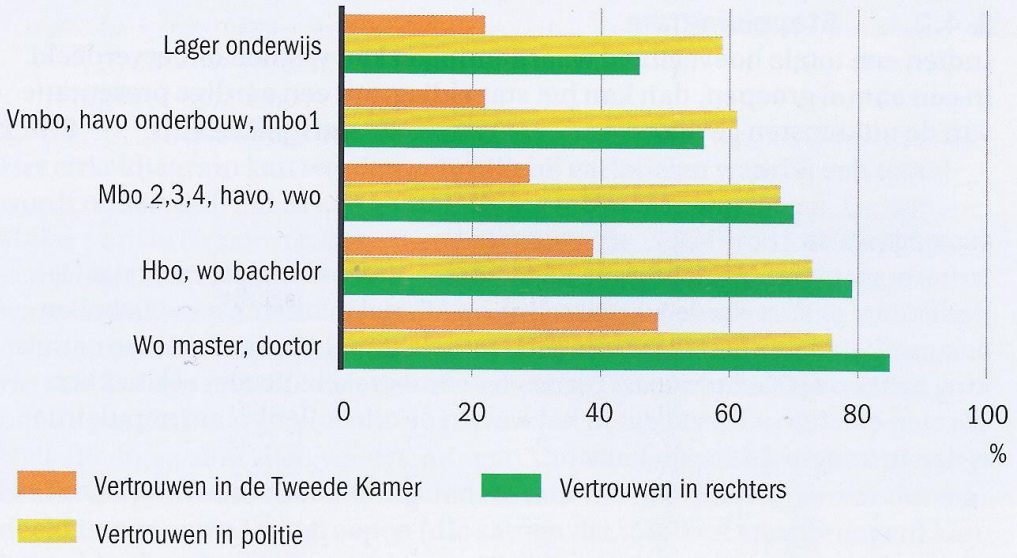
$$P(NR|L) = \frac{P(NR \text{ and } L)}{P(L)} = 0.20$$

Beschrijvende Statistiek: Diagrammen (1.4.1 – 1.4.3)

VOORBEELD 1.14

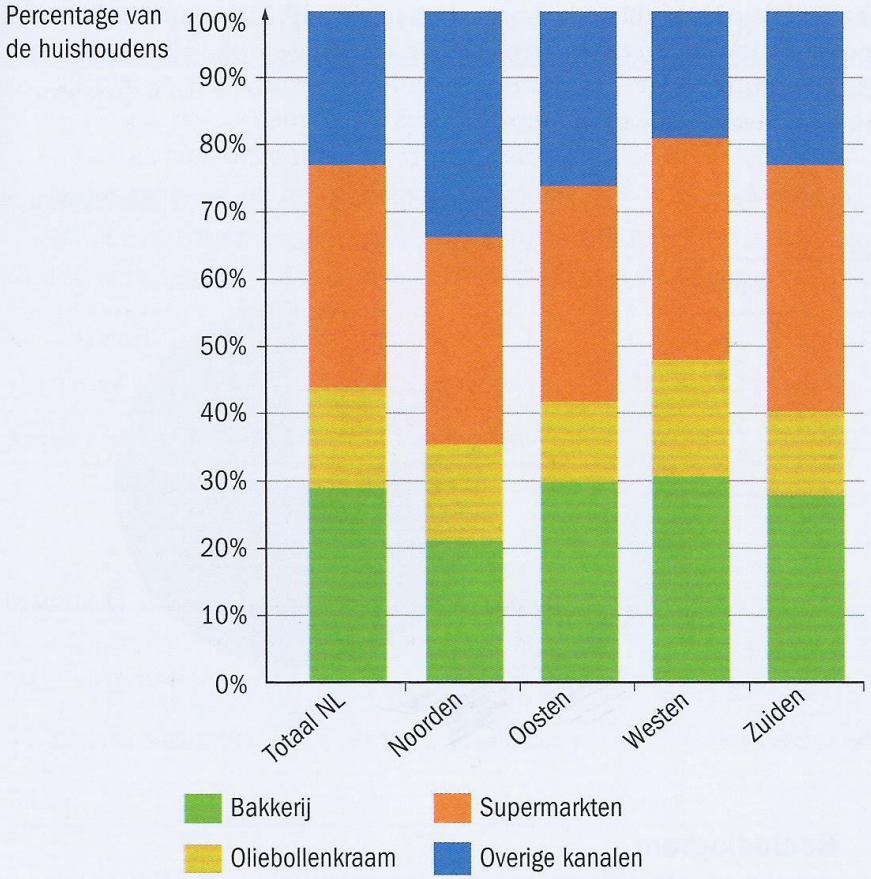
In figuur 1.6 is voor een vijftal opleidingsniveaus aangegeven hoeveel vertrouwen men heeft in de Tweede Kamer, de politie en rechters. De grafiek is te karakteriseren als een horizontaal samengesteld staafdiagram. Merk op, dat in de legenda de betekenis van de gebruikte arceringen wordt vermeld. In z'n algemeenheid kun je concluderen dat hogere opleidingsniveaus meer vertrouwen hebben in deze instanties dan lage opleidingsniveaus.

FIGUUR 1.6 Vertrouwen in autoriteiten naar opleidingsniveau



Bron: Trends in Nederland 2015

FIGUUR 1.8 Waar worden oliebollen / appelbeignets gekocht?

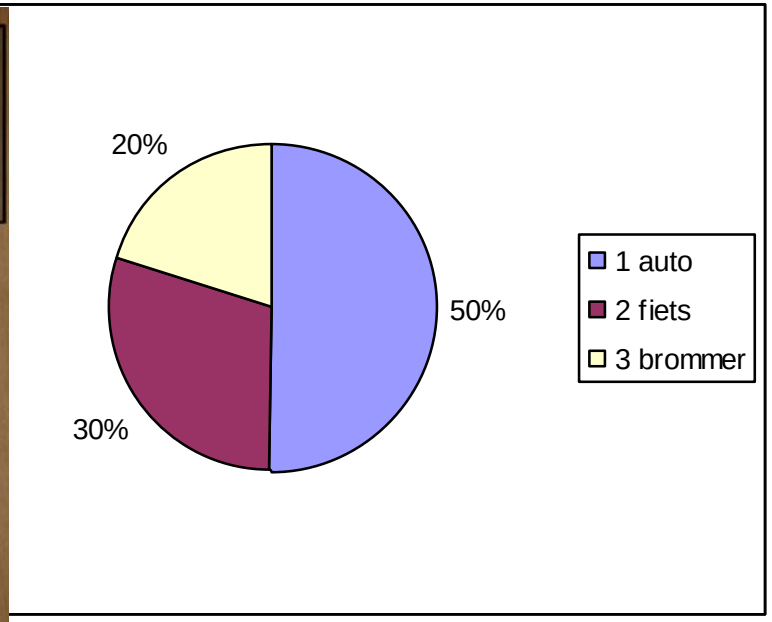


Bron: GfK jaargids 2008

Beschrijvende Statistiek: Stapeldiagram Absoluut of relatief?

Beschrijvende Statistiek: Cirkeldiagram (*Pie chart*)

- Gebruikt om een totaal in deelgroepen te verdelen. De totale oppervlakte komt overeen met 100%.
- Het diagram laat fracties van het totaal zien, maar zegt niets over hoeveelheden. Dit kan worden verholpen met data labels.

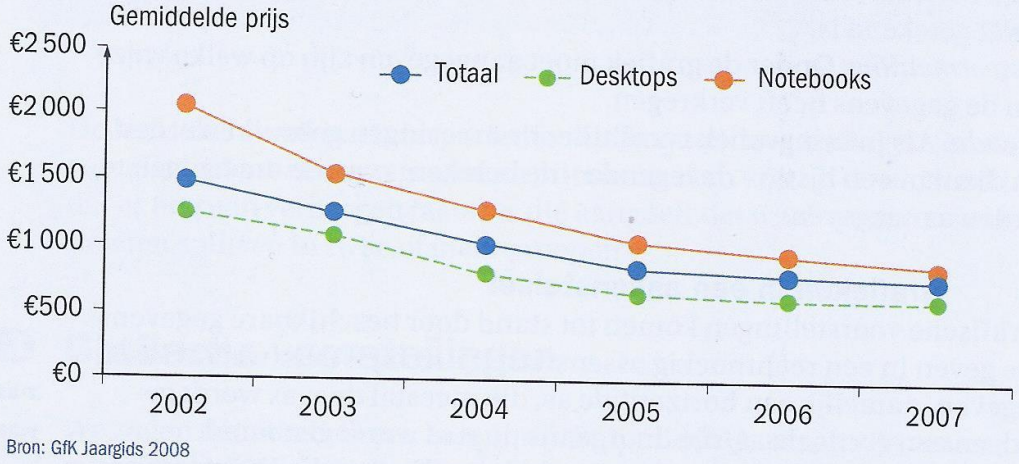


Beschrijvende Statistiek: Lijndiagrammen

VOORBEELD 1.10

In figuur 1.2 is het prijsverloop van computers (desktops, notebooks) weergegeven voor een aantal jaren.

FIGUUR 1.2 Gemiddelde prijzen computers Nederland, van GfK Panelmarkt in Nederland

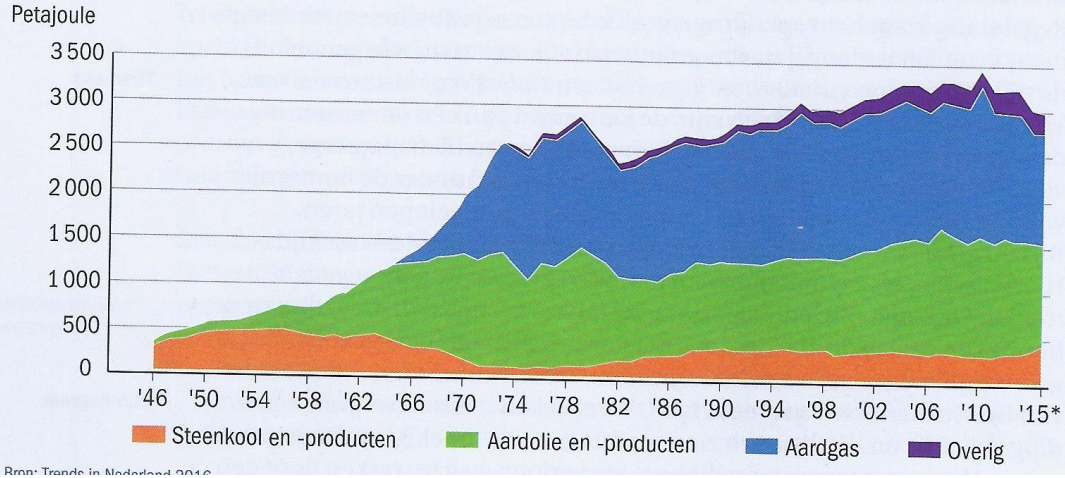


Bron: GfK Jaargids 2008

VOORBEELD 1.11

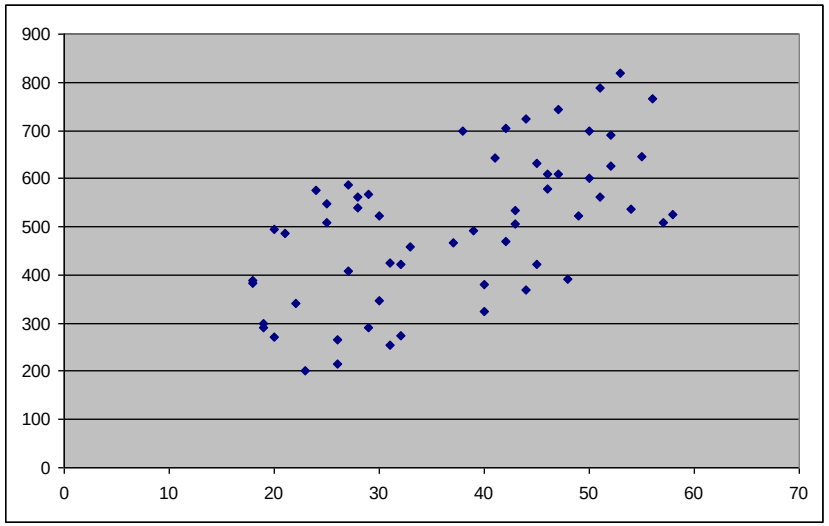
In de grafiek van figuur 1.3 zien we een zogenoemd gestapeld lijndiagram. De bovenste lijn geeft het totale energieaanbod in Nederland aan. Zichtbaar is gemaakt hoe dit totaal is opgebouwd. Zo kun je zien dat het belang van aardgas de laatste paar jaar terugloopt.

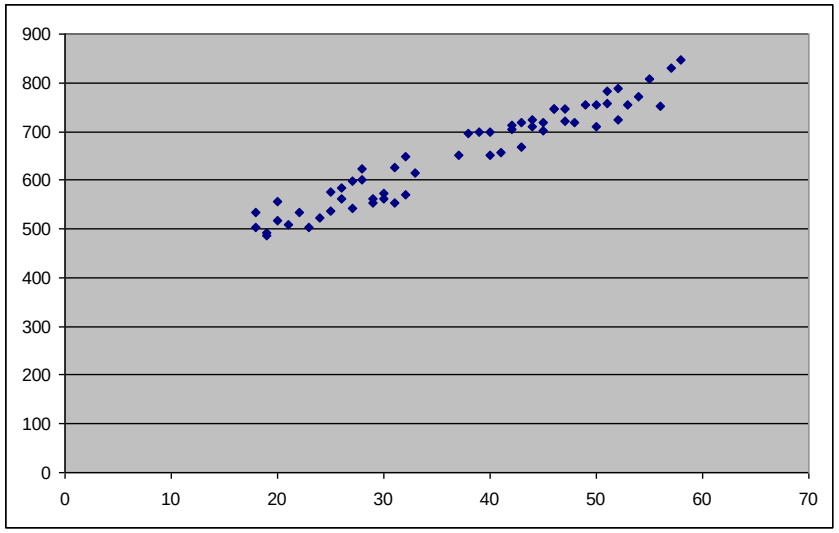
FIGUUR 1.3 Energieaanbod naar energiedrager



Vaak gebruikt voor de ontwikkeling van een variabele in de tijd.

Beschrijvende Statistiek: Correlatiediagrammen (puntenwolk)





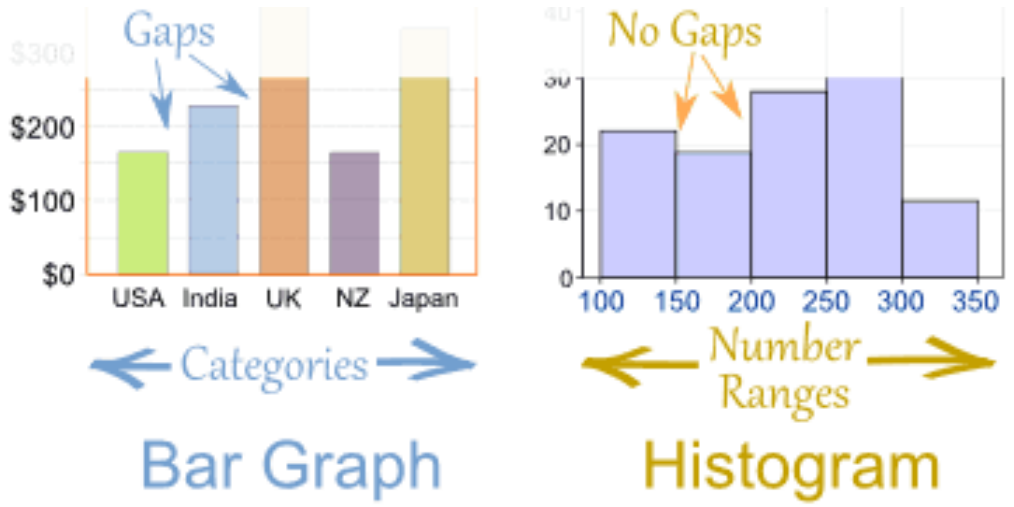
In beide diagrammen is er een relatie tussen de twee variabelen, maar rechts is de correlatie veel sterker.

Beschrijvende Statistiek: Histogram of staafdiagram?

In staafdiagrammen en histogrammen staat de hoogte van de staaf voor de waarde (frequentie of percentage).

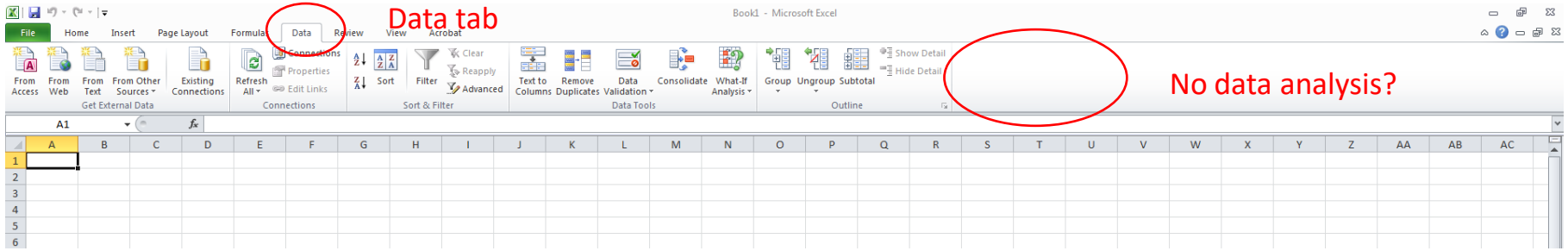
Een staafdiagram wordt gebruikt voor **categorische (kwalitatieve) data**. Een histogram is bedoeld voor **continue** data in aansluitende intervallen (bv. gewicht, tijd).

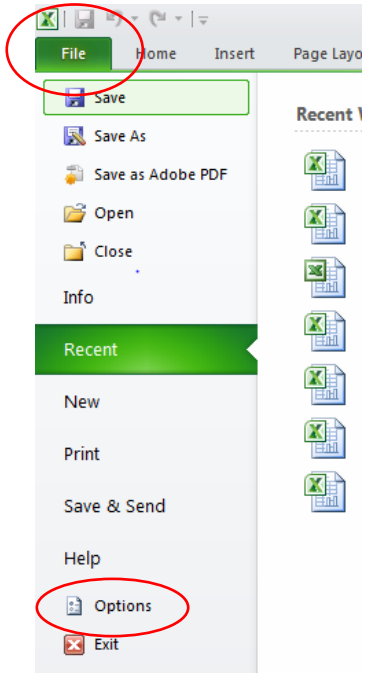
In een staafdiagram staan de balken los van elkaar, in een histogram tegen elkaar.

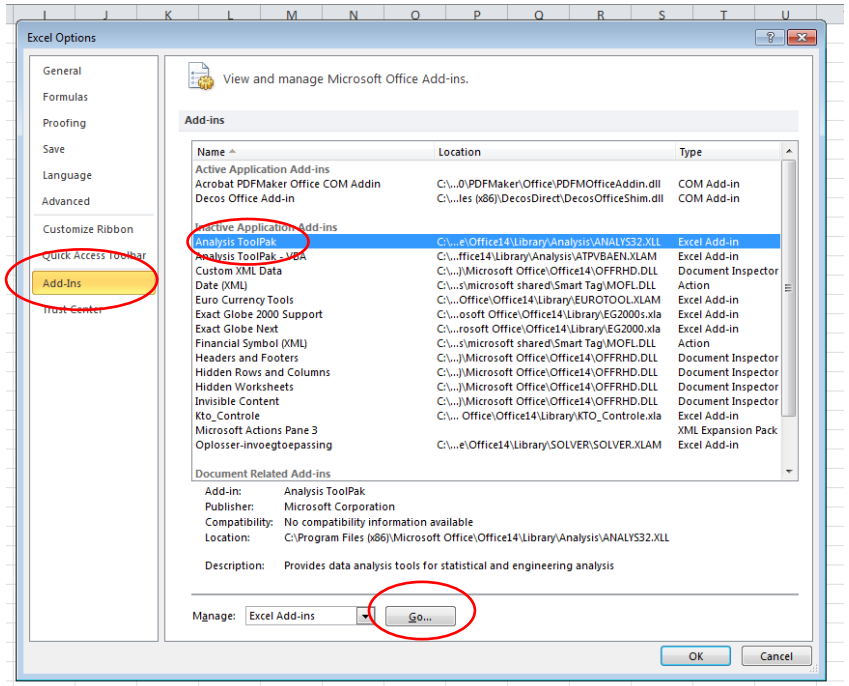


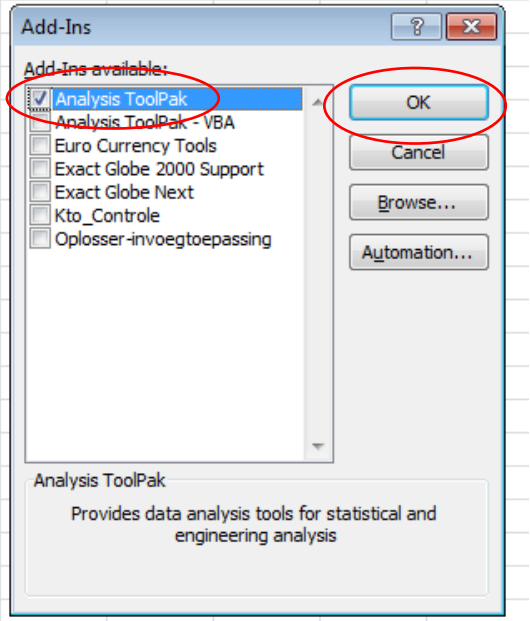
Beschrijvende Statistiek: Histogram in Excel

Excel geeft altijd een losse balken, dat kan wel worden veranderd in een histogram

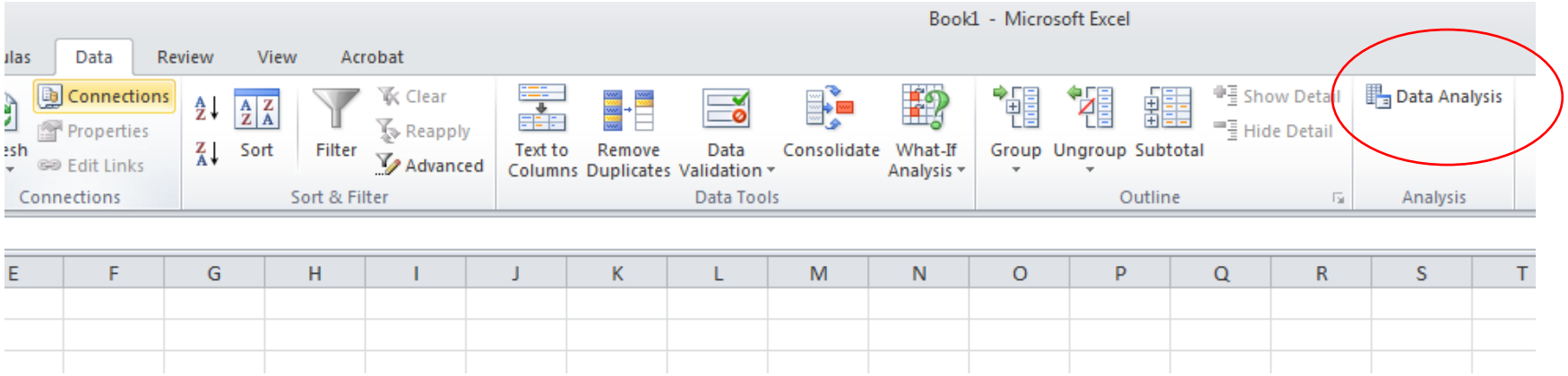


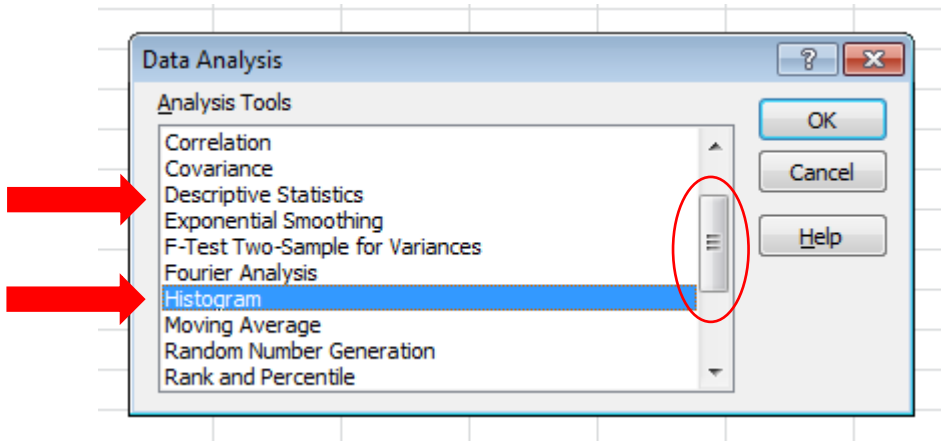




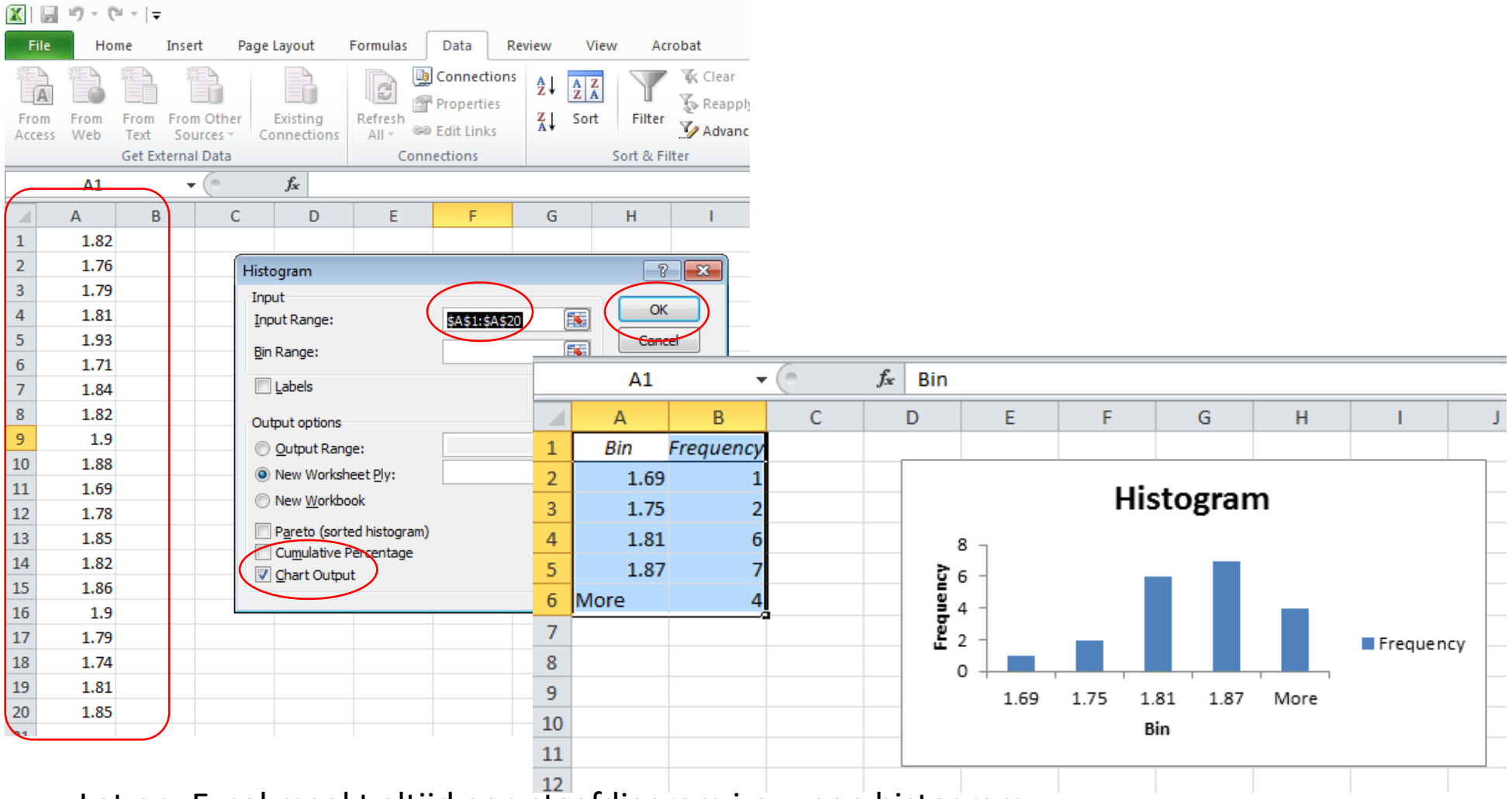


Beschrijvende Statistiek: Histogram in Excel





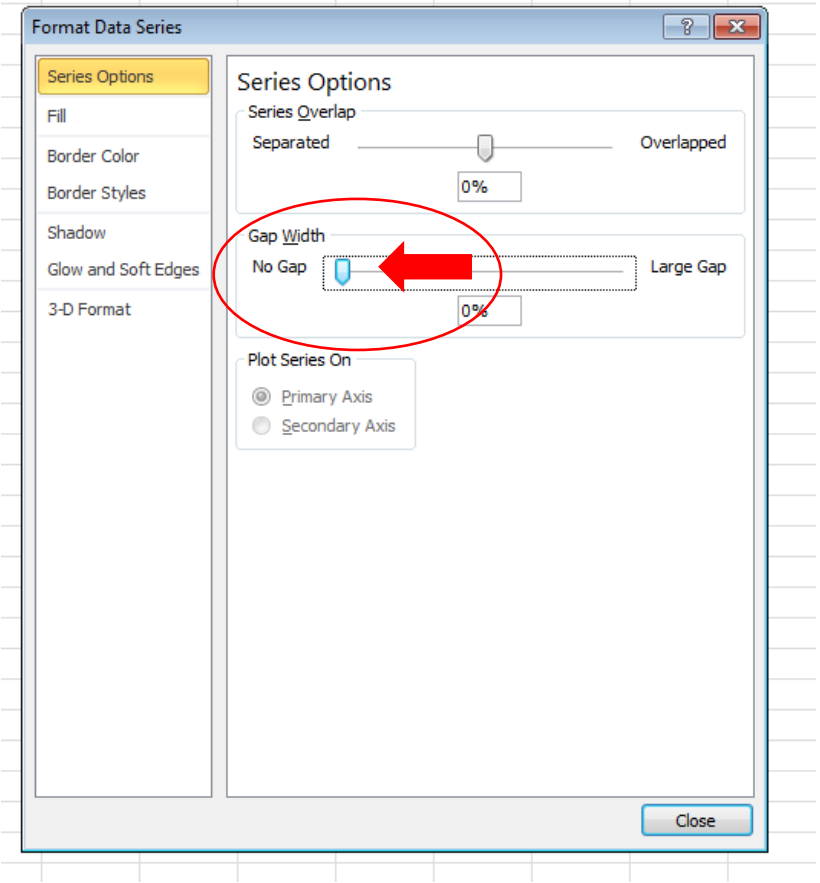
Beschrijvende Statistiek: Histogram in Excel

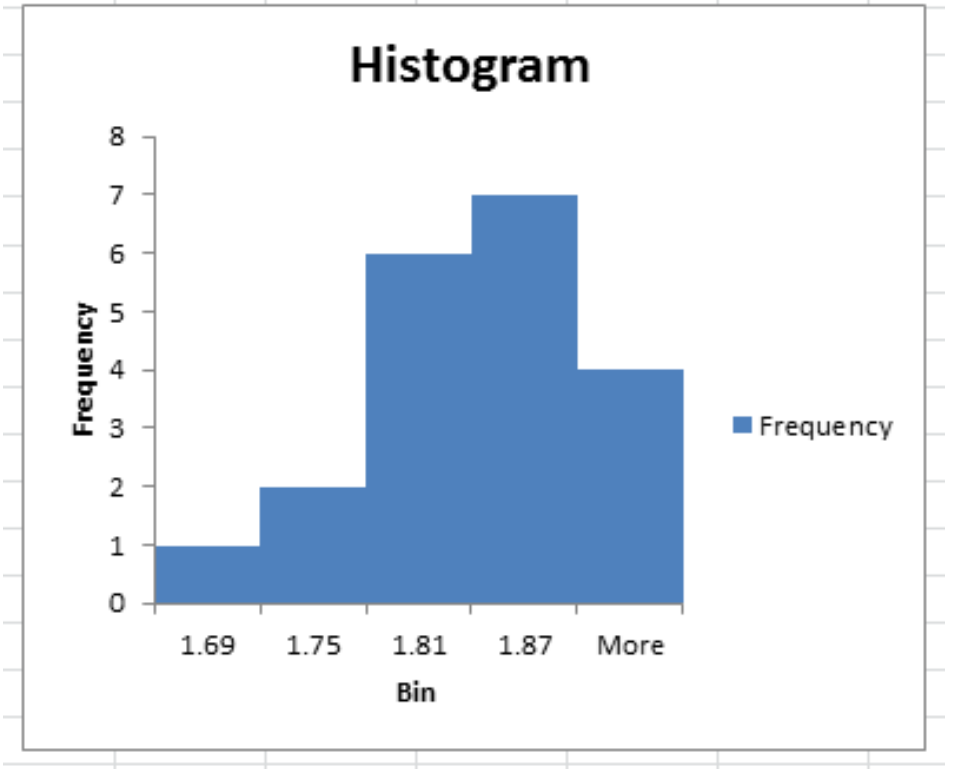


Let op: Excel maakt altijd een staafdiagram i.p.v. een histogram

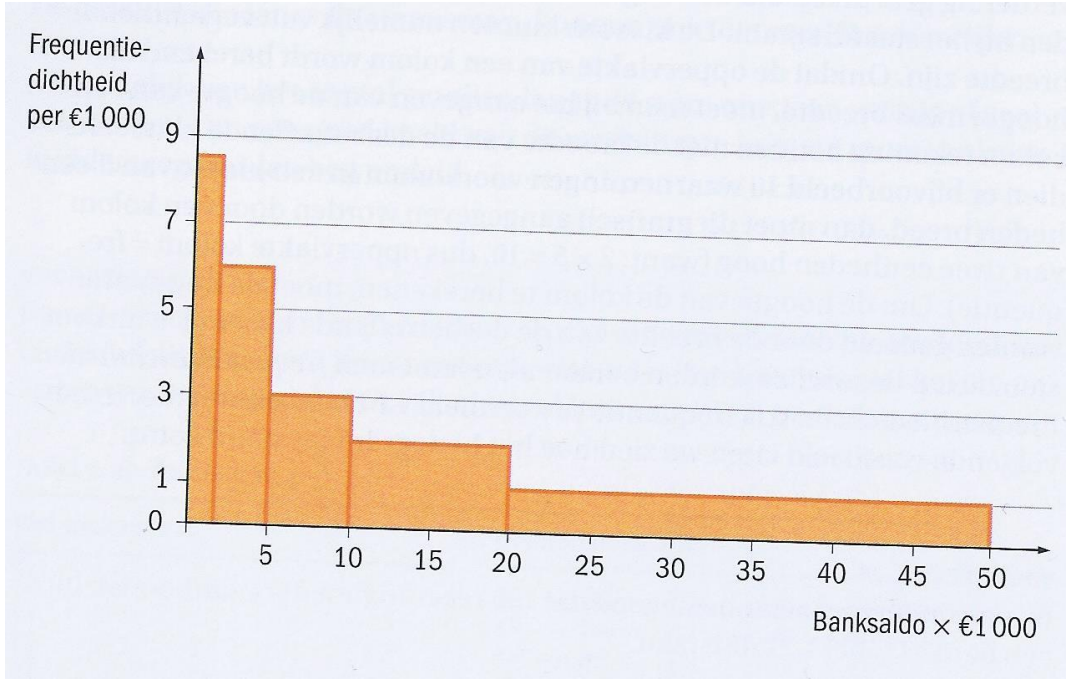
Beschrijvende Statistiek: Histogram (1.5.2)

Hoe verander je een staafdiagram in een histogram:
Rechts-click **op een staaf** in het diagram
Links-Click op Format Data Series en verander de Gap width in 0





Beschrijvende Statistiek: Histogrammen (1.4.1 – 1.4.3)



Als de kolommen niet dezelfde breedte hebben wordt in de hoogte niet de frequentie gezet, maar de **frequentiedichtheid** (de frequentie gedeeld door de breedte van de kolom.

Hierdoor wordt de **oppervlakte** van de kolom evenredig met de frequenties, zodat die optisch beter te vergelijken zijn.

Beschrijvende Statistiek: Maatstaven voor ligging (2.1 – 2.2)

Waar ligt het “midden” van de verdeling? Wat is het zwaartepunt?

(Rekenkundig) **gemiddelde**
Kwartielen (waaronder de **mediaan**)
Modus (top)

Beschrijvende Statistiek: Maatstaven voor ligging - Gemiddelde

$$\text{rekenkundig gemiddelde} = \frac{\text{som waarden}}{\text{aantal waarden}}$$

- Alleen voor numerieke gegevens (interval/ratio).
- Gevoelig voor uitschieters (uitbijters).

  Vb: Gemiddelde huizenprijs in wijk

- Er bestaan ook andere gemiddeldes:
  - meetkundig (bv. voor winststijgingen, groeifactoren, seizoensindices)
  - harmonisch (bv. voor snelheden, verbruik)

Beschrijvende Statistiek: Maatstaven voor ligging - Gemiddelde

20	27	33	36	38	43	49	54	60

(Rekenkundig) gemiddelde =
(20+27+33+36+38+43+49+54+60) / 9 = 360 / 9 = 40

- x_i *:* data, $i = 1,2,\ldots, n$
- n *:* aantal gegevens

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Beschrijvende Statistiek: Maatstaven voor ligging - Mediaan

Mediaan =$X_{me} = Q_2$= middelste waarneming van op volgorde geplaatste scores.

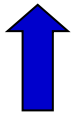
Geschikt voor gegevens vanaf ordinaal niveau.

Deelt gegevens op: 50% is kleiner en 50% is groter dan de mediaan.

20	27	33	36	38	43	49	54	60

Mediaan = middelste van - naar grootte gerangschikte - waarnemingen = 38

20	27	33	36	38	43	49	54



Als n even dan

Mediaan = gemiddelde van middelste 2 waarnemingen

= (36 + 38) / 2 = 37

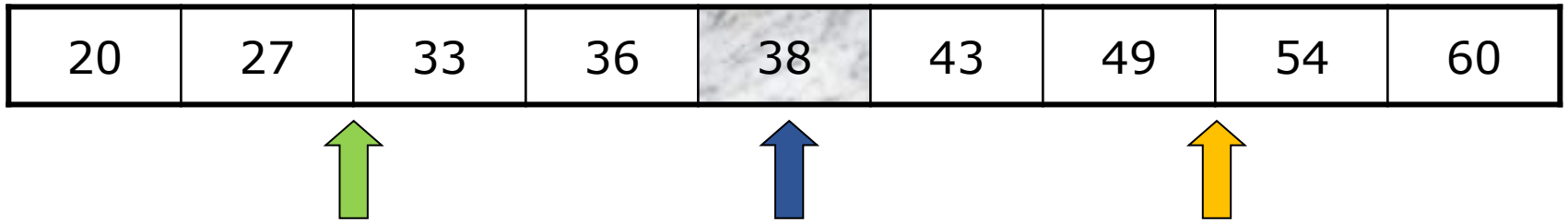
Beschrijvende Statistiek: Maatstaven voor ligging - Kwartielen

Eerste kwartiel = Q_1 = middelste waarneming van eerste 50% van op volgorde geplaatste scores.

- Geschikt voor gegevens vanaf ordinaal niveau.
- Deelt gegevens op: 25% is kleiner en 75% is groter dan Q_1

Derde kwartiel = Q_3 = middelste waarneming van laatste 50% van op volgorde geplaatste scores.

- Geschikt voor gegevens vanaf ordinaal niveau.
- Deelt gegevens op: 75% is kleiner en 25% is groter dan Q_3



$X_{me} = Q_2 = 38$

$Q_1 = \frac{27 + 33}{2} = 30$

$Q_3 = \frac{49 + 54}{2} = 51{,}5$

Beschrijvende Statistiek: Maatstaven voor ligging - Modus

Modus =X_{mo}= meest voorkomende score (bij hoogste frequentie).

Geschikt voor alle soorten gegevens.

20	30	30	40	40	50	50	50	50

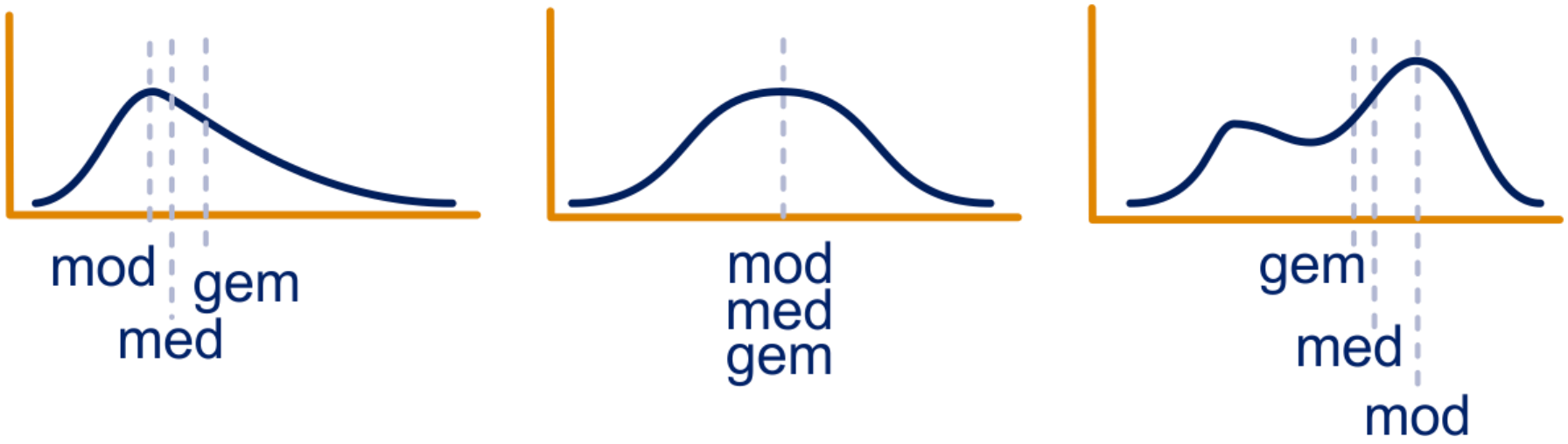
Modus = X_{mo}= 50

Een modus hoef niet te bestaan, er zijn verdelingen met twee hoogste pieken (bimodaal)

Beschrijvende Statistiek: Keuze centrummaat

- Afhankelijk van soort gegevens.
- Afhankelijk van uitschieters in verdeling.

Voorbeeld: het modale inkomen zegt meer dan het gemiddeld inkomen.



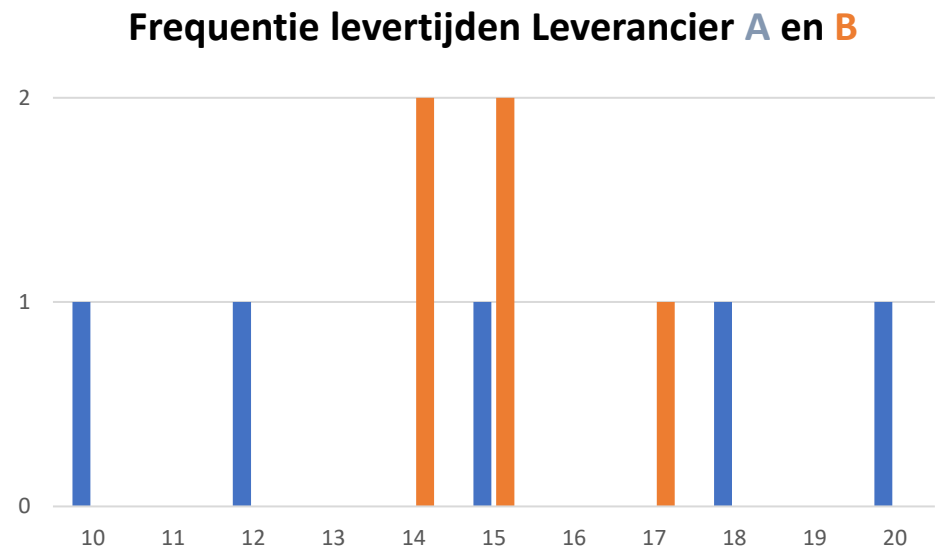
Beschrijvende Statistiek: Maatstaven voor spreiding (2.4)

Levertijden leverancier A: 10, 12, 15, 18, 20 dagen

Levertijden leverancier B: 14, 14, 15, 15, 17 dagen

a) Bepaal de gemiddelde levertijd van beide leveranciers (12,5 dagen).

b) Welke leverancier kies je? Waarom?



Beschrijvende Statistiek: Maatstaven voor spreiding (2.4)

- Zegt iets over de spreiding van de waarnemingen t.o.v. het centrum (interval of ratio meetniveau).
- Liggen de gegevens dicht bij elkaar of juist ver van elkaar af?
- Maat voor *onzekerheid* in de gegevens!

  - **Variatiebreedte/range**
  - Bij gemiddelde:
    - **Gemiddelde absolute afwijking** (G.A.A. = MAD)
    - **Standaarddeviatie**
    - **Variatiecoëfficiënt**
  - Bij mediaan:
    - **Halve kwartielafstand**

Beschrijvende Statistiek: Maatstaven voor spreiding (2.4)

Range = verschil tussen grootste en kleinste waarneming

Voorbeeld: 36, 49, 20, 33, 54, 60, 43, 27, 38

Range = 60 – 20 = 40

Opmerkingen:

- Zeer globaal
- Voor eerste indruk
- Alleen bepaald door twee extremen, hoe de verdeling daartussen is maakt niet uit.

Beschrijvende Statistiek: Maatstaven voor spreiding – Halve kwartielafstand

20	30	40	40	50	50	50	70
	Q_1		Q_2		Q_3		

Q_2 = mediaan = (40+50)/2 = 45

Q_1 = (30+40)/2 = 35

Q_3 = (50+50)/2 = 50

Halve kwartielafstand = $\frac{Q_3-Q_1}{2} = \frac{50-35}{2} = 7{,}5$

Beschrijvende Statistiek: Maatstaven voor spreiding - MAD

- Maak de deviaties absoluut en bereken het gemiddelde.
- In formule:

$$\text{MAD} = \frac{\sum_{i=1}^{n} |d_i|}{n} = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n}$$

- In het voorbeeld:

MAD = 92/9 = 10,22

- MAD = *Mean Absolute Deviation*

X_i	d_i	$\lvert d_i \rvert$
36	-4	4
49	9	9
20	-20	20
33	-7	7
54	14	14
60	20	20
43	3	3
27	-13	13
38	-2	2
totaal	0	92

Beschrijvende Statistiek: Maatstaven voor spreiding (2.4)

- **Standaarddeviatie** = wortel uit gemiddelde *kwadratische* afwijking.
- In formule:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} {d_i}^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n}}$$

- In het voorbeeld:

$$\sigma = \sqrt{1324/9} = 12{,}1289$$

Opmerkingen:

- Belangrijkste en veel gebruikte spreidingsmaat.
- Bij kleine aantallen wel gevoelig voor uitschieters want meeteenheid wordt gekwadrateerd dus uitbijters uitvergroot.

X_i	d_i	$(d_i)^2$
36	-4	16
49	9	81
20	-20	400
33	-7	49
54	14	196
60	20	400
43	3	9
27	-13	169
38	-2	4
totaal	0	1324

Beschrijvende Statistiek: Populatie of steekproef? (1.1.2 + 2.7)

Een **populatie** is de totale verzameling van personen of objecten waarop een statistische analyse betrekking op heeft.

Je wilt vaak iets zeggen of de hele populatie (bijvoorbeeld politieke voorkeur van Nederlandse stemgerechtigden), maar de populatie is vaak te groot om helemaal te onderzoeken. Daarom gebruik je een willekeurig deel daarvan (vaak veel kleiner), een steekproef, om uit resultaten van de steekproef conclusies trekken over de hele populatie.

Voorbeeld: Je wilt voor een onderzoek in kaart brengen of defensiemedewerkers tevreden zijn over hun arbeidsvoorwaarden. Je populatie bestaat dan uit ca. 55.000 defensiemedewerkers (op een bepaald tijdstip). Dat zijn er teveel om persoonlijk te bevragen, dus je kiest willekeurig 100 defensiemedewerkers uit die representatief zijn (militair/burger, rang, leeftijd, geslacht, etc.) voor de hele organisatie (dus niet 100 cadetten). Dat is je steekproef.

Als je voor elk element van de populatie een getal $x_i, i = 1, \dots, N(= 100)$ kunt bepalen (interval- of ratioschaal), dan kun je voor de hele populatie het populatiegemiddelde μ en de populatiestandaarddeviatie σ uitrekenen:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Beschrijvende Statistiek: Populatie of steekproef? (1.1.2 + 2.7)

Als je voor elk element van de populatie een getal $x_i, i = 1, \ldots, N$ kunt bepalen (interval- of ratioschaal), dan kun je voor de hele populatie het **populatiegemiddelde** μ en de **populatiestandaarddeviatie** σ uitrekenen:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Dit kun je ook doen voor de steekproef (voor het gemak even $x_i, i = 1, \ldots, n$) en dan het **steekproefgemiddelde** $\bar{x}$ en de **steekproefstandaarddeviatie** s uitrekenen (let op, er wordt een ander symbool gebruikt!):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s = \sqrt{\frac{1}{\boldsymbol{n-1}}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Dit zijn dan goede benaderingen voor populatiegemiddelde en –standaarddeviatie. **Let op het subtiele verschilletje in de formule voor s.** Dit verschil tussen populatie- en steekproef word vaak verwarrend gevonden.

In Excel zijn er twee functies voor: STDEVP en STDEVS (*Population* en *Sample*).