Hoofdstuk 1

Combinatieleer

1.1 Telproblemen

1.1.1 Tellen door middel van een boomdiagram

Om op een gestructureerde manier aantallen te tellen kunnen we gebruik maken van een schematische voorstelling die we een **boomdiagram** noemen.
Voorbeeld: Een punt bevindt zich in de oorsprong van een gegradueerde rechte. Het verplaatst zich telkens over één eenheid in positieve of in negatieve zin. De beweging houdt op ofwel na vijf verplaatsingen ofwel wanneer de punten met 3 of -2 worden bereikt.

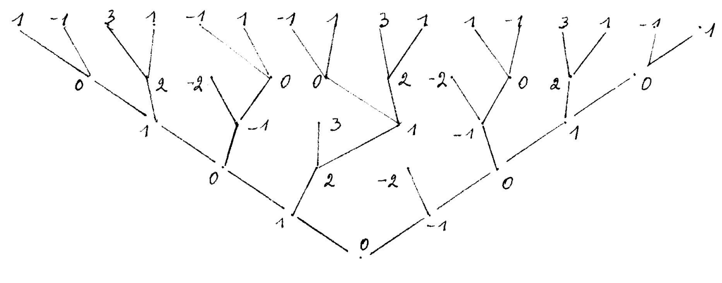
a. Construeer een boomdiagram met de verschillende mogelijke verplaatsingen;

b. Op hoeveel manieren kan p het punt met absis -2 bereiken?

c. Op hoeveel manieren kan p het punt met absis 3 bereiken?

d. Waar kan p zich bevinden na drie verplaatsingen?

e. Waar kan p zich bevinden na vijf verplaatsingen?



TAAK ♣ **1** Twee biljartspelers spelen een tornooi. Hij die het eerst twee opeenvolgende matchen wint, of die in het totaal drie matchen wint, is overwinnaar. Geef het verloop van het tornooi door het opstellen van een boomdiagram. Hoeveel matchen moeten er ten minste worden gespeeld om de overwinnaar te kennen? En hoeveel matchen ten hoogste?

Figuur 1.1: boomdiagram voor bijartspelers

♣ **2** Stel dat men in een maatschappij de gezinnen verplicht van bij de geboorte van een jongen te stoppen met kinderen krijgen en bij de geboorte van een meisje door te gaan met kinderen krijgen tot maximum vier kinderen. Hoeveel gezinssamenstellingen zijn er mogelijk?

Figuur 1.2: boomdiagram voor de geboorten van jongens en meisjes

♣ **3** Schrijf alle nummers van drie cijfers, die men kan vormen met de cijfers 0, 1, 2, 3, 4 en 5, als de som van de cijfers van elk nummer 5 is (construeer een boomdiagram).

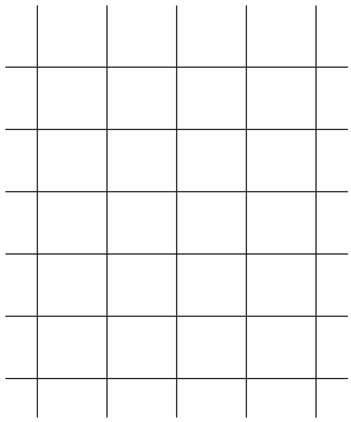
Figuur 1.3: boomdiagram voor nummers van 3 cijfers met som gelijk aan 5

1.1.2 Elementen tellen van een product van verzamelingen

1. Het product van twee verzamelingen.
   Willen we koppels vormen waarvan we voor het eerste element moeten kiezen uit de elementen van een verzameling A en voor het tweede element uit de elementen van een verzameling B dan kunnen we de mogelijkheden voorstellen dmv. een boomdiagram of van een rooster. We illustreren met voorbeelden:
   Voorbeelden:

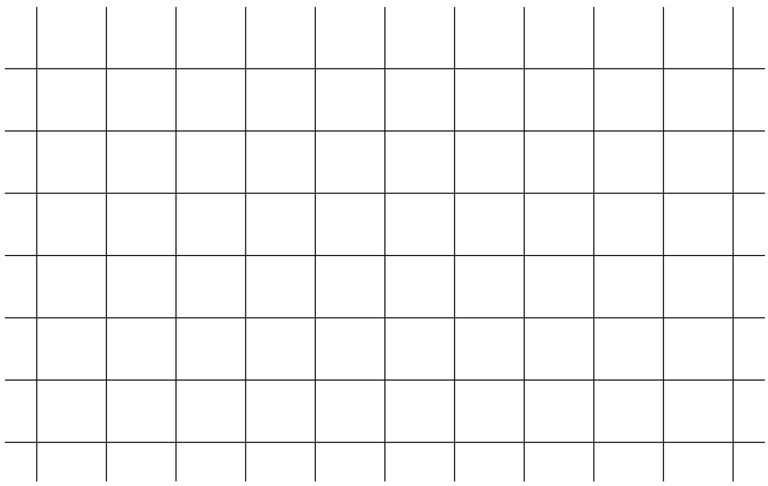
   - Op een juke-box staan twee soorten toetsen, lettertoetsen met de letters a, b, c en d en nummertoetsen met de cijfers 1,2,3,4 en 5. Op hoeveel manieren kun je op deze juke-box een plaatje kiezen?
     OPLOSSING: We hebben 4 mogelijkheden voor de keuze van de letter en 5 mogelijkheden voor de keuze van het cijfer. In het totaal hebben we $4 \times 5 = 20$ mogelijkheden om een letter en een cijfer te kiezen. We zeggen dat we de **mogelijkheden met elkaar combineren**.
     In het geval van het voorbeeld moeten er twee elementen gekozen worden. Om die reden kunnen we een andere voorstelling van de verschillende mogelijkheden hanteren. We maken een zogenaamd **roosterdiagram**. We kiezen twee snijdende rechten x en y in het vlak. De elementen a, b, c en d van A duiden we aan op de rechte x en de elementen 1, 2, 3, 4 en 5 van B op y. De 20 koppels met als eerste element een letter en als tweede element een cijfer kunnen nu in het vlak voorgesteld worden door roosterpunten.

Figuur 1.4: boomdiagram voor het kiezen van een plaatje



Tabel 1.1: roosterdiagram voor het kiezen van een plaatje

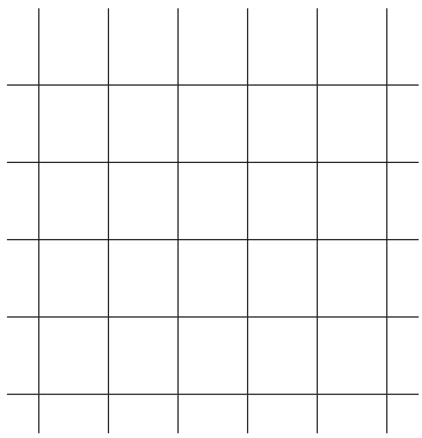
- We gooien twee dobbelstenen. Hoeveel mogelijkheden zijn er?
  OPLOSSING: Het is hier meer aangewezen de mogelijke uitkomsten voor te stellen op een roosterdiagram. We verkrijgen hier 36 roosterpunten.



Tabel 1.2: roosterdiagram voor het gooien van twee dobbelstenen

- Hoeveel mogelijkheden heeft men om twee ballen te trekken zonder terugleggen uit een vaas met 2 rode en 3 groene ballen?
  OPLOSSING: Bij de eerste trekking hebben we 5 mogelijkheden om een bal te trekken, bij de tweede trekking hebben we echter slechts 4 mogelijkheden meer om een bal te trekken. De tweede trekking is afhankelijk van de eerste trekking. Welke bal we ook bij de eerste trekking gekozen hebben, steeds hebben we nog de keuze uit 4 ballen bij de tweede trekking. De tweede verzameling is steeds een andere verzameling maar steeds met 4 elementen. We hebben 5 mogelijkheden voor de keuze van de eerste bal en 4 mogelijkheden voor de keuze van de tweede bal. In het totaal hebben we 5×4 mogelijkheden. We kunnen ook hier een roosterdiagram maken maar de koppels waarvoor het eerste element gelijk is aan het tweede (op de hoofddiagonaal) mogen niet meegeteld worden.

Figuur 1.5: boomdiagram voor het trekken van 2 ballen uit een vaas met 5 ballen



Tabel 1.3: roosterdiagram voor het trekken van 2 ballen

Het product van twee verzamelingen A **en** B is de verzameling van alle koppels waarvan het eerste element behoort tot A en het tweede element behoort tot B. We noteren:

$$A \times B = \{(x, y) : x \in A \wedge y \in B\}$$

Besluit voor het aantal elementen: *Het aantal elementen van het product van twee verzamelingen A en B met een eindig aantal elementen is gelijk aan het product van het aantal elementen van A en het aantal elementen van B.* Met symbolen:

$$|A \times B| = |A|.|B|$$

TAAK ♣ **4** Van een gezin van twee ouders en vier kinderen worden één ouder en één kind uitgenodigd om gratis naar een film te gaan kijken. Op hoeveel manieren kan het gezin hiervan gebruik maken?

Figuur 1.6: boomdiagram voor gezin van 4 kinderen en 2 ouders

♣ **5** Hoeveel natuurlijke getallen van twee cijfers bestaan er, gekozen uit de cijfers 2,5,6,8 en 9 ?

♣ **6** Van twee personen weet men dat ze jarig zijn in de maand november. Iemand moet van alle twee de verjaardag raden. Hoeveel mogelijkheden heeft hij?

♣ **7** Hoeveel wedstrijden moeten gespeeld worden in een competitie met 4 voetbalploegen als er een heen- en terugwedstrijd moet gespeeld worden.

Oplossingen: 155: $2 \times 4 = 8$; 156: $5 \times 5 = 5^2 = 25$; 6: 30^2; 8: $4 \times 3 = 12$.

Figuur 1.7: boomdiagram voor natuurlijke getallen van 2 cijfers

Figuur 1.8: diagram voor personen die jarig zijn in novenber

Figuur 1.9: diagram voor de competitie met 4 voetbalploegen

2. Het product van drie verzamelingen.
Het product van drie verzamelingen A, B en C is de verzameling van alle drietallen waarvan het eerste element behoort tot A, het tweede tot B en het derde tot C:

$$A \times B \times C = \{(x, y, z) : x \in A \wedge y \in B \wedge z \in C\}$$

We tonen aan dat

$$|(A \times B \times C)| = |A| \cdot |B| \cdot |C|.$$

We beschouwen nu ook het product van de twee verzamelingen $A \times B$ en C, d.i. de verzameling van alle koppels waarvan het eerste element een koppel is van $A \times B$ en het tweede element een element van C:

$$(A \times B) \times C = \{((x, y), z) : x \in A \wedge y \in B \wedge z \in C\}$$

Werken we met eindige verzamelingen dan geldt:

$$|(A \times B) \times C| = |(A \times B)|.|C| = (|A|.|B|).|C| = |A|.|B|.|C|$$

Dit laatste steunt op de associatieve eigenschap van het product van natuurlijke getallen.

De verzamelingen $A \times B \times C$ en $(A \times B) \times C$ zijn verschillende verzamelingen maar we zullen aantonen dat ze wel hetzelfde aantal elementen bevatten.
De verzameling $A \times B \times C$ is een verzameling van drietallen en de verzameling $(A \times B) \times C$ is een verzameling van koppels.

Tussen beide verzamelingen bestaat een bijectie:

$$A \times B \times C \stackrel{bijectie}{\longrightarrow} (A \times B) \times C : (x, y, z) \mapsto \Big((x, y), z\Big).$$

Aangezien we met eindige verzamelingen werken hebben beide verzamelingen hetzelfde aantal elementen.

$$|(A \times B \times C)| = |(A \times B) \times C| = |A|.|B|.|C|$$

Voorbeelden:

- Enkele atleten besluiten samen een vereniging op te richten. Op de stichtingsvergadering zijn negen atleten aanwezig: twee hoogspringers, vier hardlopers en drie discuswerpers. Men wil een bestuur vormen dat uit één beoefenaar van elk van deze sporttakken bestaat. Op hoeveel manieren is dit mogelijk?
Oplossing: We hebben 2 mogelijkheden voor de hoogspringers, 4 mogelijkheden voor de hardloper en 3 mogelijkheden voor de discuswerper. In het totaal zijn er $2 \cdot 4 \cdot 3 = 24$ mogelijkheden. We combineren de mogelijkheden voor de hoogspringer, de mogelijkheden voor de hardloper en de mogelijkheden voor de discuswerper.

Figuur 1.10: boomdiagram voor atleten

- Men heeft drie mogelijke wegen om van een plaats A naar een plaats B te fietsen, vier mogelijkheden om van B naar C te rijden en twee mogelijkheden om van C naar D te rijden. Op hoeveel mogelijke manieren kan men van A naar D te rijden?
  Oplossing: Het is duidelijk dat er in totaal $3 \times 4 \times 2$ mogelijkheden zijn. Elke weg van A naar B kunnen we combineren met elke weg van B naar C en met elke weg van C naar D. In plaats van een boomdiagram te maken, is het eenvoudiger een zogenaamd **wegendiagram** te maken.

Figuur 1.11: wegendiagram voor de mogelijke fietsroutes

OPGAVEN — **8** Van drie personen weet men dat ze jarig zijn in de maand juni. Iemand moet van alle drie de verjaardag raden. Hoeveel mogelijkheden heeft hij?

Figuur 1.12: boomdiagram voor 3 personen die jarig zijn in juni

$\boxed{\textbf{TAAK}}$ ♣ **9** Hoeveel getallen van drie cijfers kan men vormen

a. als het eerste cijfer een 5 is;

b. als het laatste cijfer een 6 moet zijn;

c. als ze moeten deelbaar zijn door 2;

d. als het eerste cijfer geen 9 mag zijn en het tweede cijfer verschillend moet zijn van nul.

Figuur 1.13: boomdiagram voor getallen van drie cijfers

♣ **10** Een nummerslot van een fiets bestaat uit drie ringen. Op elke ring komen de cijfers 0 tot en met 9 voor. Hoeveel mogelijke cijfercodes zijn er?

Oplossingen: **??**: 30^3; 9: (a) 10^2; (b) 9.10; (c) 9.10.5; (d) 8.9.10; 10: $10^3 = 1000$.

Figuur 1.14: boomdiagram voor het nummerslot van een fiets

3. Het product van vier en meer verzamelingen.

Met een bewijs door volledige inductie kunnen we aantonen dat het aantal elementen van het product van n verzamelingen gelijk is aan het product van het aantal elementen van elke verzameling. De verzamelingen zijn eindige verzamelingen.

$$\forall n \in \mathrm{N} : |A_1 \times A_2 \times \ldots \times A_n = |A_1|.|A_2|.\ldots.|A_n|$$

Voorbeeld 1: De nummerplaten van Belgische wagens bestaan uit drie letters gevolgd door drie cijfers. Hoeveel wagens kunnen van een Belgische nummerplaat voorzien worden?

Oplossing: Voor de eerste letter hebben we 26 mogelijkheden, alsook voor de tweede en de derde letter. Voor het eerste cijfer hebben we 10 mogelijkheden, alsook voor het tweede en derde cijfer. In totaal hebben we $26^3.10^3 = 17576000$ mogelijkheden.

Figuur 1.15: boomdiagram voor nummerplaten van Belgische wagens

Voorbeeld 2:

a. Hoeveel getallen van drie cijfers kunnen we vormen met de cijfers 0,1,2,4,5,7,9.
b. Hoeveel van die getallen beginnen met 1? En met 4?
c. Als we de getallen rangschikken volgens stijgende orde, welk is dan het honderdvijftigste getal?
d. Welk is het rangnummer van het getal 752?

OPLOSSING: (a): $6.7.7 = 294$; (b): 49; (c): 502; (d) 227.

$\boxed{\textbf{TAAK}}$ ♣ **11** Voor de afloop van een voetbalwedstrijd zijn er drie mogelijkheden. Men stelt nu een tabel op van de uitslagen van twaalf wedstrijden. Hoeveel tabellen zijn er mogelijk? Beantwoord dezelfde vraag als je weet dat in de laatste wedstrijd de thuisploeg zal winnen?

Figuur 1.16: boomdiagram voor een voetbalwedstrijd

♣ **12** Hoeveel verschillende worpen zijn er mogelijk met vier dobbelstenen van een verschillende kleur?

Figuur 1.17: boomdiagram voor 4 worpen met een dobbelsteen

♣ **13** In een bibliotheek zijn vijf wetenschappelijke verhandelingen, tien detectiveverhalen, achttien romans en één wiskundeboek. Iemand wil van iedere soort een werk lezen. Op hoeveel manieren kan hij zijn lectuurlijst opstellen?

OPLOSSINGEN: 11: $3^{12} = 531441$; 3^{11}; 12: 6^4; 13: $5.10.18.1 = 900$.

Figuur 1.18: boomdiagram voor 5 wetenschappelijke verhandelingen

1.1.3 Tellen van elementen met verschillende eigenschappen

Hier kunnen we om elementen te tellen, gebruik maken van de verzamelingsleer. We illustreren met voorbeelden.
Voorbeelden:

- Hoeveel bytes (8-tallen waarvan de elementen 1 of 0 zijn) zijn er met een 0 op de eerste plaats of een 0 op de tweede plaats.
  OPLOSSING: Alle 8-tallen die een 0 hebben op de eerste plaats vormen een verzameling A met 2^7 elementen en alle 8-tallen die op de tweede plaats een 0 hebben vormen een verzameling B met eveneens 2^7 elementen. Deze twee verzamelingen A en B hebben gemeenschappelijke elementen nl. de 8-tallen met een 0 op de eerste plaats en een 0 op de tweede plaats. Deze 8-tallen vormen de verzameling $A \cap B$ met 2^6 elementen. In $|A| + |B| = 2^7 + 2^7$ zijn $|A \cap B| = 2^6$ 8-tallen twee keer geteld. Het gevraagde aantal is dus

$$|A \cup B| = |A| + |B| - |A \cap B| = 2^7 + 2^7 - 2^6 = 192.$$

Figuur 1.19: venndiagram voor de bytes

- Aan 150 leerlingen wordt gevraagd of ze in hun vrije tijd naar een muziekschool, een sportclub of naar een jeugdbeweging gaan. 50 leerlingen gaan naar een muziekschool, 65 naar een sportclub en 60 naar een jeugdbeweging. 25 leerlingen gaan zowel naar een muziekschool als naar een een sportclub, 20 naar een sportclub en een jeugdbeweging en 15 zowel naar een muziekschool als naar een jeugdbeweging. 10 leerlingen gaan zowel naar de muziekschool als naar een sportclub als naar een jeugdbeweging.
  Hoeveel leerlingen gaan
  1. naar minstens 1 van de drie;
  2. naar geen enkele van de drie;
  3. naar juist 2 van de drie;
  4. niet naar een sportclub.

Figuur 1.20: venndiagram voor de activiteiten van de leerlingen

OPLOSSING: Uit het venn-diagram kunnen we afleiden dat

1. 125 leerlingen minstens 1 van de drie activiteiten kiezen;
2. 150-125=25 leerlingen aan geen enkele activiteit deelnemen;
3. 30 leerlingen juist 2 van de drie activiteiten kiezen;
4. 150-65=85 leerlingen gaan niet naar een sportclub.

TAAK ♣ **14** Bij een enquête is aan 112 leerlingen gevraagd of ze de vorige dag tv gekeken hebben. Het resultaat is het volgende:

1. 20 lln hebben geen tv gekeken;
2. 20 lln keken 's morgens;
3. 48 lln keken 's middags;
4. 67 lln keken 's avonds;
5. 25 lln keken 's avonds en 's middags;
6. 12 lln keken 's morgens en 's middags;
7. 16 keken 's morgens en 's avonds.

Hoeveel leerlingen keken

1. 's morgens, 's middags en 's avonds?
2. 's middags en 's avonds maar niet 's morgens?

Figuur 1.21: venndiagram voor het kijken naar tv

1.2 Groeperingen

1.2.1 Herhalingsvariaties

1.2.1.1 Probleemstelling

We beschouwen een probleem die we op twee verschillende manieren kunnen formuleren.

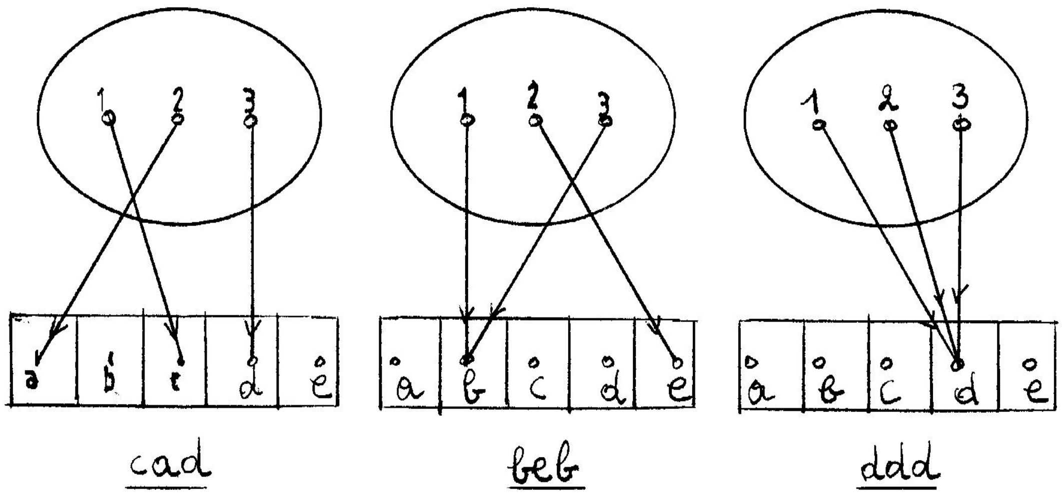
1. Hoeveel 3-tallen kunnen we vormen met elementen gekozen uit 5 verschillende elementen a, b, c, d en e?
   Voor het element op de eerste plaats kunnen we kiezen uit een verzameling $N = \{a, b, c, d, e\}$ van 5 elementen, voor het element op de tweede plaats kunnen we weer kiezen uit dezelfde verzameling N, alsook voor het element op de derde plaats. Het aantal mogelijkheden is dus:
   $$|N \times N \times N| = |N|^3 = 5^3$$
   Een 3-tal van 3 elementen gekozen uit een verzameling van 5 elementen waarbij eenzelfde element mag herhaald worden en de volgorde waarin de elementen geplaatst worden een rol speelt noemen we een **herhalingsvariatie van 5 elementen in groepjes van 3**. Het aantal herhalingsvariaties stellen we voor door
   $$\overline{V}_5^3 = 5^3.$$

2. Op hoeveel manieren kan men drie verschillende elementen 1, 2 en 3 plaatsen in 5 verschillende hokjes waarbij meerdere elementen in eenzelfde hokje mogen geplaatst worden?
   Voor het element 1 hebben we 5 mogelijkheden om een hokje te kiezen, voor het element 2 hebben we weer 5 mogelijkheden vermits er meerdere elementen in eenzelfde hokje mogen plaatsnemen, voor 3 hebben we weer 5 mogelijkheden.
   Zo een plaatsing van 3 verschillende elementen in 5 verschillende hokjes is dus ook een herhalingsvariatie van 5 elementen in groepjes van 3. Het aantal is dus $\overline{V}_5^3 = 5^3$.

Opmerkingen:

- Een plaatsing van 3 verschillende elementen in 5 hokjes is een **afbeelding** van een verzameling $P = \{1, 2, 3\}$ met 3 elementen in een verzameling $N = \{a, b, c, d, e\}$ van 5 elementen.
  Inderdaad, bij een afbeelding vertrekt uit elk element van P juist één pijl en komen in N 3 pijlen toe waarbij er eventueel meerdere pijlen in één element mogen toekomen.

- De verzameling van de afbeeldingen van P in N noteren we als N^P. Het aantal afbeeldingen van P in N is:
  $$|N^P| = 5^3$$



1.2.1.2 Definitie

Een **herhalingsvariatie van n elementen in groepjes van p** is een p-TAL (volgorde) van p elementen (herhaling) gekozen uit n elementen. Hierbij mag eenzelfde element verschillende keren herhaald worden. In een p-tal speelt de volgorde waarin de elementen staan een rol.
Een herhalingsvariatie van n elementen in groepjes van p correspondeert met een plaatsing van p verschillende voorwerpen (volgorde) in n vakjes. Hierbij mogen verschillende voorwerpen in één vakje geplaatst worden (herhaling).

Het aantal herhalingsvariaties van n elementen in groepjes van p noteren we

$$\overline{V}_n^p = n^p.$$

1.2.1.3 Toepassing: Aantal deelverzamelingen van een verzameling

Een deelverzameling van een verzameling met n elementen correspondeert met een plaatsing van n elementen in twee vakjes. In het eerste vakje leggen we de elementen van de deelverzameling en in het andere vakje de overige elementen. Het aantal deelverzamelingen van de verzameling met n elementen correspondeert dus met het aantal plaatsingen van de n elementen in de twee vakjes. Het aantal deelverzamelingen van een verzameling met n elementen is gelijk aan 2^n.

We kunnen ook als volgt redeneren: men heeft n voorwerpen, men heeft voor elk van de voorwerpen twee mogelijkheden, kiezen of niet kiezen, dus 2^n mogelijkheden in totaal.

Figuur 1.22: aantal deelverzamelingen

TAAK ♣ **15** Op hoeveel manieren kunnen tien personen een warenhuis binnengaan, als er vier ingangen zijn?

♣ **16** Op hoeveel manieren kunnen we een test invullen als hij dertig vragen omvat en als er op elke vraag vier antwoorden mogelijk zijn?

♣ **17** Een dobbelsteen wordt twee keer achtereen opgeworpen. Hoeveel uitkomsten zijn er mogelijk? Hoeveel mogelijkheden zijn er waarbij de som van het aantal ogen op beide dobbelstenen gelijk is aan zeven, groter is dan acht.

♣ **18** Hoeveel mogelijkheden zijn er voor de samenstelling van een gezin van vier kinderen?

19 De capaciteit van het resultatenbord van een oude computer is zestien tekens. Als je weet dat enkel de tekens 0 en 1 gebruikt worden, bereken dan het aantal verschillende “woorden” die we met een dergelijke computer kunnen vormen?

Oplossingen: 15 $4^{10} = 1048576$; 16 $4^{30} = 1,152921505.10^{18}$; 17 $6^2 = 36; 6; 10$; 18 $2^4 = 16$; 19 $2^{16} = 65536$.

1.2.2 Variaties en permutaties

1.2.2.1 Probleemstelling

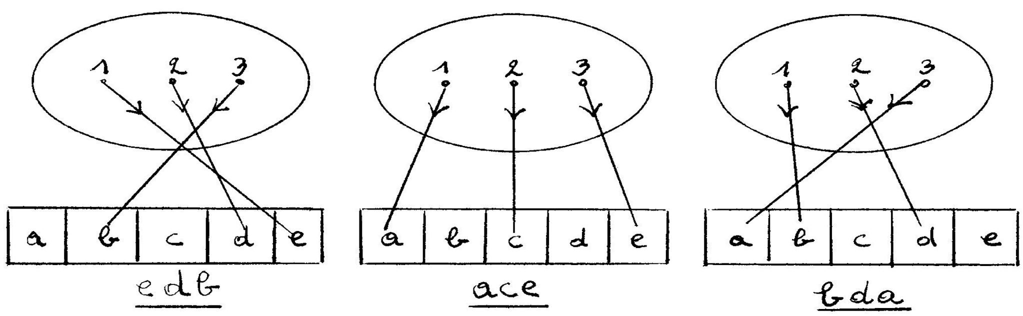
1. Hoeveel 3-tallen van 3 verschillende elementen kunnen we vormen met elementen gekozen uit 5 verschillende elementen a, b, c, d en e?
   Voor het element op de eerste plaats kunnen we kiezen uit een verzameling N van 5 elementen, voor het element op de tweede plaats kunnen we kiezen uit een verzameling M van 4 elementen en voor het element op de derde plaats kunnen we kiezen uit een verzameling L met 3 elementen. De verzamelingen M en L zijn voor elke keuze van het eerste element telkens anders maar hun aantal elementen zijn resp. steeds 4 en 3. Het aantal mogelijkheden is dus

   $$|N \times M \times L| = |N| \cdot |M| \cdot |L| = 5.4.3 = 60.$$

   Een 3-tal van drie verschillende elementen gekozen uit een verzameling van 5 elementen waarbij de volgorde van de elementen een rol speelt noemen we een **variatie van 5 elementen in groepjes van 3**.
   Het aantal variaties stelllen we voor door

   $$V_5^3 = 5 \cdot 4 \cdot 3.$$

2. Op hoeveel manieren kan men 3 verschillende elementen 1, 2 en 3 plaatsen in 5 verschillende hokjes waarbij niet meer dan één element per hokje mag geplaatst worden.
   Voor het element 1 hebben we 5 mogelijkheden om een hokje te kiezen, voor element 2 hebben we slechts 4 mogelijkheden vermits al één hokje bezet is door element 1. Voor het element 3 hebben dan nog maar 3 mogelijkheden meer.
   Zo een plaatsing van drie verschillende elementen in 5 hokjes is dus een variatie van 5 elementen in groepjes van 3. Het aantal is dus $V_5^3 = 5 \cdot 4 \cdot 3 = 60$.



OPMERKING:
Een plaatsing van 3 verschillende elementen in 5 verschillende hokjes waarbij niet meer dan één element per hokje mag geplaatst worden, is een **injectie** van een verzameling $P = \{1, 2, 3\}$ van 3 elementen in een verzameling $N = \{a, b, c, d, e\}$ van 5 elementen.
Inderdaad, bij een injectie vertrekt uit elk element van P juist één pijl en komt er in elk element van N hoogstens één pijl toe.

1.2.2.2 Definities

Een **variatie van n elementen in groepjes van** p is een p-TAL (volgorde) van p verschillende elementen (geen herhaling) gekozen uit n elementen. In een p-tal speelt de volgorde waarin de elementen staan een rol. Een element mag niet herhaald worden.
Een variatie van n elementen in groepjes van p correspondeert met een plaatsing van p verschillende voorwerpen (volgorde) in n vakjes, met hoogstens één voorwerp per vakje (geen herhaling).

Het aantal variaties van n elementen in groepjes van p noteren we

$$V_n^p = n(n-1)(n-2)\dots(n-p+1).$$

Merk op dat $p \leq n$ moet zijn.

Een **permutatie van n elementen** is een variatie van n elementen in groepjes van n. Bij één enkele permutatie komt elk element van N juist één keer voor en de volgorde waarin de elementen geplaatst worden is van belang.
Het aantal permutaties van n elementen is gelijk aan het aantal mogelijkheden om n verschillende voorwerpen te plaatsen in n vakjes, met juist één voorwerp per vakje of het aantal mogelijkheden om n verschillende voorwerpen naast elkaar te plaatsen op een rij.

Het aantal permutaties van n elementen noteren we

$$n(n-1)(n-2)...(n-n+1) = n! \quad (n \text{ faculteit})$$

In het bijzonder is $0! = 1$.

OPMERKING: Een permutatie komt overeen met een bijectie van de verzameling P in de verzameling N. Het aantal elementen van P is dan gelijk aan het aantal elementen van N.

ANDERE SCHRIJFWIJZE VOOR HET AANTAL VARIATIES:
De verkorte notatie $n!$ voor het product van de n opeenvolgende natuurlijke getallen 1 tot en met n geeft ons de mogelijkheid de formule voor een variatie op de volgende manier te schrijven:

$$V_n^p = \frac{n.(n-1).(n-2).\dots.(n-p+1).(n-p).\dots.2.1}{(n-p).(n-p-1).\dots.2.1} = \frac{n!}{(n-p)!}$$

TAAK ♣ **20** Op hoeveel manieren kunnen drie personen een warenhuis binnengaan, als er vier ingangen zijn en als er hoogstens één persoon van één ingang gebruik mag maken?

♣ **21** Op hoeveel manieren kunnen zeventien leerlingen plaats nemen in een lokaal met vijfentwintig stoelen?

♣ **22** Hoeveel getallen van vijf verschillende cijfers zijn er? Hoeveel van deze getallen

a. beginnen met 1?
b. eindigen op 5?
c. bevatten het cijfer 5?
d. bevatten de cijfers 4 en 5 niet?
e. bevatten het cijfer 4 maar niet het cijfer 3?

♣ **23** In een vergadering van 220 personen moet een bestuur van 4 personen gekozen worden uit 10 kandidaten. Ieder aanwezige moet stemmen voor 4 kandidaten in voorkeurvolgorde. Op hoeveel manieren kan elke stemgerechtigde zijn stembiljet geldig invullen?

♣ **24**
a. Op hoeveel manieren kunnen vijf jongens en vier meisjes naast elkaar op een rij zitten?
b. Op hoeveel manieren kunnen ze naast elkaar zitten op een rij, als de jongens zowel als de meisjes gegroepeerd wensen te blijven?
c. Op hoeveel manieren kunnen zij naast elkaar zitten als alleen de meisjes gegroepeerd wensen te blijven?
d. Op hoeveel manieren kunnen ze naast elkaar zitten als één bepaald meisje naast één bepaalde jongen moet zitten?
e. Op hoeveel manieren kunnen ze naast elkaar zitten als één bepaald meisje niet naast één bepaalde jongen mag zitten?
f. Beantwoord dezelfde vragen als de personen plaats nemen aan een ronde tafel?

♣ **25** Op hoeveel manieren kunnen vijf verschillende geschiedenisboeken, vier verschillende scheikundeboeken, vier verschillende wiskundeboeken naast elkaar op een boekenrek geplaatst worden, als de boeken per vak gegroepeerd moeten blijven? En als alleen de wiskundeboeken gegroepeerd moeten blijven?

26 Hoeveel vlaggen met drie vertikale strepen van verschillende kleur kan men maken als men beschikt over vijf verschillende kleuren?

27 Op hoeveel manieren kunnen we een voetbalelftal vormen, als we over 25 jongens beschikken en als iedere jongen een welbepaalde opdracht moet krijgen?

28 Op hoeveel manieren kunnen dertig mensen de beschikbare vierentwintig plaatsen in een tram bezetten?

29 Op hoeveel manieren kunnen vijf lampen van verschillende kleur naast elkaar aan een snoer worden opgehangen?

30 Op hoeveel manieren kunnen twaalf personen naast elkaar aan een ronde tafel zitten?

31 Op hoeveel manieren kunnen vier Russen, twee Amerikanen en drie Duitsers plaats nemen aan een ronde tafel als de personen van een zelfde nationaliteit naast elkaar wensen te zitten?

32 a. Hoeveel getallen van vier verschillende cijfers kunnen we vormen met de cijfers 1,2,4,5,7?

b. Bereken hun som.

c. Hoeveel van de getallen beginnen met 5, met 54, met 542?

d. Als je die getallen in stijgende volgorde rangschikt, welk is dan het vijftigste getal? Welk is het rangnummer van het getal 4251?

e. Hoeveel getallen zijn er waarin de cijfers 2,4 en 5 naast elkaar voorkomen in die volgorde?

f. Hoeveel getallen zijn er waarin de cijfers 2,4,en 5 naast elkaar in een willekeurige volgorde voorkomen?

OPLOSSINGEN; 20 $V_4^3 = 24$; 21 $V_{25}^{17} = 3,8470263.10^{20}$;
22 9.9.8.7.6; (a) 9.8.7.6; (b) 8.8.7.6; (c) 9.8.7.6 + 4.8.8.7.6; (d) 7.7.6.5.4; (e) 8.7.6.5 + 4.7.7.6.5.
23 $V_{10}^4 = 5040$;
24 (a) $9! = 362880$; (b) $5!.4!.2 = 5760$; (c) $6!.4! = 17280$; (d) $8!.2 = 80640$;
(e) $9! - 8!.2 = 282240$; (f) $8! = 40320$; $5!.4! = 2880$; $5!.4!$; $7!.2 = 10080$; $8! - 7!.2 = 30240$.
25 $5!.4!.4!.3! = 414720$; $4!.10! = 87091200$; 26 $V_5^3 = 60$; 27 $V_{25}^{11} = 1,779251443.10^{14}$;
28 $V_{30}^{24} = 3,684067497.10^{29}$;
29 $5! = 120$;
30 $11! = 39916800$;
31 $4!.2!.3!.2! = 576$;
32 (a) $V_5^4 = 5! = 120$; (b) 506616; (c) 24; 6; 2; (d) 4 127; 57; (e) 4; (f) $3!.4 = 24$.

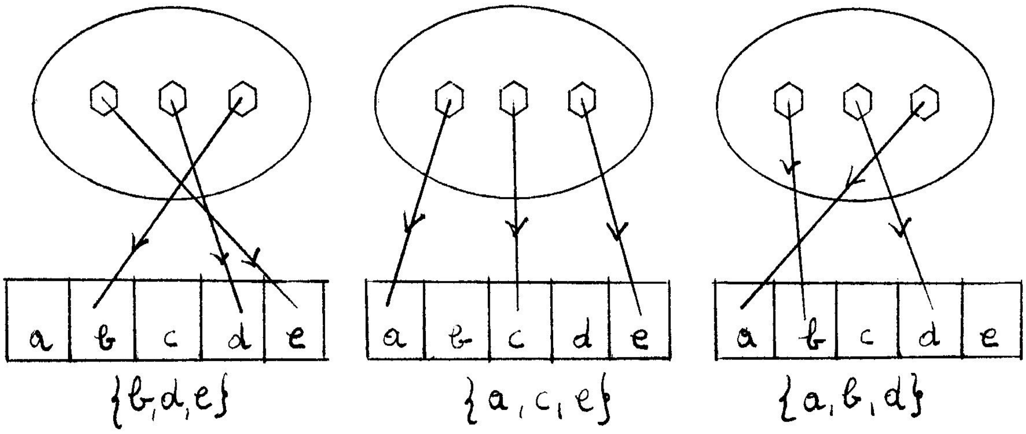
1.2.3 Combinaties

1.2.3.1 Probleemstelling:

Op hoeveel manieren kunnen we 3 identieke elementen (geen volgorde) plaatsen in 5 verschillende hokjes met hoogstens één element per hokje (geen herhaling)? Anders geformuleerd, op hoeveel manieren kan men 3 verschillende hokjes kiezen uit de 5 hokjes om de identieke elementen in te plaatsen?
Een verzameling van 3 elementen gekozen uit een verzameling met 5 elementen noemen we een **combinatie van 5 elementen in groepjes van 3**.

Om het aantal combinaties te bepalen, leggen we een verband met het aantal variaties. We veronderstellen een ogenblik dat de drie voorwerpen verschillend zijn (de volgorde dus wel een rol speelt). Bij elke keuze van 3 hokjes uit de 5, dus bij elke combinatie horen 3! variaties. We stellen het aantal combinaties voor door C_5^3. Er geldt

$$C_5^3.3! = V_5^3 \Longleftrightarrow C_5^3 = \frac{V_5^3}{3!}$$



1.2.3.2 Definitie

Een **combinatie van n elementen in groepjes van p** is gelijk aan een VERZAMELING (geen volgorde en geen herhaling) van p verschillende elementen uit n verschillende elementen. De volgorde speelt hier geen rol en een element wordt niet herhaald.
Een combinatie van n elementen in groepjes van p komt overeen met een plaatsing van p IDENTIEKE VOORWERPEN (geen volgorde) in n vakjes, met hoogstens één voorwerp per

vakje (geen herhaling).

Het aantal combinaties van n elementen in groepjes van p is

$$C_n^p = \frac{V_n^p}{p!} = \frac{n!}{(n-p)!p!}$$

Een andere notatie voor aantal combinaties van n elementen in groepjes van p is

$$\binom{n}{p}$$

1.2.3.3 Eigenschappen van combinaties

STELLING 1.1

$$\forall n \in N : \binom{n}{0} = \binom{n}{n}$$

De ledige verzameling is de enige deelverzameling met nul elementen van elke verzameling; de deelverzameling met n elementen van een verzameling met n elementen is eveneens enig.

STELLING 1.2

$$\forall n \in N : \binom{n}{1} = n$$

Het aantal deelverzamelingen met één element, die singletons zijn, van een verzameling met n elementen is gelijk aan n.

STELLING 1.3

$$\forall n, p \in N : n < p \Longrightarrow \binom{n}{p} = 0$$

STELLING 1.4

$$\forall n, p \in N : n > p \Longrightarrow \binom{n}{p} = \binom{n}{n-p}$$

STELLING 1.5

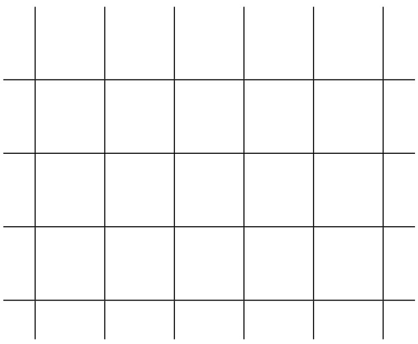
$$\forall n \in N : 2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n-1} + \binom{n}{n}$$

Het totaal aantal deelverzamelingen van een verzameling met n elementen is gelijk aan de som van het aantal deelverzamelingen met nul elementen, het aantal deelverzamelingen met één element, het aantal deelverzamelingen met twee elementen, . . . , het aantal deelverzamelingen met $n-1$ elementen en het aantal deelverzamelingen met n elementen.

INLEIDING TOT DE FORMULE VAN STIFEL-PASCAL:
Sommige steden hebben een rechthoekig stratenpatroon. We stellen deze straten voor op een roosterdiagram. Op hoeveel manieren kan men van het punt $(0,0)$ naar het punt $(4,3)$ gaan als men de kortst mogelijke weg volgt?
OPLOSSING:

1. Eerste manier van oplossen.
   Dit vraagstukje is een toepassing van combinaties. Er moeten 7 routes doorlopen worden waarvan 4 naar rechts (R) en 3 naar boven (B). We geven enkele 7-tallen als mogelijkheden: RRBBRBR, RRRBBBR, BBRBRRR. Om een mogelijkheid te bekomen, moeten we van de 7 plaatsen er 4 kiezen voor R, de rest is voor B ofwel moeten we er 3 kiezen voor de B en de rest is dan voor R. Dit kan op
   $$\binom{7}{4} = \binom{7}{3} = 35$$

2. Tweede manier van oplossen.
   We kijken eerste wat het aantal mogelijke routes zijn om van $(0,0)$ naar $(1,1)$ te gaan. Zo zijn er 2 mogelijkheden. We schrijven dat aantal bij het roosterpunt $(1,1)$. Er is 1 mogelijkheid om van $(0,0)$ naar $(2,0)$ te gaan. Het is dan gemakkelijk in te zien dat er $2+1=3$ mogelijkheden zijn om van $(0,0)$ naar $(2,1)$ te gaan. We schrijven 3 bij het roosterpunt $(2,1)$. Analoog zijn er 3 mogelijkheden om naar $(1,2)$ te gaan. Er zijn dan $3+3=6$ mogelijkheden om naar $(2,2)$ te gaan. Zo kunnen we het rooster verder aanvullen. Tot we $20+15=35$ mogelijkheden verkrijgen om naar $(4,3)$ te gaan. Eigenlijk is de laatste som uitgedrukt met combinaties:
   $$\binom{6}{3} + \binom{6}{4} = \binom{7}{4} = 15 + 20 = 35$$



Tabel 1.4: stratenpatroon —- Venndiagram voor het bewijs van de stelling van Stifel-Pascal

Deze beschouwing kunnen we veralgemenen in de volgende stelling:

STELLING 1.6

$$\forall n, p \in N : n > p \Longrightarrow \begin{pmatrix} n \\ p \end{pmatrix} = \begin{pmatrix} n-1 \\ p \end{pmatrix} + \begin{pmatrix} n-1 \\ p-1 \end{pmatrix}$$

Dit is de formule van Stifel-Pascal.

Bewijs: We beschouwen een bepaald element a van een verzameling A met n elementen. Het totaal aantal deelverzamelingen met p elementen van A is de som van het aantal deelverzamelingen met p elementen die het element a bevatten en het aantal deelverzamelingen met p elementen die het element a niet bevatten.
Het aantal deelverzamelingen met p elementen die het element a bevatten, is C_{n-1}^{p-1} en het aantal deelverzamelingen met p elementen die het element a niet bevatten is C_{n-1}^{p}.

1.2.3.4 De driehoek van Pascal

De driehoek van Pascal is een tabel waar we de waarden van alle combinaties kunnen op invullen zonder ze rechtstreeks te berekenen maar door te steunen op hun eigenschappen en dan vooral op de formule van Stifel-Pascal.

	0	1	2	3	4	5	6	7	8	9	10
0											
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											

OPGAVEN — **33** Bepaal n als
(a) $\binom{n}{8} = \binom{n}{6}$; (b) $\binom{n}{5} = 21$; (c) $\binom{n}{9} = \binom{n}{4}$

34 Bewijs:
(a) $\binom{n}{p} = \frac{n}{p}\binom{n-1}{p-1}$ (b) $\binom{n}{p} = \frac{n-p+1}{p}\binom{n}{p-1}$

(c) Uit (b) volgt: $\sum_{i=1}^{n} \frac{i\binom{n}{i}}{\binom{n}{i-1}} = \frac{n(n+1)}{2}$

35 Vereenvoudig
(a) $\frac{1}{(n+1)!} - \frac{1}{n!} + \frac{1}{(n-1)!}$; (b) $\frac{1}{p!(n-1)!} + \frac{1}{n!(p-1)!}$; (c) $\frac{n!}{p!(n-p)!} + \frac{(n-1)!}{(p-1)!(n-p)!}$

36 Toon aan dat $\frac{(n+2)!}{(n-1)!}$ deelbaar is door 6.

37 Hoeveel producten van drie verschillende factoren van één cijfer kunnen we vormen met de cijfers 1 tot en met 7?

38 Op hoeveel manieren kunnen we zeven kaarten trekken uit een spel van tweeënvijftig kaarten als de getrokken kaart niet telkens wordt teruggestoken?

39 Hoeveel diagonalen heeft een convexe n-hoek? In hoeveel nieuwe punten snijden de diagonalen elkaar? Hoeveel van deze punten liggen binnen de veelhoek en hoeveel erbuiten (geen drie diagonalen zijn concurrent tenzij in een hoekpunt)? Van welke veelhoek is het aantal zijden gelijk aan het aantal diagonalen?

40 Een test omvat twaalf vragen waarvan er tien moeten beantwoord worden. Hoeveel keuzen zijn er

a. als er tien willekeurige vragen mogen beantwoord worden?
b. als er op de eerste drie moet geantwoord worden?

$\boxed{\textbf{TAAK}}$ ♣ **41** Van vier punten in het vlak zijn er geen drie collineair. Hoeveel rechten worden door die vier punten bepaald (de zijden van een volledige vierhoek)?

♣ **42** Je trekt vijf kaarten uit een spel van tweeënvijftig kaarten.

a. Op hoeveel manieren kun je precies drie heren trekken?
b. Op hoeveel manieren kun je juist vier ruiten bekomen?

♣ **43** Op hoeveel manieren kunnen we een raad samenstellen die bestaat uit drie ministers, vier senatoren en vier professoren, gekozen uit vijf ministers, tien senatoren en vijftien professoren?

♣ **44** Gegeven is een roosterdiagram. Op hoeveel manieren kan men van $(0,0)$ naar $(3,3)$ gaan door alleen rechts en naar boven te verschuiven. De punten $(2,0)$, $(3,0)$ en $(3,1)$ moeten hierbij vermeden worden.

♣ **45** Tot de vriendenkring van een gezin behoren twaalf personen.

a. Op hoeveel manieren kan dit gezin vijf onder hen op een feestje uitnodigen?

b. Op hoeveel manieren kan dit gebeuren als twee onder hen met elkaar gehuwd zijn en niet alleen wensen te komen?

c. Op hoeveel manieren kan dit gebeuren als twee onder hen helemaal niet op elkaar gesteld zijn en elkaar niet wensen te ontmoeten op een feestje?

♣ **46** Op hoeveel manieren kunnen we vier identieke bollen verschillend kleuren als we beschikken over tien verschillende kleuren?

Oplossingen:
37 $C_7^3 = 35$; 38 $C_{52}^7 = 133784560$; 41 $C_4^2 = 6$;
39 $C_n^2 - n = \frac{1}{2}n(n-3)$; $C^2_{n(n-3)/2} - nC^2_{n-3}$; C_n^4; $C^2_{n(n-3)/2} - nC^2_{n-3} - C_n^4$; vijfhoek;
40 (a) $C_{12}^{10} = 66$ (b) $C_9^7 = 36$;
41 6 42 (a) $4C_{48}^2 = 4512$; (b) $C_{39}^1.C_{13}^4 = 27885$;
43 $C_5^3.C_{10}^4.C_{15}^4 = 2866500$; 44 $C_6^3 - 4 - 2 = 14$ 45 (a) $C_{12}^5 = 792$; (b) $C_{10}^3 + C_{10}^5 = 372$; (c) $2C_{10}^4 + C_{10}^5 = C_{12}^5 - C_{10}^3 = 672$.
46 C_{10}^4.

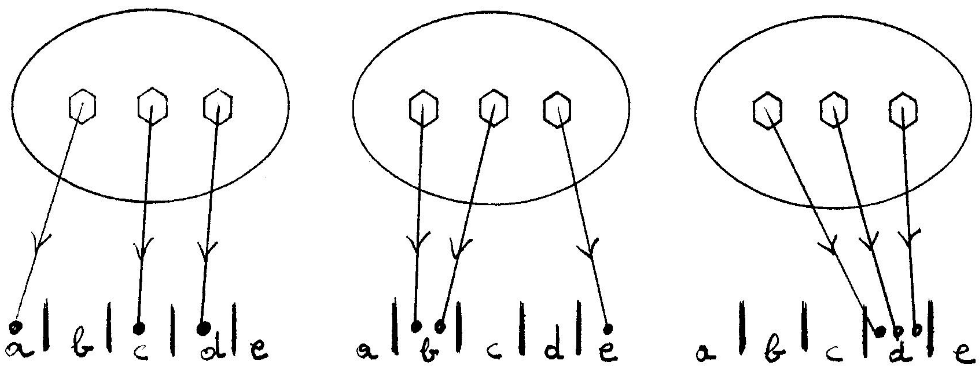
1.2.4 Herhalingscombinaties

1.2.4.1 Probleemstelling

Wat is het aantal mogelijkheden om drie elementen te kiezen uit vijf verschillende elementen waarbij eenzelfde element wel meerdere malen mag gekozen worden? Of anders geformuleerd, hoeveel mogelijkheden zijn er om drie identieke voorwerpen (geen volgorde) te plaatsen in vijf verschillende vakjes als elk vakje wel meer dan één voorwerp mag bevatten (herhaling)?

Bij zo een **plaatsing** speelt alleen de keuze van de vakjes een rol en het aantal keren dat zo een vakje gekozen wordt. Om het aantal plaatsingen te berekenen voorzien we $3+5-1=7$ plaatsen, dit is de som van het aantal voorwerpen en het aantal scheidingswanden van de aanpalende vakjes. We kennen drie van de zeven plaatsen toe aan de voorwerpen, op de overblijvende vier plaatsen zetten we de vier scheidingswanden van de vijf vakjes. Zo een keuze correspondeert met juist één plaatsing van de drie identieke voorwerpen in vijf verschillende vakjes.

Omgekeerd, een plaatsing van de drie voorwerpen in de vijf vakjes correspondeert met juist één keuze van drie plaatsen van de zeven plaatsen voor de drie voorwerpen. Het aantal mogelijkheden om drie plaatsen te kiezen uit zeven plaatsen, waarbij de volgorde geen rol speelt is gelijk aan het aantal combinaties van zeven elementen in groepjes van drie. Het aantal mogelijkheden om drie identieke voorwerpen te plaatsen in vijf vakjes, waarbij in elk vakje meer dan één voorwerp mag bevatten, is dus gelijk aan het aantal $C_7^3 = \overline{C}_5^3$.



1.2.4.2 Definitie

Een **herhalingscombinatie van n elementen in groepjes van p** is een GROEPJE (herhaling en geen volgorde) van p elementen gekozen uit n verschillende elementen, waarbij eenzelfde element meerdere malen herhaald mag worden. De volgorde speelt hier geen rol. Een herhalingscombinatie van n elementen in groepjes van p komt overeen met een plaatsing van p IDENTIEKE VOORWERPEN (geen volgorde) in n vakjes waarbij in elk vakje meer dan één voorwerp (herhaling) mag geplaatst worden.

Het aantal herhalingscombinaties van n elementen in groepjes van p is gelijk aan het aantal combinaties van $n + p - 1$ elementen in groepjes van p. We noteren

$$\overline{C}_n^p = C_{n+p-1}^p$$

$\boxed{\textbf{TAAK}}$ ♣ **47** Op hoeveel manieren kunnen acht identieke balletjes in vijf verschillende bakjes gelegd worden?

♣ **48** Hoeveel stenen bevat een dominospel?

♣ **49** Op hoeveel manieren kan een vader twintig stukken van een euro verdelen over de spaarpotjes van zijn drie kindjes?

♣ **50** Op hoeveel manieren kunnen we zes identieke bollen verven met behulp van vier verschillende kleuren?

OPLOSSINGEN:
47 $C_{12}^{8} = 495$; 48 $C_{8}^{2} = 28$; 49 $\overline{C}_{3}^{20} = C_{22}^{20} = 231$; 50 $C_{9}^{6} = 84$.

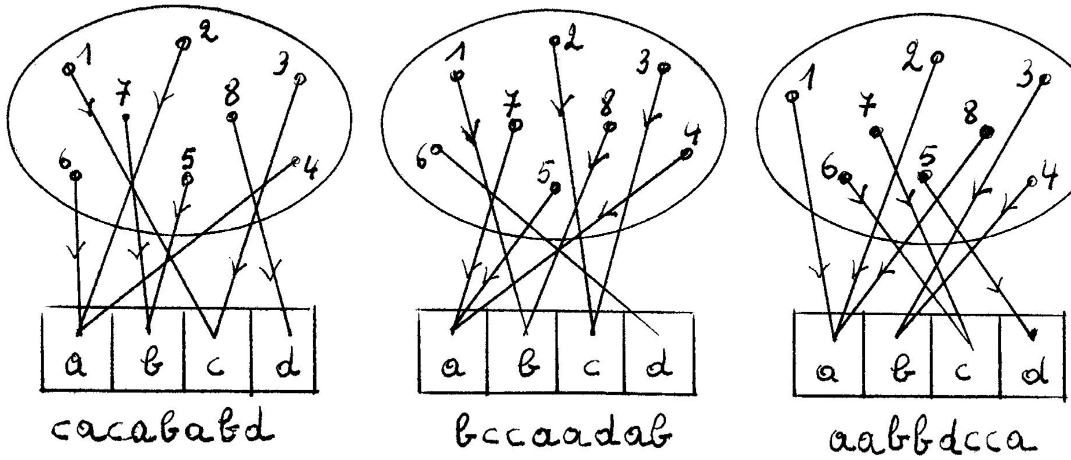
1.2.5 Herhalingspermutaties

1.2.5.1 Probleemstelling

1. Op hoeveel manieren kunnen we achttallen vormen met drie keer het element a, twee keer het element b, twee keer het element c en één keer het element d.
   We kunnen eerst voor de drie identieke elementen a drie plaatsen van de acht uitkiezen. Daarvoor hebben we $C_8^3 = 56$ mogelijkheden. Voor de twee identieke elementen b hebben we nog $8-3$ plaatsen om uit te kiezen. Dit kunnen we op $C_5^2 = 10$ manieren. Voor de plaatsing van de twee identieke voorwerpen hebben we nog 3 plaatsen over. Er zijn hiervoor $C_3^2 = 3$ mogelijkheden. Het element d neemt plaats op de overblijvende plaats. Al deze mogelijkheden met elkaar gecombineerd geeft $56 \times 10 \times 3 \times 1 = 1680$. Eén van deze 1680 mogelijkheden noemen we een herhalingspermutatie van acht elementen in groepjes van drie, twee, twee en één. Het aantal noteren we door $P_8^{3,2,2,1} = C_8^3.C_5^2.C_3^2.C_1^1$.

2. Op hoeveel manieren kunnen we acht verschillende voorwerpen verdelen over vier vakjes a, b c en d waarbij in vakje a drie voorwerpen moeten gelegd worden, in vakje b twee voorwerpen, in vakje c ook twee en in d één voorwerp?
   We beschouwen acht vakjes, drie keer vakje a, twee keer vakje b, twee keer vakje c en één keer vakje d. Om een verband te leggen tussen herhalingspermutaties en permutaties, veronderstellen we een ogenblik dat die acht vakjes verschillende zijn. We hebben dan 8! mogelijkheden om de acht verschillende voorwerpen te plaatsen in de acht vakjes. Bij een herhalingspermutatie is de volgorde waarbij de drie voorwerpen die in de drie vakjes a liggen van geen belang. Analoog voor b, c en d. Bij elke herhalingspermutatie horen dus $3! \cdot 2! \cdot 2! \cdot 1!$ permutaties. Zij x het aantal herhalingspermutaties van acht in groepjes van drie, twee, twee, één dan geldt:

$$x \cdot 3! \cdot 2! \cdot 2! \cdot 1! = 8!.$$



1.2.5.2 Definitie

Een herhalingspermutatie van n elementen in groepjes van p_1, p_2,... p_k is een n-tal (volgorde) gekozen uit k elementen, waarbij we p_1 keer het eerste element kiezen, p_2 keer het tweede element ... p_k keer het $k$de element (herhaling) waarbij $p_1 + p_2 + \cdots + p_k = n$.
Een herhalingspermutatie van n elementen in groepjes van p_1, p_2,... p_k komt overeen met een verdeling van n verschillende voorwerpen (volgorde) in k vakjes, waarbij we in het eerste vakje p_1 verschillende voorwerpen, in het tweede vakje p_2 verschillende voorwerpen ... en in het $k$de vakje p_k verschillende voorwerpen plaatsen (herhaling).

Het aantal herhalingspermutatie van n elementen in groepjes van p_1, p_2,... p_k is gelijk aan

$$P_n^{p_1,p_2,\dots,p_k} = C_n^{p_1}.C_{n-p_1}^{p_2}\dots.C_{n-p_1-p_2-\dots-p_{k-1}}^{p_k}$$

$$\Updownarrow$$

$$P_n^{p_1,p_2,\dots,p_k} = \frac{n!}{p_1!.p_2!.\dots.p_{k-1}!.p_k!}$$

$\boxed{\textbf{TAAK}}$ ♣ **51** Op hoeveel manieren kunnen we achtereenvolgens tien bollen nemen uit een vaas met vijf witte, drie rode en twee blauwe bollen?

♣ **52** Hoeveel anagrammen kunnen we vormen van het woord “vandaan”?

♣ **53** Van dertien gelijke lampen zijn er twee geel, zes paars, vier oranje en één groen. Hoeveel verschillende signalen kunnen we bekomen door die lampen naast elkaar in serie te schakelen?

♣ **54** Hoeveel manieren zijn er om bij het opgooien van vijf dobbelstenen drie keer een zes en twee keer een drie te bekomen?

OPLOSSINGEN: 51 $P_{10}^{5,3,2} = 2520$; 52 $P_7^{1,3,2,1} = 420$; 53 $P_{13}^{2,6,4,1} = 180180$; 54 $P_5^{3,2} = 10$.

HERHALINGSOEFENINGEN

OPGAVEN — **55** Een klas heeft per week twintig lesuren, acht uur wiskunde, vier uur Nederlands, vier uur Frans en vier uur Engels.

a. Hoeveel verschillende lesroosters zijn er mogelijk?

b. Hoeveel lesroosters zijn er mogelijk als elk van de vijf schooldagen begint met een uur wiskunde?

56 Hoeveel termen telt een volledig homogene veelterm van de vijfde graad in drie onbepaalden?

57 Hoeveel getallen van tien cijfers kunnen we vormen in het binair stelsel?

58 Hoeveel "woorden" van vijf verschillende letters kunnen we vormen met acht gegeven medeklinkers en twee gegeven klinkers

a. als ieder woord moet bestaan uit drie medeklinkers en twee klinkers;

b. als ieder woord drie bepaalde letters moet bevatten die

- van elkaar gescheiden zijn;
- samen moeten blijven en in een willekeurige volgorde geplaatst mogen worden;
- samen moeten blijven en in een gegeven volgorde geplaatst moeten worden?

59 Op hoeveel manieren kunnen we twaalf kaarten aan vier spelers uitdelen als elke speler drie kaarten krijgt?

60 Op hoeveel manieren kunnen we $(2m + 1)$ voorwerpen in m groepen van twee voorwerpen plaatsen, wanneer een voorwerp ongebruikt blijft?

61 Voor een schaakcompetitie moet elke school een ploeg van 3 spelers en een kapitein afvaardigen. Onze school heeft 10 geschikte kandidaten waarvan er 3 in aanmerking komen om kapitein te zijn. Op hoeveel manieren kan men een ploeg samenstellen rekening houdend met het feit dat de 3 kandidaten-kapitein ook spelers kunnen zijn?

62 Op hoeveel manieren kunnen 5 getrouwde paren aan een ronde tafel plaatsnemen? In hoeveel gevallen zit elke man naast zijn eigen vrouw?

63 Op hoeveel manieren kunnen achtereenvolgens 3 kaarten worden getrokken uit een spel van 52 speelkaarten als de getrokken kaart telkens wordt teruggestoken?

64 Hoeveel getallen van 3 cijfers kan men vormen in het hexadecimaal stelsel? (16-tallig)

65 Hoeveel getallen van 3 verschillende cijfers zijn

a. onderling ondeelbaar met 10?

b. onderling deelbaar met 10? (Twee getallen zijn onderling ondeelbaar als de ggd = 1)

66 Op hoeveel manieren kunnen 10 personen een verschillende verjaardag hebben?

67 Bij hoeveel nummers van 4 cijfers staan de cijfers in dalende volgorde?

68 Hoeveel driehoeken kun je vormen met de hoekpunten van een zevenhoek? Dezelfde vraag voor het geval waarbij geen enkele zijde van de driehoek een zijde van de zevenhoek mag zijn.

69 De huidige wereldbevolking bedraagt ongeveer 5,7 miljard. Om elke mens een verschillend nummer toe te kennen, heeft men in het decimaal stelsel 10 posities nodig. Hoeveel posities heeft men nodig als men letters i.p.v. cijfers zou gebruiken?

70 Hoeveel relaties zijn er van een verzameling met n elementen naar een verzameling met m elementen?

71 Op een fuif zijn 30 jongens en 40 meisjes aanwezig. Hoeveel verschillende danspaartjes kunnen er die avond gevomd worden?

72 Op een vragenlijst staan 20 vragen te beantwoorden met waar of vals. Een leerling moet minstens 12 vragen juist beantwoorden om geslaagd te zijn. Op hoeveel verschillende manieren kan een leerling die geslaagd is zijn antwoordformulier ingevuld hebben?

73 Hoeveel pincodes van vier cijfers kan men vormen men juist twee verschillende cijfers?

74 * Hoeveel antireflexieve permutaties (permutaties zonder lus) bezit een verzameling met 1,2,3 of 4 elementen?

Oplossingen:
55: a. $P_{20}^{8,4,4,4}$; b. $P_{15}^{3,4,4,4}$;
56: $\overline{C}_3^5 = 21$;
57: 2^9;
58: a. $C_5^2.V_8^3.V_2^2 = C_8^3.C_2^2.5! = 6720$; b. $C_7^2.3!.2! = V_7^2.3! = 252$; $C_7^2.3!.3! = V_7^2.3!3 = 756$; $3!C_7^2 = 3V_7^2$;
59: $P_{12}^{3,3,3,3,} = 369600$;
60: $\frac{2m+1}{m!}P_{2m}^{2,2,\ldots,2}$ (m keer 2);
61: $3.C_9^3$; 62: $9!$ en $4!.2^5$; 63: 52^3; 64: 15.16^2;
65: (a) $8.8.4$; (b) $8.8.5 + 9.8 = 9.9.8 - 8.8.4$;
66: V_{365}^{10};
67: C_{10}^4;

68: $C_7^3 = 35$; aantal driehoeken met 1 zijde gemeen met de 7-hoek: 7×3; aantal driehoeken met 2 zijden gemeen met de 7-hoek: 7 Antwoord op het 2de deel: $35 - (21 + 7) = 7$;
69: 5,7 miljard $= 5,7 \cdot 10^9 < 26^n$ met n het aantal posities als we met 26 letters werken.

$$\begin{aligned} 5,7 \cdot 10^9 < 26^n & \iff \log 5,7 \cdot 10^9 < log 26^n \\ & \iff n > \tfrac{9+\log 5,9}{\log 26} \iff n > 6,89... \end{aligned}$$

Antwoord: We hebben 7 posities nodig om elke mens een verschillend nummer te geven met 26 letters.
??: met 1 teken kunnen we 2 letters voorstellen, nl. / en . met twee tekens kunnen we 2^2 letters voorstellen, nl. //; /.; ./; .., met drie tekens k78unnen we 2^3 letters voorstellen, met vier tekens kunnen we 2^4 letters voorstellen,

$$2 + 2^2 + 2^3 + 2^4 = 30$$

We moeten maar 26 letters voorstellen dus moeten we minimum 4 tekens na elkaar gebruiken.
70: Het aantal relaties van een verzameling A met n elementen naar een verzameling B met m elementen is gelijk aan het aantal deelverzamelingen van de productverzameling $A \times B$. Antwooord: $2^{n.m}$;
71: $30 \cdot 40$;
72:

$$C_{20}^{12} + C_{20}^{13} + C_{20}^{14} + C_{20}^{15} + C_{20}^{16} + C_{20}^{17} + C_{20}^{18} + C_{20}^{19} + C_{20}^{20}$$

Er zijn C_{20}^{12} manieren om 12 juiste antwoorden te kiezen uit 20 juiste antwoorden, er zijn C_{20}^{13} manieren om 13 juiste antwoorden te kiezen uit 20 juiste antwoorden enz...
74: Beredenering voor 4 elementen: We plaatsen 1 op de plaats van 2 en vormen dan alle antireflexieve permutaties van de drie overige elementen (=2), waarbij 2 de rol speelt van 1. Zo zijn er twee mogelijkheden. Dit zijn dan alle antireflexieve permutaties, waarbij 1 op de tweede plaats staat met uitzondering van de antireflexieve permutaties waarbij de 2 op de eerste plaats staat (omdat 2 de rol speelt van de 1). Het aantal antireflexieve permutaties waarbij 2 op de eerste en 1 op de tweede plaats staat is gelijk aan het aantal antirelfexieve permutaties van 3 en 4 en dit is gelijk aan 1. Het aantal antireflexieve permutaties waarbij 1 op de tweede plaats staat is gelijk aan $2 + 1 = 3$. Op analoge wijze hebben we 3 antireflexieve permutaties waatbij 1 op de derde plaats en 3 waarbij 1 op de vierde plaats staat. Het totaal aantal antireflexieve permutaties van 4 elementen is

$$A_4 = 3.A_3 + 3.A_2$$

Veralgemening voor n elementen:

$$A_n = (n-1)(A_{n-1} + A_{n-2})$$

Uit deze formule kunnen we een formule afleiden waarbij we A_n uitdrukken in functie van enkel n en A_{n-1}.

$$A_n = (n-1)A_{n-1} + (n-1)A_{n-2}$$
$$\Updownarrow$$
$$A_n - nA_{n-1} = A_{n-1} + (n-1)A_{n-2}$$
$$\Updownarrow$$
$$A_n - nA_{n-1} = -(A_{n-1} - (n-1)A_{n-2})$$
$$\Updownarrow$$
$$A_n - nA_{n-1} = -(-(A_{n-2} - (n-2)A_{n-3}))$$
$$\Updownarrow$$
$$A_n - nA_{n-1} = (-1)^2(A_{n-2} - (n-2)A_{n-3})$$

$$\Updownarrow$$

$$A_n - nA_{n-1} = (-1)^3(A_{n-3} - (n-3)A_{n-4})$$

$$\Updownarrow$$

$$\vdots$$

$$A_n - nA_{n-1} = (-1)^k(A_{n-k} - (n-k)A_{n-k-1})$$

Stel $k = n - 2$ dan is

$$A_n - nA_{n-1} = (-1)^{n-2}(A_2 - 2A_1)$$

Hieruit volgt de betrekking

$$A_n - nA_{n-1} = (-1)^n$$

$$\Updownarrow$$

$$A_n = nA_{n-1} + (-1)^n \tag{1.1}$$

Hieruit kunnen we een formule afleiden waarbij A_n nog enkel functie is van n. Passen we de laatste formule toe op A_{n-1}

$$A_{n-1} = (n-1)A_{n-2} + (-1)^{n-1}$$

en substitueren deze in 1.1. We doen het ook voor A_{n-1} enz. Na wat rekenwerk vinden we de volgende uitdrukking:

$$A_n = \frac{n!}{2!} - \frac{n!}{3!} + \frac{n!}{4!} - \cdots + (-1)^{n-1}\frac{n!}{(n-1)!} + \frac{n!}{n!}(-1)^n \tag{1.2}$$

Toepassing: Neem 5 gehuwde paren. De mannen moeten loten voor een vrouw. Hoeveel mogelijkheden zijn er waarbij geen enkele man zijn eigen vrouw geloot heeft?
Oplossing:

$$A_5 = \frac{5!}{2!} - \frac{5!}{3!} + \frac{5!}{4!} - \frac{5!}{5!} = 44$$

Opmerking: Het aantal permutaties waarbij tenminste één element op zijn eigen plaats staat is gelijk aan

$$n! - A_n = \frac{n!}{1!} - \frac{n!}{2!} + \frac{n!}{3!} - \cdots + (-1)^{n-1}\frac{n!}{(n-2)!} + \frac{n!}{(n-1)!}(-1)^n + \frac{n!}{n!}(-1)^{n+1}$$

Delen we beide leden door $n!$ (het totaal aantal permutaties), dan bekomen we de kans dat er voor n elementen minstens één zijn eigen plaats inneemt.

$$\frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{(n-2)!} + \frac{1}{(n-1)!}(-1)^n + \frac{1}{n!}(-1)^{n+1}$$

Laten we in de laatste uitdrukking n naar $+\infty$ gaan dan bekomen we de reeksontwikkeling van $y = 1 - e^x$ met $x = -1$.

De reekssom is gelijk aan $1 - \frac{1}{e} = 0,632120558...$

Berekenen we achtereenvolgens van 2, 3, 4, 5, 6 enz. elementen de kans dat er minstens één op zijn eigen plaats staat, dan bekomen we:

- voor 2 elementen 0,5
- voor 3 elementen 0.6666...

- voor 4 elementen 0,625
- voor 5 elementen 0,6333...
- voor 6 elementen 0,6319444...

De waarde voor n gaande naar $+\infty$ wordt al heel vlug goed benaderd. De kans dat op de wereld van alle gehuwde paren (als we ze kriskras door elkaar op de aarde neerzetten) er minstens één zijn eigen partner ontmoet is buiten de verwachtingen tamelijk groot, meer dan 0,5.

1.3 Het binomium van Newton

We berekenen het volgend product:

$$\begin{aligned}(x+a).(x+b).(x+c).(x+d) &= x^4 + (a+b+c+d)x^3 \\ &\quad +(ab+ac+ad+bc+bd+cd)x^2 \\ &\quad +(abc+abd+acd+bcd)x + abcd\end{aligned}$$

Het aantal termen van de coëfficiënt $a+b+c+d$ van x^3 is gelijk aan het aantal deelverzamelingen met één element uit de verzameling $\{a,b,c,d\}$ d.i. $C_4^1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix} = 4$. Het aantal termen van de coëfficiënt $ab+ac+ad+bc+bd+cd$ van x^2 is gelijk aan het aantal deelverzamelingen met twee elementen uit de verzameling $\{a,b,c,d\}$ d.i. $C_4^2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 6$. Het aantal termen van de coëfficiënt $abc+abd+acd+bcd$ van x is gelijk aan het aantal deelverzamelingen met drie elementen uit de verzameling $\{a,b,c,d\}$ d.i. $C_4^3 = \begin{pmatrix} 4 \\ 3 \end{pmatrix} = 4$. Stellen we nu $a=b=c=d$, dan verkrijgen we:

$$(x+a)^4 = x^4 + \begin{pmatrix} 4 \\ 1 \end{pmatrix} ax^3 + \begin{pmatrix} 4 \\ 2 \end{pmatrix} a^2x^2 + \begin{pmatrix} 4 \\ 3 \end{pmatrix} a^3x + a^4$$

We kunnen dit nieuw merkwaardig product veralgemenen voor de n-de macht van een tweeterm.

STELLING 1.7 (Het binomium van Newton)

$$\begin{aligned}(x+a)^n &= x^n + \begin{pmatrix} n \\ 1 \end{pmatrix} ax^{n-1} + \begin{pmatrix} n \\ 2 \end{pmatrix} a^2x^{n-2} \\ &\quad + \cdots + \begin{pmatrix} n \\ n-1 \end{pmatrix} a^{n-1}x + \begin{pmatrix} n \\ n \end{pmatrix} a^n\end{aligned}$$

Met verkorte notatie:

$$(x+a)^n = \sum_{k=0}^{n} \binom{n}{k} a^k x^{n-k}$$

We noemen dit merkwaardig product, **het binomium van Newton**. De coëfficiënten $\binom{n}{k}$ noemen we de **binomiaalcoëfficiënten**

Bewijs: We voeren het bewijs door volledige inductie

1. De formule is geldig voor $n = 1$.

2. Als de formule geldt voor n dan geldt ze ook voor $n+1$:

$$\begin{aligned}
(x+a)^{n+1} &= (x+a).(x+a)^n = (x+a).\sum_{k=0}^{n} \binom{n}{k} a^k x^{n-k} \\
&= x.\sum_{k=0}^{n} \binom{n}{k} a^k x^{n-k} + a.\sum_{k=0}^{n} \binom{n}{k} a^k x^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} a^k x^{n-k+1} + \sum_{k=0}^{n} \binom{n}{k} a^{k+1} x^{n-k} \\
&= \binom{n}{0} x^{n+1} + \sum_{k=1}^{n} \binom{n}{k} a^k x^{n-k+1} + \sum_{k'=1}^{n+1} \binom{n}{k'-1} a^{k'} x^{n-k'+1} \; (k' = k+1) \\
&= \binom{n}{0} x^{n+1} + \sum_{k=1}^{n} \binom{n}{k} a^k x^{n-k+1} + \sum_{k'=1}^{n} \binom{n}{k'-1} a^{k'} x^{n-k'+1} + \binom{n}{n} a^{n+1} \\
&= \binom{n}{0} x^{n+1} + \sum_{k=1}^{n} (\binom{n}{k} + \binom{n}{k-1}) a^k x^{n-k+1} + \binom{n}{n} a^{n+1} \\
&= \binom{n}{0} x^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} a^k x^{n-k+1} + \binom{n}{n} a^{n+1} \\
&= \binom{n+1}{0} x^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} a^k x^{n-k+1} + \binom{n+1}{n+1} a^{n+1} \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} a^k x^{n+1-k}
\end{aligned}$$

Opmerkingen:

* De $(m+1)$de term in de ontwikkeling ($m \in \mathrm{N}$) is

$$T_{m+1} = \binom{n}{m} a^m x^{n-m}$$

* De ontwikkeling

$$(x+a)^n = \sum_{k=0}^{n} \binom{n}{k} a^k x^{n-k}$$

heeft $n+1$ termen. Is n even dan heeft de ontwikkeling een middelste term, nl. $T_{\frac{n}{2}+1}$. Is n oneven dan is er geen middelste term.

* Het binomium van Newton kunnen we gebruiken om, samen met de formule van de Moivre voor de macht van een complex getal, formules van goniometrie op te stellen zoals voor $\cos 3x$, $\sin 3x$, $\cos 4x$, $\sin 4x$, enz.

We stellen de formules op voor $\cos 4x$ en $\sin 4x$:
Met de formule van de Moivre is

$$(\cos x + i \sin x)^4 = \cos 4x + i \sin 4x \tag{1.3}$$

Met het binomium van Newton is

$$\begin{aligned}(\cos x + i \sin x)^4 &= \cos^4 x + 4\cos^3 x.i.\sin x + 6\cos^2 x.i^2 \sin^2 x \\ &\quad +4\cos x.i^3 \sin^3 x + i^4 \sin^4 x \\ &= \cos^4 x - 6\cos^2 x.\sin^2 x + \sin^4 x \\ &\quad +i4(\cos^3 x.\sin x - \cos x.\sin^3 x)\end{aligned} \tag{1.4}$$

Door de definitie van gelijkheid van twee complexe getallen volgt uit 1.3 en 1.4 dat

$$\cos 4x = \cos^4 x - 6\cos^2 x.\sin^2 x + sin^4 x$$

en

$$\sin 4x = 4(\cos^3 x.\sin x - \cos x.\sin^3 x)$$

OPGAVEN — 75 Werk uit:

1. $(\frac{a}{2} + 2b)^5$
2. $(\frac{a}{3} + 3)^7$
3. $(2a^2 - \frac{3}{a^3})^5$
4. $(x^{2/3} + y^{3/5})^4$
5. $(3+\sqrt{2})^4 - (3-\sqrt{2})^4$
6. $(a+\sqrt{a^2-1})^6 + (a-\sqrt{a^2-1})^6$

76 Bereken door het binomium van Newton toe te passen a. $(1,01)^4$; b. $(0,95)^5$; c. $(x+iy)^7$; d. $(a+b+c)^3$.

77 Bereken de term in y^{12} in de ontwikkeling van $(y^2-2y)^{10}$.

78 Bereken de 58-ste term in de ontwikkeling van $(x+2a)^{100}$ en de middelste term in de ontwikkeling van $(1+3x)^8$.

79 Schrijf 3^n als een lineaire combinatie van de binomiaalcoëfficiënten.

80 * Bewijs dat:

a. $\begin{pmatrix} n+m \\ p \end{pmatrix} = \sum_{k=0}^{p} \begin{pmatrix} n \\ p-k \end{pmatrix} . \begin{pmatrix} m \\ k \end{pmatrix}$;

b. $\sum_{k=0}^{p}(-1)^k \begin{pmatrix} n \\ k \end{pmatrix} . \begin{pmatrix} n \\ p-k \end{pmatrix} = 0$ als p oneven is;

c. $n.2^{n-1} = \sum_{k=1}^{n} k. \begin{pmatrix} n \\ k \end{pmatrix}$.

81 Aan welke voorwaarden moeten n en p voldoen opdat

a. $\begin{pmatrix} n \\ p \end{pmatrix} < \begin{pmatrix} n+1 \\ p \end{pmatrix}$?

b. $\begin{pmatrix} n \\ p \end{pmatrix} < \begin{pmatrix} n \\ p+1 \end{pmatrix}$?

82 Bereken in de ontwikkeling van $(1+a)^n$ met a een strikt positief getal het quotiënt van een term met zijn voorgaande. Voor welke twee termen is dit quotiënt groter dan 1?

83 Bereken de grootste term van de ontwikkeling $(1+\pi)^{1000}$.

OPLOSSINGEN: 77 $k=8$, $11520y^{12}$
81 (a) $n \geq p > 0$; (b) $n-2p > 1$;
78 $5,49.10^{45}x^{43}a^{57}$; $5670x^4$;
82 $\frac{n-k}{k+1}a$;
83 $\begin{pmatrix} 1000 \\ 759 \end{pmatrix} .\pi^{759}$.

HUISTAAK K1

1. Als $C_n^p = C_n^{p+1}$ en $4C_n^p = 5C_n^{p-1}$ bepaal dan de waarden van n en p.
2. Voor welke waarde van n is de zevende term in de ontwikkeling

$$(\sqrt[3]{x^7} - \frac{1}{2\sqrt{x}})^n$$

   van de vierde graad in x. Bepaal deze zevende term.
3. We vormen viertallen van vier verschillende letters met de letters a, b, c, d, α, β en γ.
   a. Hoeveel dergelijke viertallen zijn er mogelijk?
   b. Hoeveel viertallen bevatten precies twee Griekse letters?
   c. Hoeveel viertallen beginnen met a en bevatten precies twee Griekse letters?
4. Voor een klas van 18 leerlingen worden drie supplementaire cursussen ingericht, nl. technologie, estetica en Duits. Bereken voor deze klas het totaal aantal verschillende keuzemogelijkheden als elke leerling
   a. 1 supplementaire cursus moet volgen;
   b. 2 supplementaire cursussen moet volgen;
   c. ten hoogste 1 supplementaire cursus moet volgen;
   d. ten minste 1 supplementaire cursus moet volgen.
5. Op hoeveel manieren kunnen 4 meisjes en 4 jongens op een rij zitten als
   a. ze afwisselend m-j of j-m moeten zitten?
   b. 1 bepaald meisje naast 1 bepaalde jongen moet zitten?
   c. 1 bepaald meisje niet naast 1 bepaalde jongen wil zitten?

   Beantwoord dezelfde vragen als de personen plaats nemen aan een ronde tafel.
6. Het morse-alfabet bevat twee tekens: punt en streep. Hoeveel tekens moet je minstens achter elkaar gebruiken om elke letter van het alfabet voor te kunnen stellen?
7. Dertig zesdejaars van Nieuwen-Bosch gaan op Romereis. In het hotel waar ze logeren worden hen 3 identieke tweepersoonskamers, 4 identieke driepersoonskamers en 3 identieke vierpersoonskamers aangeboden. Op hoeveel manieren kunnen die leerlingen een kamer toegewezen krijgen?

1.4 Wiskunde-Cultuur

STIFEL Michaël was een Duits wiskundige en een Lutherse predikant van 1487 tot 1567. In zijn hoofdwerk "Arithmetica integra" (1544) geeft hij een behandeling van negatieve en irrationale getallen, een vereenvoudigde leer van de vierkantsvergelijkingen en een uitvoerige driehoek van PASCAL (1623-1662); echter niet zoals in CARDAMUS' "Art magna" (1545), de oplossing van de derde- en vierdegraadsvergelijkingen. Hij was één van de eersten die de tekens + en - gebruikte en een onbekende door een letter aanduidde. Hij geloofde in de "gematria" en concludeerde o.m. dat het getal 666 van het beest in de openbaring van Johannes betrekking had op de paus Leo X. "Gematria " komt van het Hebreeuws en betekent "berekening", de berekening van de getalwaarde van Hebreeuwse woorden en het trekken van conclusies uit deze berekeningen. Volgens sommige geleerden bevatten de Hebreeuwse woorden een sleutel voor het dieper verstaan van de bijbel.

Hoofdstuk 2

Wiskundige kansberekening

2.1 Basisbegrippen van de kanstheorie

Het universum is de verzameling van alle mogelijke uitkomsten van het onderzoek of experiment. We duiden dit aan door $\Omega = \{\omega_1, \omega_2 \ldots, \}$. Het universum kan eindig zijn, aftelbaar oneindig of niet aftelbaar oneindig. Is het universum eindig dan is $|\Omega| = N$ met N een natuurlijk getal.

Een gebeurtenis is een deelverzameling van het universum.

> Een leerling kijkt tussen de 2 en de 3 uur TV per dag is de gebeurtenis $]2,3[\subset [0,12]$
>
> Bij het opgooien van één dobbelsteen is het gooien van een even aantal ogen een gebeurtenis, nl. $\{2,4,6\} \subset \{1,2,3,4,5,6\}$;
>
> Een gezin van twee kinderen dat bestaat uit één jongen en één meisje is de gebeurtenis. $\{(\mathrm{m},\mathrm{j}),(\mathrm{j},\mathrm{m})\} \subset \{(\mathrm{m},\mathrm{m}),(\mathrm{m},\mathrm{j}),(\mathrm{j},\mathrm{m}),(\mathrm{j},\mathrm{j})\}$

Een gunstige uitkomst voor een gebeurtenis is een element van de gebeurtenis.

De zekere gebeurtenis is het universum zelf.

De onmogelijke gebeurtenis is de ledige verzameling.

Een enkelvoudige gebeurtenis is een deelverzameling met één element van het universum (singletons van de uitkomsten).

2.2 Het begrip kans

We beschouwen een aftelbaar universum $\Omega = \{\omega_1, \omega_2, \ldots\}$.

Een **kansfunctie** P is een afbeelding die een gebeurtenis afbeeldt op een reëel getal en die voldoet aan de hierna volgende axioma's.
De **kans op een gebeurtenis** A is het beeld van de gebeurtenis A voor de kansfunctie P.
We noteren $\mathbf{P}(A)$.

(P1.) De kans $\mathbf{P}(A)$ is een positief reëel getal. Met symbolen: $\forall A \subset \Omega : \mathbf{P}(A) \geq 0$;

(P2.) De kans op de zekere gebeurtenis is gelijk aan 1. Met symbolen: $\mathbf{P}(\Omega) = 1$;

(P3.) De kans op de unie van twee disjuncte gebeurtenissen A en B is gelijk aan de som van de kans op A en de kans op B.
Met symbolen: $\forall A \subset \Omega, B \subset \Omega : A \cap B = \emptyset \Longrightarrow \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.

Gevolg van de axioma's:
Enkelvoudige gebeurtenissen zijn disjuncte gebeurtenissen.
Elke gebeurtenis is de unie van enkelvoudige gebeurtenissen.
Volgens axioma (P3) is de kans op een gebeurtenis gelijk aan de som van de kansen op de enkelvoudige gebeurtenissen $\{\omega_i\}$ waarvan ze de unie is.

$$\mathbf{P}(A) = \sum_{\omega_i \in A} \mathbf{P}(\omega_i)$$

De kansfunctie is dus volledig bepaald door de kansen op alle enkelvoudige gebeurtenissen.
In het bijzonder is de som van de kansen op alle enkelvoudige gebeurtenissen gelijk aan 1.

$$\mathbf{P}(\Omega) = \sum_{\omega_i \in \Omega} \mathbf{P}(\omega_i) = 1$$

Voorbeelden:

a. We beschouwen het experiment "Opgooien van een duimspijker" met $\Omega = \{\text{op}, \text{neer}\}$.
De kansen op de enkelvoudige gebeurtenissen zijn bvb. $\mathbf{P}(\text{op}) = \frac{2}{3}$ en $\mathbf{P}(\text{neer}) = \frac{1}{3}$.

b. Drie renners A, B en C nemen deel aan een wedstrijd op de piste. De kansen van A en B worden gelijk geacht en zij hebben dubbel zoveel kans als C.
Bereken de kans dat

   (i) A wint;

(ii) C wint;

(iii) B of C wint.

Oplossing:
$\mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) = 1$ en $\mathbf{P}(A) = \mathbf{P}(B) = 2\mathbf{P}(C)$ Uit deze twee betrekkingen volgt dat

$$4\mathbf{P}(C) + \mathbf{P}(C) = 1 \Longleftrightarrow \mathbf{P}(C) = \frac{1}{5}$$

$$\mathbf{P}(A) = \frac{2}{5}$$

en

$$\mathbf{P}(B \cup C) = \mathbf{P}(B) + \mathbf{P}(C) = \frac{3}{5}.$$

2.3 Kanswetten

2.3.1 Kansregel van Laplace

In een *eindig universum waar alle enkelvoudige gebeurtenissen dezelfde kans hebben* geldt:

$$\mathbf{P}(\omega_1) = \mathbf{P}(\omega_2) = \cdots \mathbf{P}(\omega_N) \text{ met } |\Omega| = N \tag{2.1}$$

$$\mathbf{P}(\Omega) = \sum_{1}^{N} \mathbf{P}(\omega_i) = 1 \tag{2.2}$$

Uit 2.1 en 2.2 volgt

$$\forall \omega_i \in \Omega : \mathbf{P}(\omega_i) = \frac{1}{N}$$

$$\mathbf{P}(A) = \sum_{\omega_i \in A} \mathbf{P}(\omega_i) = |A| \cdot \frac{1}{N}$$

De kansregel van Laplace:
De kans op een gebeurtenis A is gelijk aan het aantal gunstige uitkomsten voor de gebeurtenis gedeeld door het aantal even waarschijnlijke uitkomsten van het universum.

Met symbolen: $\forall A \subset (\Omega) : \mathbf{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{N}$

Voorbeelden:

a. Beschouwen we het experiment “Opgooien van een muntstuk” met uitkomsten k en m. In de veronderstelling dat de munt volledig symmetrisch is kunnen we de kans op munt gelijk stellen aan de kans op kruis.

$$\mathbf{P}(m) = \mathbf{P}(k)$$

Ook geldt er

$$\mathbf{P}(m) + \mathbf{P}(k) = 1$$

Uit de laatste betrekkingen volgt dat

$$\mathbf{P}(m) = \mathbf{P}(k) = \frac{1}{2}.$$

b. Kennen we insgelijks aan de zes enkelvoudige gebeurtenissen bij het werpen van een dobbelsteen axiomatisch gelijke kansen toe dan is

$$\forall \omega_i \in \Omega : \mathbf{P}(\omega_i) = \frac{1}{6}$$

We beschouwen bij dit experiment de gebeurtenis A: “Gooien van een even getal”.

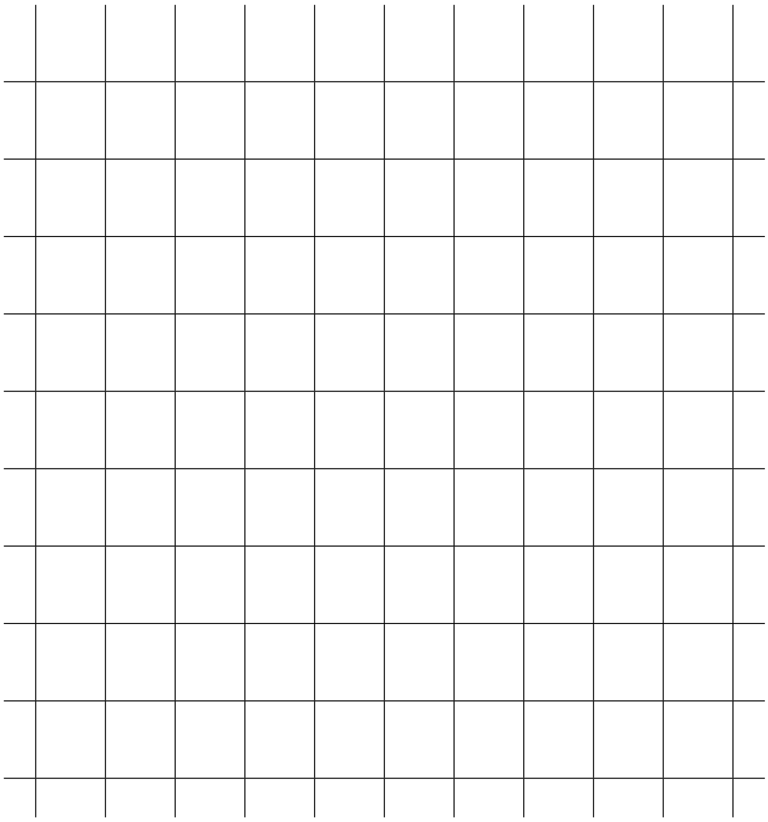
$$\mathbf{P}(A) = \frac{|A|}{N} = \frac{3}{6} = \frac{1}{2}.$$

c. We beschouwen het experiment “twee dobbelstenen opgooien” met uitkomstenverzameling $\Omega = V \times V$ met $V = \{1, 2, 3, 4, 5, 6\}$; $N = 36$. Onderstellen we dat bij beide dobbelstenen de kansen op de zes enkelvoudige gebeurtenissen gelijk zijn aan elkaar, dan zijn de kansen op de 36 mogelijke uitkomsten bij het samengesteld experiment tevens gelijk aan elkaar.

De kans op elke enkelvoudige gebeurtenis is gelijk aan $\frac{1}{36}$.

We beschouwen de gebeurtenissen:
A: “Minstens één van de dobbelstenen bevat een 6”;
B: “De som van de ogen is minstens gelijk aan 9”;
C: “De som van de ogen is gelijk aan 3”;
D: “Het aantal ogen op de tweede dobbelsteen is kleiner dan het aantal ogen op de eerste dobbelsteen”.

De kans op een gebeurtenis is gelijk aan de som van de kansen van de enkelvoudige gebeurtenissen waarvan ze de unie is.
Als we de kansregel van Laplace toepassen, is het voldoende het aantal gunstige



Tabel 2.1: roosterdiagram voor het gooien van twee dobbelstenen

uitkomsten voor de gebeurtenis te bepalen.
$|A| = 11$; $|B| = 10$; $|C| = 2$; $|D| = 15$,

$$\mathbf{P}(A) = \frac{11}{36};$$

$$\mathbf{P}(B) = \frac{10}{36} = \frac{5}{18};$$

$$\mathbf{P}(C) = \frac{2}{36} = \frac{1}{18};$$

$$\mathbf{P}(D) = \frac{15}{36} = \frac{5}{12}.$$

d. Het experiment "een balletje nemen uit een vaas met vier witte en vijf rode balletjes" heeft als universum

$$\Omega = \{w_1, w_2, w_3, w_4, r_1, r_2, r_3, r_4, r_5\}; N = 9.$$

Een gebeurtenis R is: “Een rood balletje trekken”. Het aantal gunstige uitkomsten voor R is gelijk aan 5.

$$\mathbf{P}(R) = \frac{5}{9}.$$

e. Het experiment “drie kaarten trekken uit een spel van tweeënvijftig kaarten” heeft als universum Ω waarvan de elementen alle combinaties zijn van 52 elementen in groepjes van 3.

$$N = \begin{pmatrix} 52 \\ 3 \end{pmatrix} = 22100.$$

Een gebeurtenis A is: “Drie azen trekken”. Het aantal gunstige uitkomsten voor A is gelijk aan $\begin{pmatrix} 4 \\ 3 \end{pmatrix} = 4$.

$$P(A) = \frac{4}{22100} = 0,00018.$$

f. Het experiment “van twee personen de verjaardag raden” heeft als uitkomstenverzameling Ω waarvan de elementen herhalingsvariaties zijn van 365 elementen in groepjes van 2.

$$N = 365^2.$$

Een gebeurtenis A is: “Twee personen verjaren op een verschillende dag”. Het aantal gunstige uitkomsten voor A is 365.364.

$$P(A) = \frac{365.364}{365^2} = \frac{364}{365} = 0,997.$$

2.3.2 De complementregel

STELLING 2.1 *De kans van het complement A^c in Ω van een gebeurtenis A is gelijk aan 1 min de kans van A.*

Met symbolen: $\forall A \subset \Omega : A^c = \Omega \setminus A \Longrightarrow \mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.

Voorbeeld: 50 personen bevinden zich in een zaal voor de overgang van 2003 naar 2004. Bereken de kans dat er minstens twee personen aanwezig zijn met dezelfde verjaardag.
Bij het experiment “de verjaardagen van 50 personen geven” is $N = 366^{50}$. De gebeurtenis A: “Minstens twee personen hebben dezelfde verjaardag” is het complement van de gebeurtenis A^c: “Alle personen hebben een verschillende verjaardag”. De kans op A^c is gemakkelijker te bepalen dan de kans op A. Het aantal gunstige gevallen voor A^c is V_{366}^{50}. De kans van de gevraagde gebeurtenis A is

$$\mathbf{P}(A) = 1 - \frac{V_{366}^{50}}{366^{50}} = 0,97.$$

De kans is vrij hoog om 2 mensen te hebben met dezelfde verjaardag in een groep van 50 mensen. Zo kan men gemakkelijk berekenen dat de kans nog 56,8% is in een groep van 25 personen, 25,2% in een groep van 15 personen en 7,4% in een groep van 8 personen.

GEVOLG 2.1 *De kans van de onmogelijke gebeurtenis is gelijk aan 0.*

Met symbolen: $\mathbf{P}(\emptyset) = 0$.

STELLING 2.2 *De kans op het verschil van twee gebeurtenissen A en B is gelijk aan de kans op A min de kans op de doorsnede van A en B.*

Met symbolen: $\forall A \subset \Omega, B \subset \Omega : \mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$.

STELLING 2.3 *De kans op A die een deelverzameling is van de gebeurtenis B is kleiner dan of gelijk aan de kans op B.*

Met symbolen: $\forall A \subset \Omega, B \subset \Omega : A \subset B \Longrightarrow \mathbf{P}(A) \leq P(B)$.

STELLING 2.4 *De kans op een gebeurtenis is steeds kleiner dan of gelijk aan 1.*

Met symbolen: $\forall A \subset \Omega : \mathbf{P}(A) \leq 1$.

2.3.3 De somregel

STELLING 2.5 *De kans op de unie van twee gebeurtenissen A en B is gelijk aan de som van de kansen op A en B min de kans op de doorsnede van A en B.*

Met symbolen: $\forall A \subset \Omega, B \subset \Omega : \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.
Een bijzonder geval van deze stelling is het axioma (P3):

$$A \text{ en } B \text{ zijn disjuncte gebeurtenissen} \Longleftrightarrow$$

$$\mathbf{P}(A \cap B) = 0 \Longleftrightarrow \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

Voorbeeld: Uit een spel van tweeëndertig kaarten wordt aselect een kaart getrokken. Bereken de kans op de volgende gebeurtenissen:
“Een boer of een schoppen trekken”;
“Een 7 of een 9 trekken”.

Oplossing:
Bij het experiment "een kaart trekken uit een spel van tweeëndertig kaarten" is $N = 32$. De uitkomsten zijn even waarschijnlijk. Noem de gebeurtenis A: "Een boer trekken" en B: "Een schoppen trekken".
Het aantal gunstige gevallen voor A is 4 en voor B is 8.
$A \cap B$: "Een schoppen boer trekken". Het aantal gunstige gevallen voor deze doorsnede is 1.

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) = \frac{4}{32} + \frac{8}{32} - \frac{1}{32} = \frac{11}{32}.$$

Noem C "een 7 trekken" en D: "Een 9 trekken". Het aantal gunstige gevallen voor C is 4 en voor D is 4. De gebeurtenissen C en D zijn disjunct.

$$\mathbf{P}(C \cup D) = \mathbf{P}(C) + \mathbf{P}(D) = \frac{4}{32} + \frac{4}{32} = \frac{1}{4}.$$

STELLING 2.6 *De kans op de unie van drie gebeurtenissen A, B en C is gelijk aan de som van de kansen op A, B, C en $A \cap B \cap C$ min de kansen op $A \cap B$, $B \cap C$ en $A \cap C$.*

Met symbolen: $\forall A \subset \Omega, B \subset \Omega, C \subset \Omega$:

$$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) + \mathbf{P}(A \cap B \cap C) - \mathbf{P}(A \cap B) - \mathbf{P}(B \cap C) - \mathbf{P}(A \cap C).$$

Bewijs deze stelling. *Voorbeeld:* In een hogere technische school moeten elk van de 500 studenten minstens 1 van de kursussen algebra, fysica of statistiek volgen.

1. 320 studenten volgen de kursus algebra;
2. 190 studenten volgen de kursus fysica;
3. 280 studenten volgen de kursus statistiek;
4. 80 studenten volgen de algebra en fysica;
5. 200 studenten volgen de algebra en statistiek;
6. 60 studenten volgen de fysica en statistiek.

Bereken de kans dat een willekeurige gekozen student

1. de drie kursussen volgt;

2. algebra volgt maar geen statistiek;
3. statistiek volgt en geen fysica;
4. algebra of statistiek volgt, maar geen fysica;
5. algebra volgt, maar geen fysica en geen statistiek.

Oplossing: De uitkomstenverzameling van dit experiment "Een student kiezen uit 500 studenten" heeft 500 elementen. De kans op een enkelvoudige gebeurtenis is 1/500.
De drie gebeurtenissen zijn: A: 'een student kiest algebra', F: 'een student kiest fysica' en S: 'een student kiest statistiek'

1. $\mathbf{P}(A \cap F \cap S) = 1 - \frac{320}{500} - \frac{190}{500} - \frac{250}{500} + \frac{80}{500} + \frac{200}{500} + \frac{60}{500} = \frac{50}{500} = 0,01;$
2. $\mathbf{P}(A \setminus S) = \frac{320}{500} - \frac{200}{500} = 0,24;$
3. $\mathbf{P}(S \setminus F) = \frac{280}{500} - \frac{60}{500} = 0,44;$
4. $$\begin{aligned} \mathbf{P}((A \cup S) \setminus F) &= \mathbf{P}(A \cup S) - \mathbf{P}((A \cup S) \cap F) \\ &= \mathbf{P}(A) + \mathbf{P}(S) - \mathbf{P}(A \cap S) - \mathbf{P}((A \cap F) \cup \mathbf{P}(S \cap F)) \\ &= \mathbf{P}(A) + \mathbf{P}(S) - \mathbf{P}(A \cap S) - \mathbf{P}(A \cap F) - \mathbf{P}(S \cap F)) + \mathbf{P}(A \cap S \cap F) \\ &= \frac{320}{500} + \frac{280}{500} - \frac{200}{500} - \frac{80}{500} - \frac{60}{500} + \frac{50}{500} \\ &= 0,62 \end{aligned}$$
5. $$\begin{aligned} \mathbf{P}((A \setminus (S \cup F) &= \mathbf{P}(A) - \mathbf{P}(A \cap (S \cup F)) \\ &= \mathbf{P}(A) - \mathbf{P}((A \cap S) \cup \mathbf{P}(A \cap F)) \\ &= \mathbf{P}(A) - \mathbf{P}(A \cap S) - \mathbf{P}(A \cap F)) + \mathbf{P}(A \cap F \cap S) \\ &= \frac{320}{500} - \frac{200}{500} - \frac{80}{500} + \frac{50}{500} \\ &= 0,12 \end{aligned}$$

OPGAVEN — **84** Aan 80 mensen wordt gevraagd of ze een televisietoestel, een computer of een gsm bezitten. 72 hebben een televisietoestel, 45 een computer en 28 een gsm. 22 hebben een televisietoestel en een gsm, 39 bezitten een televisietoestel en een computer en 16 hebben een computer en een gsm. 10 personen hebben zowel een televisietoestel, een computer als een gsm. Wat is de kans dat één van die mensen

Figuur 2.1: venndiagram voor het volgen van 3 cursussen

1. geen van de drie toestellen heeft;
2. enkel een gsm bezitten;
3. een televisie en een computer maar geen gsm;
4. een televisie of een computer hebben.

Oplossing: 1. 0,025; 2. 0; 3. 0,6125; 4. 0,975.

2.3.4 Voorwaardelijke kans

De **voorwaardelijke kans op een gebeurtenis** A **gegeven dat de gebeurtenis** B **is opgetreden** is de kans op de doorsnede van A en B gedeeld door de kans op B.
Met symbolen:

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

We zeggen kort: de kans op A op voorwaarde B.

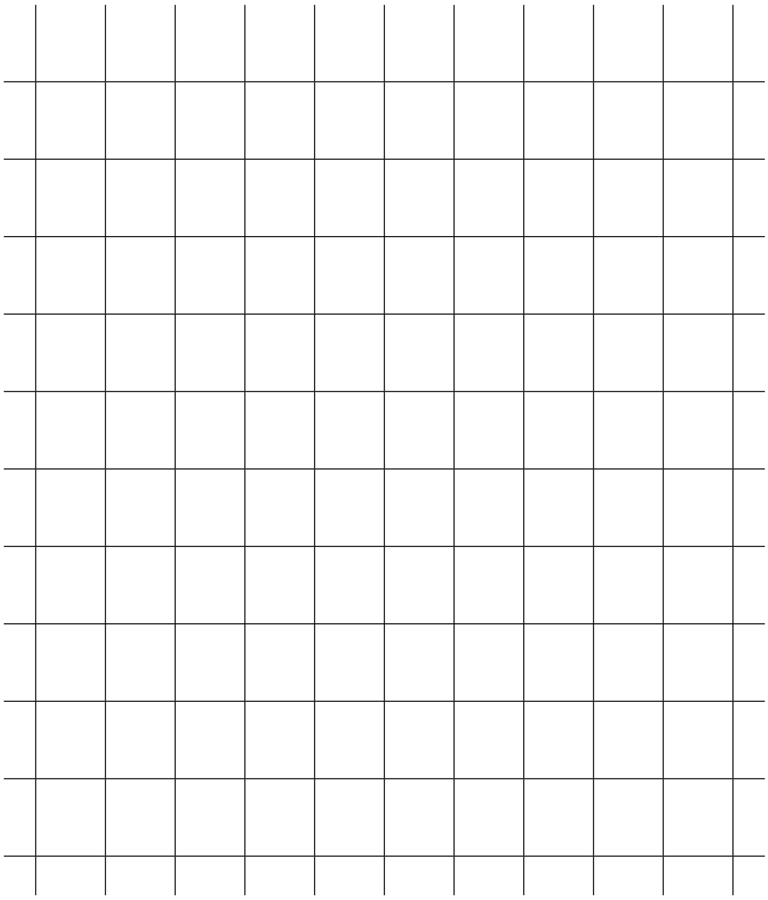
Voorbeelden :
Bij het opgooien van twee dobbelstenen beschouwen we twee gebeurtenissen, nl. A: ‘de som van de ogen groter is dan 5’ en B: ‘het verschil van de ogen gelijk is aan 1’.

We maken een roosterdiagram.

Op het roosterdiagram zien we dat er 6 uitkomsten voldoen aan de twee voorwaarden.

$$\mathbf{P}(A \cap B) = \frac{6}{36}$$

$$\mathbf{P}(A) = \frac{26}{36}$$



Tabel 2.2: roosterdiagram voor het gooien van twee dobbelstenen

Volgens de definitie van voorwaardelijke kans is:

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{6}{26} = 0,231$$

We beschouwen het experiment "twee balletjes nemen uit een vaas met vier witte en vijf rode balletjes zonder terugleggen".
W_1 en R_1 zijn de gebeurtenissen "een wit balletje trekken" resp. "een rood balletje trekken" bij de eerste trekking en W_2 en R_2 zijn de gelijknamige gebeurtenissen maar bij de tweede trekking. De kansen bij de eerste trekking zijn:

$$\mathbf{P}(W_1) = \frac{4}{9} \text{ en } \mathbf{P}(R_1) = \frac{5}{9}.$$

De kansen bij de tweede trekking zijn voorwaardelijke kansen:

$$\mathbf{P}(W_2|W_1) = \frac{3}{8} \text{ en } \mathbf{P}(W_2|R_1) = \frac{4}{8}.$$

$$\mathbf{P}(R_2|W_1) = \frac{5}{8} \text{ en } \mathbf{P}(R_2|R_1) = \frac{4}{8}.$$

We duiden deze kansen aan op het volgend boomdiagram:

•

W_1		R_1	
$W_2\|W_1$	$R_2\|W_1$	$W_2\|R_1$	$R_2\|R_1$
$W_1 \cap W_2$	$W_1 \cap R_2$	$R_1 \cap W_2$	$R_1 \cap R_2$
$\mathbf{P}(W_1 \cap W_2) = \mathbf{P}(W_1)\mathbf{P}(W_2\|W_1)$	$\mathbf{P}(W_1 \cap R_2) = \mathbf{P}(W_1)\mathbf{P}(R_2\|W_1)$	$\mathbf{P}(R_1 \cap W_2) = \mathbf{P}(R_1)\mathbf{P}(W_2\|R_1)$	$\mathbf{P}(R_1 \cap R_2) = \mathbf{P}(R_1)\mathbf{P}(R_2\|R_1)$

OPGAVEN — 85 Bewijs dat de complementregel ook geldig is voor de voorwaardelijke kans.

$$P(A^c|B) = 1 - P((A|B)$$

86 Bereken de kans dat bij het opgooien van twee dobbelstenen we een 6 gooien op voorwaarde dat de som van de ogen 9 is. (opl:1/2)

2.3.5 De productregel

Uit de definitie van voorwaardelijke kans volgt de productregel (zie het boomdiagram).

STELLING 2.7 *De kans op de doorsnede van twee gebeurtenissen A en B is gelijk aan het product van de kans op A en de kans op B op voorwaarde A.*

Met symbolen: $\forall A \subset \Omega, B \subset \Omega : \mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B|A)$

OPGAVEN — 87 Bewijs de productregel op voor drie gebeurtenissen

$$\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A) \cdot \mathbf{P}(B|A) \cdot \mathbf{P}(C|A \cap B)$$

2.3.6 Onafhankelijke gebeurtenissen

Twee gebeurtenissen A en B zijn onafhankelijke gebeurtenissen als en slechts als de kans op B op voorwaarde A gelijk is aan de kans op B, m.a.w. de kans op de doorsnede van A en B gelijk is aan het product van de kansen op A en B.

Met symbolen: A en B zijn onafhankelijke gebeurtenissen $\Longleftrightarrow$

$$\mathbf{P}(B|A) = \mathbf{P}(B) \Longleftrightarrow \mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B).$$

Hier is $\mathbf{P}(B|A) = \mathbf{P}(B)$. Dit betekent dat het optreden van B niet afhangt van het al dan niet optreden van A. Ook hangt het optreden van A niet af van het optreden van B. Inderdaad, uit $\mathbf{P}(B \cap A) = \mathbf{P}(A).\mathbf{P}(B|A) = \mathbf{P}(B).\mathbf{P}(A|B)$ en $\mathbf{P}(A).\mathbf{P}(B|A) = \mathbf{P}(A).\mathbf{P}(B)$ volgt dan $\mathbf{P}(B).\mathbf{P}(A|B) = \mathbf{P}(A).\mathbf{P}(B)$. Uit dit laatste kunnen we besluiten dat $\mathbf{P}(A) = \mathbf{P}(A|B)$ en dit betekent dat het optreden van A niet afhangt van het optreden van B.

OPGAVEN — **88** Zijn A en B twee onafhankelijke gebeurtenissen dan zijn de gebeurtenissen A en B^c en A^c en B^c eveneens onafhankelijk.

Drie gebeurtenissen A, B en C zijn onafhankelijke gebeurtenissen als en slechts als ze twee aan twee onafhankelijk zijn en de kans van de doorsnede van A, B en C gelijk is aan het product van de kansen van A, B en C.

Met symbolen: A, B en C zijn onafhankelijke gebeurtenissen als en slechts als

$$\mathbf{P}(A \cap B) = \mathbf{P}(A).\mathbf{P}(B);$$

$$\mathbf{P}(B \cap C) = \mathbf{P}(B).\mathbf{P}(C);$$

$$\mathbf{P}(C \cap A) = \mathbf{P}(C).\mathbf{P}(A);$$

en

$$\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A).\mathbf{P}(B).\mathbf{P}(C);$$

Voorbeelden:

- Wat is de kans bij "opgooien van twee dobbelstenen, een rode en een groene" dat we en even aantal ogen gooien met de rode dobbelsteen en een 1 of een 2 met de groene dobbelsteen?
  A: "gooien van een even aantal ogen met de rode dobbelsteen";
  B: "gooien van 1 of 2 met de groene dobbelsteen".
  De gebeurtenissen A en B zijn onafhankelijk:
  $\mathbf{P}(A \cap B) = \mathbf{P}(A).\mathbf{P}(B) = \frac{1}{2}\frac{1}{3} = \frac{1}{6}$.

- We beschouwen bij het experiment "opgooien van twee muntstukken", de gebeurtenissen A: "kop bij de eerste worp" en B: "gooien van juist één keer munt".
  De kansen zijn als volgt:
  $\mathbf{P}(A) = \frac{1}{2}$, $\mathbf{P}(B) = \frac{1}{2}$ en $\mathbf{P}(A \cap B) = \frac{1}{4} = \mathbf{P}(A).\mathbf{P}(B)$.
  Als het muntsuk nu 3 keer wordt opgegooid dan zijn de kansen $\mathbf{P}(A) = \frac{1}{2}$, $\mathbf{P}(B) = \frac{3}{8}$ en $\mathbf{P}(A \cap B) = \frac{1}{4} \neq \mathbf{P}(A).\mathbf{P}(B)$.
  We zien dus dat in de ene kansruimte twee gebeurtenissen onafhankelijk zijn maar in een andere kansruimte zijn ze afhankelijk.

- We beschouwen het experiment met als uitkomstenverzameling $\Omega = \{x_1, x_2, x_3, x_4, x_5\}$ met als kansverdeling: $\mathbf{P}(x_1) = \frac{2}{6}$, $\mathbf{P}(x_2) = \frac{1}{12}$, $\mathbf{P}(x_3) = \frac{1}{6}$, $\mathbf{P}(x_4) = \frac{2}{6}$ en $\mathbf{P}(x_5) = \frac{1}{12}$.
  We beschouwen de gebeurtenissen $A = \{x_3, x_4\}$ en $B = \{x_1, x_4\}$.
  $A \cap B = \{x_4\}$ en A en B zijn onafhankelijk omdat $\mathbf{P}(A).\mathbf{P}(B) = \frac{1}{2}.\frac{2}{3} = \frac{1}{3} = \mathbf{P}(A \cap B)$.
  Indien de uitkomsten even waarschijnlijk zouden zijn, dan zijn de gebeurtenissen A en B niet meer onafhankelijk. Inderdaad, $\mathbf{P}(A) = \mathbf{P}(B) = \frac{2}{5}$ en $\mathbf{P}(A).\mathbf{P}(B) = \frac{4}{25} \neq \frac{1}{5} = \mathbf{P}(A \cap B)$.

- Beschouwen we het experiment met uitkomstenverzameling $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, waarvan de uitkomsten even waarschijnlijk zijn en de gebeurtenissen $A = \{\omega_1, \omega_2\}$, $B = \{\omega_1, \omega_3\}$ en $C = \{\omega_1, \omega_4\}$.
$$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{2}$$
$$\mathbf{P}(A \cap B) = \frac{1}{4} = \mathbf{P}(A).\mathbf{P}(B);$$
$$\mathbf{P}(B \cap C) = \frac{1}{4} = \mathbf{P}(B).\mathbf{P}(C);$$
$$\mathbf{P}(C \cap A) = \frac{1}{4} = \mathbf{P}(C).\mathbf{P}(A);$$
  en
$$\mathbf{P}(A \cap B \cap C) = \frac{1}{4} \neq \mathbf{P}(A).\mathbf{P}(B).\mathbf{P}(C) = \frac{1}{8}.$$
  Dit voorbeeld toont aan dat niettegenstaande de drie gebeurtenissen twee aan twee onafhankelijk zijn de kans op de doorsnede van de drie gebeurtenissen niet gelijk is aan het product van de kansen op de drie gebeurtenissen apart.

- Beschouwen we het experiment met uitkomstenverzameling
$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\},$$
  waarvan de uitkomsten even waarschijnlijk zijn en de gebeurtenissen $A = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $B = \{\omega_2, \omega_3, \omega_4, \omega_5\}$ en $C = \{\omega_2, \omega_6, \omega_7, \omega_8\}$.
$$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{2}$$

$$\mathbf{P}(A \cap B) = \frac{3}{8} \neq \mathbf{P}(A).\mathbf{P}(B) = \frac{1}{4}$$
$$\mathbf{P}(A \cap B \cap C) = \frac{1}{8} = \mathbf{P}(A).\mathbf{P}(B).\mathbf{P}(C)$$

Dit voorbeeld toont aan dat de gebeurtenissen A en B niet onafhankelijk zijn niettegenstaande de kans op de doorsnede van de drie gebeurtenissen gelijk is aan het product van de kansen op de drie gebeurtenissen apart.

2.3.7 De wet der totale kans

STELLING 2.8 *Als $\Omega = A_1 \cup A_2 \cup \cdots A_n$ met $A_i \cap A_j = \phi$ voor alle $i \neq j$, dan geldt voor elke gebeurtenis*

$$\mathbf{P}(B) = \mathbf{P}(B \cap A_1) + \mathbf{P}(B \cap A_2) + \cdots \mathbf{P}(B \cap A_n) = \sum_{i=1}^{n} \mathbf{P}(B \cap A_i).$$

Bewijs:

$$\begin{aligned} B &= B \cap \Omega \\ &= B \cap (A_1 \cup A_2 \cup \cdots A_n) \\ &= (B \cap A_1) \cup (B \cap A_2) \cup \cdots (B \cap A_n) \end{aligned}$$

Omdat $A_1, A_2 \cdots A_n$ paarsgewijze disjunct zijn, geldt dit ook voor de gebeurtenissen $B \cap A_1, B \cap A_2, \cdots B \cap A_n$. Volgens het axioma (P3) geldt:

$$\mathbf{P}(B) = \mathbf{P}((B \cap A_1) \cup (B \cap A_2) \cup \cdots (B \cap A_n)) = \mathbf{P}(B \cap A_1) + \mathbf{P}(B \cap A_2) + \cdots \mathbf{P}(B \cap A_n).$$

Figuur 2.2: de wet der totale kans

Voorbeeld: Drie machines A, B en C produceren resp. 50, 30 en 20 percent van de totale productie. Het percentage van de slechte voortgebrachte stukken bedraagt resp. 3, 4 en 5. Als een stuk willekeurig uit de productie gekozen wordt, welk is dan de kans dat het gekozen stuk slecht is?

Oplossing:
S de gebeurtenis “een slecht stuk nemen”;

A is de gebeurtenis "een stuk nemen van machine A";
B is de gebeurtenis "een stuk nemen van machine B";
C is de gebeurtenis "een stuk nemen van machine C".
De uitkomstenverzameling bestaat uit alle stukken van de productie:
$\Omega = A \cup B \cup C$ met A, B, C paarsgewijze disjunct.
Volgens de wet der totale kans is $\mathbf{P}(S) = \mathbf{P}(S \cap A) + \mathbf{P}(S \cap B) + \mathbf{P}(S \cap C)$

$$\mathbf{P}(S \cap A) = \mathbf{P}(A).\mathbf{P}(S|A) = \frac{50}{100}.\frac{3}{100};$$

$$\mathbf{P}(S \cap B) = \mathbf{P}(B).\mathbf{P}(S|B) = \frac{30}{100}.\frac{4}{100};$$

$$\mathbf{P}(S \cap C) = \mathbf{P}(C).\mathbf{P}(S|C) = \frac{20}{100}.\frac{5}{100};$$

$$\mathbf{P}(S) = \frac{37}{1000} = 0,037 = 3,7\%$$

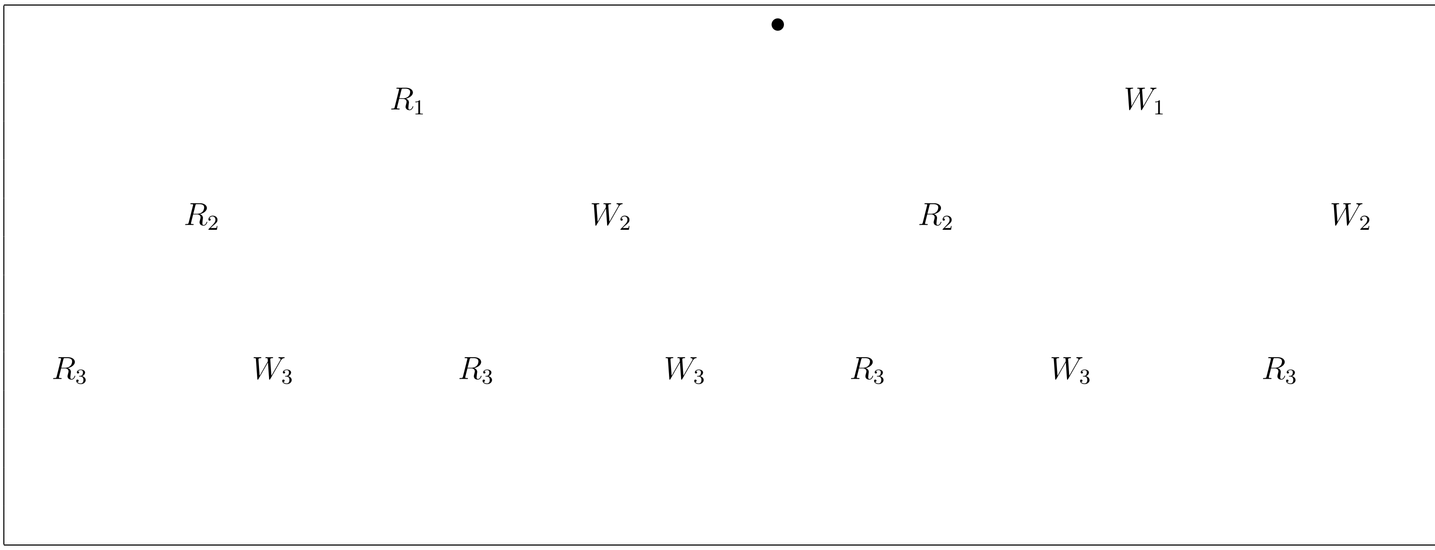
2.3.8 Uitgewerkte oefeningen op de kanswetten

- We trekken 3 ballen uit een doos met 3 rode en 5 witte ballen met terugleggen. Wat is de kans dat we

  a) bij de eerste trekking een rode bal trekken *en* dan twee witte ballen trekken? (A)
  b) juist 1 witte bal trekken? (B)
  c) juist 1 rode bal trekken? (C)
  d) juist 1 witte en juist 1 rode bal trekken? (D)
  e) tenminste 1 witte bal trekken? (E)
  f) tenminste 2 witte ballen trekken? (F)
  g) bij de eerste trekking een rode trekken *of* bij de derde trekking een witte trekken? (G)

  Beantwoord dezelfde vragen maar we trekken drie ballen zonder terugleggen.

  Oplossing: We maken een boomdiagram waarbij we bij elke trekking beschikken over de mogelijke uitkomsten 'een rode bal' met de kans van 3/8 en 'een witte bal' met een kans van 5/8. We voeren de volgende notaties in: w1 betekent witte bal bij de eerste trekking, r3 betekent rode bal bij de derde trekking enz.. . . .



Tabel 2.3: boomdiagram voor het trekken van drie ballen met terugleggen

a) De tweede trekking is onafhankelijk van de eerste en de derde trekking van de tweede trekking.

$$\mathbf{P}(A) = \mathbf{P}(R_1, W_2, W_3) = \mathbf{P}(R_1) \cdot \mathbf{P}(W_2) \cdot \mathbf{P}(W_3) = \frac{3}{8} \cdot (\frac{5}{8})^2 = 0,146$$

b)

$$\begin{aligned} \mathbf{P}(B) &= \mathbf{P}(W_1, R_2, R_3) + \mathbf{P}(R_1, W_2, R_3) + \mathbf{P}(R_1, R_2, W_3) \\ &= \mathbf{P}(W_1) \cdot \mathbf{P}(R_2) \cdot \mathbf{P}(R_3) + \mathbf{P}(R_1) \cdot \mathbf{P}(W_2) \cdot \mathbf{P}(R_3) + P(R_1) \cdot \mathbf{P}(R_2) \cdot \mathbf{P}(w3) \\ &= \frac{5}{8} \cdot (\frac{3}{8})^2 + \frac{5}{8} \cdot (\frac{3}{8})^2 + \frac{5}{8} \cdot (\frac{3}{8})^2 = 0,264 \end{aligned}$$

c)

$$\begin{aligned} \mathbf{P}(C) &= \mathbf{P}(W_1, W_2, R_3) + \mathbf{P}(R_1, W_2, W_3) + \mathbf{P}(W_1, R_2, W_3) \\ &= \mathbf{P}(W_1) \cdot \mathbf{P}(W_2) \cdot \mathbf{P}(R_3) + \mathbf{P}(R_1) \cdot \mathbf{P}(W_2) \cdot \mathbf{P}(W_3) + P(W_1) \cdot \mathbf{P}(R_2) \cdot \mathbf{P}(W_3) \\ &= \frac{3}{8} \cdot (\frac{5}{8})^2 + \frac{3}{8} \cdot (\frac{5}{8})^2 + \frac{3}{8} \cdot (\frac{5}{8})^2 = 0,439 \end{aligned}$$

d) Dit is de onmogelijke gebeurtenis. De kans om juist 1 rode en 1 witte te trekken is bijgevolg gelijk aan nul: $\mathbf{P}(D) = 0$.

e) Tenminste 1 witte bal trekken is de negatie van geen enkele witte bal trekken. Dus de kans op tenminste 1 witte bal is gelijk aan

$$\mathbf{P}(E) = 1 - \mathbf{P}(R_1, R_2, R_3) = 1 - (\frac{3}{8})^3 = 0,947$$

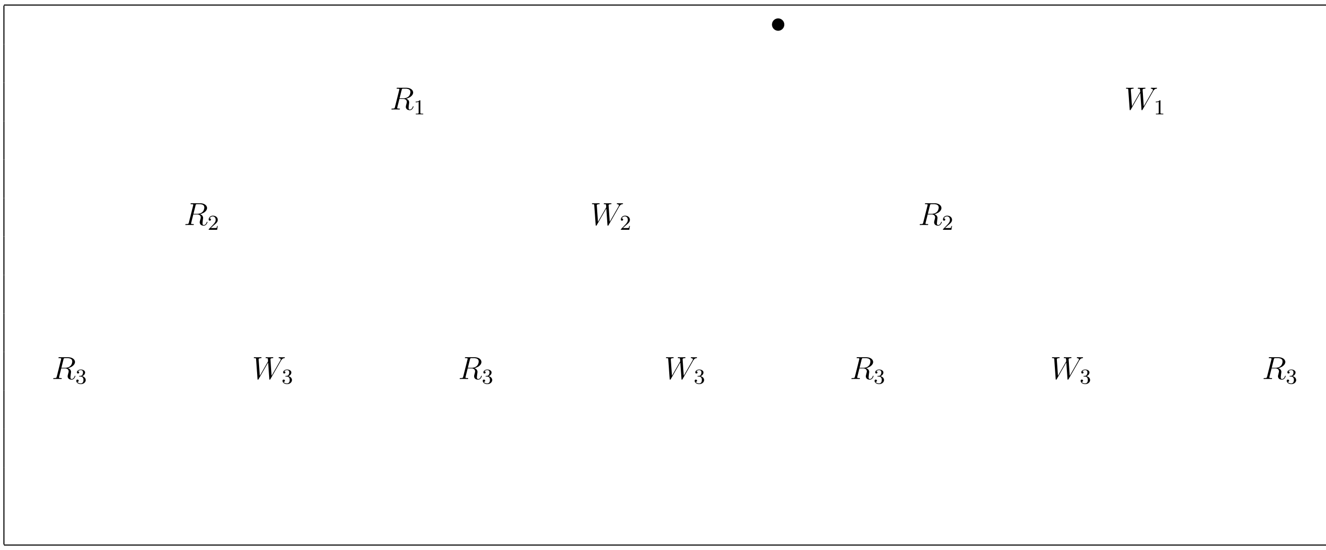
f) Tenminste 2 witte ballen trekken is juist 2 witte ballen trekken (1 rode bal) of juist drie witte ballen trekken.

$$\mathbf{P}(F) = \mathbf{P}(C) + \mathbf{P}(W_1, W_2, W_3) = \frac{3}{8} \cdot (\frac{5}{8})^2 + (\frac{5}{8})^3 = 0,684$$

g)

$$\mathbf{P}((R_1) \cup (W_3)) = \mathbf{P}(R_1) + \mathbf{P}(W_3) - \mathbf{P}((R_1) \cap (W_3))$$
$$\frac{3}{8} + \frac{5}{8} - \frac{3}{8} \cdot \frac{5}{8} = 0,766$$

Voor het trekken van drie ballen zonder terugleggen, maken we een nieuw boomdiagram.



Tabel 2.4: boomdiagram voor het trekken van drie ballen zonder terugleggen

De tweede trekking is nu afhankelijk van de eerste en de derde trekking van de eerste en de tweede trekking.

a)

$$\mathbf{P}(A) = \mathbf{P}(R_1, W_2, W_3) = \mathbf{P}(R_1) \cdot \mathbf{P}(W_2|R_1) \cdot \mathbf{P}(W_3|(R_1 \cap W_2)) = \frac{3}{8} \cdot \frac{5}{7} \cdot \frac{4}{6} = 0,179$$

b)

$$\mathbf{P}(B) = \mathbf{P}(W_1, R_2, R_3) + \mathbf{P}(R_1, W_2, R_3) + \mathbf{P}(R_1, R_2, W_3)$$
$$= \mathbf{P}(W_1) \cdot \mathbf{P}(R_2|W_1) \cdot \mathbf{P}(R_3|(W_1 \cap R_2)) + \mathbf{P}(R_1) \cdot \mathbf{P}(W_2|R_1) \cdot \mathbf{P}(R_3|(W_1 \cap W_2))$$

$$+P(R_1)\cdot\mathbf{P}(R_2|R_1)\cdot\mathbf{P}(W_3|(R_1\cap R_2))$$
$$=\frac{3}{8}\cdot\frac{5}{7}\cdot\frac{2}{6}+\frac{3}{8}\cdot\frac{2}{7}\cdot\frac{5}{6}+\frac{5}{8}\cdot\frac{3}{7}\cdot\frac{2}{6}=0.268$$

c)
$$\mathbf{P}(C)=\mathbf{P}((W_1,W_2,R_3)+\mathbf{P}(R_1,W_2,W_3)+\mathbf{P}(W_1,R_2,W_3)$$
$$=\mathbf{P}(W_1)\cdot\mathbf{P}(W_2|W_1)\cdot\mathbf{P}(R_3|(W_1\cap W_2))+\mathbf{P}(R_1)\cdot\mathbf{P}(W_2|R_1)\cdot\mathbf{P}(W_3|(R_1\cap W_2))$$
$$+P(W_1)\cdot\mathbf{P}(R_2|W_1)\cdot\mathbf{P}(W_3|(W_1\cap R_2))$$
$$=\frac{3}{8}\cdot\frac{5}{7}\cdot\frac{4}{6}+\frac{5}{8}\cdot\frac{3}{7}\cdot\frac{4}{6}+\frac{5}{8}\cdot\frac{4}{7}\cdot\frac{3}{6}=0,536$$

e)
$$\mathbf{P}(E)=1-\mathbf{P}(R_1,R_2,R_3)=1-\frac{3}{8}\cdot\frac{2}{7}\cdot\frac{1}{6}=0,982$$

f) Tenminste 2 witte ballen trekken is juist 2 witte ballen trekken of juist drie witte ballen trekken.
$$\mathbf{P}(F)=\mathbf{P}(C)+\mathbf{P}(W_1,W_2,W_3)=3\cdot(\frac{5}{8}\cdot\frac{4}{7}+\frac{3}{6})+\frac{5}{8}\cdot\frac{4}{7}\cdot\frac{3}{6}=0,715$$

g)
$$\mathbf{P}((R_1)\cup(W_3))=1-\mathbf{P}((W_1)\cap(R_3))$$
$$1-\frac{3}{8}\cdot\frac{5}{7}=0,732=\frac{3}{8}+2\cdot\frac{5}{8}\cdot\frac{3}{7}\cdot\frac{4}{6}$$

- Gegeven zijn twee vazen: een witte, die 6 groene en 4 rode bollen bevat, en een zwarte, die 3 groene en 7 rode bollen bevat. Men trekt aselect een bol uit een der vazen. Bereken de kans dat:

  a. een groene bol getrokken wordt uit de witte vaas;

  b. een groene bol getrokken wordt uit om het even welke vaas.

  Oplossing: Het verschijnsel van het trekken van een bol is hier samengesteld uit twee enkelvoudige verschijnselen: het kiezen van een vaas en het trekken van een bol uit die vaas.

  De gebeurtenis A: "het trekken uit de witte vaas"; $\mathbf{P}(A)=\frac{1}{2}$.
  De gebeurtenis B: "het trekken uit de zwarte vaas"; $\mathbf{P}(B)=\frac{1}{2}$.
  De gebeurtenis C: "het trekken van een groene bol".

De kans dat een groene bol getrokken wordt uit de witte vaas is de kans op $A \cap C$.

$$\mathbf{P}(A \cap C) = \mathbf{P}(A).\mathbf{P}(C|A)$$

De kans dat een groene bol wordt getrokken op voorwaarde dat we een bol trekken uit de witte vaas is $\mathbf{P}(C|A) = \frac{6}{10}$.
Hieruit volgt:

$$\mathbf{P}(A \cap C) = \frac{1}{2}.\frac{6}{10} = \frac{3}{10}.$$

De kans om een groene bol te trekken uit om het even welke vaas is gelijk aan het totaal aantal groene bollen gedeeld door het totaal aantal bollen omdat de kans om uit de witte vaas te trekken gelijk is aan de kans om uit de zwarte vaas te trekken.

$$\mathbf{P}(C) = \frac{9}{20}.$$

- Een vaas bevat 20 witte en 30 zwarte bollen. Bereken de kans om minstens één witte bol te trekken in drie trekkingen zonder terugleggen.

Oplossing:

$$\begin{aligned}
\mathbf{P}(\text{minstens 1 witte bol}) &= 1 - \mathbf{P}(\text{3 zwarte bollen}) \\
&= 1 - \mathbf{P}(Z_1 \cap Z_2 \cap Z_3) \\
&= 1 - \mathbf{P}(Z_1).\mathbf{P}(Z_2|Z_1).\mathbf{P}(Z_3|Z_1 \cap Z_2) \\
&= 1 - \tfrac{30}{50}.\tfrac{29}{49}.\tfrac{28}{48} = \tfrac{111}{140}
\end{aligned}$$

- Een dobbelsteen is zo geladen dat de kans waarmee een gegeven aantal ogen boven komt evenredig is met dit aantal. We beschouwen de gebeurtenissen A: "een even getal gooien"; B: "een priemgetal groter dan 1 gooien"; C: "een oneven getal gooien". *Gevraagd:*

  a. Beschrijf dit waarschijnlijkheidsmodel;
  b. Bereken $\mathbf{P}(A)$, $\mathbf{P}(B)$ en $\mathbf{P}(C)$;
  c. Bereken de kans dat men een even of een priemgetal werpt;
  d. Bereken de kans dat men een oneven priemgetal werpt;
  e. Bereken de kans dat men een even getal doch geen priemgetal werpt.

Oplossing:

a. Uitkomstenverzameling $\Omega = \{1, 2, 3, 4, 5, 6\}$.
De kansverdeling: $p_1 = p$, $p_2 = 2p$,. . ., $p_6 = 6p$

$$p_1 + p_2 + \cdots + p_6 = 1$$

Uit de twee betrekkingen volgt dat $p + 2p + \cdots + 6p = 1 \Longrightarrow 21p = 1$ of $p = \frac{1}{21}$

$$p_1 = \frac{1}{21}, p_2 = \frac{2}{21}, \ldots, p_6 = \frac{6}{21}.$$

b.

$$A = \{2,4,6\}; \mathbf{P}(A) = p_2 + p_6 + p_6 = \frac{12}{21};$$

$$B = \{2,3,5\}; \mathbf{P}(B) = p_2 + p_3 + p_5 = \frac{10}{21};$$

$$C = \{1,3,5\}; \mathbf{P}(C) = p_1 + p_3 + p_5 = \frac{9}{21}.$$

c. Even of priemgetal: $A \cup B = \{2,3,4,5,6\}$

$$\mathbf{P}(A \cup B) = p_2 + p_3 + p_4 + p_5 + p_6 = 1 - \mathbf{P}(1) = 1 - \frac{1}{21} = \frac{20}{21}$$

d. Oneven priemgetal: $B \cap C = \{3,5\}$

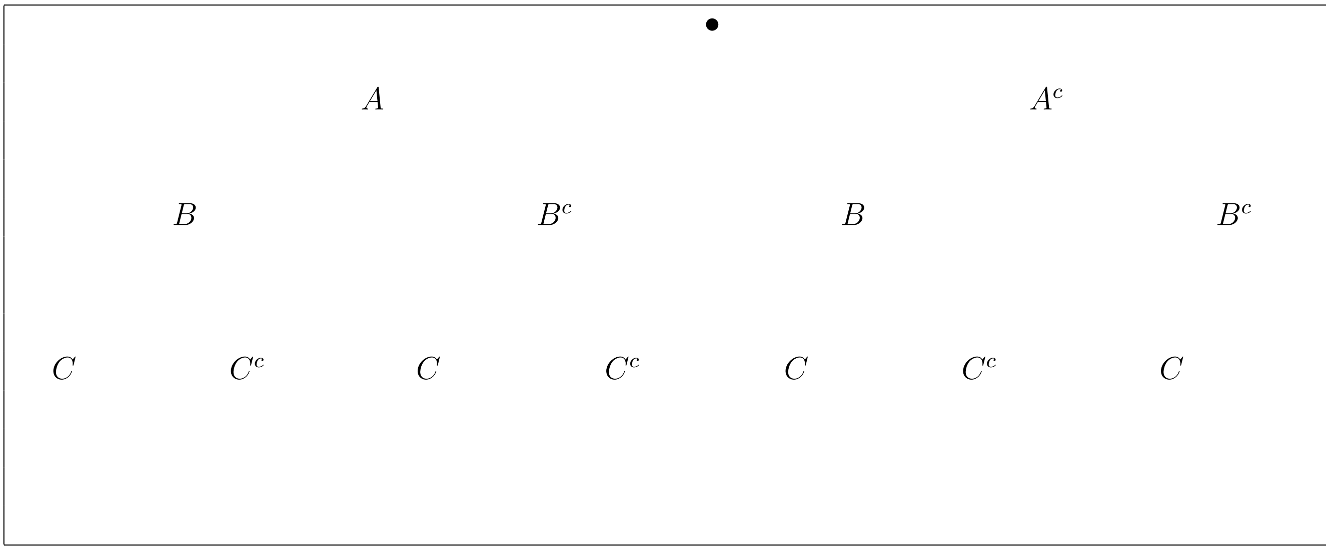
$$\mathbf{P}(B \cap C) = p_3 + p_5 = \frac{8}{21}.$$

e. Even doch geen priemgetal: $A \setminus B$

$$\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(A \cap B) = \frac{12}{21} - \frac{2}{21} = \frac{10}{21}$$

- Mevrouw D'Hollander wil dat de leerlingen van de zesdes iets actiever zijn in haar lessen L.O. en daarom organiseert zij voor die leerlingen een wedstrijd. De leerlingen moeten achtereenvolgens drie oefeningen tot een goed einde brengen, nl. een evenwichtsoefening op een balk (in de wiskundeles gedefiniëerd), een salto maken en over een bepaalde niet te onderschatten hoogte springen.
  Mevrouw D'Hollander heeft ervaring met die oefeningen en volgens haar zijn de kansen om die oefeningen tot een goed einde te brengen resp. $0,8$; $0,1$ en $0,5$. Een foutief uitgevoerde oefening betekent 4 strafpunten. Onderstel dat we 'weigering' buiten beschouwing laten. Wat is de kans voor een leerling om juist 4 strafpunten te krijgen?

  (i) In eerste instantie onderstellen we dat het al dan niet goed uitvoeren van een oefening onafhankelijk is van het al dan niet goed uitvoeren van een andere oefening;



Tabel 2.5: boomdiagram van de turnles 1

(ii) In tweede instantie onderstellen we een meer realistische situatie dat een leerling die een oefening foutief uitvoerde meer kans heeft om ook de volgende oefening foutief uit te voeren. We doen dit als volgt: als een bepaalde oefening foutief uitgevoerd is, verhogen we de kans om de volgende oefening ook foutief uit te voeren met $0,1$.

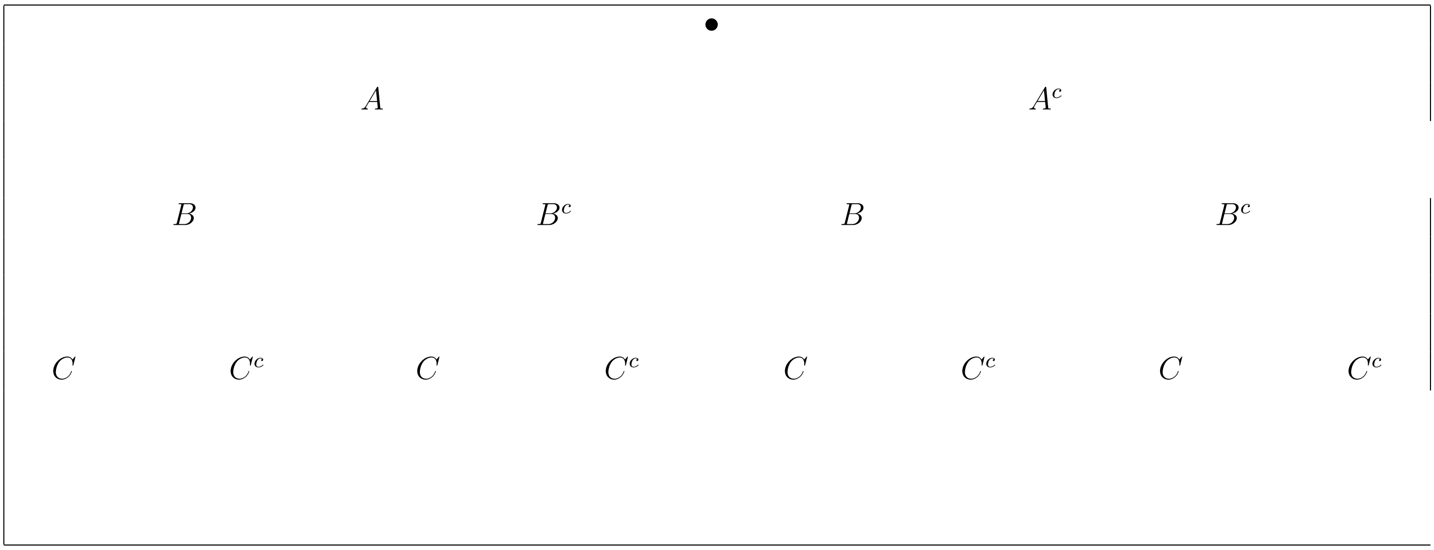
OPLOSSING:

(i) We beschouwen de gebeurtenissen
A: "slagen in de evenwichtsoefening";
B: "de salto kunnen maken";
C: "slagen in het hoogspringen".

We moeten de kans berekenen dat een leerling juist één oefening niet tot een goed einde brengt. Daarvoor zijn er drie mogelijkheden, nl. $A \cap B \cap C^c$, $A \cap B^c \cap C$ en $A^c \cap B \cap C$. Deze drie gebeurtenissen zijn onafhankelijk van elkaar. We kunnen ze best voorstellen op een boomdiagram (zie fig.2.5). De kans op juist vier strafpunten lezen we af op het boomdiagram:

$$\mathbf{P}(A \cap B \cap C^c) + \mathbf{P}(A \cap B^c \cap C) + \mathbf{P}(A^c \cap B \cap C)$$

$$= 0,8 \cdot 0,1 \cdot 0,5 + 0,8 \cdot 0,9 \cdot 0,5 + 0,2 \cdot 0,1 \cdot 0,5 = 0,41$$



Tabel 2.6: boomdiagram van de turnles 2

(ii) We kunnen de verhoging van de kansen gemakkelijk aanduiden op een boomdiagram (zie fig.2.6). De evenwichtsoefening mag niet fout lopen want dan hebben we geen kans meer nog een salto te maken. Dus in dat geval zeker minstens 8 strafpunten.
De kans op juist vier strafpunten lezen we af op het boomdiagram:

$$\mathbf{P}(A \cap B \cap C^c) + \mathbf{P}(A \cap B^c \cap C) + \mathbf{P}(A^c \cap B \cap C)$$

$$= 0,8 \cdot 0,1 \cdot 0,5 + 0,8 \cdot 0,9 \cdot 0,4 + 0,2 \cdot 0 \cdot 0,4 = 0,328.$$

De kans op juist vier strafpunten is hier uiteraard kleiner dan in (i).

- In Groot-Brittannië nemen drie personen deel aan een verkiezing. We beschouwen de gebeurtenissen A, B en C dat resp. de eerste, de tweede en de derde persoon stemt voor de Labour partij en we onderstellen dat $\mathbf{P}(A) = 0,6$; $\mathbf{P}(B) = 0,5$ en $\mathbf{P}(C) = 0,4$. De personen stemmen onafhankelijk van elkaar.
Wat is de kans dat er

  (i) juist één persoon stemt voor Labour?
  (ii) tenminste één persoon stemt voor Labour?

  Oplossing:

  (i) De enige Labour-stem kan afkomstig zijn van elk van de drie personen. Bijgevolg is de kans op juist één Labour-stem gelijk aan

$$\mathbf{P}(A \cap B^c \cap C^c) + \mathbf{P}(A^c \cap B \cap C^c) + \mathbf{P}(A^c \cap B^c \cap C).$$

De gebeurtenissen A, B^c en C^c zijn onafhankelijk dus geldt er

$$\mathbf{P}(A \cap B^c \cap C^c) = \mathbf{P}(A).\mathbf{P}(B^c).\mathbf{P}(C^c) = (0,6).(0,5).(0,6) = 0,18$$

Op analoge wijze is:

$$\mathbf{P}(A^c \cap B \cap C^c) = \mathbf{P}(A^c).\mathbf{P}(B).\mathbf{P}(C^c) = (0,4).(0,5).(0,6) = 0,12$$

$$\mathbf{P}(A^c \cap B^c \cap C) = \mathbf{P}(A^c).\mathbf{P}(B^c).\mathbf{P}(C) = (0,4).(0,5).(0,4) = 0,08$$

De gevraagde kans is 0,38.

(ii) De gebeurtenis "minstens één stem is een Labour-stem" is $A \cup B \cup C$. We kunnen beter de kans berekenen van het complement

$$\mathbf{P}(A \cup B \cup C) = 1 - \mathbf{P}((A \cup B \cup C)^c) = 1 - \mathbf{P}(A^c \cap B^c \cap C^c).$$

Aangezien A, B en C onafhankelijke gebeurtenissen zijn geldt:

$$\mathbf{P}(A^c \cap B^c \cap C^c) = \mathbf{P}(A^c).\mathbf{P}(B^c).\mathbf{P}(C^c).$$

De kans op minstens één Labour-stem is

$$\mathbf{P}(A \cup B \cup C) = 1 - (0,4).(0,5).(0,6) = 0,88.$$

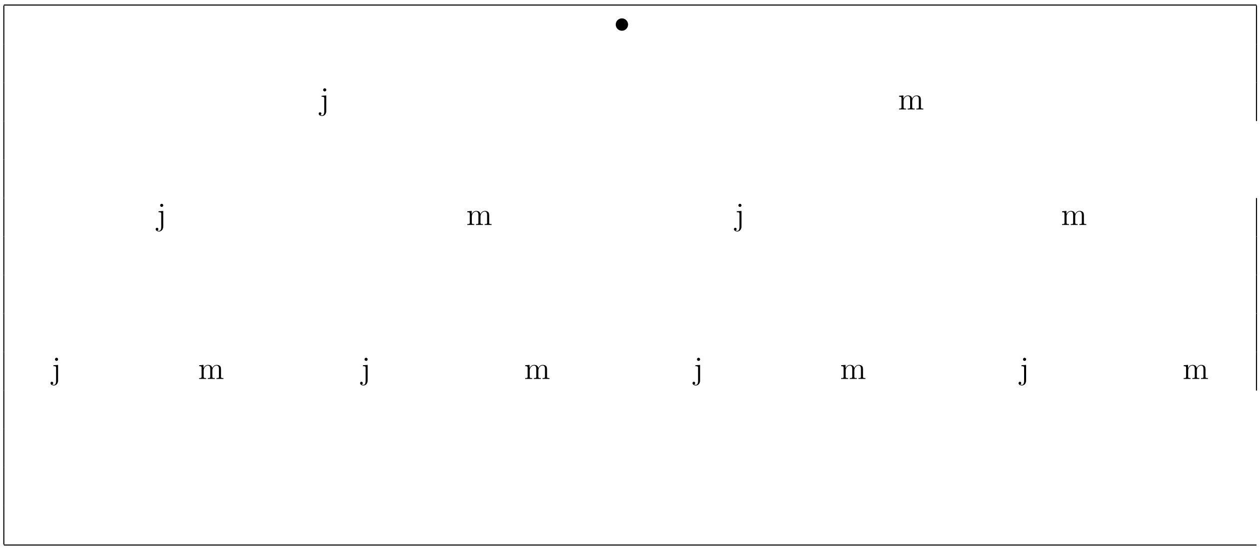
- Experimenteel werd vastgesteld dat de kans op een meisje 0,49 en de kans op een jongen is 0,51. Wat is de kans dat een gezin is samengesteld uit 2 meisjes en 1 jongen?

  Oplossing: Bij een geboorte zijn er twee mogelijke uitkomsten, nl. een jongen en een meisje. Beschouwen we nu gezinnen van 3 kinderen dan kunnen we de wiskundige kans bereken op bvb. 2 meisjes en 1 jongen. Daartoe maken we gebruik van een boomdiagram (zie figuur 2.7).

  Op het boomdiagram zien we dat we voor een gezin van 3 kinderen 8 mogelijke uitkomsten hebben, dit zijn de 2^3 mogelijke drietallen die we kunnen vormen met elementen gekozen uit een verzameling van 2 elementen nl. de verzameling $\{\mathrm{m}, \mathrm{j}\}$ (herhalingsvariatie van 2 elementen in groepjes van 3). We plaatsen bij elke tak de kans op een jongen en de kans op een meisje Voor de gebeurtenis '2 meisjes en 1 jongen' zijn er van de 8 uitkomsten 3 gunstig. Uit het boomdiagram leiden we af dat de kans op 2 meisjes en 1 jongen gelijk is aan

$$\mathbf{P}((m,m,j) \cup (m,j,m) \cup (j,m,m)) = 0,51 \cdot 0,49 \cdot 0,49 + 0,49 \cdot 0,51 \cdot 0,49 + 0,49 \cdot 0,49 \cdot 0,51$$

$$= 30,51 \cdot (0,49)^2 = 0,122.$$



Tabel 2.7: boomdiagram voor een gezin van 3 kinderen

2.3.9 De regel van Bayes.

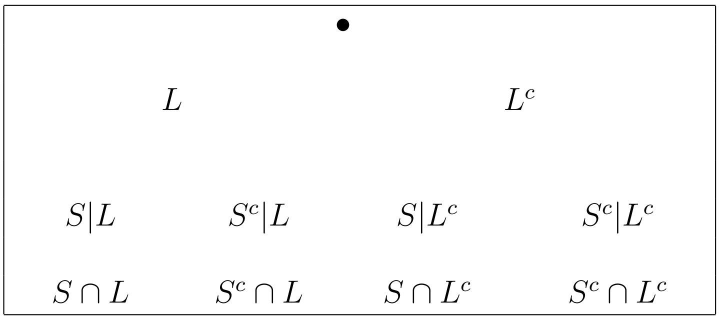
De *Regel van Bayes* geeft een methode aan om de waarschijnlijkheid van een bepaalde oorzaak te zoeken. We geven een voorbeeldje. In Nieuwen Bosch worden onder de middag extra lessen wiskunde gegeven ter voorbereiding op de toelatingsproef geneeskunde. Tachtig procent van de leerlingen die van plan zijn dit examen af te leggen volgt die lessen. Uit de statistieken blijkt nu dat een leerling die de lessen gevolgd heeft 55% kans heeft om te slagen, terwijl 65% van diegenen die deze lessen links laat liggen niet slaagt. Suna van Nieuwen Bosch doet het examen mee en slaagt. Hoe groot is de kans dat zij die lessen volgde?

Je voelt aan dat dit een averechtse vraag is. Normaal zou je verwachten dat we vragen naar de waarschijnlijkheid dat Suna slaagt, maar neen, we onderstellen dat Suna slaagt en vragen naar de kans van een bepaalde oorzaak — hier het feit van de lessen gevolgd te hebben.

We tekenen een boomdiagram voor een leerling We stellen de gebeutenissen voor:
L: “een leerling volgt de lessen”;
S: “een leerling slaagt”.

Volgens de wet der totale kans is

$$\begin{aligned}\mathbf{P}(S) &= \mathbf{P}(S \cap L) + \mathbf{P}(S \cap L^c)\\ &= \mathbf{P}(L)\mathbf{P}(S|L) + \mathbf{P}(L^c)\mathbf{P}(S|L^c)\\ &= 0,8 \cdot 0,55 + 0,2 \cdot 0,35 = 0,51\end{aligned}$$



De kans dat Suna de lessen volgde op voorwaarde dat ze slaagt is

$$\mathbf{P}(L|S) = \frac{\mathbf{P}(L \cap S)}{\mathbf{P}(S)} = \frac{\mathbf{P}(L)\mathbf{P}(S|L)}{\mathbf{P}(S)} = \frac{0,8 \cdot 0,55}{0,51} = 0,86$$

Een voorbeeld uit de medische wereld:
In de populatie Ω is A een deelpopulatie met een bepaalde ziekte en A^c de deelpopulatie die de ziekte niet heeft. Men wenst nu deze populatie te testen op deze ziekte. We noemen '+' de deelpopulatie die positief regeert en '-' de deelpopulatie die negatief reageert. Meestal is het zo dat de testen niet 100% nauwkeurig werken: er zijn vals-positieven en vals-negatieven. Twee belangrijke vragen worden dan gesteld:

(1) Als een persoon bij een test positief reageert, wat is dan de kans dat hij ook werkelijk de ziekte heeft?

(2) Als een persoon bij een test negatief reageert, wat is dan de kans dat hij ook werkelijk de ziekte niet heeft?

Een foutieve conclusie is van tweeërlei aard:

Type I: de persoon is ziek en men geeft hem een negatieve testuitslag;

Type II: de persoon is niet ziek en men geeft hem een positieve testuitslag.

Bij de huidige Elisa-testen voor antistoffen tegen het Aids-virus is

$$\mathbf{P}(+|A) = \mathbf{P}(-|A^c) = 0,99$$

De kans dat een persoon van deze populatie Aids heeft is $\mathbf{P}(A) = 0,006$.
Volgens de wet der totale kans is

$$\mathbf{P}(+) = \mathbf{P}(+\cap A) + \mathbf{P}(+\cap A^c) = \mathbf{P}(A)\mathbf{P}(+|A) + \mathbf{P}(A^c)\mathbf{P}(+|A^c) = 0,006 \cdot 0.99 + 0,994 \cdot 0,01 = 0,01588$$

De voorspellende waarde van een positieve test is:

$$\mathbf{P}(A|+) = \frac{\mathbf{P}(A \cap +)}{\mathbf{P}(+)} = \frac{\mathbf{P}(A)\mathbf{P}(+|A)}{\mathbf{P}(+)} = \frac{0,006 \cdot 0.99}{0,01588} = 0,374$$

Dit cijfer is laag. Ongeveer 60% van de populatie die positief reageert heeft de ziekte niet en maakt zich dus nodeloos zorgen.

Nog een voorbeeldje, een doordenkertje. Een president Tom van een verre staat heeft beslist dat van drie gevangenen A, B en C er twee geëxecuteerd zullen worden. De cipier Kristof mag er twee lukraak kiezen, maar mag nog niet bekend maken wie het zal zijn. Enige tijd worden de gevangenen in volledige afzondering geplaatst en A vraagt aan Kristof om één naam te noemen van de twee anderen die geëxecuteerd zullen worden.
Aangezien A niet kan communiceren met B of C en aangezien A weet dat minstens één van B of C zal geëxecuteerd worden, kan het volgens Kristof geen kwaad om het A te vertellen. Hij zegt tegen A: "B zal geëxecuteerd worden".
Vooraleer Kristof sprak, wist A dat zijn overlevingskans 1/3 was. Volgens A is zijn overlevingskans nu gestegen tot 1/2, aangezien hij weet dat C ofwel hijzelf zal overleven.
Redeneert A juist of heeft Kristof gelijk dat de informatie die hij aan A gaf niet bruikbaar is voor A?

Oplossing: We noemen A, B, C de gebeurtenissen waarbij resp. A, B en C geëxecuteerd worden. Dan geldt oorspronkelijk

$$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = 2/3.$$

Nadat A met Kristof had gesproken, berekent hij zijn kans om te sterven als volgt:

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = 1/2.$$

Maar A had de kans moeten berekenen dat hij sterft op voorwaarde dat Kristof zegt dat B zal sterven. We duiden die gebeurtenis aan door ZB.

$$\begin{aligned}\mathbf{P}(A|ZB) &= \frac{\mathbf{P}(A \cap ZB)}{\mathbf{P}(ZB)} \\ &= \frac{\mathbf{P}(A) \cdot \mathbf{P}(ZB|A)}{\mathbf{P}(ZB)} \\ &= \frac{2/3 \times 1/2}{1/2} \\ &= 2/3\end{aligned}$$

In geval A geëxecuteerd wordt dan is de andere ofwel B ofwel C. Dus de kans dat Kristof B zegt is 1/2.

In geval A niet geëxecuteerd wordt, dan worden B en C geëxecuteerd en Kristof kan B of C vernoemen met dezelfde kans, nl. 1/2.
Bijgevolg is:

$$\begin{aligned}\mathbf{P}(ZB) &= \mathbf{P}(ZB|A) \cdot \mathbf{P}(A) + \mathbf{P}(ZB|A') \cdot \mathbf{P}(A') \\ &= (1/2 \times 2/3) + (1/2 \times 1/3) \\ &= 1/2\end{aligned}$$

Kristof had dus gelijk.

OPGAVEN — 89 Wat is de kans dat bij het opgooien van twee dobbelstenen het eerst gegooide getal niet kleiner is dan het tweede op voorwaarde dat de som van de gegooide getallen acht is?

90 Uit een urne die vijf witte en negen rode bollen bevat worden aselect en zonder terugleggen vier bollen getrokken. Bereken de kans dat de vier bollen rood zijn.

91 We gooien twee keer een dobbelsteen op. Bereken de kans dat de som van de geworpen aantallen gelijk is aan vijf. Welke is de som met de grootste kans?

92 Een klas is samengesteld uit acht meisjes waarvan er vier een bril dragen en uit twaalf jongens waarvan er zes een bril dragen. We kiezen lukraak een leerling uit deze klas. Bereken de kans dat de gekozen leerling een jongen is of een leerling die een bril draagt.

93 Een urne bevat vijftig identieke bollen genummerd van 1 tot en met vijftig. We trekken aselect een bol. Bereken de kans dat het getrokken nummer een getal is dat deelbaar door zes of door negen.

94 Een urne bevat 100 identieke bollen waarvan er veertig rood en zestig wit zijn.

a. We trekken aselect een bol uit de urne. Bereken de kans dat we een rode bol trekken; dat we een witte bol trekken.
b. We trekken aselect drie bollen uit de urne zonder terugleggen. Bereken de kans dat we twee rode en een witte bol trekken, dat we tenminste twee witte bollen trekken.

95 Vier gehuwde paren nemen deel aan een spel. De dames loten om een partner uit hun echtgenoten. Bereken de kans dat ten minste één dame haar echtgenoot niet geloot heeft.

96 In de paardesport noemt men tiercé het spel dat erin bestaat te voorspellen welke drie paarden de eerste drie plaatsen zullen wegkapen. Onderstel dat er twaalf paarden deelnemen aan een bepaalde wedstrijd en dat men niets weet over de prestaties van deze paarden in het verleden. Bereken de kans om het tiercé te winnen

a. in de juiste volgorde;
b. als de volgorde niet in aanmerking komt.

97 Een geldstuk wordt drie keer opgeworpen. Bereken de kans dat we

a. ten minste twee keer kop gooien;

b. ten hoogste twee keer munt gooien.

98 Een gezin telt vier kinderen. Bereken de kans dat

a. de kinderen van hetzelfde geslacht zijn;

b. ten hoogste twee jongens voorkomen.

99 Een kinderleidster heeft voor haar kleuters een kersttombola ingericht. Zij heeft de volgende prijzen samengebracht: tien zakjes snoep en tien ballen met elk een waarde van vijftien frank, negen zakjes snoep en zes zakjes knikkers met elk een waarde van twintig frank, zes zakjes snoep en vier ballen met elk een waarde van vijfentwintig frank. Elk van de vijfenveertig tombolanummers hebben dezelfde kans. Bereken de kans om

a. snoep te winnen;

b. geen knikkers te winnen;

c. iets van minstens twintig frank te winnen;

d. iets van vijfentwintig frank of snoep te winnen;

e. een bal of iets van minstens twintig frank en geen knikkers te winnen.

100 Tien personen nemen plaats aan een ronde tafel. Wat is de kans dat twee bepaalde personen naast elkaar zullen zitten?

101 Bij een pokerspel wordt uit een goed geschud stel speelkaarten aan elke speler vijf kaarten gegeven. Bereken de kans dat een speler de volgende kaarten krijgt:

a. een reeks full royal (10,boer,dame,heer en aas van dezelfde soort);

b. poker (vier kaarten met dezelfde waarde);

c. een reeks full (vijf opeenv. kaarten van dezelfde soort, niet (a));

d. een full (vijf kaarten van dezelfde soort maar niet (a) of (c));

e. een reeks (vijf opeenv. kaarten maar niet van dezelfde soort).

102 In een school is 70% van de leerlingen extern, 25% van de externen en 10% van de internen kozen Engels als tweede taal. Als een leerling willekeurig gekozen wordt, wat is dan de kans dat hij extern is, als hij Engels als tweede taal gekozen heeft?

103 Twee cijfers worden willekeurig gekozen zonder terugleggen uit negen cijfers 1 tot en met 9. Bereken de kans dat, als de som van de getallen even is, de twee getallen oneven zijn.

104 Bereken bij het opgooien van drie muntstukken de kans om drie keer kruis te gooien

a. als de eerste opgooi kruis geeft;

b. als één van de opgegooide muntstukken kruis geeft.

105 Trek een kaart uit een stel speelkaarten en gooi daarna een dobbelsteen. Hoe groot is de kans dat je een schoppen trekt en een vier gooit?

106 Je gooit een muntstuk drie keer. Wat is de kans om

a. twee keer kop en één keer munt te hebben?

b. zo te gooien dat het aantal keer kop het aantal keer munt overtreft?

107 Een loterij bevat 10 000 loten en 500 prijzen en een andere loterij 20 000 loten en 900 prijzen. Hoe groot is de kans om minstens één prijs te hebben als je van elke loterij één lot koopt?

108 Een bakje bevat vier rode en vijf groene knikkers en een ander zes blauwe en zeven zwarte. Je neemt uit het eerste bakje twee knikkers en uit het tweede één knikker. Bepaal de kans op

a. twee rode en één blauwe;

b. een rode en één zwarte;

c. twee groene of twee rode en één blauwe;

d. een groene, één zwarte en één blauwe.

109 Men kiest één van de 4 cijfers 1, 2, 3, 4; en daarna onder de drie overblijvende kiest men nog één. Bepaal de kans om een even cijfer te kiezen.

a. de eerste maal;

b. de tweede maal;

c. beide malen.

110 Men gooit met twee dobbelstenen.

a. Wat is de kans dat ze beide op 4 liggen als gegeven is dat de som van het aantal ogen 7 of 8 is;

b. Wat is de kans dat beide een twee aangeven als gegeven is dat hun som 7 of 4 is.

111 Welk is het meest waarschijnlijk verschijnsel:

1. Met één dobbelsteen geen enkele 6 gooien in 4 worpen of met twee dobbelstenen geen enkel paar zessen gooien in 24 worpen?
2. Met één dobbelsteen tenminste één 6 gooien in 4 worpen of met twee dobbelstenen tenminste één paar zessen gooien in 24 worpen?

112 In een schroevenfabriek fabriceren de machines A, B en C resp. 25%, 35% en 40% van de totale productie. Van hun producten zijn er resp. 5%, 4% en 2% defect. Men kiest willekeurig een schroef en deze is slecht. Hoe groot is de kans dat deze gemaakt is door A of B of C?

113 Een vaas V_1 bevat 2 zwarte bolletjes en 3 witte, een andere vaas V_2 bevat 4 zwarte en 5 witte. Men neemt willekeurig een bolletje uit vaas V_1 en werpt dit in vaas V_2. Men trekt daarna een bolletje uit V_2 en dit blijkt zwart te zijn. Bereken de kans dat er een wit bolletje werd verwisseld van vaas.

114 Een brouwerij beschikt over drie vulmachines, die resp. 50%, 35% en 15% van het totaal aantal flessen vullen. De proportie foutief gevulde flessen bedraagt voor deze drie machines resp. 0,4%' 0,8% en 1%. Bereken voor de ganse productie het % slecht gevulde flessen.

115 Een doos bevat 6 goede en 4 slechte lampen. Men kiest er 2 uit. Men test er 1 van en deze wordt goed bevonden. Welk is de kans dat de andere lamp ook goed is.

116 Bereken de kans om uit een spel van 52 kaarten achtereenvolgens 3 ruiten te trekken zonder teruglegging.

117 We beschouwen drie identieke vazen. De eerste vaas bevat 30 witte en 20 zwarte bollen, de tweede vaas bevat 20 witte en 30 zwarte bollen en de derde vaas bevat 10 witte en 40 zwarte bollen. Bereken de kans om geblinddoekt uit

a. de eerste vaas een witte bol te trekken;

b. om het even welke vaas een witte bol te trekken;

c. de eerste vaas te trekken als de getrokken bol wit is.

118 Een statistisch onderzoek heeft uitgemaakt dat 90% van de roller-studenten mislukken, van de doodblokkers mislukken 20% wegens overspanning en van de regelmatige werkers mislukken er slechts 10%. Welk is de kans dat een mislukt student tot de roller-groep behoort, aangenomen dat elke groep op de universiteit even sterk vertegenwoordigd is?

119 Statistieken hebben uitgemaakt dat we gemiddeld in België tijdens de maand juni 15 dagen zonnig weer hebben, 10 dagen bewolkte hemel en 5 dagen regen. Zonder enige andere informatie zouden we voor een bepaalde dag bvb. 15 juni, het weer moeten voorspellen. De weerkundigen vergissen zich ook regelmatig. Ook hiervan werden statistieken opgesteld: voor 10 dagen mooi weer werd 8 maal mooi en 2 maal bewolkt weer voorspeld; voor 10 dagen bewolkt weer werd 3 maal mooi weer voorspeld, 5 maal bewolkt en 2 maal regen voorspeld en voor 10 regendagen werd 6 maal regen, 3 maal bewolkt en 1 maal mooi voorspeld. Onderstel dat de weerkundige mooi weer voorspelt voor de 15de juni, bereken dan de kans dat

a. het mooi weer zal zijn;

b. het bewolkt zal zijn;

c. het zal regenen.

120 Volgens de sterftecijfers heeft een persoon van 43 jaar 0,7 en een persoon van 52 jaar 0,5 kans om nog 20 jaar te leven.
Bereken

a. de kans dat na 20 jaar beide personen nog leven;

b. de kans dat minstens één van beide na 20 jaar nog leeft.

121 Twee personen A en B spelen 12 keer schaak. A wint 6 spelen, B wint er 4 en 2 spelen eindigen in remise. Zij komen overeen nog 3 keer te spelen.
Bereken de kans dat

a. A de 3 spelen wint;

b. juist 2 spelen eindigen in remise;

c. B minstens één spel wint.

122 Men werpt met twee dobbelstenen, een witte en een zwarte. We beschouwen twee gebeurtenissen A: "de witte dobbelsteen komt op 1" en B: "de som van het aantal ogen is 7".
Bereken: $\mathbf{P}(A)$, $\mathbf{P}(B)$, $\mathbf{P}(A \cap B)$, $\mathbf{P}(A|B)$ en $\mathbf{P}(B|A)$.

123 90% van de personen die kanker hebben reageren positief op de kankertest en 5% van de personen die geen kanker hebben reageren ook positief. Veronderstel dat in een bepaald ziekenhuis 1% kanker heeft. Als een persoon van het ziekenhuis positief reageert op de kankertest, wat is dan de kans dat hij kanker heeft.

124 In een verkiezing voor gemeenteraadslid heeft kandidaat I 45% van de stemmen, kandidaat II krijgt 40% van de stemmen en kandidaat III krijgt 15%. Na de verkiezingen komt men vijf kiezers tegen, wat is de kans dat twee gestemd hebben voor kandidaat I, twee voor kandidaat II en één voor kandidaat III.

125 In een fabriek staan drie machines opgesteld die alle vijzen produceren. Machine A is de oudste en produceert slechts één vijfde van de totale productie, machine B is de nieuwste en neemt de helft van de productie voor zijn rekening. Machine C doet de rest. Van alle vijzen die machine A produceert zijn er twee procent slechte bij, terwijl dit voor machine B slechts $0,8$ procent is. Anderhalve procent van de productie van machine C is slecht. De directeur neemt nu een willekeurige vijs en ziet dat ze slecht is. Wat is de kans dat ze van de oudste machine afkomstig is?

126 Tien procent van de kankerpatiënten in een bepaald ziekenhuis sterft reeds de eerste dag van hun verblijf. Vijftig procent sterft binnen de week en negentig procent binnen de maand. Voor AIDS-patiënten is dit respectievelijk vijftien, zestig en vijfennegentig, terwijl voor mensen met hartziekten dit respectievlijk twee, zestien en tweeëndertig is. Andere ziekten komen niet voor in deze bepaalde afdeling, maar de AIDS-patiënten zijn in de meerderheid: ze maken 45% uit van alle zieken; de kanker-patiënten 30% . Op een dag sterft een pas binnengebrachte patiënt. Wat is de kans dat het van AIDS was? Een andere sterft binnen de maand, maar na de eerste dag. Wat is de kans dat hij een hartziekte had? Nog een ander leeft na een maand nog. Wat is de kans dat hij kanker had?

127 Tom is de oudste van zes kinderen in een zeer rijke familie. Voor de eerste communie van hun kinderen hebben de ouders drie gouden armbanden en drie gouden uurwerken gekocht. Elk voorwerp steken ze in een lade van drie kasten en elke kast bevat zo twee laden. In beide laden van kast A zit nu een gouden uurwerk en in beide laden van kast C een gouden armband. Daar de ouders van Tom niet weten wat Tom verkiest, mag Tom naar eigen willekeur één lade openen op haar eerste communiefeestje en wat erin zit krijgt ze. Dat doet ze en ze vindt een gouden uurwerk. De vraag is niet hoe laat het is, maar wel wat de kans is dat Tom een lade van kast B opentrok?

128 Thomas drinkt alle dagen voor hij slapen gaat een borreltje. Hij heeft dertig flessen whisky en twintig flessen cognac geërfd van zijn overgrootvader en wil die op die manier soldaat maken. Alle avonden kiest hij lukraak een fles en drinkt. Hij weet echter dat hij van een borrel whisky in tachtig procent van de gevallen de volgende dag met hoofdpijn zal wakker worden, terwijl dit van een cognacje slechts in drie gevallen op tien gebeurt. Vanmorgen had Thomas barstende hoofdpijn. Wat is de kans dat hij gisterenavond een whiskietje heeft gedronken?

OPLOSSINGEN:

89. $\frac{3}{5}$;

90. $\binom{9}{4} : \binom{14}{4} = 0,125874126$;

91. $\frac{4}{36} = \frac{1}{9}$ en $\frac{6}{36} = \frac{1}{6}$ voor som 7;

92. $\frac{12}{20} + \frac{10}{20} - \frac{6}{20} = \frac{4}{5}$;

93. $\frac{8}{50} + \frac{5}{50} - \frac{2}{50} = \frac{11}{50}$;

94. a. $\frac{60}{100}$ voor een witte en $\frac{40}{100}$ voor een rode; b. $60.\binom{40}{2} : \binom{100}{3} = 0,289424861$ voor twee rode en een witte bol;
"tenminste twee witte bollen trekken" is de negatie van "één witte en geen witte bollen trekken": dus $1 - 60.\binom{40}{2} : \binom{100}{3} - \binom{40}{3} : \binom{100}{3} = 0,649474335$;

95. "tenminste één dame heeft haar echtgenoot geloot" is de negatie van "elke dame heeft haar echtgenoot geloot" dus 1 - 1/4! = 23/24;

96. a. $\frac{1}{V_{12}^3} = \frac{1}{1320}$; b. $1 \div \binom{12}{3} = 3!/V_12^3$;

97. a. $\frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$; we kunnen ook onmiddellijk het resultaat geven want tenminste twee keer kop komt voor in de helft van de gevallen;
b. "ten hoogste twee keer munt gooien" is de negatie van "juist drie keer munt gooien" dus $1 - \frac{1}{8}$;

98. a. $\frac{2}{16} = \frac{1}{8}$; b. $\frac{1}{16} + \frac{4}{16} + \frac{6}{16} = \frac{11}{16} = 1 - \frac{1}{16} - \frac{4}{16}$;

99. a. $\frac{25}{45}$; b. $1 - \frac{6}{45} = \frac{39}{45}$; c. $\frac{25}{45}$; d. $\frac{10}{45} + \frac{25}{45} - \frac{6}{45} = \frac{29}{45}$; e. $\frac{29}{45}$;

100. $\frac{2.8!}{9!}$;

101. a.$4 \div \binom{52}{5} = \frac{15}{13}.\frac{4}{51}.\frac{3}{50}.\frac{2}{49}.\frac{1}{48}$; b. $13.48 \div \binom{52}{5}$; c. $8.4 \div \binom{52}{5}$; d. $(4.\binom{13}{5} - 4 - 8.4) \div \binom{52}{5}$; e. $(9.4^5 - 9.4) \div \binom{52}{5}$;

102. $\frac{35}{41}$;

103. $\frac{5}{8}$;

104. a. $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$; b. $\frac{1}{7}$;

105. $\frac{1}{4} \cdot \frac{1}{6} = \frac{1}{24}$;

106. a. $\frac{3}{8}$; b. $\frac{4}{8} = \frac{1}{2}$;

107. $\frac{371}{4000} = 1 - \frac{9500}{10000} \cdot \frac{19100}{20000}$;

108. a. $\frac{1}{6} \cdot \frac{6}{13}$; b. $\frac{13}{18} \cdot \frac{7}{13} = \frac{7}{18}$; c. $\frac{4}{9} \cdot \frac{6}{13} = \frac{24}{117}$; d. 0;

109. $\frac{1}{2}$; $\frac{1}{2}$; $\frac{1}{6}$;

110. $\frac{1}{11}$; $\frac{1}{9}$;

111. a. $(\frac{5}{6})^4 < \frac{35}{36})^{24}$, b. $1 - (\frac{5}{6})^4 > 1 - (\frac{35}{36})^{24}$;

112. $\frac{25}{69}$; $\frac{28}{69}$; $\frac{16}{69}$;

113. $\frac{6}{11}$;

114. $0,63\%$;

115. $5/9$;

116. $\frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50}$;

117. a. $\frac{3}{5}$; b. $\frac{6}{15}$; c. $\frac{1}{2}$;

118. $\frac{3}{4}$;

119. a. $\frac{24}{31}$; b. $\frac{6}{31}$; c. $\frac{1}{31}$;

120. a. $0,35$; b. $0,85$;

121. a. $\frac{1}{8}$; b. $\frac{5}{36}$; c. $\frac{19}{27}$;

122. $\frac{1}{6}$; $\frac{1}{6}$; $\frac{1}{36}$; $\frac{1}{6}$; $\frac{1}{6}$;

123. $\frac{2}{13}$.

124. $P_5^{2,2,1} \cdot 0,45^2 \cdot 0,40^2 \cdot 0,15$

125. $\frac{2\times 2}{2\times 2+0,8\times 5+1,5\times 3} = \frac{8}{25}$;

126.
* $\frac{45\times 15}{45\times 15+30\times 10+25\times 2} = \frac{27}{41}$,

* $\frac{30\times 25}{25\times 30+30\times 80+45\times 80} = \frac{1}{9}$,

* $\frac{30\times 10}{30\times 10+45\times 5+25\times 68} = \frac{12}{89}$;

127. $\frac{1}{3}$; 128. $\frac{8\times 6}{8\times 6+3\times 4} = \frac{4}{5}$;

2.4 Wiskunde-Cultuur

BAYES Thomas is een Engels predikant en wiskundige van de 18de eeuw, omtrent wiens leven we vrijwel niets weten. LAPLACE heeft zijn theorie over de omgekeerde waarschijnlijkheden (waarschijnlijkheden a posteriori) gered van de vergetelheid. Laplace heeft zijn werk na zijn dood in 1763-64 gepubliceerd.

HUISTAAK K2

1. De kans dat een vrouw jonger dan 30 jaar een universitair diploma had in 1970 was 0,2 en de kans dat ze getrouwd was bedroeg 0,35. De kans dat zo een vrouw getrouwd was als zij een universitair diploma had was 0,43 . Als een vrouw jonger dan 30 getrouwd was, wat is dan de kans dat ze een universitair diploma had.

2. Een urne I bevat 10 witte en 3 rode bollen. Een urne II bevat 3 witte en 5 rode bollen. Twee bollen worden van urne I naar urne II overgeplaatst en dan wordt een bol uit urne II genomen.
   Bereken de kans dat de twee bollen die overgeplaatst worden

   a. allebei wit zijn;
   b. allebei rood zijn;
   c. ene wit en ene rood;

   Wat is de kans dat een witte bol wordt getrokken uit de urne II?

3. In een schoenenfabriek produceert machine I 60% van de schoenen, waar van 10% slecht zijn en machine II produceert 40% van de schoenen, waarvan 20% slecht zijn.
   Bereken de kans dat

   a. als een schoen slecht wordt bevonden, hij uit machine I komt;
   b. als een schoen goed wordt bevonden, hij uit machine I komt;
   c. als twee schoenen goed zijn en één slecht, ze alledrie uit machine I komen.

4. Suna werpt twee keer met een stuk van 1 EUR en Stephanie werpt driemaal met een stuk van 2 EUR. Wat is de kans dat ze met het een-eurostuk evenveel keer kop hebben als met het twee-eurostuk? Een kwestie dat ze goede vriendinnen blijven.

5. Een sleutel van een gesloten klaslokaal in blok D is één van 12 sleutels. Suna neemt willekeurig twee sleutels uit de 12 mee naar het klaslokaal. Wat is de kans dat Suna de deur van het klaslokaal zal kunnen openen zonder te moeten terugkeren?

6. Op een ochtend bellen 6 mensen naar een loodgieter, die ze willekeurig gekozen hebben in een telefoongids uit drie loodgieters. Wat is de kans dat de eerste loodgieter drie oproepen krijgt, de tweede loodgieter twee oproepen en de derde loodgieter één oproep?

Hoofdstuk 3

Kansverdelingen

3.1 Toevalsveranderlijken

Een toevalsveranderlijke, stochastische veranderlijke of random veranderlijke (s.v.) X over het universum van een experiment is een observeerbare grootheid waarvan bij het uitvoeren van het experiment de waarde van het toeval afhangt en zich voordoet met een bepaalde kans.

Een discrete toevalsveranderlijke is een toevalsveranderlijke over een aftelbaar universum $\Omega = \{\omega_1, \omega_2 \ldots\}$ die de volgende functie is

$$X : \Omega \longrightarrow \mathbb{R}, \omega \longmapsto X(\omega).$$

Een beeld van X noemen we een waarde van de toevalsveranderlijke. We stellen de verschillende waarden van de toevalsveranderlijke voor door $x_1, x_2, \cdots$.
De verzameling van alle uitkomsten die dezelfde waarde x_i opleveren voor X is een bepaalde gebeurtenis.

Bij de gezinnen van 2 kinderen beschouwen we als toevalsveranderlijke X 'het aantal jongens'. We krijgen

$$\begin{aligned} x_1 &= X((\mathrm{m}, \mathrm{m})) = 0 \\ x_2 &= X((\mathrm{m}, \mathrm{j})) = X((\mathrm{j}, \mathrm{m})) = 1 \\ x_3 &= X((\mathrm{j}, \mathrm{j})) = 2 \end{aligned}$$

De mogelijke waarden van de toevalsveranderlijke X zijn de 3 waarden 0, 1 en 2 aan ($m = 3$ en $N = 4$).

Bij het werpen van twee dobbelstenen beschouwen we als toevalsveranderlijke de som van de ogen. We krijgen

$$
\begin{aligned}
&x_1 = X((1,1)) = 2\\
&x_2 = X((1,2)) = X((2,1)) = 3\\
&x_3 = X((1,3) = X((2,2)) = X((3,1)) = 4\\
&\vdots\\
&x_{10} = X((5,6)) = X((6,5)) = 11\\
&x_{11} = X((6,6)) = 12
\end{aligned}
$$

De mogelijke verschillende waarden van de toevalsveranderlijke X zijn de 11 waarden 2,3,4,...11 en 12 aan ($m = 11$ en $N = 36$).

Bij het experiment 'opgooien van 3 muntstukken' krijgt men 1 EUR als men munt werpt en moet men 1 EUR betalen als men kop werpt. Men beschouwt als toevalsveranderlijke X 'het bedrag dat men wint of verliest na de 3 worpen'. We krijgen

$$
\begin{aligned}
&x_1 = X((\mathrm{m},\mathrm{m},\mathrm{m})) = 3\\
&x_2 = X((\mathrm{m},\mathrm{m},\mathrm{k})) = X((\mathrm{m},\mathrm{k},\mathrm{m})) = X((\mathrm{k},\mathrm{m},\mathrm{m})) = 1\\
&x_3 = X((\mathrm{m},\mathrm{k},\mathrm{k})) = X((\mathrm{k},\mathrm{k},\mathrm{m})) = X((\mathrm{k},\mathrm{m},\mathrm{k})) = -1\\
&x_4 = X((\mathrm{k},\mathrm{k},\mathrm{k})) = -3
\end{aligned}
$$

De mogelijke waarden van de toevalsveranderlijke zijn de 4 waarden -3, -1, 1 en 3 aan ($m = 4$ en $N = 8$).

Een continue toevalsveranderlijke is een toevalsveranderlijke over een niet aftelbaar universum.

Wordt zonder richten een schot afgevuurd door een buis die afgesloten is door een schietschijf dan kunnen alle punten op de schietschijf getroffen worden. Het universum bestaat uit de oneindig veel punten van de schietschijf die niet aftelbaar zijn;

De tijd die een Vlaamse leerling slijt voor TV op een dag is een getal van het interval $[0, 24]$. Het universum bestaat uit alle punten van het interval $[0, 24]$ die niet aftelbaar zijn;

Aan een draad met lengte l wordt getrokken tot hij breekt. Men kan niet voorzien waar de draad zal afbreken. De breuk kan zich voordoen in elk punt van de oneindig veel punten van de draad. Het universum bestaat uit de oneindig veel punten van de draad die niet aftelbaar zijn.

3.2 Discrete kansverdelingen

De **kansverdeling van een discrete toevalsveranderlijke** X is de functie

$$x_i \longmapsto \mathbf{P}(X = x_i)$$

waarvoor geldt

$$0 \leq \mathbf{P}(X = x_i) \leq 1 \qquad \text{en} \qquad \sum_{i=1}^{m} \mathbf{P}(X = x_i) = 1$$

Een **uniforme discrete kansverdeling** is een kansverdeling waarvoor de kansen van de verschillende waarden x_i van de toevalsveranderlijke gelijk zijn aan elkaar.
Het experiment "Opgooien van een dobbelsteen" met als toevalsveranderlijke " aantal ogen" levert een uniforme kansverdeling op. Teken het staafdiagram rechts hieronder.

x_i	1	2	3	4	5	6
$\mathbf{P}(X = X_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

of

x_i	$\mathbf{P}(X = x_i)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
$\sum$	1

3.2.1 De Bernoulli kansverdeling

Bij veel experimenten stelt men belang in één bepaalde uitkomst, waarvan we het optreden ervan een **succes** noemen. Treedt de beoogde uitkomst niet op dan spreken we van een **mislukking**. Zo bvb. bij het opgooien van een dobbelsteen kunnen we het optreden van een zes succes noemen en het optreden van elke andere uitkomst is dan een mislukking. Komt men aan verkeerslichten bij groen licht dan is dit een succes, bij oranje en rood is dit een mislukking. De geboorte van een meisje kan je als een succes zien en de geboorte van een jongen als een mislukking. In sommige culturen is dat wel het omgekeerde, zoals in China en in India.

Een **Bernoulli experiment** is een enkelvoudig experiment waar men slechts twee uitkomsten beschouwt: succes en mislukking.

$$\Omega = \{s, m\}$$

De toevalsveranderlijke van een Bernoulli experiment is "het aantal keer succes".

$$X(s) = 1 \qquad \text{en} \qquad X(m) = 0$$

De kans op succes stellen we voor door $\mathbf{p}$. De kans op mislukking is gelijk aan $1 - \mathbf{p}$ die we gelijk stellen aan $\mathbf{q}$.

$$\begin{array}{l} \mathbf{P}(1) = \mathbf{p} \\ \mathbf{P}(0) = \mathbf{q} \end{array} \quad \text{met } \mathbf{p} + \mathbf{q} = 1$$

Voorbeeld: Bij het opgooien van een dobbelsteen is de kans op succes, dus op het gooien van een zes gelijk aan 1/6 en de kans op mislukking is dan 5/6.

$$\mathbf{p} = 1/6 \text{ en } \mathbf{q} = 5/6.$$

De Bernoulli kansverdeling is

x_i	0	1
$\mathbf{P}(X = x_i)$	$\frac{5}{6} = 0,833$	$\frac{1}{6} = 0,167$

of

x_i	$\mathbf{P}(X = x_i)$
0	$5/6 = 0,833$
1	$1/6 = 0,167$

Teken hieronder het staafdiagram.

3.2.2 De binomiale kansverdeling

3.2.2.1 Definitie van de binomiale kansverdeling

Een veel voorkomende situatie is dat het Bernoulli experiment herhaalde malen wordt uitgevoerd en dan is de vraag hoeveel keer succes zal ik hebben. Bij dit experiment is de toevalsveranderlijke X het aantal keer succes bij het n keer uitvoeren van het Bernoulli experiment. De mogelijke waarden voor X zijn de natuurlijke getallen $i \in \{0, 1, \cdots n\}$.
De **binomiale verdeling** geeft de kansen van de waarden van deze toevalsveranderlijke.

3.2.2.2 Berekening van de waarden van binomiale kansverdeling

Inleidend voorbeeld:
Stephanie wordt als slavin verkocht op de slavenmarkt. Normaal krijgt iedere mogelijke koper de kans een slavin gratis binnen te rijven door met een dobbelsteen juist 4 zessen te gooien in juist 10 worpen. Door het aanzienlijk succes van Stephanie (er zijn maar weinig slaven die latijn kennen) besluit de sluwe verkoper een andere regel te hanteren, waarvan de mogelijke kopers denken dat hij evenwaardig is. Om Stephanie gratis te kunnen krijgen moet men met twee dobbestenen juist 2 keer dubbel zes gooien in 5 worpen. Dit is ook 4 zessen in het totaal, en dat zijn ook 10 worpen in het totaal, redeneren de kopers. Is het nu moeilijker of gemakkelijker of hetzelfde om Stephanie gratis te mogen meenemen?

We beschouwen eerst het Bernoulli experiment 'gooien van één dobbelsteen' waarbij gooien van een zes succes is. Hier is $\mathbf{p} = \frac{1}{6}$.
Het experiment '10 keer gooien van één dobbelsteen' is het 10 keer herhalen van het Bernoulli experiment. Een element van de uitkomstenverzameling is bvb. $(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)$ waarbij zes gegooid wordt op de tweede, derde, zevende en negende plaats. Volgens de productregel is de kans op deze uitkomst

$$\mathbf{P}(\{(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)\}) = \mathbf{q} \cdot \mathbf{p} \cdot \mathbf{p} \cdot \mathbf{q} \cdot \mathbf{q} \cdot \mathbf{q} \cdot \mathbf{p} \cdot \mathbf{q} \cdot \mathbf{p} \cdot \mathbf{q} \cdot = \mathbf{p}^4 \cdot \mathbf{q}^6 = \left(\frac{1}{6}\right)^4 \cdot \left(\frac{5}{6}\right)^6$$

De toevalsveranderlijke X bij dit experiment is 'het aantal keer een zes gooien'.
X neemt de gehele waarden aan tussen 0 en 10.
Bvb. $X((0, 1, 1, 0, 0, 0, 1, 0, 1, 0)) = 4$. We kijken nu hoeveel uitkomsten de waarde 4 oplevert voor de toevalsveranderlijke. Je kan nu die 4 zessen op allerlei mogelijke plaatsen gooien.
Het aantal mogelijkheden om 4 plaatsen te kiezen uit de 10 is precies $C_{10}^4 = \binom{10}{4}$.
Dit is dus het aantal uitkomsten waarbij we precies 4 keer succes hebben.
Al deze uitkomsten hebben dezelfde kans tot optreden.
De kans om Stephanie gratis mee te krijgen is

$$\mathbf{P}(X = 4) = \binom{10}{4} \cdot (\frac{1}{6})^4 \cdot (\frac{5}{6})^6 = 0,054266.$$

We beschouwen nu het Bernoulli experiment 'gooien van twee dobbelstenen' waarbij gooien van een dubbele zes succes is. Hier is $\mathbf{p} = \frac{1}{36}$.
Het experiment '5 keer gooien met twee dobbelstenen' is het 5 keer herhalen van het beschouwde Bernoulli experiment. Een element van de uitkomstenverzameling is bvb. $(0, 0, 1, 0, 1)$ waarbij $(6, 6)$ gegooid wordt op de derde en de vijfde plaats. Volgens de productregel is de kans op deze uitkomst

$$\mathbf{P}(\{(0, 0, 1, 0, 1)\}) = \mathbf{q} \cdot \mathbf{q} \cdot \mathbf{p} \cdot \mathbf{q} \cdot \mathbf{p} \cdot = \mathbf{p}^2 \cdot \mathbf{q}^3 = (\frac{1}{36})^2 \cdot (\frac{35}{36})^3$$

De toevalsveranderlijke Y bij dit experiment is 'het aantal keer een dubbele zes'.
Y neemt de gehele waarden aan tussen 0 en 5.
Bvb. $Y((0, 0, 1, 0, 1)) = 2$. Je kan nu die 2 dubbele zessen op allerlei plaatsen gooien. Het aantal mogelijkheden om 2 plaatsen te kiezen uit de 5 is $C_5^2 = \binom{5}{2}$. Dit is het aantal uitkomsten waarbij we precies 2 keer succes hebben. Al deze uitkomsten hebben dezelfde kans tot optreden.
De kans om voor Stephanie niet te moeten betalen is gelijk aan

$$\mathbf{P}(Y = 2) = \binom{5}{2} \cdot (\frac{1}{36})^2 \cdot (\frac{35}{36})^3 = 0,007091.$$

De sluwe handelaar heeft dus bijna achtmaal meer kans om rijk te worden van Stephanie!

Algemeen

Men onderstelt dat een Bernoulli experiment n maal uitgevoerd wordt terwijl de kans op succes voor het één keer uitvoeren van het Bernoulli experiment gelijk is aan $\mathbf{p}$, $0 \leq \mathbf{p} \leq 1$.
De toevalsveranderlijke X is “het aantal keer succes bij n keer uitvoeren van eenzelfde Bernoulli experiment”.
De kans om juist i keren succes te hebben is dan

$$\mathbf{P}(X = i) = \binom{n}{i} \cdot \mathbf{p}^i \cdot (1 - \mathbf{p})^{n-i}.$$

Volgens het binomium van Newton geldt

$$\sum_{i=0}^{n} \mathbf{P}(X = i) = \sum_{i=0}^{n} C_n^i \cdot \mathbf{p}^i \cdot (1 - \mathbf{p})^{n-i} = (\mathbf{p} + (1 - \mathbf{p}))^n = 1.$$

Dit is ook de reden waarvoor we deze verdeling de **binomiaalverdeling** noemen.
De toevalsveranderlijke van de binomiale kansverdeling voor het n keer uitvoeren van een Bernoulli experiment met de kans op succes gelijk aan $\mathbf{p}$, duiden we aan door

$$X \sim \mathrm{B}(n; \mathbf{p})$$

$\boxed{\text{Berekening met de computer}}$

De volgende kansen kunnen we berekenen met Excel onder 'functies statistiek'. ONWAAR betekent niet-cumulatief en WAAR betekent cummulatief:

$$\begin{aligned}\mathbf{P}(X=i) &= \binom{10}{i}\cdot(\tfrac{1}{6})^i\cdot(\tfrac{5}{6})^{10-i} = \text{COMBINATIES}(10;i)\cdot(\tfrac{1}{6})^i\cdot(\tfrac{5}{6})^{10-i}\\ &= \text{BINOMIALE.VERD}(i;10;1/6;\text{ONWAAR})\end{aligned}$$

$$\mathbf{P}(X\le i) = \text{BINOMIALE.VERD}(i;10;1/6;\text{WAAR})$$

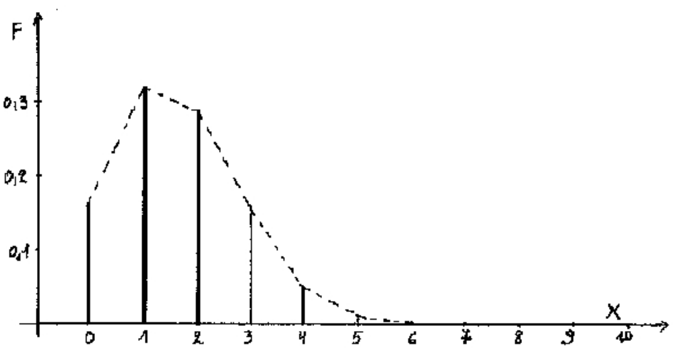
$$\mathbf{P}(X< i) = \mathbf{P}(X\le i-1) = \text{BINOMIALE.VERD}(i-1;10;1/6;\text{WAAR})$$

$$\mathbf{P}(X> i) = 1-\mathbf{P}(X\le i) = 1-\text{BINOMIALE.VERD}(i;10;1/6;\text{WAAR})$$

$$\mathbf{P}(X\ge i) = 1-\mathbf{P}(X\le i-1) = 1-\text{BINOMIALE.VERD}(i-1;10;1/6;\text{WAAR})$$

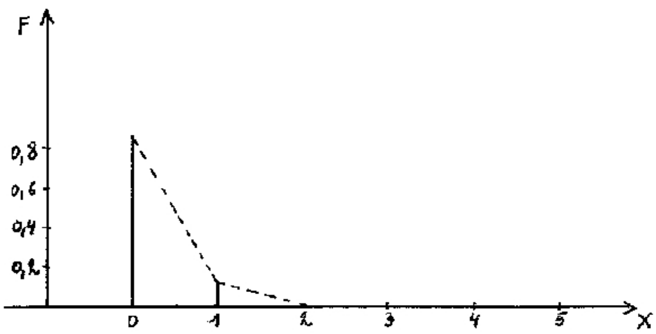
Vervolledig nu de volgende tabellen door gebruik te maken van EXCEL.
De kansverdeling voor $X\sim \text{B}(10;1/6)$:

i	$\mathbf{P}(X=i)$	$\mathbf{P}(X=i)$	$\mathbf{P}(X\le i)$	$\mathbf{P}(X<i)$	$\mathbf{P}(X>i)$	$\mathbf{P}(X\ge i)$
0	$(\frac{5}{6})^{10}$					
1	$10\frac{1}{6}\cdot(\frac{5}{6})^9$					
2	$45(\frac{1}{6})^2\cdot(\frac{5}{6})^8$					
3	$120(\frac{1}{6})^3\cdot(\frac{5}{6})^7$					
4	$210(\frac{1}{6})^4\cdot(\frac{5}{6})^6$					
5	$252(\frac{1}{6})^5\cdot(\frac{5}{6})^5$					
6	$210(\frac{1}{6})^6\cdot(\frac{5}{6})^4$					
7	$120(\frac{1}{6})^7\cdot(\frac{5}{6})^3$					
8	$45(\frac{1}{6})^8\cdot(\frac{5}{6})^2$					
9	$10(\frac{1}{6})^9\cdot\frac{5}{6}$					
10	$(\frac{1}{6})^{10}$					



De kansverdeling voor $X \sim \mathrm{B}(5; 1/36)$:

i	$\mathbf{P}(X = i)$	$\mathbf{P}(X = i)$	$\mathbf{P}(X \leq i)$	$\mathbf{P}(X > i)$	$\mathbf{P}(X < i)$	$\mathbf{P}(X \geq i)$
0	$(\frac{35}{36})^5$					
1	$5\frac{1}{36} \cdot (\frac{35}{36})^4$					
2	$10(\frac{1}{36})^2 \cdot (\frac{35}{36})^3$					
3	$10(\frac{1}{36})^3 \cdot (\frac{35}{36})^2$					
4	$5(\frac{1}{36})^4 \cdot \frac{35}{36}$					
5	$(\frac{1}{36})^5$					



OPGAVEN — 129 Teken zelf de kansverdelingen voor $\mathbf{p} = 1/6$ met $n = 4$ en $\mathbf{p} = 1/36$ met $n = 24$.

Verdere toepassingen en voorbeelden

1. Om de verkoop van een bepaald artikel te stimuleren heeft de fabrikant in 1/4 van de pakjes een ticketje mee verpakt dat recht geeft op een geschenk. Een huisvrouw heeft twee pakjes van het bewuste artikel gekocht. Bereken de kans op 0, 1, 2 geschenken.

   OPLOSSING: $X \sim \mathrm{B}(2; 1/4)$.
   Dus $\mathbf{P}(X = 0) = (\frac{3}{4})^2 = 0,5625$, $\mathbf{P}(X = 1) = 2.\frac{3}{4}.\frac{1}{4} = 0,3750$ en
   $\mathbf{P}(X = 2) = (\frac{1}{4})^2 = 0,0625$.

2. In een stad zijn alle verkeerslichten zo geregeld dat zij 0,4 van de tijd op groen licht staan en 0,6 van de tijd op oranje of rood. Bij doortocht van de stad moet men 9 verkeerslichten voorbijrijden (de lichten zijn onafhankelijk van elkaar ingesteld). We berekenen de kans dat men

   (a) 5 maal zal kunnen doorrijden;

   (b) 7 of meer maal zal kunnen doorrijden;

(c) minder dan 3 maal zal kunnen doorrijden.

OPLOSSING: $X \sim \mathrm{B}(9; 0,4)$ als je groen licht als een succes beschouwd.

(a) De kans op 5 maal groen licht is $\mathbf{P}(X = 5) = C_9^5.(0,4)^5.(0,6)^4 = 0,1672$.
Controleer met de computer.

(b) de kans op 7 of meer maal groen licht is
$\mathbf{P}(X \geq 7) = \mathbf{P}(7) + \mathbf{P}(8) + \mathbf{P}(9) = 0,0212 + 0,0035 + 0,0003 = 0,0250$.
Met de computer maak je gebruik van de cumulatieve binomiale kansverdeling:
$\mathbf{P}(X \geq 7) = 1 - \mathbf{P}(X < 7) = 1 - \mathbf{P}(X \leq 6) = 0,0250$.

(c) de kans op minder dan 3 maal groen licht is
$\mathbf{P}(X < 3) = \mathbf{P}(0) + \mathbf{P}(1) + \mathbf{P}(2) = 0,0101 + 0,0605 + 0,1612 = 0,2318$.
$\mathbf{P}(X < 3) = \mathbf{P}(X \leq 2) = 0,2318$ (met de computer).

3. Levensverzekeringen.
Een verzekeringsmaatschappij sluit met 10 personen een levensverzekering af. De kans dat elk van deze personen na 30 jaar nog zullen leven is 0,6. Bereken de kans dat na 30 jaar

(a) alle verzekerden nog in leven zijn;

(b) minstens 3 personen nog in leven zijn;

(c) nog precies 4 personen in leven zijn.

(d) nog tussen de 3 en de 7 personen in leven zijn.

OPLOSSING:
$X \sim \mathrm{B}(10; 0,6)$ als in leven zijn een succes is.

(a) $\mathbf{P}(X = 10) = (0,6)^{10} = 0,0060$;

(b) Dit is het complement van "hoogstens 2 personen zijn nog in leven".
Dus $\mathbf{P}(X \geq 3) = 1 - (\mathbf{P}(X = 2) + \mathbf{P}(X = 1) + \mathbf{P}(X = 0))$
$= 1 - (0,0106 + 0,0016 + 0,0001) = 0,9877$;
$\mathbf{P}(X \geq 3) = 1 - \mathbf{P}(X < 3) = 1 - \mathbf{P}(X \leq 2) = 0,9877$ (met de computer)

(c) $\mathbf{P}(X = 4) = C_{10}^4.(0,6)^4.(0,4)^6 = 0,1115$,

(d) $\mathbf{P}(3 \leq X \leq 7) = \mathbf{P}(X \leq 7) - \mathbf{P}(X < 3) = \mathbf{P}(X \leq 7) - \mathbf{P}(X \leq 2) = 0,820416\cdots$ (met de computer).

OPGAVEN — **130** Teken zelf voor de drie bovenstaande voorbeelden de kansverdelingen d.i. voor $\mathbf{p} = 0,4$ met $n = 9$, $\mathbf{p} = 0,6$ met $n = 10$ en $\mathbf{p} = 1/4$ met $n = 2$.

131 Een verzekeringsmaatschappij verzekert 5 personen, allen van dezelfde leeftijd. Volgens de sterftecijfers is de kans 2/3 dat een persoon van hun leeftijd nog 30 jaar leeft. Bereken de kans dat na 30 jaar

a. 5 nog in leven zullen zijn;

b. tenminste 3 nog in leven zullen zijn;

c. tenminste 1 nog in leven zal zijn.

132 Op een vrachtwagen werden 6 banden geplaatst. Deze werden genomen uit een zeer grote partij banden waarvan achteraf bleek dat 1/3 der banden aan de vereiste kwaliteitsvoorwaarden niet voldoen. Bereken de kans dat op de wagen 0, 1, 2, ..., 6 slechte banden werden geplaatst.

133 Een lampenfabriek produceert 3% defekte lampen. Bereken de kans dat een onderzoek in een steekproef van 100 lampen (a) 0; (b) 1; (c) 2; (d) 3; (e) 4; (f) 5; (g) meer dan 5; (h) tussen 0 en 4; (i) 0, 1 of 2 lampen defekt zullen zijn.

134 In een bevolkingsgroep hebben 1% van de mensen hooikoorts. Hoeveel lukraak gekozen personen moet men onderzoeken opdat

a. men met meer dan 90% waarschijnlijkheid

b. men met meer dan 99% waarschijnlijkheid

minstens één persoon met hooikoorts zou vinden.

Oplossingen:

131 (a) 0,131687; (b) 0,790123457; (c) 0,164609053; (d) 0,995884;

132 $n = 6$ en $\mathbf{p} = 1/3$; De kansen zijn: $\mathbf{P}(X = 0) = 0,088$; $\mathbf{P}(X = 1) = 0,264$; $\mathbf{P}(X = 2) = 0,329$; $\mathbf{P}(X = 3) = 0,220$; $\mathbf{P}(X = 4) = 0,082$; $\mathbf{P}(X = 5) = 0,016$; $\mathbf{P}(X = 6) = 0,001$.

133 (a) 0,047552508; (b) 0,147069612; (c) 0,225152463; (d) 0,22747; (e) 0,1706; (f) 0,1013; (g) 0,080837; (h) 0,597; (i) 0,42;

134 (a) $n = 230$; (b) $n = 459$.

3.2.3 De Poisson kansverdeling.

3.2.3.1 Definitie en berekening van de Poisson kansverdeling

De **Poisson kansverdeling** is een kansverdeling die geschikt is om een bepaald type van vraagstukken op te lossen. Meestal is in deze problemen de kans op succes evenredig met de tijd. Toepassingsvoorwaarde voor de formule van Poisson:

Succes treedt gemiddeld μ keer op per tijdseenheid;

De kans op succes is recht evenredig met de tijd;

Succes kan geen twee keer tegelijk op hetzelfde tijdstip optreden maar kan wel optreden op elk moment.

Zijn deze voorwaarden vervuld dan is de kans dat succes i keer optreedt gedurende t tijdseenheden gelijk aan

$$\mathbf{P}(X = i) = \frac{e^{-\mu t} \cdot (\mu t)^i}{i!}.$$

Deze formule kan bewezen worden met de differentiaalrekening.

Berekening met de computer

Zoals voor de binomiale verdeling zit ook de Poisson kansverdeling en zijn cumulatieve in Excel.

$$\mathbf{P}(X = i) = \text{POISSON}(i; \mu t; \text{ONWAAR}).$$

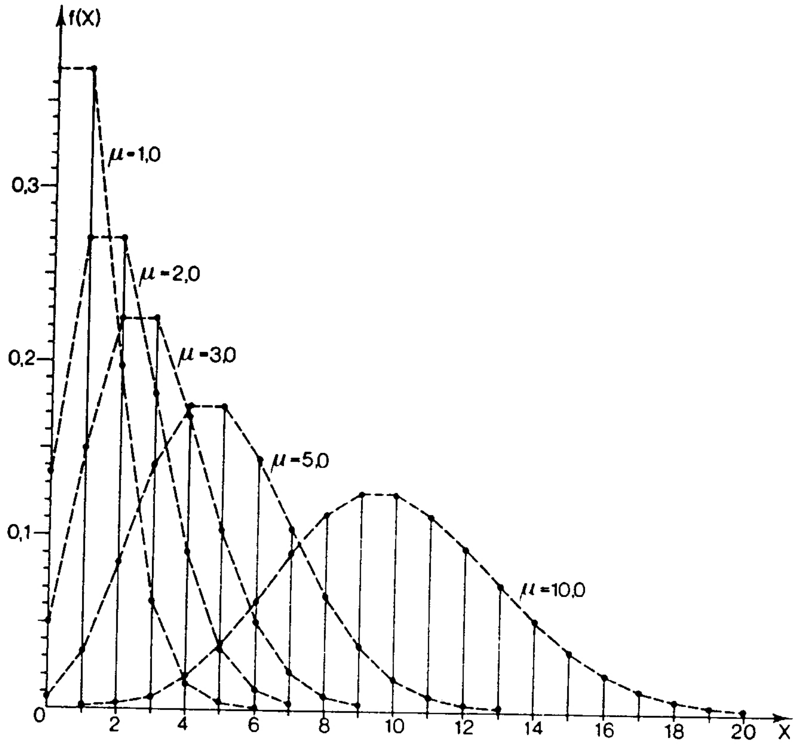
$$\mathbf{P}(X \leq i) = \text{POISSON}(i; \mu t; \text{WAAR}).$$

Voorbeeldjes

- Kristof heeft een advertentie in de krant gezet met zijn telefoonnummer bij en hij wacht nu geduldig op reactie. De kans dat iemand belt is evenredig met de tijd. Kristof zit gedurende acht uren aan de telefoon en heeft tot dusver veertig reacties gehad. Wat is de kans dat hij gedurende het volgende uur geen enkele telefoon meer krijgt?

  Meestal kiezen we onze tijdseenheid zó dat $t = 1$. Kiezen we in ons voorbeeld de tijdseenheid gelijk aan een uur, dan zien we dat Kristof gemiddeld vijf telefoons per uur krijgt. De kans om geen enkele te krijgen in één uur is dan

  $$\mathbf{P}(X = 0) = e^{-5} \cdot 5^0/0! = 0,0067.$$



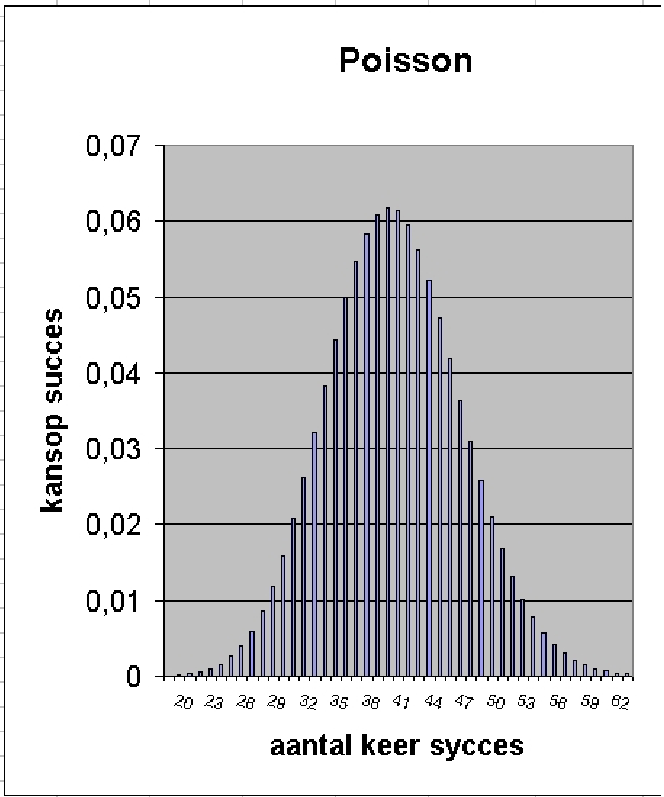
Figuur 3.1: de Poisson kansverdelingen

- Een ander typisch vraagstukje om op te lossen met de Poisson verdeling is het volgende. Suna woont aan een drukke baan dicht bij de Moerbrug over de vaart Gent-Brugge. Vermits ze vroeg klaar is met de voorbereiding van haar examen wiskunde, amuseert ze zichzelf met auto's tellen die passeren. Na juist één uur heeft Suna reeds 2500 wagens geteld, die op één rijvak voorbijgesnord zijn. Wat is de kans dat ze gedurende de volgende minuut er minder dan 20 ziet voorbijrazen op datzelfde rijvak?

  Daar geen twee wagens tegelijk kunnen passeren op één rijvak en er op elk moment een auto kan passeren, is dit een Poisson vraagstukje. We hebben de logische keuze om ofwel de tijdseenheid gelijk te nemen aan 1 uur en dan is $t = \frac{1}{60}$, ofwel nemen we de tijdseenheid onmiddellijk gelijk aan 1 minuut en dan is $t = 1$. In allebei de gevallen is $\mu \cdot t = \frac{2500}{60} = \frac{125}{3}$. De gevraagde kans is dus

$$\mathbf{P}(X < 20) = \mathbf{P}(X \leq 19) = \sum_{i=0}^{19} e^{-\frac{125}{3}} \cdot \frac{\left(\frac{125}{3}\right)^i}{i!} = 0,00007$$

			Tijdsenheid=		1 minuut			
20	8,2044E-05							
21	0,00016279		gemi	41,66667			28,8	54,6
22	0,00030831		standaardafwijking=		6,45			
23	0,00055853							
24	0,00096966							
25	0,00161611							
26	0,00258991							
27	0,00399678							
28	0,00594759							
29	0,00854539							
30	0,01186859							
31	0,01595241							
32	0,02077137							
33	0,02622648							
34	0,03214029							
35	0,03826225							
36	0,04428501							
37	0,04987051							
38	0,05468257							
39	0,05842155							
40	0,06085578							
41	0,06184531							
42	0,06135447							
43	0,05945201							
44	0,05629925							
45	0,05212893							
46	0,04721824							
47	0,04186014							
48	0,03633693							
49	0,03089875							
50	0,02574896							
51	0,02103673							
52	0,01685635							
53	0,01325185							
54	0,01022519							
55	0,00774636							
56	0,00576366							
57	0,0042132							
58	0,00302673							
59	0,00213752							
60	0,00148439							
61	0,00101393							
62	0,0006814							
63	0,00045066							
64	0,0002934							



Suna besluit dat er geen minuut voorbijgaat zonder dat er twintig of meer wagens voorbij haar deur passeren.

OPGAVEN — **135** Tussen 14u en 16u komen in de telefooncentrale van een grote firma gemiddeld 2,5 oproepen per minuut binnen. Bereken de kans dat in een bepaalde minuut tussen 14u en 16u (a) 0; (b) 1; (c) 2; (d) 3; (e) 4 of minder; (f) meer dan 6 oproepen toekomen.

136 Van een grote vervoerfirma worden gemiddeld 2 wagens per jaar door verkeersongevallen vernield. Bereken de kans dat het eerstvolgend jaar (a) 2; (b) meer dan 4 wagens worden vernield.

Oplossingen:

135 (a) 0,082085; (b) 0,20521; (c) 0,25652; (d) 0,21376; (e) 0,8912; (f) 0,01429; 136 (a) 0,271; (b) 0,0526;

3.2.4 Gemiddelde waarde en variantie bij een discrete kansverdeling

$\boxed{\text{Definities}}$

Zij X een discrete toevalsveranderlijke met waarden $x_1, x_2, \ldots$

De **gemiddelde waarde** of de **verwachtingswaarde** $\mathrm{E}(X)$ is de som van de producten van de verschillende mogelijke waarden x_i van X met hun resp. kans

$$\mathrm{E}(X) = x_1 \cdot \mathbf{P}(X = x_1) + x_2 \cdot \mathbf{P}(X = x_2) + \cdots = \sum_i x_i. \cdot \mathbf{P}(X = x_i)$$

Een maat voor de spreiding van de verschillende waarden van de toevallige veranderlijke rond de gemiddelde waarde is de **variantie**. De variantie is de gemiddelde waarde van het kwadraat van de afwijkingen van de verschillende mogelijke waarden x_i van X t.o.v. de gemiddelde waarde $\mathrm{E}(X)$:

$$\mathrm{Var}(X) = \mathrm{E}\left(x_i - \mathrm{E}(X)\right)^2 = \sum_i (x_i - \mathrm{E}(X))^2 \mathbf{P}(X = x_i)$$

Merk op dat steeds geldt dat $\mathrm{Var}(X) \geq 0$.

De **standaardafwijking** is de positieve vierkantswortel uit de variantie.

$$\sqrt{\mathrm{Var}(X)}$$

$\boxed{\text{Eigenschap}}$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$$

Bewijs:

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_i (x_i - \mathrm{E}(X))^2 \mathbf{P}(X = x_i) \\
&= \sum_i (x_i^2 - 2x_i\mathrm{E}(X) + (\mathrm{E}(X))^2)\mathbf{P}(X = x_i) \\
&= \sum_i (x_i^2 \mathbf{P}(X = x_i) - 2x_i\mathrm{E}(X)\mathbf{P}(X = x_i) + (\mathrm{E}(X))^2\mathbf{P}(x_i)) \\
&= \sum_i x_i^2 \mathbf{P}(X = x_i) - 2\mathrm{E}(X)\sum_i x_i \mathbf{P}(X = x_i) + (\mathrm{E}(X))^2 \sum_i \mathbf{P}(X = x_i) \\
&= \mathrm{E}(X^2) - 2\mathrm{E}(X)^2 + (\mathrm{E}(X))^2 \cdot 1 \\
&= \mathrm{E}(X^2) - (\mathrm{E}(X))^2
\end{aligned}$$

Voorbeelden: We hernemen een aantal experimenten uit het vorig hoofdstuk.

- Een doos bevat 10 lampen waarvan twee slechte. We nemen er willekeurig 2 uit. De toevalsveranderlijke X is "het aantal genomen slechte stukken".

x_i	$\mathbf{P}(X = x_i)$	$x_i \cdot \mathbf{P}(X = x_i)$	$(x_i - \mathrm{E}(X))^2 \cdot \mathbf{P}(X = x_i)$	$x_i^2 \cdot \mathbf{P}(X = x_i)$
0	$\frac{28}{45}$	0	$(\frac{2}{5})^2 \cdot \frac{28}{45}$	0
1	$\frac{16}{45}$	$\frac{16}{45}$	$(\frac{3}{5})^2 \cdot \frac{16}{45}$	$\frac{16}{45}$
2	$\frac{1}{45}$	$2 \cdot \frac{1}{45}$	$(\frac{8}{5})^2 \cdot \frac{1}{45}$	$4 \cdot \frac{1}{45}$
$\sum$	1	$\mathrm{E}(X) = \frac{2}{5} = 0,4$	$\mathrm{Var}(X) = \frac{64}{225} = 0,28$	0,44

  Gemiddeld hebben we 0,4 slechte lampen als we twee lampen trekken.
  $\mathrm{Var}(X) = 0,44 - 0,4^2 = 0,28$ waaruit volgt dat de standaardafwijking gelijk is aan is $\sqrt{0,28} = 0,53$.

- In een gezin van 4 kinderen beschouwen we als toevalsveranderlijke het aantal jongens. We krijgen de volgende kansverdeling:

x_i	$\mathbf{P}(X = x_i)$	$x_i \cdot \mathbf{P}(X = x_i)$	$(x_i - \mathrm{E}(X))^2 \mathbf{P}(X = x_i)$	$x_i^2 \cdot \mathbf{P}(X = x_i)$
0	$\frac{1}{16}$	0	$4 \cdot \frac{1}{16}$	0
1	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$
2	$\frac{6}{16}$	$2 \cdot \frac{6}{16}$	0	$4 \cdot \frac{6}{16}$
3	$\frac{4}{16}$	$3 \cdot \frac{4}{16}$	$\frac{4}{16}$	$9 \cdot \frac{4}{16}$
4	$\frac{1}{16}$	$4 \cdot \frac{1}{16}$	$4 \cdot \frac{1}{16}$	$16 \cdot \frac{1}{16}$
$\sum$	1	$\mathrm{E}(X) = 2$	$\mathrm{Var}(X) = 1$	5

  Gemiddeld hebben we 2 jongens in een gezin van 4 kinderen.
  $\mathrm{Var}(X) = 5 - 2^2 = 1$ waaruit volgt dat de standaardafwijking gelijk is aan 1.

- Men gooit met twee dobbelstenen. De toevalsveranderlijke is "de som van de ogen". We krijgen de volgend kansverdeling:

x_i	$\mathbf{P}(X = x_i)$	$x_i \cdot \mathbf{P}(X = x_i)$	$(x_i - \mathrm{E}(X))^2 \mathbf{P}(X = x_i)$	$x_i^2 \cdot \mathbf{P}(X = x_i)$
2	$\frac{1}{36}$	$2 \cdot \frac{1}{36}$	$25 \cdot \frac{1}{36}$	
3	$\frac{2}{36}$	$3 \cdot \frac{2}{36}$	$16 \cdot \frac{2}{36}$	
4	$\frac{3}{36}$	$4 \cdot \frac{3}{36}$	$9 \cdot \frac{3}{36}$	
5	$\frac{4}{36}$	$5 \cdot \frac{4}{36}$	$4 \cdot \frac{4}{36}$	
6	$\frac{5}{36}$	$6 \cdot \frac{5}{36}$	$\frac{5}{36}$	
7	$\frac{6}{36}$	$7 \cdot \frac{6}{36}$	0	
8	$\frac{5}{36}$	$8 \cdot \frac{5}{36}$	$\frac{5}{36}$	
9	$\frac{4}{36}$	$9 \cdot \frac{4}{36}$	$4 \cdot \frac{4}{36}$	
10	$\frac{3}{36}$	$10 \cdot \frac{3}{36}$	$9 \cdot \frac{3}{36}$	
11	$\frac{2}{36}$	$11 \cdot \frac{2}{36}$	$16 \cdot \frac{2}{36}$	
12	$\frac{1}{36}$	$12 \cdot \frac{1}{36}$	$25 \cdot \frac{1}{36}$	
$\sum$	1	$\mathrm{E}(X) = 7$	$\mathrm{Var}(X) = 5,83$	54,83

  De gemiddelde som bij het opgooien van 2 dobbelstenen is 7.
  $\mathrm{Var}(X) = 54,83 - 7^2 = 5,83$ waaruit volgt dat de standaardafwijking gelijk is aan $\sqrt{5,83} = 2,41$.
  Vul zelf de laatste kolom in.

- Bij het experiment 'opgooien van 3 muntstukken' krijgt men 1 EUR als men munt werpt en moet men 1 EUR betalen als men kop werpt. Men beschouwt als toevalsveranderlijke het bedrag dat men wint of verliest na de 3 worpen.
  We krijgen de volgende kansverdeling:

x_i	$\mathbf{P}(X = x_i)$	$x_i \cdot \mathbf{P}(X = x_i)$	$(x_i - \mathrm{E}(X))^2 \mathbf{P}(X = x_i)$	$x_i^2 \cdot \mathbf{P}(X = x_i)$
-3	$\frac{1}{8}$	$-3 \cdot \frac{1}{8}$	$9 \cdot \frac{1}{8}$	
-1	$\frac{3}{8}$	$-1 \cdot \frac{3}{8}$	$\frac{3}{8}$	
1	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	
3	$\frac{1}{8}$	$3 \cdot \frac{1}{8}$	$9 \cdot \frac{1}{8}$	
$\sum$	1	$\mathrm{E}(X) = 0$	$\mathrm{Var}(X) = 3$	

  Gemiddeld heeft men noch winst noch verlies als men 3 muntstukken opgooit. De standaardafwijking is 1,7. Vul zelf de laatste kolom in.

- De roulette.
  Bij de roulette kan men een inzet doen op 37 nummers, van 0 tot 36; 18 nummers zijn rood; 18 zijn zwart; 0 is wit. Wanneer men een bepaald bedrag op een nummer inzet, en dit nummer wint, dan krijgt men 36 maal dit bedrag terug, dus heeft men een winst van 35 maal de inzet. Bij een inzet van 100 EUR bv. kan de winst 3500 EUR bedragen; dit doet zich voor met een kans van 1/37. Men kan ook 100 EUR verliezen met een kans van 36/37.
  De toevalsveranderlijke is hier "het bedrag dat men wint of verliest als men op een bepaald nummer inzet".
  De kansverdeling is:

x_i	$\mathbf{P}(X = x_i)$	$x_i \cdot \mathbf{P}(X = x_i)$
-100	$\frac{36}{37}$	$-100 \cdot \frac{36}{37}$
3500	$\frac{1}{37}$	$3500 \cdot \frac{1}{37}$
$\sum$	1	$\mathrm{E}(X) = -\frac{100}{37}$

  Gemiddeld verliest men 1/37 van zijn inzet.

  Men kan echter ook op een kleur, rood of zwart spelen. Wint de kleur waarop men heeft ingezet, dan trekt men het dubbel van zijn inzet terug.
  De kansverdeling is:

x_i	$\mathbf{P}(X = x_i)$	$x_i \cdot \mathbf{P}(X = x_i$
-100	$\frac{19}{37}$	$-100 \cdot \frac{19}{37}$
100	$\frac{18}{37}$	$100 \cdot \frac{18}{37}$
$\sum$	1	$\mathrm{E}(X) = -\frac{100}{37}$

  Dus hier ook is het gemiddelde verlies 1/37 van de inzet. Andere spelcombinaties geven hetzelfde resultaat.

- Het samennemen van monsters.
  Veronderstel dat van een bepaalde bevolkingsgroep geweten is dat ongeveer 5% van zijn leden door suikerziekte is aangetast. Door een bloedanalyse wil men de zieke personen opsporen. Dit kan gebeuren door een bloedmonster van ieder persoon afzonderlijk te onderzoeken of door samengegoten bloedmonsters te testen. Stel dat we k personen van de bevolkingsgroep willen testen en dat we de k bloedmonsters samenvoegen. Is de uitslag negatief, dan is men zeker dat er geen zieke tussen die k personen is, en men heeft $k-1$ proeven uitgespaard. Is de uitslag echter positief, dan moet men herbeginnen en ditmaal de k monsters afzonderlijk onderzoeken. Men verliest alsdan 1 proef. De vraag is nu hoeveel monsters er moeten samengevoegd worden om zoveel mogelijk proeven uit te sparen.

  De toevalsveranderlijke X is "het aantal te verrichten proeven als k monsters samengenomen worden".
  De kansverdeling is:

x_i	$\mathbf{P}(X = x_i$	$x_i \cdot \mathbf{P}(X = x_i)$
1	$(0,95)^k$	$(0,95)^k$
$k+1$	$1-(0,95)^k$	$(k+1)\cdot(1-(0,95)^k)$
$\sum$	1	$k - k(0,95)^k + 1$

  Het gemiddeld aantal proeven is : $k - k(0,95)^k + 1$.
  Bij individueel onderzoek zijn er k proeven nodig; dus hebben we uitgespaard:

$$k - k + k(0,95)^k - 1 = k(0,95)^k - 1$$

  De uitsparing per eenheid is functie van het aantal monsters k:

$$y = \frac{k(0,95)^k - 1}{k} = (0,95)^k - \frac{1}{k}.$$

  Als we de grafiek tekenen van deze functie zien we dat ze een maximum bereikt in de omgeving van 5.
  We kunnen het maximum ook bepalen door de functie af te leiden:

$$y' = (0,95)^k \ln(0.95) + \frac{1}{k^2}.$$

  Om het nulpunt van y' te bepalen kijken we voor welke waarde van k de grafieken van functies $y = (0,95)^k \ln(0.95)$ en $y = -\frac{1}{k^2}$ elkaar snijden. We zien dat ze elkaar snijden dicht bij de waarde $k = 5$. Als we 5 monsters samenvoegen krijgen we de maximale besparing van $57,4\%$.

OPGAVEN — 137 In België heeft men gemiddeld op 11 dagen: 2 dagen zon, 6 dagen overtrokken en 3 dagen regen. Een ijroomfirma wint 50 EUR per dag bij zonnig weer, 10 EUR bij overtrokken weer en verliest 15 EUR bij regenweer. Bereken voor deze firma de gemiddelde winst per dag.

138 Thomas koopt een doos met 10 spaarlampen. Bij vergissing werden 4 afgekeurde lampen mee verpakt. Thomas neemt lukraak 3 lampen uit de doos. Bereken het gemiddeld aantal slechte lampen dat Thomas bij deze trekking. Bereken ook de standaardafwijking bij deze trekking.

139 Suna en Stephanie spelen een spel en gooien samen 100 keer met een dobbelsteen. Suna zegt tegen Stephanie als we geen 6 gooien dan win jij en als we een 6 gooien dan win ik. Maar aangezien jij meer kans hebt om te winnen krijg jij telkens 1 dollar als je wint maar als ik win moet je 6 dollar geven. Vooralleer Stephanie beslist om mee te spelen berekent ze vlug nog het volgende:

1. de gemiddelde winst of verlies bij 1 keer gooien en bij 100 keer gooien;
2. de standaardafwijking van winst of verlies bij 1 keer gooien en bij 100 keer gooien;
3. haar kans om minstens 50 dollar te winnen;
4. haar kans om minder dan 16 dollar te winnen;
5. de kans dat Stephanie meer wint dan Suna.

OPLOSSINGEN: 137. $\mathrm{E}(X) = 18,64$, $\mathrm{Var}(X) = 223,14$, 138. $\mathrm{E}(X) = 1,2$, $\mathrm{Var}(X) = 0,56$, 139. 1. $\mathrm{E}(X) = -1/6$, $\mathrm{E}(Y) = -16,7$; $\mathrm{Var}(X) = 6,81$, $\mathrm{Var}(Y) = 680,5$; 2. $0,0038$; 3. $0,870$; 4. $0,287$.

3.2.4.1 $\boxed{\text{BASISREGELS}}$ van de gemiddelde waarde en de variantie

STELLING 3.1 *Als de waarden van de toevalsveranderlijke X vermeerderd of vermenigvuldigd worden met een constant bedrag dan geldt:*

$$E(rX + s) = rE(X) + s$$

$$Var(rX + s) = r^2\, Var(X)$$

Voorbeelden:
Gooi je één maal een dobbelsteen en neemt de toevalsveranderlijke waarden aan tussen 1 en 6, dan is de verdeling rechthoekig met gemiddelde en variantie:

$$\mathrm{E}(X) = 3,5$$

$$\mathrm{Var}(X) = \frac{1}{6}(2,5^2 + 1,5^2 + 0,5^2 + 0,5^2 + 1,5^2 + 2,5^2) = 2,917.$$

- Tekent men op ieder zijvlak van de dobbelsteen 2 ogen bij, dan vergroot iedere waarde van X met 2. De gemiddelde waarde vermeerdert met 2 maar de variantie blijft onveranderd (dit geeft een verschuiving van de verdeling).

$$\mathrm{E}(X+2) = \mathrm{E}(X) + 2 = 3.5 + 2 = 5.5$$

$$\mathrm{Var}(X+2) = \mathrm{Var}(X) = 2,917.$$

- Verdrievoudigt men het aantal ogen op ieder zijvlak van de dobbelsteen dan verdrievoudigt X. De gemiddelde waarde en de standaardafwijking worden ook verdrievoudigd. De variantie wordt vermenigvuldigd met het kwadraat van 3 (dit geeft een uitrekking van de verdeling in de richting van de x-as).

$$\mathrm{E}(3X) = 3\mathrm{E}(X) = 3 \cdot 3.5 = 10.5$$

$$\mathrm{Var}(3X) = 9\mathrm{Var}(X) = 9 \cdot 2,917 = 26.253$$

STELLING 3.2 *Als twee toevalsveranderlijken X en Y bij elkaar opgeteld worden dan geldt:*
De gemiddelde waarde van de som is de som van de gemiddelde waarden.

$$E(X \pm Y) = E(X) \pm E(Y)$$

Als bovendien X en Y onafhankelijk zijn dan is de variantie van de som gelijk aan de som van de varianties.

$$Var(X \pm Y) = Var(X) + Var(Y)$$

Voorbeelden:

- We gooien nu met een blauwe en een rode dobbelsteen.
De toevalsveranderlijke X neemt de waarden aan tussen 1 en 6 op de blauwe dobbelsteen en de toevalsveranderlijke Y neemt de waarden aan tussen 1 en 6 op de rode dobbelsteen.
Als we de som maken van de twee rechthoekige verdelingen is $X+Y$ de toevalsveranderlijke die waarden $x+y$ aanneemt van 2 tot en met 12 (zie tabel hieronder).

$x_i \setminus y_j$	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

$x+y$	$\mathbf{P}(X+Y=x+y)$	$(x+y)\cdot\mathbf{P}(X+Y=x+y)$
2	$\frac{1}{36}$	$2\cdot\frac{1}{36}$
3	$\frac{2}{36}$	$3\cdot\frac{2}{36}$
4	$\frac{3}{36}$	$4\cdot\frac{3}{36}$
5	$\frac{4}{36}$	$5\cdot\frac{4}{36}$
6	$\frac{5}{36}$	$6\cdot\frac{5}{36}$
7	$\frac{6}{36}$	$7\cdot\frac{6}{36}$
8	$\frac{5}{36}$	$8\cdot\frac{5}{36}$
9	$\frac{4}{36}$	$9\cdot\frac{4}{36}$
10	$\frac{3}{36}$	$10\cdot\frac{3}{36}$
11	$\frac{2}{36}$	$11\cdot\frac{2}{36}$
12	$\frac{1}{36}$	$12\cdot\frac{1}{36}$
$\sum$	1	$\mathrm{E}(X+Y)=7$

Uit de tabel leiden we af dat de basisregel geldig is:

$$\mathrm{E}(X)+\mathrm{E}(Y)=3,5+3,5=7=\mathrm{E}(X+Y)$$

Ga zelf de basisregel na voor de variantie.

- Kristof draait met een geluksrad verdeeld in vier sectoren I, II, III en IV. De uitkomstenverzameling is $\Omega=\{\mathrm{I, II, III, IV}\}$ met de volgende kansverdeling: $\mathbf{P}(I)=\mathbf{P}(\text{ III})=\frac{1}{8}$, $\mathbf{P}(\text{ II})=\frac{1}{2}$ en $\mathbf{P}(\mathrm{IV})=\frac{1}{4}$. Kristof speelt twee maal met het geluksrad. De eerste keer wint hij het bedrag in EUR dat gelijk is aan 8 keer de kans om op die sector terecht te komen en de tweede keer wint Kristof niets als hij terecht komt op I of II en anders wint hij 1 EUR. Het spel van Kristof levert twee toevalsveranderlijken X en Y op waarvan de waarden in de volgende tabel genoteerd staan.

ω	$X(\omega)$	$Y(\omega)$
I	1	0
II	4	0
III	1	1
IV	2	1

De kansverdelingen van X en Y zijn:

x_i	$\mathbf{P}(X=x_i)$
0	0
1	1/4
2	1/4
3	0
4	1/2

en

y_i	$\mathbf{P}(Y=y_i)$
0	5/8
1	3/8
2	0
3	0
4	0

$\mathrm{E}(X) = 0 + 1/4 + 2/4 + 0 + 2 = 11/4 = 2,75$ en $\mathrm{E}(Y) = 0 + 3/8 + 0 + 0 + 0 = 0.375$
De toevalsveranderlijke $X + Y$ geeft ons de winst na twee maal spelen.
In de tabel hieronder vinden we alle mogelijke waarden $x + y$ van $X + Y$.

$x_i \setminus y_i$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	5
2	2	3	4	5	6
3	3	4	5	6	7
4	4	5	6	7	8

We berekenen de kansen van enkele waarden van $X + Y$:

$$\mathbf{P}(X + Y = 0) = \mathbf{P}(X = 0) \cdot \mathbf{P}(Y = 0) = 0 \cdot \frac{5}{8} = 0$$

$$\mathbf{P}(X + Y = 1) = \mathbf{P}(X = 0) \cdot \mathbf{P}(Y = 1) + \mathbf{P}(X = 1) \cdot \mathbf{P}(Y = 0) = 0 \cdot \frac{3}{8} + \frac{1}{4} \cdot \frac{5}{8} = \frac{5}{32}$$

$$\vdots$$

De kansverdeling van $X + Y$ is:

$x + y$	$\mathbf{P}(X + Y = x + y)$	$(x + y) \cdot \mathbf{P}(X + Y = x + y)$
0	0	0
1	5/32	5/32
2	1/4	1/2
3	3/32	9/32
4	5/16	5/4
5	3/16	15/16
6	0	0
7	0	0
8	0	0
$\sum$	1	3,125

We zien dat $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y) = 2,75 + 0,375 = 3,125$.
Ga zelf de basisregel na voor de variantie.

.

3.2.4.2 De kengetallen bij de binomiale kansverdeling.

We kunnen ons nu ook afvragen wat de gemiddelde waarde van de binomiale verdeling is, met andere woorden, als we het experiment een aantal keren herhalen, hoeveel keer zullen we dan gemiddeld succes hebben? Stel dat we een dobbelsteen 600 maal gooien. Is het dan niet logisch om te beweren dat we gemiddeld toch zo een honderd keer zes zullen gooien? Als we dus een enkelvoudig experiment waarbij de kans op succes $\mathbf{p}$ is n maal uitvoeren, dan zullen we gemiddeld $n\mathbf{p}$ keer succes hebben.

Bewijs: Zij X de toevalsveranderlijke van een Bernoulli experiment met uitkomsten 0 (mislukking) en 1(succes) met de kans op succes gelijk aan $\mathbf{p}$.
Het gemiddelde en de variantie van X zijn:

$$\mathrm{E}(X) = 0 \cdot (1 - \mathbf{p}) + 1 \cdot \mathbf{p} = \mathbf{p}$$

$$\mathrm{Var}(X) = 0^2 \cdot (1 - \mathbf{p}) + 1^2 \cdot \mathbf{p} - \mathbf{p}^2 = \mathbf{p}(1 - \mathbf{p})$$

Maken we de som van n keer dit Bernoulli experiment dan neemt de nieuwe toevalsveranderlijke waarden aan tussen 0 en n. Deze waarden zijn het aantal keer succes bij het uitvoeren van n keer het Bernoulli experiment, dus de waarden van de toevalsveranderlijke Y van de binomiale verdeling $\mathrm{B}(n; \mathbf{p})$.

Volgens de basisregel zijn de gemiddelde waarde en de variantie:

$$\mathrm{E}(Y) = \mathrm{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{E}(X_i) = n\mathbf{p}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}((X_i) = n\mathbf{p}(1 - \mathbf{p})$$

Dit betekent dat we ongeveer twee keer op drie voor n proeven tussen de $n\mathbf{p} - \sqrt{n\mathbf{p}(1 - \mathbf{p})}$ en $n\mathbf{p} + \sqrt{n\mathbf{p}(1 - \mathbf{p})}$ keer succes hebben. □

.

$\boxed{\text{BESLUIT}}$: Bij een binomiale kansverdeling, waarvan de kans op succes gelijk is aan $\mathbf{p}$ bij n proeven

(i) de gemiddelde waarde is $\mathrm{E}(X) = n.\mathbf{p}$

(ii) de standaardafwijking is $\sqrt{Var(X)} = \sqrt{n.\mathbf{p}.(1 - \mathbf{p})}$

3.2.4.3 De kengetallen bij de Poisson kansverdeling

De gemiddelde waarde en variantie bij een Poisson verdeling is

$$\mathrm{E}(X) = \mu \cdot t.$$

$$\mathrm{Var}(X) = \mu \cdot t \Longrightarrow \sqrt{\mathrm{Var}(X)} = \sqrt{\mu \cdot t}.$$

3.2.5 Benadering van een binomiale verdeling door een Poisson verdeling

Voorwaarden om een binomiale verdeling te benaderen door een Poisson verdeling is dat:

$$n \geq 100 \text{ en } \mathbf{p} \leq 0,05$$

In dit geval doen we n proeven per tijdseenheid. Op geen enkel moment kan men meer dan één keer succes hebben en omdat n zeer groot is, kan bij benadering succes elk moment optreden.
Bovendien is het benaderen van een binomiale verdeling door een Poissonverdeling soms handig als we noch n noch $\mathbf{p}$ kennen, maar wel het gemiddeld aantal keer succes per tijdseenheid.

Dat dit in de praktijk voorkomt bewijst het volgende voorbeeldje.

Op een baan tussen twee steden gebeuren er jaarlijks gemiddeld 10 ongevallen. Bereken de kans dat er het komende jaar juist 10 ongevallen zullen zijn. Bereken ook de kans dat er op een bepaald jaar eens geen enkel ongeval zal zijn.

Duidelijkerwijs is dit een binomiaalverdeling. Bij elke rit met de wagen heb je ofwel een accident (succes) ofwel geen (geen succes). Het aantal proeven per jaar is zeker meer dan 100; we mogen inderdaad onderstellen dat er meer dan honderd wagens per jaar rijden op die baan. Anderszijds mogen we ook gerust onderstellen dat de kans op een ongeval kleiner is dan $0,05$ (anders zouden er slechts 200 wagens per jaar rijden en ook dat is een te optimistische onderstelling). Dus we passen Poisson toe met $\mu = 10$ en bekomen:

$$\mathbf{P}(X = 10) = e^{-10} \cdot \frac{10^{10}}{10!} = 0,125 \text{ en } \mathbf{P}(X = 0) = e^{-10} = 0,000045.$$

Voorbeelden

In de volgende voorbeelden worden de binomiale verdelingen benaderd door een Poisson verdeling en worden ze tevens rechtstreeks berekend. We vergelijken de resultaten.

1. Kwaliteitscontrole.
   Aan een productieband worden gemiddeld per 1000 afgewerkte producten 5 defecte stukken gefabriceerd. Het uittesten van alle stukken kost te veel en men verpakt goede en slechte stukken samen in kisten van 200. Bereken de kans dat

   (a) een kist 3 of meer slechte stukken bevat;

   (b) een kist 2 slechte stukken bevat.

   Oplossing: $X \sim \mathrm{B}(n; \mathbf{p}) = \mathrm{B}(200; 0,005)$ en $n\mathbf{p} = 1$

   (a) $\mathbf{P}(X \geq 3) = 1 - \mathbf{P}(X < 3) = 1 - \mathbf{P}(X \leq 2) =$
   $1 - (\mathbf{P}(X = 0) + \mathbf{P}(X = 1) + \mathbf{P}(X = 2)) = 1 - e^{-1}(1 + 1 + 1/2) = 0,0803$;
   met de binomiale verdeling: $\mathbf{P}(X \geq 3) = 1 - \mathbf{P}(X \leq 2) = 0,0798$.

   (b) $\mathbf{P}(X = 2) = \frac{e^{-1}}{2} = 0,1839$;
   met de binomiale verdeling: $\mathbf{P}(X = 2) = 0,1844$

2. Schadelijke nevenreacties bij gebruik van antibiotica.
   Een antibioticum heeft bij 1 op 1000 patiënten een schadelijke nevenreactie. Bereken de kans dat op 2000 inspuitingen

   (a) 3 patiënten de nadelige nevenreactie ondergaan;

   (b) er meer dan 5 patiënten de nevenreactie ondergaan.

   Oplossing: $X \sim \mathrm{B}(n; \mathbf{p}) = \mathrm{B}(2000; 0,001)$ en $n\mathbf{p} = 2$

   (a) $\mathbf{P}(X = 3) = e^{-2} \cdot \frac{2^3}{3!} = 0,18045$;
   met de binomiale verdeling: $\mathbf{P}(X = 3) = 0,18054$

   (b) $\mathbf{P}(X \leq 5)$
   $= \mathbf{P}(X = 0) + \mathbf{P}(X = 1) + \mathbf{P}(X = 2) + \mathbf{P}(X = 3) + \mathbf{P}(X = 4) + \mathbf{P}(X = 5)$
   $= e^{-2}(1 + 2 + 2 + 8/3 + 2/3 + 4/15) = 0,9834$.
   $\mathbf{P}(X > 5) = 1 - \mathbf{P}(X \leq 5) = 1 - 0,9834 = 0,01656$
   $\mathbf{P}(X > 5) = 1 - \mathbf{P}(X \leq 5) = 0,01651$ (met de binomiale verdeling)

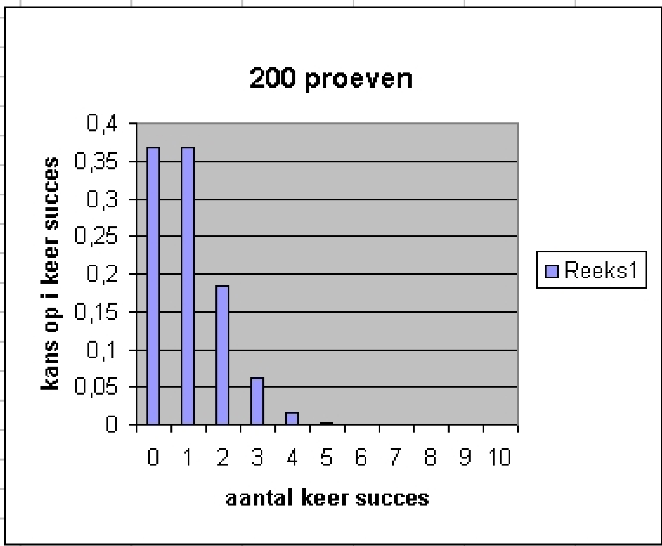
OPGAVEN — 140 Gemiddeld worden 1 op 2000 huizen per jaar door brand vernield. Bereken de kans dat in een gemeente met 6000 huizen er 4 door brand worden vernield in een bepaald jaar. Maak gebruik van de binomiale verdeling en benader door de Poissonverdeling.

Oplossingen:

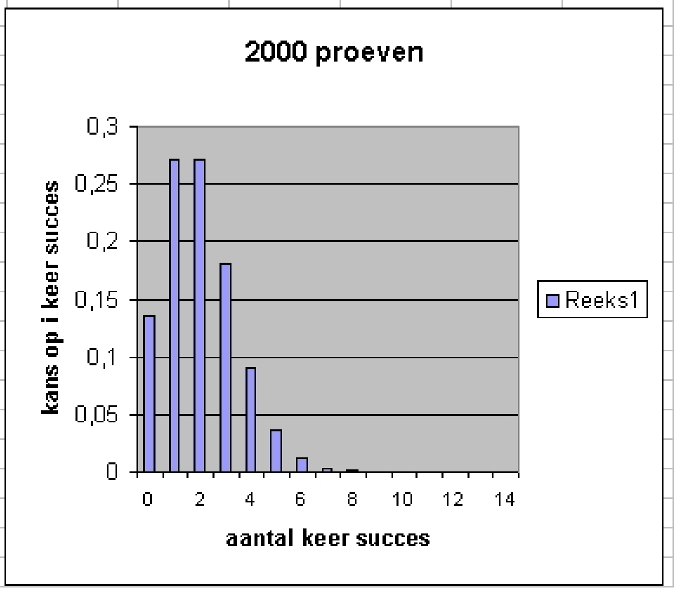
140. Binomiale: $0,168073$, Poisson: $0,168031$;

Binomiale verdelingen

		n=200	kans op succes=	0,005		
0	0,366958					
1	0,368802	gemiddelde=	1		-0,99	2,99
2	0,184401	standaardafwijking=	0,997			
3	0,061158					
4	0,015136					
5	0,002982					
6	0,000487					
7	6,78E-05					
8	8,22E-06					
9	8,81E-07					
10	8,46E-08					



		n=2000	kans op succes=	0,001
0	0,1352			
1	0,270671	gemiddelde=	2	
2	0,270806	standaardafwijking=	1,4	
3	0,180537			
4	0,090223			
5	0,036053			
6	0,012			
7	0,003422			
8	0,000853			
9	0,000189			
10	3,77E-05			
11	6,82E-06			
12	1,13E-06			
13	1,73E-07			
14	2,46E-08			



3.3 Continue kansverdelingen

3.3.1 Definitie van een continue kansverdeling

Soms zijn er echter oneindig veel mogelijke uitkomsten. Voorbeelden daarvan zijn "het schieten op een schijf", "het trekken aan een draad tot hij doorbreekt", "de vertragingstijd van een trein". In deze gevallen zullen we spreken van een **continue kansverdeling**. We geven nog een voorbeeld om te illustreren hoe we ons een voorstelling kunnen maken van zo'n kansverdeling.

Stephanie wordt elke dag gevoerd naar school. Op een dag is de wagen buiten gebruik en moet ze de bus nemen. Er is een bus om de twintig minuten weet ze, maar wanneer precies is haar onbekend. Dus ze gaat op goed geluk af naar de bushalte. Het is duidelijk dat ze evengoed 2 min als 18 min als 7 min als 5 min 33,3 sec kan wachten. Er zijn zo oneindig veel mogelijkheden — als men een continue tijdsopmeting onderstelt. De mogelijke uitkomsten liggen in het interval $[0, 20]$. Hier is het zinloos om van de kans te spreken dat Stephanie juist 3 min 16,678 sec zal moeten wachten op de bus, want die is nagenoeg nul. Juist zoals de kans dat ze *juist* 1 min bijvoorbeeld zal moeten wachten nul is. Juister is het te spreken van de kans dat Stephanie tussen de 3 min en de 5 min zal moeten wachten, want die is één op tien. Dus we zien dat we in zulke gevallen met *intervallen* moeten werken. De kansverdelingsfunctie zoals we die voorheen getekend hebben zou hier ongeveer de constante nul-functie opleveren, doch het is handiger om hier een functie te beschouwen waarvoor de oppervlakte boven een interval I en onder de grafiek precies de kans weergeeft dat een uitkomst in het interval I gelegen is. In ons voorbeeld met Stephanie zou dit dus een functie zijn die overal nul is, uitgezonderd boven het interval $[0, 20]$ waar de constante functiewaarde $1/20$ aangenomen wordt.

$$\mathbf{P}(3 \leq X \leq 5) = \int_3^5 \frac{1}{20} dx = \frac{1}{20}(5-3) = \frac{1}{10}$$

Voor deze functie geldt

$$\int_{-\infty}^{+\infty} \frac{1}{20} dx = 1$$

en bijvoorbeeld is

$$\mathbf{P}(X = 1) = \int_1^1 \frac{1}{20} dx = 0$$

De functie $f(x) = \frac{1}{20}$ wordt een **dichtheidsfunctie** genoemd.

Algemeen: Een functie $f(x)$ is een **dichtheidsfunctie van de continue toevalsveranderlijke** X, die alle waarden kan aannemen in het interval $[a,b]$ als en slechts als

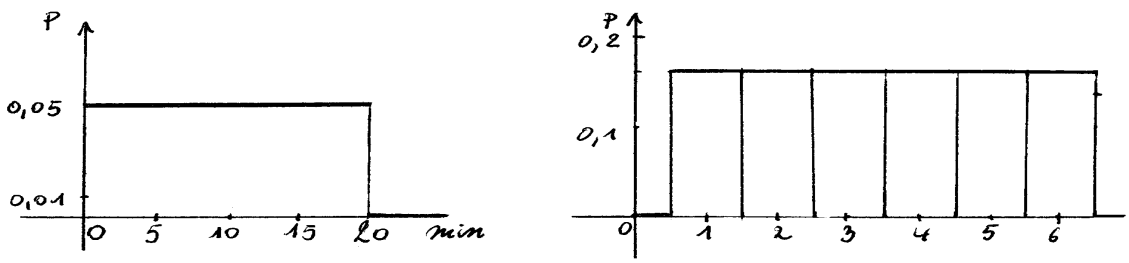
$$\begin{cases} \forall x \in [a,b] : f(x) \geq 0 \\ \\ \int_{-\infty}^{+\infty} f(x) = 1 \end{cases} \quad \text{met } a \in \mathbb{R} \cup \{-\infty\} \text{ en } b \in \mathbb{R} \cup \{+\infty\}$$

De kans dat X een waarde aanneemt gelegen in het interval $[x_1, x_2] \subset [a,b]$ is dan:

$$\mathbf{P}(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)\mathrm{d}x$$

Doordat we nu oppervlakten handig met integralen kunnen berekenen zal men proberen om ook een discrete verdeling te benaderen door een continue functie f zó dat de oppervlakte onder de grafiek van f en boven een bepaald interval op de X-as juist de kans weergeeft dat de uitkomst in dit interval ligt. De grenzen van dit interval moeten dan op een bepaalde manier goed gekozen worden. We geven een voorbeeld.

Stel dat je de discrete kansverdeling van het gooien met een dobbelsteen wilt benaderen door een continue functie. Dan zal je om redenen van symmetrie en voor de eenvoud de bewering *de uitkomst is i* moeten vervangen door *de uitkomst ligt tussen $i - \frac{1}{2}$ en $i + \frac{1}{2}$*. Je neemt dus de eindpunten van de intervallen precies in het midden tussen twee opeenvolgende mogelijke uitkomsten. Dit is een algemene regel die we later vaak zullen toepassen. In ons voorbeeldje is de kansverdelingsfunctie dan een functie die overal nul is, uitgezonderd in het interval $[\frac{1}{2}, \frac{13}{2}]$ waar het de waarde $\frac{1}{6}$ aanneemt. Andere meer ingewikkelde oplossingen zijn natuurlijk ook mogelijk, maar dit is de eenvoudigste en verkiezen we dus boven alle andere (continuïteitscorrectie).



Figuur 3.2: continue kansverdeling en discrete kansverdeling benaderd door een continue

3.3.2 De kengetallen bij een continue kansverdeling

$\boxed{\text{Definitie}}$

Zij X een continue toevalsveranderlijke met waarden $x \in [a,b]$ ($a < b$ en $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$) en zij $f(x)$ de **dichtheidsfunctie** van X dan is

$$\mathrm{E}(X) = \int_a^b x \cdot f(x)\mathrm{d}x$$

en

$$\mathrm{Var}(X) = \int_a^b (x - \mathrm{E}(X))^2 \cdot f(x)\mathrm{d}x.$$

$\boxed{\text{Eigenschap}}$

$$\mathrm{Var}(X) = \int_a^b x^2 \cdot f(x)\mathrm{d}x - (\mathrm{E}(X))^2.$$

Bewijs deze eigenschap zelf.

De basisregels voor gemiddelde en variantie gelden ook voor continue kansverdelingen.

Voorbeeld: Zij X een continue s.v. met waarden in het interval $[0,1]$ en met dichtheidsfunctie

$$f(x) = \begin{cases} 2x & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

1. Bewijs dat f werkelijk een dichtheidsfunctie is;

2. Bepaal de kengetallen bij deze continue kansverdeling.

OPLOSSING:

1. $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_0^1 2x\mathrm{d}x = 1$;

2. $$\mathrm{E}(X) = \int_0^1 2x\mathrm{d}x = \int_0^1 x \cdot 2x\mathrm{d}x = \int_0^1 2x^2 = \frac{2}{3}$$

$$\mathrm{Var}(X) = \int_0^1 x^2 \cdot 2x\mathrm{d}x - (\frac{2}{3})^2 = \int_0^1 2x^3\mathrm{d}x - (\frac{2}{3})^2 = \frac{1}{18}$$

3.3.3 De normale kansverdeling

Een *normale kansverdeling* is zoals de benaming zelf zegt, de verdeling die normaal de beste benadering is voor de kansbeschrijving bij de meeste experimenten. Een normale verdeling is de belangrijkste verdeling in de theorie van de waarschijnlijkheid en de statistiek. De reden is dat, als men een bepaalde waarschijnlijkheidsverdeling heeft bij een bepaald experiment, men zeer veel keer dat experiment herhaald en men alle uitkomsten optelt, dan is de bekomen verdeling benaderend een normale verdeling.
Bijvoorbeeld, gooi je één maal een dobbelsteen dan neemt de toevalsveranderlijke waarden aan tussen 1 en 6. De kansen op de verschillende uitkomsten zijn gelijk aan elkaar. Gooi je 100 maal een dobbelsteen en tel je de bekomen resultaten op, dan kan de waarde van die som liggen tussen 100 en 600, maar natuurlijk zul je meer kans hebben op 350 dan op 100 of 600. De toevalsveranderlijke die we hier beschouwen bij het 100 keer opgooien van een dobbelsteen, heeft een kansverdeling die bij benadering een normale verdeling is.

In het voorgaande voorbeeld is de oorspronkelijke verdeling rechthoekig en we maken de som van 100 rechthoekige verdelingen. Alle verdelingen hoeven echter niet dezelfde wet te volgen. Toch bekomen we benaderend een normale verdeling. Bijvoorbeeld, in de foutentheorie zijn er verschillende oorzaken die fouten opleveren, alle met hun eigen wetten. De som van al deze foutjes is de fout op het eindresultaat en is dus benaderend normaal verdeeld.

3.3.3.1 Definitie van een normale kansverdeling

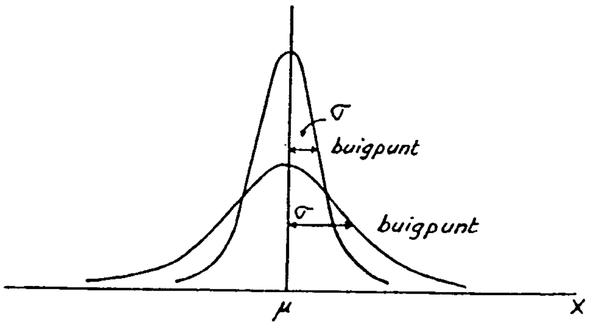
Een continue toevalsveranderlijke X heeft een **normale kansverdeling met parameters μ en σ^2** ($-\infty < \mu < +\infty$ en $0 < \sigma < +\infty$) indien haar dichtheidsfunctie gegeven wordt door:

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2}.$$

Met symbolen: $X \sim \mathrm{N}(\mu; \sigma^2)$

Opmerking:
Bewijs met graphmatica dat $f(x)$ inderdaad een dichtheidsfunctie is.
Het wiskundig bewijs (in hogere wiskunde) bestaat erin te bewijzen dat

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\mathrm{d}x = \sqrt{2\pi} \iff \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\mathrm{d}x = 1 \tag{3.1}$$



Figuur 3.3: normale kansverdelingen

STELLING 3.3 *De grafiek van de dichtheidsfunctie $f(x)$ van een toevalsveranderlijke $X \sim N(\mu; \sigma^2)$ vertoont de volgende eigenschappen:*

(1) De rechte $x = \mu$ is een as van symmetrie voor de grafiek van f;

(2) De x-as is een horizontale asymptoot voor de grafiek van $f(x)$;

(3) f bereikt juist één relatief maximum voor $x = \mu$;

(4) De grafiek van f bezit twee buigpunten, nl. voor $x = \mu \pm \sigma$.

Bewijs: Bewijs deze stelling zelf.

De grafiek van een normale dichtheidsfunctie wordt wel eens de "**klok van Gauss**" genoemd. Het gedeelte tussen ruwweg $\mu - 2\sigma$ en $\mu + 2\sigma$ is het middengedeelte, de rest bestaat uit de twee staarten.

STELLING 3.4 *De symbolen μ en σ zijn respectievelijk de gemiddelde waarde en de standaardafwijking van de normale verdeling*

Bewijs: Steun op 3.1 en gebruik de substitutie $z = \frac{x-\mu}{\sigma}$ om aan te tonen dat $\mathrm{E}(X) = \mu$ en $\mathrm{Var}(X) = \sigma^2$.

3.3.3.2 Berekening met de gecumuleerde normale kansverdeling

Is een toevalsveranderlijke $X \sim \mathrm{N}(\mu;\sigma^2)$ dan is de kans dat ze een waarde aanneemt gelegen in het interval $[x_1, x_2]$ gelijk aan:

$$\mathbf{P}(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2} \mathrm{d}x.$$

Deze integraal is echter niet te berekenen door middel van de methodes die we gezien hebben. We voeren de berekeningen uit met de computer met de gecumuleerde normale verdelingsfunctie.

$$\mathbf{P}(X \leq x') = \int_{-\infty}^{x'} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2} \mathrm{d}x = \text{NORM.VERD}(x;\mu;\sigma;\text{WAAR}).$$

Hieruit volgt dat

$$\begin{aligned}\mathbf{P}(x_1 \leq X \leq x_2) &= \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2} \mathrm{d}x \\ &= \textstyle\int_{x_1}^{-\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2} \mathrm{d}x + \int_{-\infty}^{x_2} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2} \mathrm{d}x \\ &= \textstyle\int_{-\infty}^{x_2} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2} \mathrm{d}x - \int_{-\infty}^{x_1} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2} \mathrm{d}x \\ &= \text{NORM.VERD}(x_2;\mu;\sigma;\text{WAAR}) - \text{NORM.VERD}(x_1;\mu;\sigma;\text{WAAR}).\end{aligned}$$

Voorbeelden

1. Procent defecte stukken.
   De doormeter van wrijvingsringen door een machine vervaardigd is normaal verdeeld met een gemiddelde van 14,04 mm en een standaardafwijking van 0,14 mm. De toegelaten speling rond het gemiddelde bedraagt 0,18 mm. Bereken het % defecte stukken door de machine voortgebracht.

   OPLOSSING:
   De doormeter in mm is de toevalsveranderlijke $X \sim \mathrm{N}(14,04; 0,14^2)$ daar de fout op een stuk van vele factoren afhangt. Wegens de symmetrie van de normale dichtheidsfunctie is de fractie defecte stukken gelijk aan

   $$2 \cdot \mathbf{P}(X \leq (14,04 - 0,18)) = 2 \cdot \mathbf{P}(X \leq 13,86) = 0,1990.$$

   Het gevraagde percentage is dus $19,9\% \sim 20\%$.

2. Om een last te dragen gebruikt men in parallel twee kabels, één uit een reeks met een gemiddeld draagvermogen van 4000 kg en standaardafwijking van 400 kg en een tweede uit een reeks met een gemiddeld draagvermogen van 3000 kg en standaardafwijking van 300 kg. Bereken de kans dat het draagvermogen van de twee kabels samen meer dan 8000 kg bedraagt.

   OPLOSSING:
   Het draagvermogen van de twee kabels is een toevalsveranderlijke X met

   $$\mathrm{E}(X) = 4000 + 3000 = 7000 \qquad \text{en} \qquad \mathrm{Var}(X) = 400^2 + 300^2 = 250000$$

   $X \sim \mathrm{N}(7000; 250000)$

   $$\mathbf{P}(X \geq 8000) = 1 - \mathbf{P}(X \leq 8000) = 0,0228$$

3. Automatische vulling van flessen.
   In een fabriek worden flessen automatisch gevuld. De inhoud van de gevulde flessen heeft een normale verdeling met een gemiddelde van 500 cc en een standaardafwijking van 5 cc. De fabriek wenst te weten wat precies de inhoud kan zijn van 90% van de flessen met een inhoud rond 500 cc.

   OPLOSSING:
   De inhoud in cc is de toevalsveranderlijke $X \sim \mathrm{N}(500; 5^2)$. We zoeken de waarde van x waarvoor

   $$\mathbf{P}(X \leq x) = 0,05.$$

   Daarvoor kunnen we gebruik maken van de *inverse functie van de gecumuleerde normale kansverdelingsfunctie.*

   $$x = \mathrm{NORM.INV}(0,05; 500; 5) = \cdots$$

   Het gevraagde interval is $[x, 2\mu - x] = [\cdots\cdots ; \cdots\cdots]$.
   De inhoud van de 90% van de flessen die een inhoud hebben rond 500 cc is gelegen tussen $\cdots\cdots$ cc en $\cdots\cdots$ cc.

OPGAVEN — **141** Men schakelt 4 akkumulatoren in serie. Zij werden genomen uit een reeks met een gemiddelde spanning van 15 V en een standaardafwijking van 0.2 V. Bereken de kans dat de gecombineerde spanning 60.4 V of meer bedraagt.

142 Men weegt eerst een vat vloeistof en men vindt 103 g met een standaardafwijking van 1 g. Men weegt daarna het ledige vat en men vindt 43 g met een standaardafwijking van 1 g. Bereken de kans dat het gewicht van de vloeistof ligt tussen 59 g en 61 g.

OPLOSSINGEN: 141. 0,1587; 142. 0,5205.

3.3.4 Benadering van een discrete kansverdeling

3.3.4.1 Enkele voorbeelden

1. "Waar of vals"-vragen.
   In een massale test werden 10 "waar of vals-vragen" gesteld. Elke kandidaat bekomt aldus een cijfer tussen 0 en 10. De gemiddelde uitslag bedraagt 6,7 en de standaardafwijking 1,2.
   Bereken:

   (a) het procent kandidaten dat 6 punten behaalt;

   (b) het theoretische interval $[a, b]$ symmetrisch om 6,7 waartussen 80% van de uitslagen vallen.

   Oplossing:

   (a) De toevalsveranderlijke X is hier discontinu, om de continue normale verdeling hierop toe te passen, is het nodig de veranderlijke als continu te beschouwen $X \sim \mathrm{N}(6,7; 1,2^2)$. Dus 6 halen betekent in feite tussen de $5,5$ en $6,5$ halen. We verkrijgen voor de kans om 6 te halen

   $$\mathbf{P}(5,5 \leq X \leq 6,5) = \mathbf{P}(X \leq 6,5) - \mathbf{P}(X \leq 5,5) = 0,2754$$

   dus $27,5\%$ behaald 6/10.

   (b) Bij de gevraagde 80% vinden we a en b als volgt:

   $$\mathbf{P}(X \leq a) = 0,1 \Longleftrightarrow a = \mathrm{NORM.INV}(0,1; 6,7; 1,2) = 5,164$$

   en

   $$\mathbf{P}(X \leq b) = 0,9 \Longleftrightarrow b = \mathrm{NORM.INV}(0,9; 6,7; 1,2) = 8,236.$$

   Practisch betekent dit dat ongeveer 80% van de uitslagen 5, 6, 7 of 8 zijn.

2. Voor de som van honderd worpen met een dobbelsteen neemt de toevalsveranderlijke waarden aan tussen 100 en 600. De nieuwe kansverdeling is de som van 100 rechthoekige verdelingen en is bij benadering een normale verdeling. De gemiddelde waarde en variantie berekenen we met de basisregels:

   $$\mathrm{E}(\sum_{i=1}^{100} X_i) = \sum_{i=1}^{100} \mathrm{E}(X_i) = 100 \cdot 3,5 = 350$$

   $$\mathrm{Var}(\sum_{i=1}^{100} X_i) = \sum_{i=1}^{100} \mathrm{Var}(X_i) = 100 \cdot 2,91 = 291,7$$

dus de standaardafwijking is $\sqrt{291,7} = 17$.
De toevalsveranderlijke is discreet maar we benaderen ze door de continue toevalsveranderlijke $X \sim \mathrm{N}(350; 291,7)$.
De kans dat we de som 250 werpen wordt nu benaderd door:

$$\mathbf{P}(249,5 \leq X \leq 250,5) = \mathbf{P}(X \leq 250,5) - \mathbf{P}(X \leq 249,5) = 7,2 \cdot 10^{-10}$$

en de kans dat we een waarde bekomen tussen 330 en 400 is benaderend:

$$\mathbf{P}(329,5 \leq X \leq 400,5) = \mathbf{P}(X \leq 400,5) - \mathbf{P}(X \leq 329,5) = 0,99819 - 0,12507 = 0,87312.$$

3.3.4.2 Benadering van de binomiaalverdeling voor grote n

Men mag de binomiale verdeling benaderen door de normale met zelfde gemiddelde waarde en standaardafwijking. Dit levert een goed resultaat op van zodra $n \cdot \mathbf{p} > 5$ voor $\mathbf{p} \leq \frac{1}{2}$ en $n \cdot (1 - \mathbf{p}) > 5$ voor $\mathbf{p} \geq \frac{1}{2}$. Door de regel is de benadering juist op zeker twee beduidende cijfers op voorwaarde dat je in het midden blijft en niet te veel in de staart. Daar heb je zekerheid over de eerste twee cijfers na de komma, maar die zijn toch altijd 00 of 99.

Voorbeelden:

1. Als eerste voorbeeldje komen we even terug op de kans op juist tien ongevallen in een bepaald jaar op de weg tussen de twee steden. We lossen het vraagstuk op door de Poisson verdeling te benaderen door de normale verdeling $X \sim \mathrm{N}(10; 10)$. Dit kan want $n\mathbf{p} > 5$ en $p \leq 0,5$. De kans dat er in een bepaald jaar juist 10 ongevallen gebeuren is
$$\mathbf{P}(9,5 \leq X \leq 10,5) = 0,1255$$
hetgeen overeenkomt met de waarde die we vonden zonder te benaderen. De kans dat er dat jaar geen enkel ongeval gebeurt is
$$\mathbf{P}(-0,5 \leq X \leq 0,5) = 0,00034.$$
We zien dus dat de benadering in de staarten slechter is dan in het midden. Dit kan je ook zien als je het vraagstukje in verband met Suna's hobby van auto's tellen oplost door benadering door de normale verdeling. Doe dit zelf.

2. Produktieuitval.
Een machine produceert per dag 500 stuks. Het uitvalspercentage bedraagt gemiddeld 20%. Bereken de kans dat op een bepaalde dag er minder dan 80 slechte stukken werden vervaardigd.

   OPLOSSING:
Dit is een binomiaalverdeling met $n = 500$ en $\mathbf{p} = 0,2$. We benaderen deze door de

normale verdeling $X \sim \mathrm{N}(100; 80)$. Dit kan want $n \cdot \mathbf{p} = 100 > 5$ en $\mathbf{p} \leq 0,5$. De gevraagde kans is dus

$$\mathbf{P}(X \leq 79,5) = \cdots$$

3. Als we tweehonderd keer met de dobbelsteen werpen dan is de toevalsveranderlijke binomiaal verdeeld met

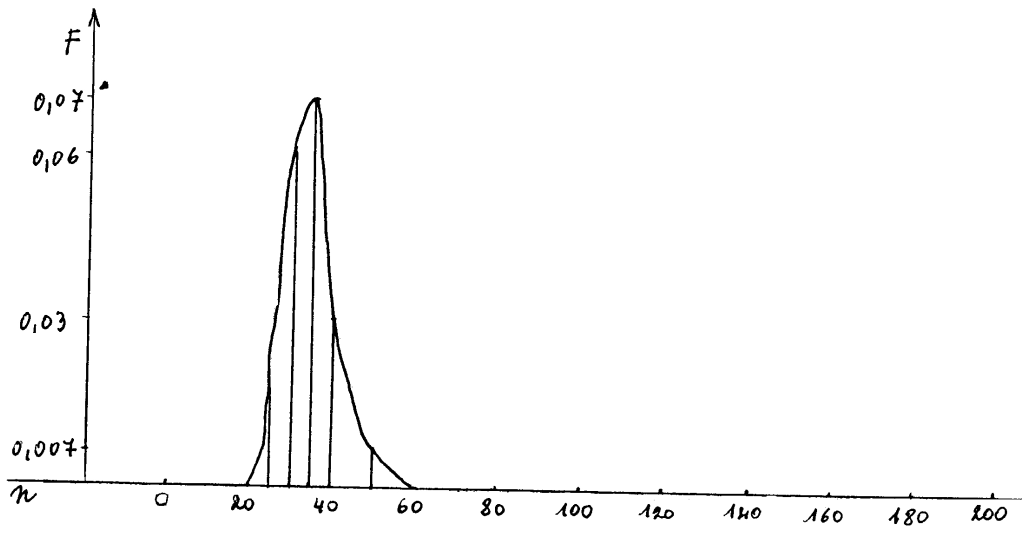
$$\mathrm{E}(X) = n \cdot \mathbf{p} = \frac{200}{6} = \frac{100}{3}$$

$$\mathrm{Var}(X) = n \cdot \mathbf{p} \cdot (1 - \mathbf{p}) = \frac{200}{6} \cdot \frac{5}{6} = \frac{250}{9}$$

Deze binomiale verdeling kan benaderd worden door een normale verdeling $X \sim \mathrm{N}(\frac{100}{3}; \frac{250}{9})$. Dit kan want $\mathbf{p} = 0,16 \cdots \leq 0,5$ en $n \cdot \mathbf{p} > 5$.
De kans dat we tussen de vijftig en de honderd keer een 6 gooien wordt benaderd gegeven door

$$\mathbf{P}(50,5 < X < 99,5) = \mathbf{P}(X \leq 99,5) - \mathbf{P}(X \leq 50,5) = 0,00056$$

Bereken deze kans tevens rechtstreeks met de binomiale verdeling en vergelijk.



Figuur 3.4: de binomiale kansverdeling voor $\mathbf{p} = 1/6$ en n=200

4. Aantal malen groen licht aan verkeerslichten.
Een autobuslijn loopt over 15 kruispunten waar verkeerslichten zijn geplaatst. Aangenomen dat de kans om bij groen licht aan te komen 0,5 bedraagt, vraagt men de kansverdeling te berekenen om 0, 1, 2, ..., 15 maal bij groen licht aan te komen.

Maak deze berekening door middel van de binomiale verdeling en door middel van de normale verdeling.

Oplossing:

(a) Volgens de binomiale verdeling is de kans om i maal groen licht te hebben $C_{15}^{i}/2^{15}$. Voor $i = 0, 1, 2$ geeft dit respectievelijk $0,00003052$, $0,0004578$ en $0,0032043$. Reken de andere waarden zelf uit.

(b) Volgens de normale verdelingswet moeten we $\mu = 15 \times 0,5 = 7,5$ nemen en $\sigma = \sqrt{15 \times 0,5 \times 0,5} = 1,9365$ ($n \cdot \mathbf{p} > 5$ en $\mathbf{p} \leq 0,5$).
$X \sim \mathrm{N}(7,5; 3,75)$
De kansen om 0, 1 en 2 groene verkeerslichten tegen te komen zijn resp.

$$\mathbf{P}(-0,5 \leq X \leq 0,5) = \cdots$$

$$\mathbf{P}(0,5 \leq X \leq 1,5) = \cdots$$

$$\mathbf{P}(1,5 \leq X \leq 2,5) = \cdots$$

We zien dus dat de benadering in de staart relatief slecht is, maar dit wordt beter en beter als we naar het midden toe komen.

5. Thomas moet van de dokter pilletjes nemen. Het gewicht van de pilletjes is normaal verdeeld met gemiddelde waarde 3 gram en standaardafwijking $0,2$ gram. Door de ernst van de situatie moet Thomas 10 pilletjes per dag slikken gedurende 10 dagen. Hij koopt daarom een maxibox van 100 pillen. Het doosje zelf weegt 27 gram, de bijsluiter 19 gram.

   (a) Bepaal de kans dat Thomas's maxibox meer weegt dan 350 gram.

   (b) Bepaal de kans dat, van de tien pilletjes die Thomas de eerste dag inneemt, er juist 4 zijn waarvan het gewicht begrepen is tussen $2,86$ gram en $3,16$ gram.

   (c) Bepaal benaderend de kans dat van de honderd pilletjes in 't totaal er tussen de dertig en de veertig minder wegen dan $2,86$ gram.

   Oplossing:

   (a) We moeten dus de kans zoeken dat de som van de gewichten van honderd Thomas-pilletjes groter is dan $350 - 27 - 19 = 304$ gram.
   Het gewicht van een pilletje is de toevalsveranderlijke $X \sim N(3; 0,2^2)$.
   We maken de som van 100 keer de normale kansverdeling X en we verkrijgen de normale verdeling $Y = \sum_{i=1}^{100} X_i$ waarvoor geldt:

$$\mathrm{E}(Y) = \mathrm{E}(\textstyle\sum_{i=1}^{100} x_i) = \sum_{i=1}^{100} \mathrm{E}(x_i) = 100 \cdot 3 = 300$$
$$\mathrm{Var}(Y) = \mathrm{Var}(\textstyle\sum_{i=1}^{100} x_i) = \sum_{i=1}^{100} \mathrm{Var}(x_i) = 100 \cdot 0,2^2 = 2^2$$

$Y \sim \mathrm{N}(300; 4)$

$$\mathbf{P}(Y > 304) = 1 - \mathbf{P}(Y \leq 304) = 1 - 0,97725 = 0,02275.$$

De kans dat de maxibox van Thomas meer weegt dan 350 gram is 2,3%.

(b) De kans dat een pilletje tussen de $2,86$ en $3,16$ gram weegt is

$$\mathbf{P}(2,86 \leq X \leq 3,16) = \mathbf{P}(X \leq 3,16) - \mathbf{P}(X \leq 2,86) = 0,546181.$$

De kans dat juist vier pilletjes op tien zulk een gewicht hebben wordt gegeven door de binomiaalverdeling $Y \sim \mathrm{B}(10; 0,546)$.

$$C_{10}^{4} \cdot (0,546181)^4 \cdot (1 - 0,546181)^6 = 0,16325.$$

(c) De kans dat een pilletje minder weegt dan $2,86$ gram is

$$\mathbf{P}(X \leq 2,86) = 0,241964.$$

De kans dat tussen de dertig en de veertig daaraan voldoen wordt gegeven door de binomiaalverdeling $Y \sim B(100; 0,241964)$.

$$\mathbf{P}(31 \leq Y \leq 39) = \cdots$$

Benaderend is deze binomiaalverdeling een normale verdeling

$$V \sim \mathrm{N}(100\mathbf{p}; 100\mathbf{pq}) = \mathrm{N}(24,1964; 4,282726^2)$$

$$\mathbf{P}(31 \leq V \leq 39) = 0,9998215 - 0,9294896 = 0,0703319.$$

De uitkomsten hoeven niet op zoveel cijfers nauwkeurig gegeven te worden, maar we doen het omdat je dan gemakkelijker kunt controleren of jouw uitkomst juist is. Als we afronden, dan bekom je misschien door afronding van ook jouw uitkomst per ongeluk hetzelfde, terwijl het misschien verkeerd is.

OPGAVEN — **143** Een krantenjongen verliest geld als hij minder dan 50 kranten per dag verkoopt. De kans dat hij minder dan 50 kranten per dag verkoopt is 0,1. Wat is de kans dat de krantenjongen in de maand juni niet meer dan 3 dagen verlies lijdt.

OPLOSSING: 143. met de binomiale verd: 0,6474, met de normale verd: 0,6198.

3.3.5 De standaard normale kansverdeling

Voor elke waarde van μ en σ^2 krijgen we een andere normale verdeling. In deze paragraaf zien we dat alle normale verdelingen herleid kunnen worden naar één enkele normale verdeling die onafhankelijk is van μ en σ^2.

3.3.5.1 Definitie van de standaard normale kansverdeling

De **standaard normale verdeling** is de normale verdeling met gemiddelde 0 en variantie 1. Haar dichtheidsfunctie is gegeven door (zie 3.1):

$$f : \mathbb{R} \longrightarrow \mathbb{R} : z \longmapsto \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot z^2}$$

STELLING 3.5 *Elke normale kansverdeling $X \sim N(\mu;\sigma^2)$ is te herleiden naar de standaard normale kansverdeling $Z \sim N(0;1)$ mits de substitutie $Z = \frac{X-\mu}{\sigma}$ toe te passen.*

Bewijs: Er bestaat een affiene transformatie die de grafiek van een normale dichtheidsfunctie afbeeldt op de grafiek van de standaard normale dichtheidsfunctie. Om die affiene transformatie te bepalen brengen we het voorschrift van de normale dichtheidsfunctie in een andere gedaante.

$$y = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2}$$

$$\Updownarrow$$

$$y \cdot \sigma = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot(\frac{(x-\mu)}{\sigma})^2}.$$

De vergelijking komt in de gedaante

$$y' = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot x'^2}$$

mits het toepassen van de volgende transformatieformules:

$$\begin{cases} x' &= \frac{(x-\mu)}{\sigma} \\ y' &= y \cdot \sigma \end{cases}$$

Stellen we $x' = z$ en $X' = Z$ dan kunnen we hieruit besluiten dat de dichtheidsfunctie van Z gelijk is aan $\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot z^2}$, dus

$$Z \sim \mathrm{N}(0;1).$$

□

Opmerking: De formules 3.3.5.1 zijn de transformatieformules van een affiene transformatie, dit is hier de verschuiving met vector $(-\mu, 0)$ gevolgd door de lineaire transformatie met matrix

$$\begin{pmatrix} \frac{1}{\sigma} & 0 \\ 0 & \sigma \end{pmatrix}$$

Als $\sigma > 1$ dan is deze lineaire transformatie de samenstelling van een inkrimping in de richting van de x-as met factor σ gevolgd door een uitrekking in de richting van de y-as met factor σ . Is $\sigma < 1$ dan is het de samenstelling van een uitrekking in de richting van de x-as gevolgd door een inkrimping in de richting van de y-as met factor $\frac{1}{\sigma}$.
Voor de 8-uurs brengen we deze formules in matrixgedaante.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma} & 0 \\ 0 & \sigma \end{pmatrix} \cdot \begin{pmatrix} x-\mu \\ y \end{pmatrix}$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma} & 0 \\ 0 & \sigma \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{1}{\sigma} & 0 \\ 0 & \sigma \end{pmatrix} \cdot \begin{pmatrix} -\mu \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma} & 0 \\ 0 & \sigma \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -\frac{\mu}{\sigma} \\ 0 \end{pmatrix}$$

Deze transformatieformules drukken de samenstelling uit van de lineaire transformatie gevolgd door de verschuiving met vector $(-\frac{\mu}{\sigma}, 0)$ in deze volgorde.

3.3.5.2 De gecumuleerde standaard normale kansverdeling

We noemen $\Psi(z)$ de primitieve functie van $\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot z^2}$ waarvoor geldt

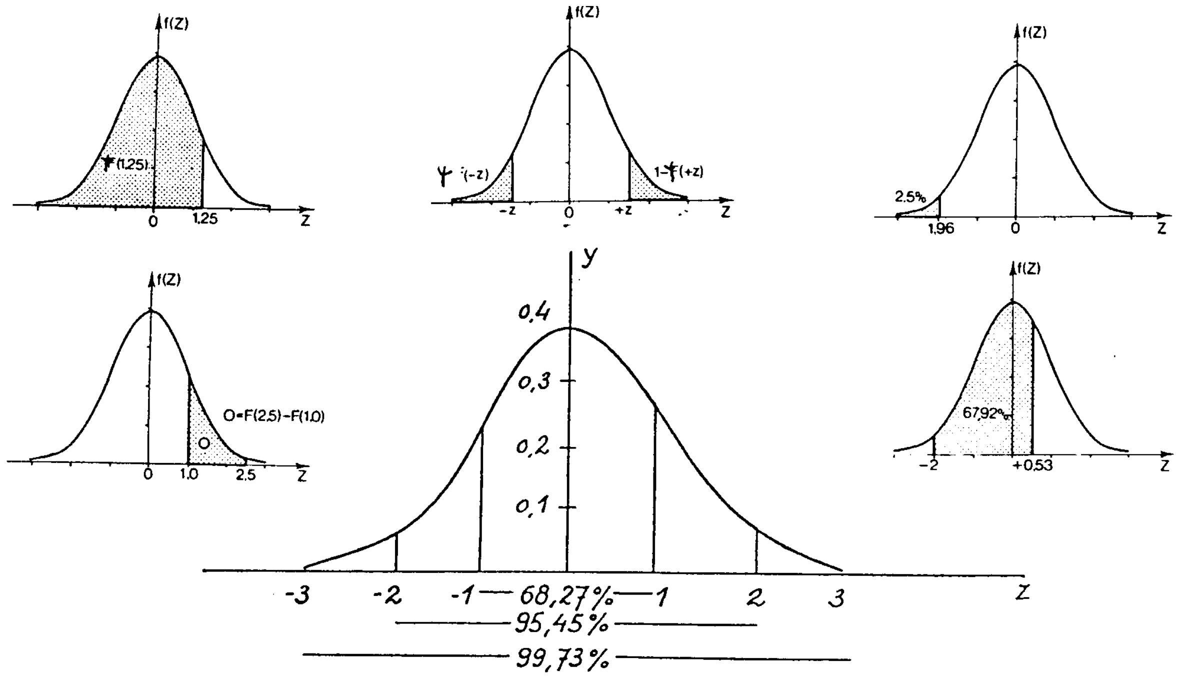
$$\Psi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot t^2} \mathrm{d}t,$$

d.w.z. dat $\Psi(z)$ de oppervlakte voorstelt onder de grafiek van standaard normale dichtheidsfunctie van $-\infty$ tot z. De functie $\Psi(z)$ is de **gecumuleerde standaard normale kansverdelingsfunctie**. Er geldt:

$$\mathbf{P}(Z \leq z) = \Psi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot z^2} \mathrm{d}z = \text{STAND.NORM.VERD}(z),$$

Is de kans p gegeven dan kan men z bepalen door gebruik te maken van de inverse van de gecumuleerde standaardnormale verdeling.

$$\Psi(z) = p \Longleftrightarrow z = \Psi^{-1}(p) = \text{STAND.NORM.INV}(p)$$



Figuur 3.5: de standaard normale verdeling

3.3.5.3 Standaardeenheden

Voor de waarneming x van een gebeurtenis die een normale verdeling volgt met gemiddelde waarde μ en standaardafwijking σ, wordt de grootheid $\frac{x-\mu}{\sigma}$ wel eens het aantal **standaardeenheden** van x genoemd. Het geeft aan in objectieve termen hoever de waarneming zich van het gemiddelde bevindt. Hoe verder het zich bevindt, hoe groter in absolute waarde het aantal standaardeenheden, hoe zeldzamer de gebeurtenis. Op die manier kunnen we soms waarnemingen uit totaal verschillende universa objectief met elkaar vergelijken. We geven een voorbeeld.

In een bepaalde volksgroep is de gemiddelde lengte van de jongens van 10 jaar $1,40$ m en de standaardafwijking bedraagt 6cm. Voor jongens van 18 jaar bedragen deze getallen respectievelijk $1,74$ m en 8 cm. Een bepaalde jongen A van 10 jaar meet $1,42$ m en een andere jongen B van 18 jaar meet $1,765$ m. Wie van beide is relatief de grootste?

We mogen ervan uitgaan dat lengten van mensen normaal verdeeld zijn, daar dit van zovele factoren afhankelijk is. Om beide lengten te kunnen vergelijken drukken wij ze uit in standaardeenheden. Voor jongen A is dit $\frac{1{,}42-1{,}40}{0{,}06} = 0,33$ en voor jongen B is dit $\frac{1{,}765-1{,}74}{0{,}08} = 0,31$. Hieruit volgt dat A relatief groter is dan B.

Verdere toepassingen

1. Een examen van wiskunde gaf op 100 punten een gemiddelde uitslag van 72. De standaardafwijking bedraagt 15 punten. Men veronderstelt dat de uitslagen normaal verdeeld zijn (ze zijn ook van zovele factoren afhankelijk).

   Een student A behaalt 60 punten;
   een student B behaalt 93 punten;
   het resultaat van student C bedraagt in standaardeenheden -1;
   het resultaat van student D bedraagt in standaardeenheden 1,6;
   Het resultaat van student E is zo dat 30% van de studenten een beter resultaat behaalden.

   Men vraagt:

   (a) de resultaten van A en B in standaardeenheden;
   (b) het percentage studenten met een resultaat tussen deze van A en B;
   (c) het percentage studenten met een resultaat minder dan deze van A;
   (d) het percentage studenten met een resultaat beter dan deze van B;
   (e) het resultaat van C, D en E.

   Oplossing: De examenresultaten is een normaalverdeelde statistische veranderlijke $X \sim N(72; 15^2)$.

   (a) Voor A is de standaardmaat $\frac{60-72}{15} = -0,8$. Voor B is dit $\frac{93-72}{15} = 1,4$.
   (b) $\mathbf{P}(-0,8 < Z < 1,4) = \Psi(1,4) - \Psi(-0,8) = 0,7073$, dus 71%,
   (c) $\mathbf{P}(Z < -0,8) = 1 - \Psi(0,8) = 1 - 0,7881 = 0,2119$, dus 21%;
   (d) $\mathbf{P}(1,4 < Z) = 1 - \Psi(1,4) = 1 - 0,9192 = 0,0808$, dus 8%.
   (e) Het resultaat van C is $\mu - \sigma = 72 - 15 = 57$.
   Het resultaat van D is $\mu + 1,6 \times \sigma = 72 + 1,6 \times 15 = 96$.
   Om het resultaat van E te vinden, bepalen we a zó dat $\mathbf{P}(X < a) = 70\%$.

   $$\text{NORM.INV}(0,7; 72; 15) = 80$$

2. Automatische vulling van flessen.
   In een fabriek worden flessen automatisch gevuld. De inhoud van de gevulde flessen heeft een normale verdeling rond de inhoud waarop de machine wordt ingesteld. De standaardafwijking bedraagt 5 cc. De fabriek wenst dat 90% van de flessen een inhoud hebben van minstens 500 cc. Op welke maat moet de machine ingesteld worden?

OPLOSSING:
De machine wordt ingesteld op de gemiddelde waarde μ, voorlopig onbekend, maar zó dat 90% van de flessen een inhoud hebben groter of gelijk aan 500 cc, of dus zó dat 10% van de flessen een inhoud heeft van minder dan 500 cc. Bijgevolg is

$$\Psi(\frac{500-\mu}{5}) = 0,1 \Longrightarrow \frac{500-\mu}{5} = \Psi^{-1}(0,1) = -1,282$$

dus $\mu = x - \sigma z = 500 - 5 \times (-1,28) = 506,4$. De machine moet dus ingesteld worden op 506,4 cc.

OPGAVEN — **144** Bereken het oppervlak onder de normaalkromme

a. tussen $z = 0$ en $z = 1,4$;

b. tussen $z = -0,65$ en $z = 0$;

c. tussen $z = -0,48$ en $z = 2,23$;

d. tussen $z = 0,81$ en $z = 1,94$;

e. links van $z = -0,56$;

f. rechts van $z = -1,26$;

g. rechts van $z = 2,97$ en links van $z = -1,42$.

145 Bereken de waarde van z als gegeven is de oppervlakte onder de normaalkromme

a. tussen 0 en z is $0,3770$;

b. links van z is 0,8621;

c. tussen $-1,5$ en z is 0,0217;

146 Op een examen is op 100 punten het gemiddelde 78 en de standaardafwijking bedraagt 10 punten.

a. Bepaal de uitslagen in standaardeenheden van twee studenten A en B die resp. 93 en 62 punten behalen;

b. Bepaal het aantal punten behaald door de studenten C en D van wie de uitslagen in standaardeenheden $-0,6$ en 1,2 bedragen.

147 Bereken het gemiddelde en de standaardafwijking van een examen waarbij de uitslagen 70% en 88% overeenstemmen met standaardcijfers van $-0,6$ en $1,4$.

148 Op het eindexamen is de gemiddelde uitslag 73% en de standaardafwijking 9. De top 10% van de geslaagden krijgen een eervolle vermelding op het diploma. Welk is de minimale uitslag om deze vermelding te verkrijgen?

149 Een reeks temperatuurmetingen zijn normaal verdeeld. Welk procent metingen zullen verschillen opleveren ten opzichte van het gemiddelde

1. meer dan een halve standaardafwijking;
2. minder dan 3/4 van de standaardafwijking.

Oplossingen:
144. .
145.
146. (a) 1,5; -1,6; (b) 72; 90;
147. 75,4; 9;
148. 84,53%. 149. 61,71%., 54,67%.

3.3.5.4 Tot slot nog een voorbeeld...

De lengte van een pasgeboren meisje wordt beïnvloed door zovele factoren, dat de verdeling ervan normaal zal zijn. Uit metingen blijkt dat het Belgisch pasgeboren meisje gemiddeld 49,4 cm groot is en 90% van alle Belgische pasgeboren meisjes kleiner zijn dan 52,2 cm. Gevraagd wordt

1. zoek de standaardafwijking van de kansverdeling van de lengte van pasgeborene Belgische meisjes;
2. als Stephanie geboren is, waren van al de pasgeboren meisjes in dat hospitaal dat jaar er precies 33% kleiner dan zij. Hoe groot was lieve Stephanie?
3. Suna was exact 50 cm groot toen ze geboren werd. Tom daarentegen boeide zodanig het vroedpersoneel dat men vergat haar te meten. Maar haar moeder wist zeker dat ze niet groter was dan 51 cm. Wat is de kans dat kloeke Tom groter was dan kleine Suna bij haar geboorte?
4. Toen Suna geboren werd, werd er juist op datzelfde ogenblik in hetzelfde hospitaal Tom geboren. De lengte van pasgeboren jongens is ook normaal verdeeld met gemiddelde waarde 50,2 en standaardafwijking $1,56$. Wat is de kans dat snoezig Sunake groter was dan Tom?
5. In een bepaalde week werden in dat ziekenhuis 9 meisjes geboren. Bereken de kans dat de grootste baby groter is dan 51 cm.

Oplossing: We noemen X de toevalsveranderlijke voor de lengte van pasgeboren meisjes.

1. $z = \Psi^{-1}(0,9) = 1,282 \Longrightarrow z = \frac{52,2-49,4}{\sigma} = 1,282 \Longrightarrow \sigma = \frac{2,8}{1,282} = 2,184.$

$$X \sim \mathrm{N}(49.4; 2.184^2)$$

2. Stel dat Stephanie x cm was dan is $\mathbf{P}(X < x) = 0,33 \Longrightarrow x = 48,439$.
Stephanie was 48,4 cm bij haar geboorte (Stephanie is nochtans een groot meisje geworden).

3. We moeten de kans berekenen dat een meting tussen 50 en 51 cm ligt.

$$\mathbf{P}(50 \leq X \leq 51) = \mathbf{P}(X \leq 51) - \mathbf{P}(X \leq 50) = 0,159958.$$

Tom heeft dus ongeveer 16% kans dat ze groter is dan Suna bij haar geboorte.

4. We noemen Y de toevalsveranderlijke voor de lengte van pasgeboren jongentjes.
We moeten de kans berekenen dat $X > Y$, of dus dat $Y - X < 0$.

$$X \sim \mathrm{N}(49.4; 2.18^2) \text{ en } Y \sim \mathrm{N}(50.2; 1,56^2)$$

Volgens de basisregels is $Y - X$ normaal verdeeld en er geldt:

$$\mathrm{E}(Y - X) = \mathrm{E}(Y) - \mathrm{E}(X) = 50,2 - 49,4 = 0,8$$
$$\mathrm{Var}(Y - X) = \mathrm{Var}(Y) + \mathrm{Var}(X) = 1,56^2 + 2,184^2 = 2,684^2.$$

$$Y - X \sim N(0,8; 2,684^2).$$
$$\mathbf{P}(Y - X < 0) = 0,382852.$$

De kans dat lief Sunatje groter was dan Suna is dus 38,3%.

5. Als de grootste baby groter is dan 51 cm is minstens 1 baby groter dan 51 cm en dat is de negatie van alle baby's zijn kleiner dan 51 cm.
De kans dat één baby kleiner is dan 51 cm is

$$\mathbf{P}(X < 51) = 0,77$$

We beschouwen nu de toevalsveranderlijke $Y \sim \mathrm{B}(9; 0.77)$.
De kans dat elke baby van die 9 baby's kleiner is dan 51 cm is

$$\mathbf{P}(Y = 9) = 0,905$$

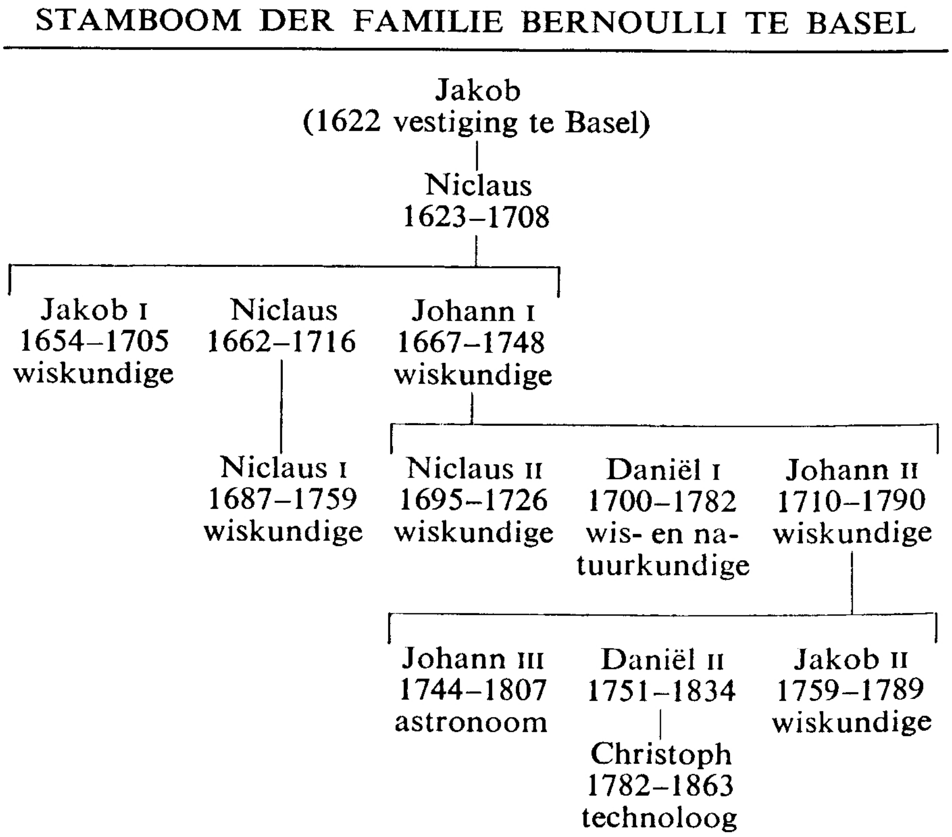
De gevraagde kans is $90,5\%$.

3.4 Wiskunde-Cultuur

De kansrekening is eigenlijk onstaan uit een vraagstukje in verband met de binomiale verdeling. In 1654 deed de Fransman De Méré de volgende proefneming. Hij wierp met één dobbelsteen viermaal en door dit een aantal keren te doen probeerde hij de kans te weten te komen om minstens éénmaal zes te gooien. Hij deed nu hetzelfde met twee dobbelstenen en vierentwintig worpen. Daarmee bekwam hij ongeveer de kans om minstens éénmaal dubbelzes te gooien in vierentwintig worpen. Hij dacht dat beide kansen gelijk moesten zijn omdat hij in het tweede geval zesmaal zoveel worpen deed terwijl er in het eerste geval zesmaal zoveel gunstige uitslagen zijn. Maar zijn proeven waren hiermee in strijd. Hij bekwam een grotere kans in het eerste geval. Daarom schreef hij een brief naar de beroemde Wiskundige PASCAL (1623-1662) om dit theoretisch te verklaren. Pascal deed dat en zo werd de kansrekening geboren.

Het vraagje van de Méré is een toepassing op de binomiaalverdeling. Reken zelf uit dat de waarschijnlijkheid om minstens éénmaal zes te gooien bij vier worpen met een zelfde dobbelsteen gelijk is aan $1-(\frac{5}{6})^4 = 0,518$, terwijl de kans om minstens éénmaal dubbelzes te gooien bij 24 worpen met twee dobbelstenen gelijk is aan $1-(\frac{35}{36})^{24} = 0,491$ (zie opgave 111).

Bazel, in Zwitserland, reeds in 1263, een vrije Rijksstad geworden, was reeds lange tijd het middelpunt van wetenschappelijk leven. Wij hoeven slechts aan Erasmus te denken, die "grote ster" die in Rotterdam rees, en ging in Bazel onder. Evenals in de Hollandse steden bloeiden ook in Bazel kunsten en wetenschappen onder de bescherming van rijke koopmansfamilies. Een van deze families waren de Bernoulli's. In de 17de eeuw overgekomen uit Antwerpen, nadat deze stad blijvend Spaans was geworden. Deze familie heeft vanaf de laatste jaren van de 17de eeuw tot op heden, opnieuw in ieder geslacht mannen van de wetenschap voortgebracht. Het is moeilijk in de geschiedenis van de wetenschappen nog een andere familie te vinden, die op wetenschappelijk gebied zulke hoge prestaties heeft geleverd.
Deze wetenschappelijke belangstelling begon bij de twee broers Jacob en Johann. Jacob de oudste begon met theologie, Johann met medicijnen te studeren, doch toen LEIBNIZ' artikelen (1646-1716) in de "Acta Eruditorum" verschenen, besloten beide zich op de wiskunde toe te leggen. Zo werden ze de eerste leerlingen van betekenis die Leibniz kreeg. Jacob BERNOULLI (1654-1705) was een onderzoeker van de nog nieuwe waarschijnlijkheidsrekening, waarvoor hij zijn "Ars conjectandi" schreef. In het eerste deel van het boek vinden we HUYGENS' opstel over spelen van geluk; het andere gedeelte bevat een verhandeling over permutaties en combinaties die haar hoogtepunt vindt in het meest beroemde gedeelte: het "theorema van Bernoulli" over het gedrag van binomiale kansverdelingen. In dit boek vinden we ook een discussie over de driehoek van Pascal en treffen we ook de zogenaamde getallen van Bernoulli aan.



Figuur 3.6: stamboom van de familie Bernoulli

$\boxed{\text{POISSON Siénon}}$ is een Frans wiskundige van 1781 tot 1840. Poisson behoorde samen met PONCELET (1788-1867), FOURIER (1768-1830) en CAUCHY (1789-1857) tot de leidende wiskundigen wier naam met de eerste tientallen jaren van de École Polytechnique waren verbonden. Alle vier toonden diepe belangstelling voor de toepassing van de wiskunde op de natuur- en werktuigkunde, en alle vier werden door deze belangstelling weer tot ontdekking in de "zuivere" wiskunde gevoerd. Poisson's productiviteit blijkt uit de vele manieren waarop zijn naam in onze leerboeken voorkomt: de haakjes van Poisson in de leer van de differentiaalvergelijkingen, de constante van Poisson in de elasticiteitsleer, de integraal en de vergelijking van Poisson in de potentiaaltheorie. Poisson schreef ook een boek over waarschijnlijkheidsrekening (1837). Onder de vele resultaten die dit boek bevat vinden we de "wet van Poisson" als benadering van de binomiale wet voor kleine kansen zoals we bestudeerden. De grote betekenis van deze wet voor de statistiek, o.a. van straling en verkeer, is eerst in de twinstigste eeuw begrepen.

HUISTAAK K3

1. Voor een bepaald theater weet men uit ervaring dat 20% van de personen die een plaats reserveren niet komt opdagen. Indien het theater 50 plaatsen heeft en 52 reservaties opneemt, wat is dan de kans dat iedereen een plaats zal krijgen?

2. Van 35 normale kinderen van een klas werd een psychologische test afgenomen. De gemiddelde score is 82 en de standaardafwijking is 6.
   Gevraagd:

   a. de kans dat alle scores kleiner zijn dan 94;
   b. de kans dat de laagste score minder is dan 64;
   c. het aantal kinderen met een score tussen 76 en 88.

3. Het aantal studenten dat 's middags binnenkomt in een studentenrestaurant om een maaltijd te nemen is gemiddeld 3 per minuut.
   Bereken de kans dat

   a. er meer dan 5 studenten binnenkomen tussen 12u en 12u01min;
   b. er geen enkele student binnenkomt tussen 12u03min en 12u06min.

4. Een anesthesist beschikt over twee soorten medicaties om een patiënt te verdoven. De slaapduur bij de eerste medicatie is normaal verdeeld met een gemiddelde van 6 uur en een standaardafwijking van 1 uur. De slaapduur bij de tweede medicatie is normaal verdeeld met een gemiddelde van 5 uur en een standaardafwijking van 1,5 uur.Gezien de zwakte van de patiënt zou de anesthesist graag de tweede medicatie gebruiken, maar bovendien wil hij 99% zekerheid dat de patiënt minstens 2,5 uur slaapt. Welke medicatie kan de anesthesist gebruiken?

5. In een fabriek worden dozen suiker machinaal gevuld, zodat hun gemiddeld gewicht 1 kg is en de standaardafwijking 0.050. De suiker wordt verpakt in kisten van 50 dozen. Bereken de kans dat het verschil in totaal gewicht tussen twee kisten tussen 0,5 kg en 1,5 kg bedraagt.

6. Bereken het gemiddelde en de standaardafwijking van een examen waarbij de uitslagen 50% en 75% overeenstemmen met standaardcijfers van -1,4 en 0,8.

Hoofdstuk 4

Statistiek

Een statiticus is iemand die, als hij met zijn hoofd in een brandende kachel steekt en met zijn voeten in een emmer ijs staat, verklaart: "gemiddeld genomen, voel ik me lekker".

4.1 Beschrijvende statistiek - herhaling

Een populatie is het 'grote geheel' van personen, dieren, planten of voorwerpen waarvan men iets wil onderzoeken door middel van enquêtes of experimenten.
Al naar gelang de aard van de informatie onderscheiden we twee soorten data:

Kwalitatieve of niet-numerieke data zijn waarnemingsgegevens die geen getallen zijn.
De n waarnemingsgegevens worden ingedeeld in zogenaamde **categorieën**. Elke categorie krijgt een "label". De categorieën kunnen soms wel door een bepaalde code (nummer) worden weergegeven. Het aantal categorieën die mogelijk kunnen voorkomen stellen we gelijk aan m.

De Vlaamse bevolking waarvan men de bloedgroep wil onderzoeken is de populatie. De kwalitatieve data bevat bvb. 100 Vlamingen waarvan men de bloedgroep opvraagt ($n = 100$).
De mogelijke categorieën bij het bloedgroepenonderzoek zijn: A, B, AB en O ($m = 4$).

De Vlaamse gezinnen met twee kinderen waarvan men de samenstelling jongen, meisje wil onderzoeken. De kwalitatieve data bestaat bvb. uit 8665 gezinnen met twee kinderen in het Gentse waarvan men de samenstelling opvraagt ($n = 8665$).
De mogelijke categorieën bij de samenstelling van een gezin met 2 kinderen

zijn: (meisje,meisje), (meisje,jongen), (jongen,meisje), (jongen,jongen).
Kort genoteerd: (m,m), (m,j), (j,m), (j,j) ($m = 4$).

Kwantitatieve of numerieke data zijn waarnemingsgegevens die getallen zijn. Deze getallen zijn waarden van een toevalsveranderlijke, die we voor de populatie willen beschouwen, over het universum van het experiment. Deze toevalsveranderlijke noemen we de **populatie-s.v.** (stochastische veranderlijke).

Het theoretisch model van een steekproef :
Een steekproef met grootte n uit een populatie met populatie-s.v. X is een set van n toevalsveranderlijken $X_1, X_2, \cdots X_n$ die onafhankelijk zijn en dezelfde verdeling hebben als X.

Een steekproef in praktijk :
Van zodra we in praktijk een steekproef trekken, vinden we een set van n **steekproefwaarden** $r_1, r_2, \cdots r_n$ die één van de vele realisaties is van het theoretisch model van de steekproef. Telkens we een steekproef trekken verkrijgen we andere steekproefwaarden.

Steekproefvariabiliteit is het verschijnsel van de verschillende steekproefresultaten.

Discrete kwantitatieve data zijn numerieke data die waarden zijn van een discrete populatie s.v.

De steekproefwaarden worden gerangschikt volgens grootte en eventueel ingedeeld in een eindig aantal **klassen**.
Het aantal mogelijke waarden van de toevalsveranderlijke of het aantal klassen stellen we voor door m.
We stellen de verschillende steekproefwaarden of de klassenmiddens voor door $x_1, x_2, \ldots x_m$.

Het experiment "Een dobbelsteen opgooien".
De toevalsveranderlijke X: "een 6 of geen 6 gooien" ($m = 2$).
Een steekproef in praktijk bestaat bvb. uit het honderd keer opgooien van de dobbelsteen en telkens noteert men een 6 of geen 6 ($n = 100$).

Het experiment 'opgooien van twee dobbelstenen'
De toevalsveranderlijke X: "de som van de ogen", neemt waarden aan die behoren tot $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ ($m = 11$).
Een steekproef in praktijk bestaat uit het twintig keer opgooien van twee dobbelstenen en telkens noteert men de som van de ogen ($n = 20$).

In de eerste ronde van de wiskunde olympiade de resultaten op 150 van leerlingen van de derde graad.
De toevalsveranderlijke X: "het resultaat op 150", neemt waarden aan die behoren tot $\{0, 1, 2, \cdots 149, 150\}$ ($m = 151$).
Een steekproef in de praktijk bestaat uit de scores van 11185 leerlingen van de derdegraad die in 2007 deelnamen aan de wiskunde olympiade ($n = 11185$).
Hier zal de data ingedeeld worden in klassen.

Continue kwantitatieve data zijn numerieke data die waarden zijn van een continue populatie s.v.
De steekproefwaarden worden in de praktijk afgerond (afhankelijk van de nauwkeurigheid van de meettoestellen) en worden dan in een eindig aantal **klassen** ingedeeld. De **klassenmiddens** stellen we voor door $x_1, x_2, \ldots, x_m$.

Bij jongeren tussen 15 en 18 nemen we als populatie s.v; "het aantal uren TV kijken en computer spelen per dag" die waarden in het interval $[0, 24]$.
Een steekproef in de praktijk bevat bvb. 150 leerlingen van Nieuwen Bosch tussen 15 en 18 waarvan men het aantal uren TV kijken en computerspelen per dag opvraagt ($n = 150$).

Bij achtienjarige meisjes nemen we als populatie s.v. de lichaamslengte die waarden aanneemt in het interval $[100, 200]$ (of een groter of kleiner interval eigen aan het experiment).
Een steekproef bestaat uit bvb. 30 achtienjarige meisjes van Nieuwen Bosch waarvan men de lengte opvraagt ($n = 30$).

De absolute frequentie f_i van steekproefwaarde x_i is bij discrete data het aantal keer dat de waarde x_i in de dataset voorkomt of bij gegroepeerde data het aantal waarden van de dataset die liggen in het interval waarvan x_i het midden is.

Relatieve frequentie F_i van steekproefwaarde x_i is de verhouding van de absolute frequentie van x_i tot de steekproefgrootte n.

$$F_i = \frac{f_i}{n}$$

De relatieve frequenties drukken we meestal uit in percentages.

Volgende eigenschappen voor de absolute en relatieve frequenties zijn geldig:

- *De som van de absolute frequenties van de verschillende steekproefwaarden is gelijk aan de steekproefgrootte*

$$f_1 + f_2 + \ldots + f_m = \sum_{i=1}^{m} f_i = n \tag{4.1}$$

- *De som van de relatieve frequenties van de verschillende steekproefwaarden is gelijk aan 1*

$$F_1 + F_2 + \ldots + F_m = \sum_{i=1}^{m} F_i = 1 \tag{4.2}$$

De gecumuleerde absolute frequentie cf_i van x_i is de som van de absolute frequenties van de verschillende steekproefwaarden die kleiner zijn dan of gelijk aan x_i mits deze geordende zijn volgens grootte.

De gecumuleerde relatieve frequentie CF_i van x_i is de gecumuleerde absolute frequentie gedeeld door de steekproefgrootte n.

De drie kwartielen zijn de getallen die we vinden als we de rij geordende waarnemingsgetallen verdelen in vier gelijke delen.

De mediaan is het tweede kwartiel.

De interkwartielafstand is de afstand tussen het eerste en derde kwartiel.

Het gemiddelde van de steekproefwaarden is het getal dat we bekomen door elke waarde x_i te vermenigvuldigen met zijn relatieve frequentie F_i en de bekomen producten op te tellen. In geval we werken met met klassen (gegroepeerde gegevens) dan nemen voor x_i het klassemidden.

$$\overline{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \cdots f_m \cdot x_m}{n} = F_1 \cdot x_1 + F_2 \cdot x_2 + \cdots + F_m \cdot x_m.$$

Met het sommatieteken:

$$\overline{x} = \sum_{i=1}^{m} \frac{f_i \cdot x_i}{n} = \sum_{i=1}^{m} F_i \cdot x_i.$$

De variantie van de steekproefwaarden

$$s^2 = \frac{1}{n-1}\left(f_1 \cdot (x_1 - \overline{x})^2 + f_2 \cdot (x_2 - \overline{x})^2 + \cdots + f_m \cdot (x_m - \overline{x})^2\right)$$

Met het sommatieteken:

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{m} f_i.(x_i - \overline{x})^2\right)$$

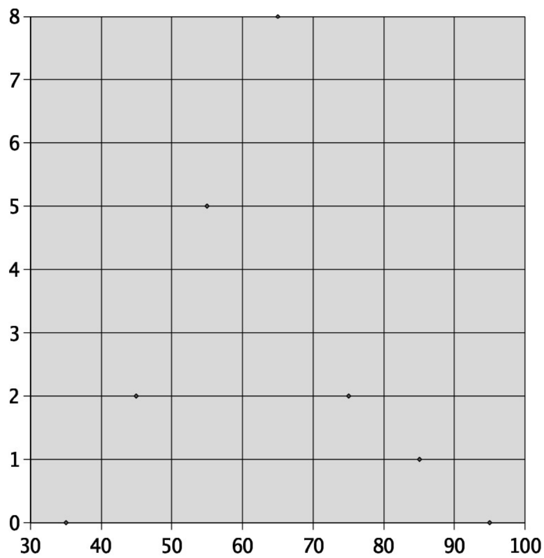
Frequentietabel is een tabel waarin we de verschillende steekproefwaarden geordend volgens grootte weergeven alsook de absolute, relatieve en gecumuleerde frequenties.

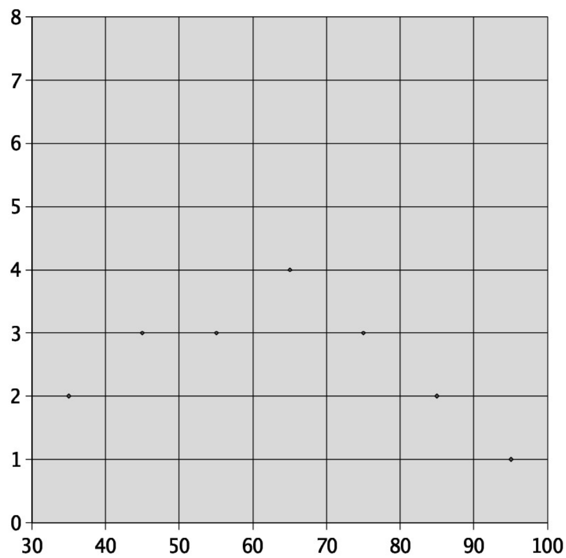
Grafische voorstellingen Voor discrete data is dat voor niet-gegroepeerde data het **staafdiagram** en voor gegroepeerde data het **histogram** of het **stengel-en-bladdiagram**.
Voor continue data is dat het **histogram** of het **stengel-en-bladdiagram**.
De **boxplot** is een rechthoekige doos die getekend wordt van het eerste kwartiel tot het derde kwartiel t.o.v. een schaal. De lijn in de doos wijst de mediaan (tweede kwartiel) aan. We illustreren de laatste begrippen met de wiskunderesultaten van 18 leerlingen. Deze data werd reeds gegroepeerd en de x_i in de twee tabellen zijn de klassemiddens. Met deze twee reeksen wiskunderesultaten tonen we aan dat het rekenkundig geiddelde iets zegt over de prestaties van de leerlingen. Maar dat alleen is niet voldoende. Het is ook belangrijk te weten hoe dicht de resultaten rond het gemiddelde liggen. Een maat voor

die spreiding is de standaardafwijking.
De twee reeksen resultaten hebben hetzelfde gemiddelde maar de tweede reeks heeft een veel kleinere spreiding dan de eerste reeks. We tekenen de histogrammen.

Klassemidden Xi	Fi	Fi	Xi*Fi	Xi-gem	(Xi-gem)^2	(Xi-gem)^2*Fi	Cfi	Cfi
35	2	11,11%	3,89	-27,22	741,05	1482,10	2	0,11
45	3	16,67%	7,5	-17,22	296,6	889,81	5	0,28
55	3	16,67%	9,17	-7,22	52,16	156,48	8	0,44
65	4	22,22%	14,44	2,78	7,72	30,86	12	0,67
75	3	16,67%	12,5	12,78	163,27	489,81	15	0,83
85	2	11,11%	9,44	22,78	518,83	1037,65	17	0,94
95	1	5,56%	5,28	32,78	1074,38	1074,38	18	1
	18	100,00%	62,22			303,59		
			Gem			17,42		

Xi	Fi	Fi	Xi*Fi	Xi-gem	(Xi-gem)^2	(Xi-gem)^2*Fi	Cfi	Cfi
35	0	0,0%	0	-27,22	741,05	0	0	0
45	2	11,1%	5	-17,22	296,6	593,21	2	0,11
55	5	27,8%	15,28	-7,22	52,16	260,80	7	0,39
65	8	44,4%	28,89	2,78	7,72	61,73	15	0,83
75	2	11,1%	8,33	12,78	163,27	326,54	17	0,94
85	1	5,6%	4,72	22,78	518,83	518,83	18	1
95	0	0,0%	0	32,78	1074,38	0	18	1
	18	100,00%	62,22			103,59		
			Gem			10,18		





Representativiteit van een steekproef betekent dat elk element van de populatie dezelfde kans moet hebben om geselecteerd te worden voor een steekproef. Zo een steekproef noemen we een **enkelvoudige aselecte steekproef** (EAS).
Om de elementen van een steekproef aselect te kiezen bestaan standaardprocedures.

- *Selectie met randomgetallen*
  Aan elk element van de populatie wordt een getal toegekend. De steekproef wordt tot stand gebracht door willekeurig getallen te trekken zoals in een loterij.
- *Gestratifieerde steekproef*
  Bij omvangrijke populaties wordt de populatie ingedeeld in een aantal homogene subpopulaties (strata). Uit elke deelgroep of stratum wordt dan op aselecte wijze een steekproef genomen, waarbij de steekproefgrootte telkens bepaald wordt in functie van de omvang van de subpopulatie.
  Voorbeeld: Bij een budgetonderzoek "waar wordt het inkomen van een gezin aan besteed?" worden de gezinnen ingedeeld in een aantal inkomenscategorieën (laag, middelmatig en hoog). Uit elk van deze categorieën wordt dan een steekproef genomen waarbij het aantal geselecteerde gezinnen in verhouding staat tot de omvang van de categorie. Zo zijn we zeker dat alle inkomensklassen evenredig vertegenwoordigd zijn in de steekproef.
- *Getrapte steekproef*
  Hier wordt de populatie ingedeeld in deelpopulaties waarna enkele van deze subpopulaties lukraak worden uitgekozen en nog in kleinere deelpopulaties worden verdeeld. Zo kunnen in een grootschalige opiniepeiling in een eerste stap een aantal gemeenten geselecteerd worden, waarna uit elke gemeente een aantal wijken worden gekozen. In een derde stap kunnen vervolgens uit elke wijk één of meerdere straten gekozen worden en tenslotte uit elke straat enkele gezinnen. Elke deelsectie kan gebeuren op basis van randomgetallen.
- *Systematische steekproef*
  In dit geval worden de steekproefelementen op systematische wijze geselecteerd uit de populatie. Dit betekent dat tussen twee opeenvolgende selecties steeds een min of meer gelijk aantal elementen wordt overgeslagen, of dat het tijdsinterval tussen twee selecties ongeveer gelijk is. Zo kan bijvoorbeeld om de 20 namen uit een register een naam geselecteerd worden, of kan van een continu bewegende productieband om de 10 minuten een product genomen worden om de kwaliteit ervan te controleren.

Het grootste gevaar voor de representativiteit bij vooral schriftelijke enquêtes is de nonrespons, aangeschreven personen die nalaten het formulier in te vullen. Indien bijvoorbeeld bij een opiniepeiling slechts 1 op 5 personen het formulier terugstuurt, dan kunnen we bezwaarlijk stellen dat die groep personen nog representatief is voor de volledige populatie. Het zijn immers de meest geïnteresseerden die gemotiveerd zijn om op de enquête in te gaan en deze groep is niet representatief.

Voorbeelden:

- Om het vakantiegedrag van de Belgische bevolking te onderzoeken, worden 6000 personen bevraagd. De Belgische bevolking is de populatie en de 6000 personen vormen een steekproef ($n = 6000$). Men kan bijvoorbeeld vragen naar het aantal keer dat men een korte vakantie genomen heeft in 2005, of naar de bestemming van de vakantie, of naar het vervoermiddel om op vakantie te gaan.
- Bij het bepalen van kijkcijfers is het technisch onmogelijk om de circa 2 500 000 Vlaamse TV-abonnees bij het onderzoek te betrekken. Daarom wordt de kijkdichtheid berekend op basis van een steekproef van een honderd abonnees.
- Om iets te weten te komen over het rookgedrag van Vlaamse meisjes van 17 jaar, kunnen we 40 zeventienjarige meisjes van Nieuwen Bosch ondervragen over hun rookgedrag. De groep van alle zeventienjarige Vlaamse meisjes is de populatie. Deze verzameling zeventienjarige grijze muizen is een steekproef ($n = 40$). Maar deze steekproef is niet representatief voor het rookgedrag van alle Vlaamse zeventienjarige meisjes.
- De dienst die verantwoordelijk is voor de kwaliteitscontrole in een firma van zuivelproducten kan moeilijk alle geproduceerde vlootjes testen op hun samenstelling. Dit zou immers de vernietiging van de productie inhouden. Noodgedwongen moet men zich beperken tot steekproeven.
- In België wordt om de 10 jaar een volkstelling georganiseerd, teneinde inzicht te verwerven in de huidige gezinssamenstelling, huisvesting, comfort en andere aspecten van sociologische of demografische aard. Aangezien elke Belg hierbij betrokken is, is de steekproef gelijk aan de populatie.

4.2 Statistische kans

De statistiek en de kansrekening zijn historisch gegroeid uit de kansspelen. Nu wordt ze op zeer veel gebieden toegepast zoals in de fysica, de biologie, de geneeskunde, de economische en sociale wetenschappen.

In onze wereld doen zich situaties voor waarvan de afloop niet met zekerheid kan bepaald worden, maar bij herhaalde malen optreden van dezelfde situatie zien we een zekere wetmatigheid.

Voorbeelden:

- Bij het opgooien van een dobbelsteen weten we niet vooraf welk cijfer we zullen verkrijgen. Per toeval krijgen we misschien een zes. Maar als we een dobbelsteen honderdmaal opgooien zijn we zeker dat we ongeveer zestien à zeventien keer een zes zullen verkrijgen. Dit aantal kunnen we wiskundig bepalen. Bij het opgooien van een dobbelsteen hebben we zes even waarschijnlijke uitkomsten 1, 2, 3, 4, 5 en 6. We zeggen dan dat we één kans op zes hebben om een 6 te gooien. Wiskundig zou dit betekenen dat als we een dobbelsteen zes keer omhoog gooien de uitkomst 6 één keer zal optreden. In de praktijk is dit echter niet zo.
  Maar hoe groter de steekproef hoe dichter het aantal keren een 6 gooien de wiskundige voorspelling één op zes zal benaderen (de wet van de grote getallen).
  Bij het experiment 'opgooien van één dobbelsteen' is $\Omega = \{1, 2, 3, 4, 5, 6\}$ en $N = 6$. We proberen nu experimenteel de kans op een 6 te bepalen door een steekproef met grootte n uit te voeren.
  We zullen dit experiment simuleren met DERIVE.

  - 1 proef met de computer: we typen

    $$\textbf{1+random(6)}$$

    de computer geeft een getal tussen 1 en 6.
  - n proeven met de computer: we typen

    $$\textbf{vector(1+random(6)}, i, 1, n\textbf{)}$$

    de computer geeft een vector met n elementen gekozen uit de elementen 1 t.e.m. 6. Dit is een herhalingsvariatie van 6 elementen in groepjes van n.
  - Het tellen van het aantal keer 6 bij n proeven met de computer: we typen

    $$\textbf{dimension(select(}x = 6, x,\textbf{vector(1+random(6)}, i, 1, n\textbf{)))}$$

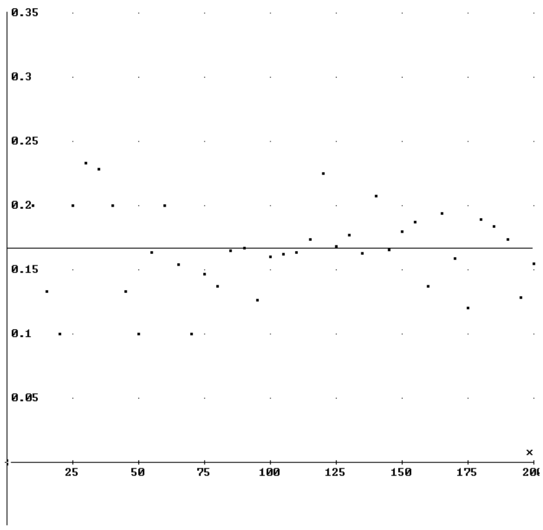
    Omdat we de computer deze telling een aantal keer willen laten uitvoeren voor verschillende waarden van n, gaan we deze uitdrukking kort noteren en we programmeren

    $$\textbf{hoeveelkeer6(}n\textbf{):=dimension(select(}x = 6, x,\textbf{vector(1+random(6)}, i, 1, n\textbf{)))}$$

Let op het aantal haakjes!

- 'hoeveelkeer6(n)' is de frekwentie van de uitkomst 6 bij n keer gooien. 'hoeveelkeer6(n)/n ' is de relatieve frekwentie van de uitkomst 6 bij n keer gooien en levert een getal op tussen 0 en 1.
- We voeren 40 steekproeven uit met grootte achtereenvolgend 5, 10, 15 enz. We noteren telkens hoeveel keer we een 6 bekomen. Zo ontstaat een rij getallen met algemene term 'hoeveelkeer6(n)/n'. De rij stellen we grafisch voor. We typen:

$$\textbf{vector}([n, \textbf{hoeveelkeer}6(n)/n], n, 1, 5, 200, 5)$$



Figuur 4.1: voer zelf de steekproeven uit en maak de grafische voorstelling van de relatieve frequenties

We zien op de grafiek dat rij convergeert naar de waarde $\frac{1}{6} = 0,1666\cdots$. Vanaf een zeker rangnummer m liggen de waarden binnen een voorafgekozen kleine basisomgeving van 0,1666.

Wiskundig kunnen we zeggen dat de

$$\mathbf{P}(X = 6) = \lim_{n \to +\infty} \frac{f}{n}$$

Deze limiet noemt men de **statistische waarschijnlijkheid** van 6 bij "het gooien van één dobbelsteen". De statistische waarschijnlijkheid van een toevalsveranderlijke is dus de relatieve frequentie van die toevalsveranderlijke voor een zeer groot aantal proeven en is een reëel getal gelegen tussen 0 en 1.

- Het tweede experiment bestaat 5 identieke muntstukken op te gooien en te kijken hoeveel keer we munt bekomen.
  - Het universum is de verzameling van $2^5 = 32$ 5-tallen ($N = 32$). De toevalsveranderlijke X kan de 6 waarden 0, 1, 2, 3, 4 en 5 aannemen ($m = 6$).

$$\begin{aligned}
&x_1 = X((\mathrm{k},\mathrm{k},\mathrm{k},\mathrm{k},\mathrm{k})) = 0\\
&x_2 = X((\mathrm{k},\mathrm{k},\mathrm{k},\mathrm{k},\mathrm{m})) = X((\mathrm{k},\mathrm{k},\mathrm{k},\mathrm{m},\mathrm{k})) = X((\mathrm{k},\mathrm{k},\mathrm{m},\mathrm{k},\mathrm{k})) = \cdots = 1\\
&x_3 = X((\mathrm{k},\ k,\mathrm{k},\mathrm{m},\mathrm{m})) = X((\mathrm{k},\mathrm{k},\mathrm{m},\mathrm{m},\mathrm{k})) = X((\mathrm{k},\mathrm{m},\mathrm{m},\mathrm{k},\mathrm{k})) = \cdots = 2\\
&x_4 = X((\mathrm{k},\mathrm{k},\mathrm{m},\mathrm{m},\mathrm{m})) = X((\mathrm{m},\ k,\mathrm{m},\mathrm{m},\ k)) = X((\mathrm{m},\mathrm{m},\mathrm{m},\mathrm{k},\mathrm{k})) = \cdots = 3\\
&x_5 = X((\mathrm{m},\mathrm{m},\mathrm{m},\mathrm{m},\mathrm{k})) = X((\mathrm{m},\mathrm{m},\mathrm{m},\mathrm{k},\mathrm{m})) = X((\mathrm{m},\mathrm{m},\mathrm{k},\mathrm{m},\mathrm{m})) = \cdots = 4\\
&x_6 = X((\mathrm{m},\mathrm{m},\mathrm{m},\mathrm{m},\mathrm{m})) = 5
\end{aligned}$$

    We zien dat de 6 waarden van de toevalsveranderlijke niet even waarschijnlijk zijn. We bepalen de statistische waarschijnlijkheid van de waarden van X.
  - 1 keer gooien met 5 identieke muntstukken: we typen:

$$\textbf{vector(random(2)}, i, 1, 5\textbf{)}$$

    de computer geeft een 5-tal met elementen gekozen uit 0 en 1, bvb. $[1, 1, 0, 0, 0]$ (0 voor munt en 1 voor kop).
    Willen we weten hoeveel keer we munt hebben dan moeten we de som maken van de elementen van het 5-tal. We definieren

$$\textbf{a:=vector(random(2)}, i, 1, 5\textbf{)}$$

    Als we

$$\Sigma(a_k, k, 1, 5)$$

    simplifiëren, krijgen we een getal tussen 0 en 5. Dit is het aantal keer munt bij 1 keer gooien met 5 muntstukken.
  - We voeren 5 steekproeven uit met grootte achtereenvolgend 6, 10, 16, 32 en 64.
    We typen

$$\textbf{vector(}\Sigma(a_k, k, 1, 5), i, 1, n\textbf{)}$$

    Als we de uitdrukking opeenvolgend simplifiëren voor $n = 6$, $n = 10$, $n = 16$, $n = 32$ en $n = 64$ dan verkrijgen we resp. een 6-tal, 10-tal, 16-tal, 32-tal en 64-tal. We geven ze opeenvolgend de namen $b6$, $b10$, $b16$, $b32$ en $b64$.
    Voorbeeld: $b6 = [2, 4, 3, 4, 3, 3,]$ en
    $b32 = [1, 1, 2, 4, 3, 2, 1, 2, 3, 2, 4, 2, 2, 4, 1, 4, 3, 4, 2, 4, 3, 3, 0, 3, 2, 1, 4, 3, 3, 3, 2, 2]$
  - In zo een n-tal tellen we het aantal keer we 0 keer munt , 1 keer munt, 2 keer munt, 3 keer munt, 4 keer munt en 5 keer munt krijgen.
    We typen

$$\textbf{vector(dimension(select(}x = j, x, b32\textbf{)),j,0,5)}$$

Als we dit simplifiëren, krijgen we bijvoorbeeld het 6-tal $[1, 5, 10, 9, 7, 0]$. Deze waarden kunnen we invullen in de vierde kolom van de tabel 4.1. Hetzelfde doen we nu met $b6$, $b10$, $b16$ en $b64$. Daartoe brengen we de bovenstaande uitdrukking naar de schrijflijn en vervangen we opeenvolgend $b32$ door de andere waarden van b. Zo kunnen we de tabel4.1 vervolledigen.

x_i	f_i	f_i	f_i	f_i	f_i
0				1	
1				5	
2				10	
3				9	
4				7	
5				0	
$\sum$	$n = 6$	$n = 10$	$n = 16$	$n = 32$	$n = 64$

Tabel 4.1: turven bij het opgooien van vijf muntstukken

– Tenslotte maken we een tabel waarbij we de relatieve frekwenties van de verschillende uitkomsten opschrijven voor de verschillende waarden van n. Bekijk nu in deze tabel de evolutie van de relatieve frekwenties naarmate het aantal proeven toeneemt. In de laatste kolom staan de kansen aangegeven.

x_i	F_i	F_i	F_i	F_i	F_i	$\mathbf{P}(X = x_i)$
0				0,03		0,03
1				0,16		0,16
2				0,31		0,31
3				0,28		0,31
4				0,22		0,16
5				0		0,03
$\sum$	1	1	1	1	1	

Tabel 4.2: relatieve frekwenties voor de verschillende uitkomsten en de wiskundige verwachting (laatste kolom)

Opmerking: Als de statistische waarschijnlijkheid van een bepaalde waarde van X gelijk is aan nul dan wil dat niet zeggen dat die waarde niet kan optreden.
Voorbeeld: Als we willekeurig een reëel getal op een gegradueerde rechte aanduiden dan is de statistische kans dat we een priemgetal aanduiden gelijk aan 0. Maar toch kan het zijn dat we een priemgetal aanduiden.

4.3 Verklarende statistiek

Voor statistisch onderzoek is het belangrijk om uit de resultaten van een EAS zoveel mogelijk betrouwbare besluiten te trekken over de desbetreffende populatie.

4.4 Karakteristieken van een populatie

Het gemiddelde en de variantie zijn **karakteristieken van de populatie** die we proberen te weten te komen door middel van een steekproef.

Het **populatiegemiddelde** met toevalsveranderlijke X stellen we voor door $E(X) = \mu$.

De **populatievariantie** met toevalsveranderlijke X stellen we voor door $\mathrm{Var}(X) = \sigma^2$.

Stel dat een set van n toevalsveranderlijken $X_1, X_2, \cdots X_n$ het theoretoisch model is van een steekproef uit een populatie met populatie-s.v. X dan is het **steekproefgemiddelde** en de **steekproefvariantie** resp.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{en} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

4.4.1 Gemiddelde en variantie van het steekproefgemiddelde

Het steekproefgemiddelde $\overline{X}$ is op zichzelf een toevalsveranderlijke die normaal verdeeld is van zodra de populatie-s.v. normaal verdeeld is.

STELLING 4.1 *Zij X een normaal verdeelde populatie-s.v. met gemiddelde waarde μ en variantie σ^2 dan is het steekproefgemiddelde $\overline{X}$ van een steekproef $X_1, X_2, \cdots X_n$ normaal verdeeld met een gemiddelde en variantie gegeven door*

$$E(\overline{X}) = \mu$$

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

Bewijs:

1. $\mathrm{E}(X_i) = \mu$ omdat X_i dezelfde verdeling heeft als X en dus ook hetzelfde gemiddelde.

$$\mathrm{E}(\overline{X}) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}n\mu = \mu$$

2. $\mathrm{Var}(X_i) = \sigma^2$ omdat X_i dezelfde verdeling heeft als X en dus ook dezelfde variantie.

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{1}{n}\sigma^2$$

Opmerking: Hoe groter de steekproef is hoe kleiner de variantie van het steekproefgemiddelde.

Voorbeeld: Een populatie bestaat uit de getallen 2,3,6,8,11. Het experiment is "het trekken van een element uit de populatie". De populatie s.v. X is de waarde van het getrokken element.
Het populatiegemiddelde en populatievariantie zijn:

$$\mathrm{E}(X) = \mu = \frac{1}{5}(2+3+6+8+11) = 6$$

$$\mathrm{Var}(X) = \sigma^2 = \frac{1}{n}((2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + 11-6)^2) = 10.8$$

Het theoretisch model van de steekproef met grootte 2 is X_1, X_2.
Een realisatie van die steekproef is bvb. $6, 11$.
We beschouwen nu alle mogelijke realisaties van de steekproef X_1, X_2 in de volgende tabel.

2,2	2,3	2,6	2,8	2,11	3,2	3,3	3,6	3,8
3,11	4,2	4,3	4,6	4,8	4,11	6,2	6,3	6,6
6,8	6,11	11,2	11,3	11,6	11,8	11,11		

De steekproefgemiddelden $\overline{x}$:

2	2.5	4	5	6.5	2.5	3	4.5	5.5
7	3	3.5	5	6	7.5	4	4.5	6
7	8.5	6.5	7	8.5	9.5	11		

De steekproefvarianties s:

0	0.5	8	18	40.5	0.5	0	4.5	12.5
32	2	0.5	2	8	24.5	8	4.5	0
2	12.5	40.5	32	12.5	4.5	0		

Het gemiddelde van het steekproefgemiddelde:

$$\mathrm{E}(\overline{X}) = \frac{1}{25}(2 + 2.5 + 4 + 5 + \cdots + 11) = 6$$

De variantie van het steekproefgemiddelde:

$$\mathrm{Var}(\overline{X}) = \frac{1}{25}((2-6)^2 + (2.5-6)^2 + \cdots (11-6)^2) = 5.4$$

Anderzijds is

$$\frac{\sigma^2}{n} = \frac{10.8}{2} = 5.4 = \mathrm{Var}(\overline{X})$$

Dit laatste is een bevestiging van de stelling 4.1.

4.4.2 Het schatten van een karakteristiek - Onvertekende schatter

Stel dat de populatie-s.v. X de lengte is van een kind bij de geboorte. We willen de gemiddelde geboortelengte kennen. Daartoe trekken we een steekproef met grootte 40, nl. $\{r_1, r_2, \cdots, r_{40}\}$.
Het gemiddelde en de variantie van die 40 lengten zijn een **schatting** van het gemiddelde μ en de variantie σ^2 van de geboortelengte van de volledige populatie.

$$\overline{x} = \frac{1}{40}\sum_{i=1}^{40} r_i \text{ en } s^2 = \frac{1}{39}\sum_{i=1}^{40}(r_i - \overline{x})^2$$

Het steekproefgemiddelde en de steekproefvariantie zijn een **schatter** van het populatiegemiddelde μ en de populatievariantie σ^2

$$\overline{X} = \frac{1}{40}\sum_{i=1}^{40} X_i \text{ en } S^2 = \frac{1}{39}\sum_{i=1}^{40}(X_i - \overline{X})^2$$

De schatter heet **onvertekend** als het gemiddelde van de schatter samenvalt met de populatiekarakteristiek die hij schat.

$$\mathrm{E}(\overline{X}) = \mu$$

$$\mathrm{E}(S^2) = \sigma^2$$

STELLING 4.2 *Zij X een populatie-s.v. met gemiddelde waarde μ en variantie σ^2 en $X_1, X_2, \cdots X_n$ een steekproef uit die populatie dan is het steekproefgemiddelde $\overline{X}$ een onvertekende schatter van μ en de steekproefvariantie S^2 een onvertekende schatter van σ^2.*

Bewijs:

1. $\mathrm{E}(\overline{X}) = \mu$. Dit is reeds bewezen (zie stelling 4.1).

2. We tonen aan dat
$$\mathrm{E}(S^2) = \sigma^2$$
.
Voor elke toevalsveranderlijke Y geldt dat (zie hfdst.3.1 op pagina 94).

$$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 \Longleftrightarrow \mathrm{E}(Y^2) = \mathrm{Var}(Y) + (\mathrm{E}(Y))^2$$

$$\begin{aligned}
S^2 &= \frac{1}{n-1}\left(\sum_{i=1}^{n}(X_i-\overline{X})^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2-2X_i\overline{X}+\overline{X}^2) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}X_i^2 - 2\overline{X}\sum_{i=1}^{n}X_i + \overline{X}^2\sum_{i=1}^{n}1 \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - 2\overline{X}\cdot n\cdot\overline{X} + \overline{X}^2\cdot n\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(S^2) &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathrm{E}(X_i^2) - n\cdot\mathrm{E}(\overline{X}^2)\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\mathrm{Var}(X_i)+(\mathrm{E}(X_i))^2) - n(\mathrm{Var}(\overline{X})+(\mathrm{E}(\overline{X}))^2)\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\sigma^2+\mu^2) - n\frac{\sigma^2}{n} - n\mu^2\right) = \frac{1}{n-1}\left(n\sigma^2+n\mu^2-\sigma^2-n\mu^2\right) \\
&= \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2
\end{aligned}$$

In het voorbeeld van paragraaf 4.4.1 hebben we reeds het gemiddelde van het steekproefgemiddelde berekend en gezien dat dit gelijk is aan het populatiegemiddelde.
Het gemiddelde van de steekproefvariantie:

$$\mathrm{E}(S^2) = \frac{1}{25}(0+0.5+8+18+\cdots+4.5+0) = \frac{270}{25} = 10.8$$

Het gemiddelde van de steekproefvariantie is aldus gelijk aan de populatievariantie.

4.4.3 Betrouwbaarheidsintervallen bij het schatten van het populatiegemiddelde

De vraag is nu: hoe nauwkeurig is de uitspraak over de gemiddelde geboortelengte. Daarom gaan we een interval aangeven waarvan we kunnen zeggen dat dit interval "met grote kans" het populatiegemiddelde μ bevat.
Het ongekend steekproefgemiddelde μ ligt met een kans van 95% in het interval

$$\left[\overline{X} - 1.96 \cdot \sqrt{\frac{S^2}{n}}\ ,\ \overline{X} + 1.96 \cdot \sqrt{\frac{S^2}{n}}\right]$$

Als we nu een steekproef trekken en de gevonden getalwaarden invullen dan verkrijgen we een 95% **betrouwbaarheidsinterval**

$$\left[\overline{x} - 1.96 \cdot \sqrt{\frac{s^2}{n}}\ ,\ \overline{x} + 1.96 \cdot \sqrt{\frac{s^2}{n}}\right]$$

Voorbeeld van intervalschatten: Stel dat we voor 40 geboortelengten gevonden hebben dat $\overline{x} = 50.3$ en $s^2 = 28$.
Het 95% betrouwbaarheidsinterval voor de gemiddelde geboortelengte is $[48.66, 51.94]$. Dit betekent dat we 95% kans hebben dat het populatiegemiddelde gelegen is tussen 48.66 en 51.94.

OPMERKING:

- We kunnen evengoed een 80%, 90%, 99%,$\cdots$ betrouwbaarheidsinterval beschouwen. Meestal wordt het 95% betrouwbaarheidsinterval gebruikt.
- De formules zijn enkel geldig voor een voldoende grote steekproef, nl. $n \geq 30$.

OPGAVEN — **150** Een steekproef van 65 studenten geeft een gemiddelde van 67.5 kg en een variantie van 8.41 kg^2. De massa wordt verondersteld normaal verdeeld te zijn. Bepaal een 95% en 99% betrouwbaarheidsinterval van het populatiegemiddelde.

151 Door middel van een groot aantal steekproeven is men tot de vaststelling gekomen dat de levensduur van een bepaald soort gloeilampen de normale verdeling volgt met een standaardafwijking $\sigma = 121$ uur. Uit een willekeurige steekproef met grootte 35 bekomt men een gemiddelde levensduur van 1100 uren. Bepaal een 95% en 99% betrouwbaarheidsinterval van het populatiegemiddelde.

152 Voor de gemiddelde lengte van de diameter van 10 tandwielen vond men 4.38 cm. De variantie bedraagt 0.36 cm^2. Hoe groot is de kans dat het populatiegelmiddelde gelegen is in het interval $[4.014, 4.746]$

153 De steekproefresultaten van een normaal verdeelde populatie-s.v. zijn $\overline{x} = 235$ en $s^2 = 462.25$. Het interval $[215.78, 254.22]$ is een 95% betrouwbaarheidsinterval van het populatiegemiddelde. Hoe groot is de steekproefomvang?

154 Onderstel dat aan 36 patiënten een experimentele behandeling gegeven wordt van zodra de allereerste symptomen van de ziekte van Alzheimer geïndentificeerd zijn. We meten de tijd tot progressie van de ziekte. Voor de 36 patiëten is de gemiddelde tijd gelijk aan 27 maanden en de standaardafwijking is gelijk aan 9 maanden. Construeer een 95% betrouwbaarheidsinterval van de gemiddelde tijd tot progressie.

OPLOSSINGEN: 150: 95%: $[66,8; 68,2]$, 99%: $[66,6; 68,4]$; 151: 95%: $[1059,9; 1140,1]$; 99%: $[1047,3; 1152,7]$; 152: $94,6\%$ 153: $n = 5$; 154: $[24,1; 29,9]$

4.5 Populatieproportie

4.5.1 Gemiddelde en variantie van de streekproefproportie

In vele gevallen wordt een populatie verdeeld in twee groepen, de groep die een bepaalde eigenschap bezit en de groep die de eigenschap niet bezit. We willen dan weten wat de verhouding is van het aantal individuen met de eigenschap tot het totaal aantal individuen in de populatie. Bijvoorbeeld het aantal gebuisden in een school, het aantal mannen in een volksgroep, het aantal afgekeurde stukken in een productie. De **populatieproportie p van een groep** is de verhouding van het aantal in die groep tot het totaal aantal in de populatie.

We beschrijven de populatie door een toevalsveranderlijke X (met waarden 0 en 1) van een Bernoulli experiment met parameter $\mathbf{p}$, die de kans voorstelt op succes, dit is in dit geval de kans dat een individu uit de populatie tot de beschouwde groep behoort.

Populatiegemiddelde en populatievarantie van de populatie-s.v. X van de Bernoulli verdeling met parameter $\mathbf{p}$ zijn resp. (zie hfdst. 3.1 op pagina 103):

$$\mu = \mathrm{E}(X) = 1 \cdot \mathbf{p} + 0 \cdot (1 - \mathbf{p}) = \mathbf{p} \text{ (populatieproportie)}$$

$$\sigma^2 = \mathrm{Var}(X) = 1^2 \cdot \mathbf{p} + 0^2 \cdot (1 - \mathbf{p}) - \mathbf{p}^2 = \mathbf{p} - \mathbf{p}^2 = \mathbf{p}(1 - \mathbf{p})$$

Het steekproefgemiddelde van een steekproef $X_1, X_2, \cdots, X_n$ noemen we de steekproefproportie

$$P = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} (\text{het aantal keer succes}).$$

Als $n\mathbf{p} \geq 5$ en $n(1 - \mathbf{p}) \geq 5$ is P is normaal verdeeld met gemiddelde en variantie resp.

$$\mathrm{E}(P) = \mathbf{p} \quad \text{en} \quad \mathrm{Var}(P) = \frac{\sigma^2}{n} = \frac{\mathbf{p}(1 - \mathbf{p})}{n}$$

De proportie $\mathbf{p}$ is meestal een onbekende parameter van de populatie.

Als we een steekproef trekken verkrijgen we een waarde voor P die we voorstellen door $\hat{p}$.

4.5.2 Betrouwbaarheidsintervallen bij het schatten van een proportie

Voorbeeld: Bij de productie van lampen zijn er steeds een aantal slechte lampen. Men neemt een steekproef van 150 lampen en men vindt 6 slechte lampen. Een schatting voor de steekproefproportie (aantal slechte lampen op de 150 lampen) is

$$\hat{p} = \frac{6}{150} = 0.053$$

Stel dat we veel steekproeven met grootte 150 uitvoeren dan zullen we steeds een waarde krijgen voor $\hat{p}$.

De onbekende populatieproportie **p** ligt met een kans van 95% in het interval

$$\left[P - 1.96 \cdot \sqrt{\frac{P(1-P)}{n}}\ ,\ P + 1.96 \cdot \sqrt{\frac{P(1-P)}{n}}\right]$$

Voor deze steekproef is het 95% betrouwbaarheidsinterval

$$\left[\hat{p} - 1.96 \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\ ,\ \hat{p} + 1.96 \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right]$$

$= [0.53-1.96\cdot 0.0183\ ,\ 0.53+1.96\cdot 0.0183] = [0.53-0.0359\ ,\ 0.53+0.0359] = [0.017\ ,\ 0.089]$

De foutenmarge is 0.036. Dit betekent dat we met 95% zeker zijn dat de werkelijke proportie **p** niet meer dan 3.6% afwijkt van de steekproefproportie 0.053.

OPMERKING: De bovenstaande formules zijn enkel geldig als

- de populatie minstens 10 keer groter is dan de steekproefgrootte.
- voor n en **p** geldt dat $np \geq 5$ en $n(1-p) \geq 5$. Dit betekent dat voor **p** dicht bij 0.5 reeds kleine steekproeven volstaan en voor **p** dicht bij 0 of 1 grotere steekproeven nodig zijn.
- Aan de formule zien we dat de variantie van de steekproefproportie kleiner wordt naarmate de steekproefgrootte n groter wordt. Dit betekent dat als n groot is, de schatter $\overline{X}$ meer kans heeft om niet veel af te wijken van de werkelijke proportie **p**.

OPGAVEN — 155 Wij willen een 95% betrouwbaarheidsinterval voor de proportie ratten die blaaskanker krijgen in een experiment waarbij de proefdieren een dieet krijgen dat rijk is aan sacharine. Het dieet werd toegekend aan 50 ratten en 5 ervan kregen blaaskanker.

156 Wij willen een 95% betrouwbaarheidsinterval voor de proportie longkankerpatiënten van minder dan 40 jaar, die nog minstens 5 jaar leven na de diagnose. Bij een steekproef van grootte 30 waren slechts 6 patiënten die na 5 jaar nog in leven waren.

OPLOSSINGEN: 155: $[0,0168; 0,1831]$; 155: $[,057; 0,343]$

4.6 Toetsen van Hypothesen

Eén van de voornaamste technieken uit de verklarende statistiek is het toetsen van hypothesen op basis van steekproefgegevens.

Bij een toetsingprobleem begint men met het maken van een onderstelling omtrent de ongekende populatiekarakteristiek of populatieproportie. Zo een onderstelling noemen we de **nulhypothese** en wordt voorgesteld door

$$\mathrm{H}_0 : \mu = \mu_0 \quad \text{of} \quad \mathrm{H}_0 : \mathbf{p} = \mathbf{p}_0$$

.
Elke hypothese verschillend van de nulhypothese wordt een **alternatieve hypothese** genoemd en wordt voorgesteld door H_1.

Vervolgens zullen we een beslissing moeten nemen, nl. de nulhypothese verwerpen of niet verwerpen.
In geval we de nulhypothese ten onrechte verwerpen, zeggen we dat we een **type I fout** maken.
Aanvaarden we echter de nulhypothese ten onrechte dan maken we een **type II fout**.
De beslissingsregel zal geconstrueerd worden op basis van bepaalde criteria die er moeten voor zorgen dat de kans op een type I fout zo klein mogelijk is want een type I fout is erger dan een type II fout (afhankelijk van de aard van het vraagstuk).
Het **significantieniveau** is de kans op een type I fout die we aanduiden door α. Meestal neemt men α gelijk aan 5%.
De kans op een type II fout stellen we voor door β.

	H_0 is juist	H_0 is verkeerd
men verwerpt H_0	foutieve beslissing type I fout (α)	juiste beslissing ($1-\beta$)
men aanvaardt H_0	juiste beslissing ($1-\alpha$)	foute beslissing type II fout (β)

Voorbeelden:

- Indien we willen nagaan of een dobbelsteen correct is, stellen we

  $$\mathrm{H}_0 : \mathbf{p} = \frac{1}{6}$$

  waarbij **p** de kans voorstelt op het gooien van een zes.
  Een mogelijke alternatieve hypothese is:

  $$\mathrm{H}_0 : \mathbf{p} \neq \frac{1}{6}$$

Stel dat we bij 10 keer gooien geen enkele keer een 6 gooien. Op basis van die steekproef zou men geneigd zijn de nulhypohese te verwerpen en besluiten dat de dobbelsteen vals is. Dit kan echter een verkeerde beslissing zijn.

- In de VS moet een jury beslissen over het al dan niet schuldig zijn van een persoon. Indien de persoon schuldig is, krijgt hij de doodstraf. De nulhypothese is "de man is onschuldig". Een type I fout wordt gemaakt indien de jury de man de doodstraf geeft terwijl hij onschuldig is.

4.6.1 Toetsen van een hypothese over een populatiegemiddelde

Voorbeelden:

1. Aan een bepaalde universiteit neemt men sinds jaren een intelligentietest af die normaal-verdeelde uitslagen oplevert met gemiddelde score van 115. Een administrator wil nu voor de nieuwe lichting studenten de hypothese toetsen dat het gemiddelde hetzelfde is als in de voorbije jaren. Hij neemt een steekproef met grootte 50 en vindt $\overline{x} = 118$ en $s^2 = 98$. Welk besluit mag de administrator trekken op significatieniveau 5%. Zijn de studenten meer of minder intelligent dan in voorgaande jaren?

   OPLOSSING:

$$\mathrm{H}_0 : \mu = 115$$
$$\mathrm{H}_1 : \mu \neq 115$$

   De populatie s.v. is "het intelligentiequotiënt van studenten":

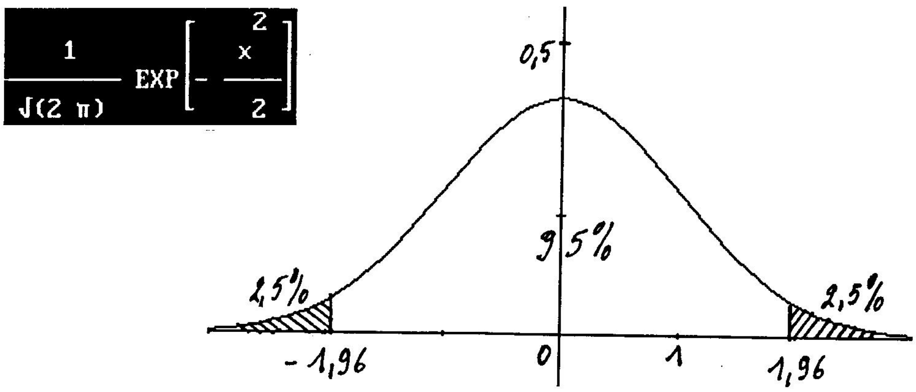
$$X \sim \mathrm{N}(115; 98) \quad \text{en} \quad \overline{X} \sim \mathrm{N}(115; \frac{98}{50})$$

   We kunnen op verschillende manieren toetsen om tot hetzelfde besluit te komen.

   - Het 95% aanvaardingsgebied voor $\frac{\overline{X}-115}{\sqrt{\frac{98}{50}}}$ die we terugvinden op de grafiek van standaardnormale dichtheidsfunctie is het interval $[-1,96 : 1,96]$.
     We drukken 118 uit in standaardeenheden:

$$\frac{118 - 115}{\sqrt{\frac{98}{50}}} = 2,14 > 1,96$$

     De waarde 2,14 valt buiten het aanvaardingsgebied. Daaruit kan de administrator besluiten dat de intelligentie van studenten niet meer gemiddeld 115 is. Duid de waarde 2,14 aan op de onderstaande tekening.



- Het 95% aanvaardingsgebied voor $\overline{X}$ is

$$\left[115 - 1,96 \cdot \sqrt{\frac{98}{50}} \quad ; \quad 115 + 1,96 \cdot \sqrt{\frac{98}{50}}\right] = [112,3; 117,7]$$

  Het steekproefgemiddelde van 118 valt er net buiten.

- Wat is de kans om, als de nulhypothese waar is, waarden te bekomen die even extreem of nog extremer zijn dan de observatie?

$$\mathbf{P}(\overline{X} \geq 118) = 1 - \mathbf{P}(\overline{X} \leq 118) = 0,016 < 0,025$$

  Deze kans noemen we de ***p*-waarde** (prob-value). Hier is de p-waarde kleiner dan 0,025. Dit betekent dat de observatie in het 5% significantiegebied ligt. Duid de p-waarde 0,016 aan op de voorgaande tekening.

Omdat het verwerpingsgebied of het significantiegebied zich op de grafiek van de dichtheidsfunctie in de twee staarten van elk 2,5% bevindt, zeggen we dat we een **tweezijdige toets** hebben uitgevoerd.
De **kritische punten** c bij significantieniveau 5% zijn de grenzen van het aanvaardingsgebied.

(a) Voor de standaardnormale kansverdeling zijn dat de waarden $-1,96$ en $1,96$;

(b) voor de verdeling van het intelligentiequotiënt zijn dat de waarden 112,3 en 117,7.

2. Meting van de bloeddruk van 60 toevalig gekozen eerstejaarsstudenten levert als resultaten (in mm Hg)

$$\sum_{i=1}^{60} x_i = 7829 \qquad ; \qquad \sum_{i=1}^{60} x_i^2 = 1035661$$

Volgens een handboek van klinische normen is de gemiddelde bloeddruk bij deze leefdheidsgroep 122,5. Is er reden om aan te nemen dat de bloeddruk bij de eerstejaarsstudenten hoger is dan deze norm.

OPLOSSING:

$$\mathrm{H}_0 : \mu = 122,5$$

$$\mathrm{H}_1 : \mu > 122,5$$

Het steekproefgemiddelde is

$$\overline{x} = \frac{1}{60}\sum_{i=1}^{60} x_i = \frac{7829}{60} = 130,5$$

De steekproefvariantie is

$$s^2 = \frac{1}{59}\sum_{i=1}^{60}(x_i - \overline{x})^2 = \frac{1}{59}\sum_{i=1}^{60} x_i^2 - \frac{60}{59}\overline{x}^2 = 239,10$$

De populatie s.v. is de bloeddruk van eerstejaarsstudenten

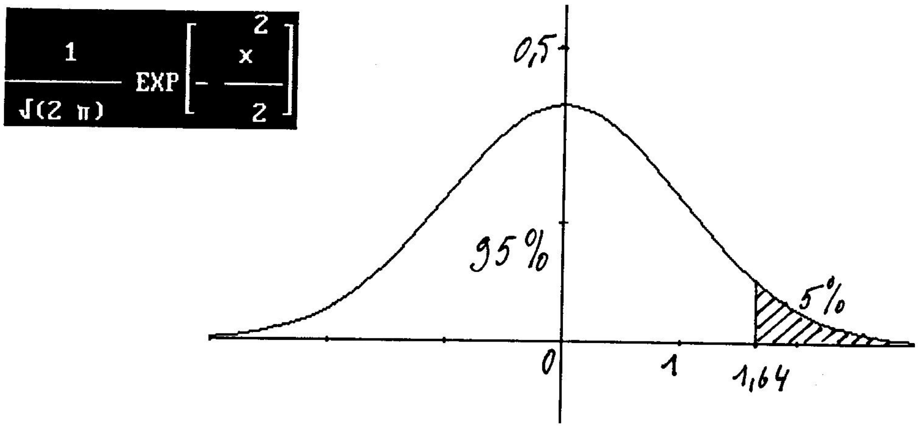
$$X \sim \mathrm{N}(122,5; 239,10) \quad \text{en} \quad \overline{X} \sim \mathrm{N}(122,5; \frac{239,10}{60}) = \mathrm{N}(122,5; 3,99)$$

De verschillende manieren om te toetsen:

- Het 95% aanvaardingsgebied voor $\frac{\overline{X}-122,5}{\sqrt{\frac{239,10}{60}}}$ die we terugvinden op de grafiek van standaardnormale dichtheidsfunctie is het interval $[-\infty; 1,645]$.
  We drukken 130,5 uit in standaardeenheden:

  $$\frac{130,5 - 122,5}{\sqrt{\frac{239,10}{60}}} = 4,00 > 1,645$$

  De waarde 4 valt buiten het aanvaardingsgebied. Duid de waarde 4 aan op de tekening.



- Het 95% aanvaardingsgebied voor $\overline{X}$ is

$$]-\infty\ ;\ \ 122,5+1,645\cdot\sqrt{\frac{239,10}{60}}]=]-\infty;125,8]$$

  Het steekproefgemiddelde van 130,5 ligt buiten het aanvaardingsgebied.
- p-waarde $= \mathbf{P}(\overline{X}\geq 130,5) = 1-\mathbf{P}(\overline{X}\leq 130,5) = 0,00 < 0,025.$

Omdat het verwerpingsgebied of het significantiegebied zich op de grafiek van de dichtheidsfunctie in de staart van 5% aan de rechterkant bevindt, zeggen we dat we een **rechts eenzijdige toets** hebben uitgevoerd.
Het **kritische punt** c bij significantieniveau 5% is de grens van het aanvaardingsgebied.

(a) Voor de standaardnormale kansverdeling is dat de waarden $1,645$;

(b) voor de verdeling van de bloeddruk is dat de waarde 125,8.

3. Een gynaecoloog beweert, bij de geboorte, meisjes gemiddeld kleiner zijn dan de aangegeven 51 cm die de norm is voor de gemiddelde lengte van meisjes bij de geboorte. Zijn collega echter verwijt hem dat zijn bewering berust op een vooroordeel en beweert dat de gemiddelde lengte wel degelijk 51 cm is. De afdeling gynaecologie doet een steekproef van 100 meisjes. Het steekproefgemiddelde en de steekproefvariantie zijn resp. $\overline{x} = 50,8$ en $s^2 = 1,6$ in cm.
   Is de gemiddelde geboortelengte van meisjes 51 cm of zijn meisjes bij de geboorte toch gemiddeld kleiner dan 51 cm?

Oplossing:

$$\mathrm{H}_0 : \mu = 51$$
$$\mathrm{H}_1 : \mu < 51$$

De populatie s.v. is "de lengte van pasgeboren meisjes":

$$X \sim \mathrm{N}(51; 1,6) \quad \text{en} \quad \overline{X} \sim \mathrm{N}(51; \frac{1,6}{100})$$

De verschillende manieren om te toetsen:

- Het 95% aanvaardingsgebied voor $\frac{\overline{X}-51}{\sqrt{\frac{1,6}{100}}}$ die we terugvinden op de grafiek van standaardnormale dichtheidsfunctie is het interval $[-1,645; +\infty]$.
  We drukken 50,8 uit in standaardeenheden:
  $$\frac{50,8 - 51}{\sqrt{\frac{1,6}{100}}} = -1,58 > -1,645$$
  De waarde -1,58 valt binnen het aanvaardingsgebied.
- Het 95% aanvaardingsgebied voor $\overline{X}$ is
  $$[51 - 1,645 \cdot \sqrt{\frac{1,6}{100}}; +\infty] = [50,79; +\infty[$$
  Het steekproefgemiddelde van 50,8 valt binnen het 95% aanvaardingsgebied. Daaruit kan de gynaecoloog besluiten dat de gemiddelde lengte van 51 cm nog steeds als norm geldt.
- p-waarde $= \mathbf{P}(\overline{X} \leq 50,8) = 0,44 > 0,05$

Omdat het verwerpingsgebied of het significantiegebied zich op de grafiek van de dichtheidsfunctie in de staart van 5% aan de linkerkant bevindt, zeggen we dat we een **links eenzijdige toets** hebben uitgevoerd.
Het **kritische punt** c bij significantieniveau 5% is de grens van het aanvaardingsgebied.

(a) Voor de standaardnormale kansverdeling is dat de waarden $-1,65$;

(b) voor de verdeling van de bloeddruk is dat de waarde 50,79

OPGAVEN — **157** Men wil het verband nagaan tussen dieet en serum cholesterol. Bij een steekproef van 23 vrouwen (30-35 jaar) wordt na een dieetkuur van 3 maanden de serum cholesterol gemeten: het gemiddelde is 176,9 mg/dl en de standaardafwijking is 28,7 mg/dl. Is dit een verbetering ten opzichte van het gemiddelde van 196 mg/dl in de algemene bevolking van vrouwen van die leeftijd? Neem aan dat serum cholesterolgehalte normaal verdeeld is.Toets op het 5% significantieniveau. Bereken ook de p-waarde.

158 Men wil nagaan of kinderen die binnen een straal van 5 km van een smeltoven wonen gemiddeld slechter presteren op een gestandardiseerde IQ test. Neem aan dat OQ normaal verdeeld is en dat de algemene bevolking van twaalfjarigen het gemiddeld IQ gelijk is aan 100 en de standaardafwijking gelijk is aan 16. Een steekproef van 25 twaalfjarigen die de voorbije jaren in de nabijheid van de smeltoven gewoond hebben levert een gemiddeld IQ op van 90. Toets op het 5% significantieniveau. Bereken ook de p-waarde.

159 De gemiddelde overlevingstijd na diagnose van een bepaalde hartkwaal is 2,5 jaar. Een nieuw medicament wordt gebruikt bij 50 dergelijke patiënten en zij worden gevolgd tot aan hun dood. men vond een gemiddelde overlevingstijd van 3,3 jaar en een standaardaking van 1,8 jaar. Heeft het gebruik van dit medicament de gemiddelde overlevingstijd verhoogd? Toets op 5% significantieniveau. bereken ook de p-waarde.

OPLOSSING: 157. p-waarde $= -3,19 < -1,645$, kritisch punt: $186,2 > 176,9$, 176,9 is significant, er is een verbetering met het dieet. 158. p-waarde $= -12,50 < -1,645$, kritisch punt: $99,9 < 90$, 90 is niet significant, er is geen daling van het gemiddeld IQ. 159. p-waarde $= 4,14 > 1,645$, kritisch punt: $2,92 < 3,3$, 3,3 is significant, het medicament heeft de gemiddelde overlevingstijd verhoogd.

4.6.2 Toetsen van een hypothese over één proportie

Bij een steekproef met grootte n beschouwen we de binomiaal verdeelde toevalsveranderlijken X met parameters n en $\mathbf{p}_0$.

$$X \sim B(n; \mathbf{p}_0) \text{ onder } \mathrm{H}_0 : \mathbf{p} = \mathbf{p}_0$$

Indien aan de voorwaarden $n\mathbf{p} > 5$ en $n(1-\mathbf{p}) > 5$ voldaan is, dan is de steekproefproportie P benaderd normale verdeeld.

$$P \sim \mathrm{N}(\mathbf{p}_0; \mathbf{p}_0(1-\mathbf{p}_0)) \text{ onder } \mathrm{H}_0 : \mathbf{p} = \mathbf{p}_0$$

Voorbeelden:

1. We beschouwen de decimale voorstelling van het getal π. Als nulhypothese zetten we voorop dat het het aantal even cijfers gelijk is aan het aantal oneven cijfers. Geven we de eerste 100 decimale cijfers van het getal dan tellen we 48 oneven en 52 even cijfers. Het is hier duidelijk dat we de nulhypothese kunnen aanvaarden.

$$\begin{aligned} \pi &= 3,14159265358979323846626433832795028841971693993751 \\ &\quad 05820974944592307816406286208998628034825342117067 \end{aligned}$$

   Maar krijgen we enkel de eerste 14 cijfers dan zijn er slechts 4 cijfers even.

$$\pi = 3,14159265358979$$

We toetsen tweezijdig.

$$X \sim \mathrm{B}(14; 1/2) \text{ onder } \mathrm{H}_0 : \mathbf{p} = 0,5$$

$$\mathrm{H}_1 : \mathbf{p} \neq 0,5$$

We berekenen de p-waarde:

$$\mathbf{P}(0) + \mathbf{P}(1) + \mathbf{P}(2) + \mathbf{P}(3) + \mathbf{P}(4)$$
$$= (0,5)^{14}(1 + 14 + 91 + 364 + 1001) = \frac{1471}{2^{14}} = 0,0897$$

p-waarde $= 0,0897 > 0,25$ betekent dat het steekproefresultaat van slechts 4 even cijfers op de 14 niet significant is. De nulhypothese dat er evenveel even als oneven decimalen zijn in het getal π mag niet verworpen worden.

$$\mathbf{P}(0) + \mathbf{P}(1) + \mathbf{P}(2) + \mathbf{P}(3) = \mathbf{P}(12) + \mathbf{P}(13) + \mathbf{P}(14) = 0,02868 > 0,025.$$

Hieruit besluiten we dat de uitkomsten die liggen in het significantiegebied met significatieniveau $\alpha = 0,05$ de uitkomsten 0, 1, 2, 12, 13 en 14 zijn.

2. Voor een loterij beweert de organisator dat de winstkans 60% is. Iemand koopt 10 lotjes en wint hiermee slechts 3 prijzen. Heeft de organisator de succeskans niet overdreven groot voorgesteld?

   We toetsen links-eenzijdig.

$$X \sim \mathrm{B}(10; 0,6) \text{ onder } \mathrm{H}_0 : \mathbf{p} = 0,6$$

$$\mathrm{H}_1 : \mathbf{p} < 0,6$$

   We berekenen de p-waarde

$$\begin{aligned}\mathbf{P}(0) + \mathbf{P}(1) + \mathbf{P}(2) + \mathbf{P}(3) &= (0,4)^{10} + 10(0,6)(0,4)^9 + 45(0,6)^2(0,4)^8 \\ &\quad +120(0,6)^3(0,4)^7 = 0,055\end{aligned}$$

   p-waarde $= 0,055 > 0,05$ betekent dat het steekproefresultaat van 3 prijzen op de 10 niet-significant is. De nulhypothese dat de winstkans 60% is, kan aanvaard worden.

3. Van een klassiek geneesmiddel tegen griep weet men sinds lang dat het 30% kans op genezing biedt. Van een nieuw geneesmiddel wordt beweerd dat het beter is. We proberen het geneesmiddel op 10 patiënten en stellen vast dat er 5 genezen. Is het nieuwe nu echt beter?

We toetsen rechts-eenzijdig.

$$X \sim \mathrm{B}(10; 0,3) \text{ onder } \mathrm{H}_0 : \mathbf{p} = 0,3$$

$$\mathrm{H}_1 : \mathbf{p} > 0,3$$

We berekenen de p-waarde

$$\begin{aligned}
&\mathbf{P}(5) + \mathbf{P}(6) + \mathbf{P}(7) + \mathbf{P}(8) + \mathbf{P}(9) + \mathbf{P}(10) = \\
&\quad 252(0,3)^5(0,7)^5 + 210(0,3)^6(0,7)^4 \\
&\quad +120(0,3)^7(0,7)^3 + 45(0,3)^8(0,7)^2 + 10(0,3)^9(0,7) + (0,3)^10 \\
&\quad = 0,1503
\end{aligned}$$

p-waarde $= 0,15 > 0,05$ betekent dat het steekproefresultaat van 5 genezingen op 10 niet-significant is. Het nieuw geneesmiddel is niet echt beter dan het klassiek geneesmiddel.

Het is pas significant vanaf 6 genezingen want dan is
p-waarde $= \mathbf{P}(6) + \mathbf{P}(7) + \mathbf{P}(8) + \mathbf{P}(9) + \mathbf{P}(10) = 0,047 < 0,05$

4. Een bakker beweert dat zijn broden in 70% van de gevallen meer wegen dan 800 gram. Je wenst dit na te gaan en koopt 12 broden bij hem. Na weging blijken slechts 6 broden meer dan 800 gram te wegen. Is de bewering van de bakker juist?

   We toetsen links-eenzijdig.

$$X \sim \mathrm{B}(12; 0,7) \text{ onder } \mathrm{H}_0 : \mathbf{p} = 0,7$$

$$\mathrm{H}_1 : \mathbf{p} < 0,7$$

   We berekenen de p-waarde:

$$\begin{aligned}
&\mathbf{P}(0) + \mathbf{P}(1) + \mathbf{P}(2) + \mathbf{P}(3) + \mathbf{P}(4) + \mathbf{P}(5) + \mathbf{P}(6) = \\
&\quad (0,3)^{12} + 12(0,3)^{11}(0,7) + 66(0,3)^{10}(0,7)^2 + 220(0,3)^9(0,7)^3 \\
&\quad +495(0,3)^8(0,7)^4 + 792(0,3)^7(0,7)^5 + 924(0,3)^6(0,7)^6 \\
&\quad = 0,1178
\end{aligned}$$

   p-waarde $= 0,1178 > 0,05$ betekent dat het steekproefresultaat van 6 broden op de 12 niet-significant is. De nulhypothese dat de 70% van de broden meer wegen dan 800 gram kan aangehouden worden.

   Het zal pas significant zijn als er slechts 5 broden op de 12 meer wegen dan 800 gram want dan is de p-waarde

$$\mathbf{P}(0) + \mathbf{P}(1) + \mathbf{P}(2) + \mathbf{P}(3) + \mathbf{P}(4) + \mathbf{P}(5) = 0,0386 < 0,05$$

5. Bij een dobbelspel ontdekt Suna dat er een aantal dobbelstenen vervalst zijn. Als men gooit met een valse dobbelsteen dan is de kans op een zes gelijk aan 1/4. Suna wil het probleem snel oplossen en ze beslist dat de leerlingen van haar klas elke dobbelsteen 10 keer moeten opgooien. Als een dobbelsteen minstens 4 keer een zes oplevert dan beslist ze dat de dobbelsteen vals is. Anders wordt de dobbelsteen als eerlijk beschouwd.

   De nulsituatie is dat de situatie waarin de dobbelsteen eerlijk is en de alternatieve hypothese is dat de dobbelsteen vervalst is.

$$X \sim \mathrm{B}(10;1/6) \text{ onder } \mathrm{H}_0 : \mathbf{p} = 1/6$$
$$Y \sim B(10;1/4) \text{ onder } \mathrm{H}_1 : \mathbf{p} = 1/4$$

   (a) De beslissingsregel staat vast. Het aanvaardingsgebied voor X is $\{0,1,2,3\}$ en het verwerpinggebied is $\{4,5,6,7,8,9,10\}$.

   De kans op een type I fout is de kans dat de dobbelsteen als vals wordt beschouwd, terwijl hij eerlijk is.

$$\alpha = \mathbf{P}(X \geq 4) = \sum_{i=4}^{10} \mathbf{P}(X = i) = 0,0697$$

   Het significantieniveau is 7,0%

   De kans op een type II fout is de kans dat de dobbelsteen als eerlijk wordt beschouwd, terwijl hij eigenlijk vals is.

$$\beta = \mathbf{P}(Y \leq 3) = 0,776$$

   De **power van de toets** is $1-\beta = 0,224$. Dit is de kans dat we de nulhypothese verwerpen (dobbelsteen is eerlijk) als hij inderdaad vals is.

   (b) Nemen we een andere beslissingsregel, bvb. met aanvaardingsgebied $\{0,1,2\}$ en verwerpinggebied $\{3,4,5,6,7,8,9,10\}$.

   Omdat we het verwerpingsgebied groter maken, maken we de kans op een type I fout groter, de kans op een type II fout wordt kleiner en de power van de test wordt groter. De kans op een type I fout is:

$$\alpha = \mathbf{P}(X \geq 3) = \sum_{i=3}^{10} \mathbf{P}(Y = i) = 0,2248$$

   Het significantieniveau is 22,5%

   De kans op een type II fout is:

$$\beta = \mathbf{P}(Y \leq 2) = \sum_{i=0}^{2} \mathbf{P}(X = i) = 0,5256$$

   De power van de toets is $1 - \beta = 0,4744$.

(c) Bepaal zelf wat er gebeurt als we i.p.v. minstens 4 keer een zes, minstens 5 keer een zes nemen ($\alpha = 0,0156$ en $\beta = 0,9219$).

(d) Bepaal zelf wat er gebeurt als we i.p.v. minstens 4 keer een zes, minstens 6 keer een zes nemen ($\alpha = 0,0024$ en $\beta = 0,9803$).

(e) Wat is de beslissingsregel als het significantieniveau hoogstens gelijk is aan $0,05$ is.
Omdat de binomiale verdeling discreet is, is het onmogelijk om voor een bepaalde beslissingsregel exact 0,05 te bekomen voor α. We kiezen dan de beslissingsregel waarvoor α hoogstens 0,05 is. Dit is in geval $X \geq 5$ en $\alpha = 0,0154$.

6. Hernemen we vorig voorbeeld. Stephanie vindt 10 keer gooien weinig om een beslissing te nemen. Stephanie geeft opdracht aan de medeleerlingen om 60 keer te gooien en de dobbelsteen als vals te beschouwen als men hierbij minstens 14 keer een zes gooit.

(a) Bepaal de kans op een type I fout, de kans op een type II fout en de power van de test.

$$\mathrm{H}_0 : \mathbf{p} = \frac{1}{6}$$
$$\mathrm{H}_1 : \mathbf{p} = \frac{1}{4}$$

Omdat $n\mathbf{p} = 10 \geq 5$ geldt

$$P_0 \sim \mathrm{N}(\frac{1}{6}; \frac{1}{60}\frac{1}{6}\frac{5}{6})$$

$$P_1 \sim \mathrm{N}(\frac{1}{4}; \frac{1}{60}\frac{1}{4}\frac{3}{4})$$

$$\alpha = \mathbf{P}(P_0 \geq \frac{14}{60}) = 0,0829$$

$$\beta = \mathbf{P}(P_1 \leq \frac{14}{60}) = 0,3828$$

$$1 - \beta = 0,6172$$

(b) Wat is de beslissingsregel als het significantieniveau hoogstens gelijk is aan $0,05$ is.
De toets is rechts eenzijdig. Het kritisch punt is

$$\frac{1}{6} + 1,645 \cdot \frac{1}{12\sqrt{3}} = 0,2458$$

De beslissingregel ligt op 15 keer een zes gooien als we 60 keer gooien.

OPGAVEN — 160 De leerlingen van de 6des hebben een munstuk 100 keer opgegooid en hebben 58 keer kop verkregen. Voor een onvervalst muntstuk verwachten we 50 keer kop op de 100 keer opgooien.

1. Mag men aanvaarden dat het geldstuk onvervalst is op significantieniveau 5%? Toets dit op de verschillende mogelijke manieren. Ga na of je hierbij mag benaderen door de normale verdeling.
2. Bepaal de kritisch punten.
3. Als je als alternatieve hypothese stelt dat de proportie 0,6 is, wat is dan de power van de test?

161 Een reismaatschappij vertelt ons bij boeking dat Noorwegen een land is met weinig regen. De kans op regen is er 0,4 zegt hij. Kristof wil dat nagaan en vertrekt stande pede naar Noorwegen met het eerste beste vliegtuig. Kristof sms-t naar huis dat van de 30 dagen er 16 dagen zijn met regen. Overdrijft de reismaatschappij?

162 Een wereldbekend doelwachter heeft zich een stevige reputatie opgebouwd dat hij 70% van de strafschoppen stopt. Om hem te testen worden 10 topvoetballers gevraagd een strafschop te geven. Hij stopt er slechts 4. Is zijn reputatie wat overroepen op significantieniveau 5%?

163 Een leverancier van bloembollen beweert dat in zijn speciale mengeling van tulpebollen er 1 witte is voor elke 3 rode. Je plant zo'n pakje van 20 en vindt volgende lente 2 witte tulpen en 18 rode. Geloof je de bewering van de leverancier?

164 Men wil door een muntstuk 4 keer op te gooien testen of het homogeen is. Men wil dit doen op significantieniveau 0,05. Kan dat?

165 Men beweert dat de helft van de volwassen bevolking zich jaarlijks laat inenten tegen griep. Een navraag bij 30 lukraak gekozen volwassenen leert ons dat hiervan 18 ingeënt werden. Is dit in overeenstemming met de bewering?

166 Iemand beweert dat 90% van de mensen niet in staat is het verschil te proeven tussen twee soorten limonade. De fabricanten vinden dat overdreven en doen een test.

a. Bij 500 lukraak gekozen personen waren er 72 die het verschil proefden. Is de bewering overdreven?
b. Wat is de conclusie voor een kleine steekproef? Stel dat bv. bij 10 lukraak gekozen personen er 3 het verschil proefden.

OPLOSSINGEN: 160. het muntstuk is niet vervalst; de kritische punten zijn 41 en 59; power van de toets is 0,65; 160. ; Uit $1,49 < 1,64$ kunnen we het besluit trekken dat de reismaatschappij niet overdrijft; 162. p-waarde $= 0,0473 < 0,05$: significant ; 163. p-waarde $= 0,0913 > 0,025$ niet significant ; 164. neen ; 165. p-waarde $= 0,1808 > 0,025$ niet significant ; 166. (a) p-waarde $= 0,00114 < 0,05$ significant, (b) p-waarde $= 0,0703 > 0,05$ niet significant.

4.6.3 Toetsen van een hypothese over twee proporties

4.6.3.1 Hypergeometrische verdeling

Tot hiertoe hebben we steekproeven gedaan in één enkele populatie bv. bij het testen van de broden van een bepaalde bakker, bij het testen van een nieuw medicament.

Hier gaan we steekproeven uitvoeren in twee verschillende populaties, bv. bij het vergelijken van twee verschillende geneesmiddelen aangewend voor dezelfde ziekte kan men het verschil in proportie genezingen tot het aantal behandelde zieken bestuderen.

We beschouwen een universum V van M elementen, waarvan er M_1 behoren tot een populatie A en M_2 behoren tot een populatie B. We trekken n elementen uit deze M zonder terugplaatsen.

We vragen ons af wat de kans is op de gebeurtenis "n elementen trekken uit M elementen waarvan er precies X elementen behoren tot de populatie A". De toevalsveranderlijke X heeft dan de zogenaamde **hypergeometrische verdeling**.

$$X \sim \mathrm{HYP}(n; M_1; M)$$

4.6.3.2 Berekening van de hypergeometrische verdeling

Het totaal aantal mogelijkheden om n elementen te kiezen uit M elementen is $\begin{pmatrix} M \\ n \end{pmatrix}$.

Men kan i elementen kiezen uit de M_1 elementen van A op $\begin{pmatrix} M_1 \\ i \end{pmatrix}$ manieren. De $n-i$ overige elementen moeten dan gekozen worden uit de M_2 elementen van B en dit kan op $\begin{pmatrix} M_2 \\ n-i \end{pmatrix}$ manieren.

Bij elke keuze van i elementen uit A hoort een keuze van $n-i$ elementen uit B.

Het aantal mogelijkheden dat we bij het trekken van n elementen uit het universum V er i komen uit A en $n-i$ komen uit B is:

$$\begin{pmatrix} M_1 \\ i \end{pmatrix} \begin{pmatrix} M_2 \\ n-i \end{pmatrix}.$$

De kans dat er op de n elementen die komen uit V er i komen uit A en $n-i$ komen uit B of de kans dat de toevalsveranderlijke X de waarde i aanneemt, is:

$$\mathbf{P}(X = i) = \frac{\begin{pmatrix} M_1 \\ i \end{pmatrix} \begin{pmatrix} M_2 \\ n-i \end{pmatrix}}{\begin{pmatrix} M \\ n \end{pmatrix}}.$$

Met Excel kunnen we al kansen van $\mathrm{HYP}(n; M_1; M)$ berekenen.
We schrijven

$$\mathbf{P}(X = i) = \mathrm{HYPERGEO.VERD}(i; n; M_1; M)$$

4.6.3.3 De gemiddelde waarde en variantie bij een hypergeometrische verdeling

1. $\boxed{\text{De gemiddelde waarde}}$ Het bewijs steunt op de volgende eigenschap van binomiaalcoëfficiënten nl.

$$i\binom{M_1}{i} = i\frac{M_1(M_1-1)!}{M_1-i)!i!} = M_1\frac{(M_1-1)!}{(M_1-1-(i-1))!(i-1)!} = M_1\binom{M_1-1}{i-1}$$

en een eigenschap van hfdst. 1 in oefening nr. 80 op pagina 41.

$$\binom{n+m}{p} = \sum_{k=0}^{p}\binom{n}{p-k}\cdot\binom{m}{k}$$

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{i=0}^{n} i\frac{\binom{M_1}{i}\binom{M_2}{n-i}}{\binom{M_1+M_2}{n}} = \frac{M_1}{\binom{M_1+M_2}{n}}\sum_{i=1}^{n}\binom{M_1-1}{i-1}\binom{M_2}{n-i} \\
&= \frac{M_1}{\binom{M_1+M_2}{n}}\sum_{i'=0}^{n-1}\binom{M_1-1}{i'}\binom{M_2}{n-i'-1} \\
&= \frac{M_1}{\binom{M_1+M_2}{n}}\binom{M_1+M_2-1}{n-1} = n\frac{M_1}{M_1+M_2} = n\frac{M_1}{M}
\end{aligned}$$

2. $\boxed{\text{De variantie}}$ Het bewijs steunt ook op oefening 80 en op de volgende eigenschap van binomiaalcoëfficiënten nl.

$$i(i-1)\binom{M_1}{i} = i(i-1)\frac{M_1(M_1-1)(M_1-2)!}{(M_1-i)!(i!} = M_1(M_1-1)\frac{(M_1-2)!}{(M_1-2-(i-2))!(i-2)!}$$

$$= M_1(M_1-1)\binom{M_1-2}{i-2}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{i=0}^{n} i^2 \frac{\begin{pmatrix} M_1 \\ i \end{pmatrix}\begin{pmatrix} M_2 \\ n-i \end{pmatrix}}{\begin{pmatrix} M_1+M_2 \\ n \end{pmatrix}} - \left(n\frac{M_1}{M_1+M_2}\right)^2 \\
&= \sum_{i=0}^{n} (i^2-i) \frac{\begin{pmatrix} M_1 \\ i \end{pmatrix}\begin{pmatrix} M_2 \\ n-i \end{pmatrix}}{\begin{pmatrix} M_1+M_2 \\ n \end{pmatrix}} + \sum_{i=0}^{n} i \frac{\begin{pmatrix} M_1 \\ i \end{pmatrix}\begin{pmatrix} M_2 \\ n-i \end{pmatrix}}{\begin{pmatrix} M_1+M_2 \\ n \end{pmatrix}} \\
&\quad - \left(n\frac{M_1}{M_1+M_2}\right)^2 \\
&= \frac{1}{\begin{pmatrix} M_1+M_2 \\ n \end{pmatrix}} \sum_{i=2}^{n} i(i-1) \begin{pmatrix} M_1 \\ i \end{pmatrix}\begin{pmatrix} M_2 \\ n-i \end{pmatrix} + n\frac{M_1}{M_1+M_2} \\
&\quad - \left(n\frac{M_1}{M_1+M_2}\right)^2 \\
&= \frac{M_1(M_1-1)}{\begin{pmatrix} M_1+M_2 \\ n \end{pmatrix}} \sum_{i=2}^{n} \begin{pmatrix} M_1-2 \\ i-2 \end{pmatrix}\begin{pmatrix} M_2 \\ n-i \end{pmatrix} + n\frac{M_1}{M_1+M_2} \\
&\quad - \left(n\frac{M_1}{M_1+M_2}\right)^2 \\
&= \frac{M_1(M_1-1)}{\begin{pmatrix} M_1+M_2 \\ n \end{pmatrix}} \sum_{i'=0}^{n-2} \begin{pmatrix} M_1-2 \\ i' \end{pmatrix}\begin{pmatrix} M_2 \\ n-i'-2 \end{pmatrix} + n\frac{M_1}{M_1+M_2} \\
&\quad - \left(n\frac{M_1}{M_1+M_2}\right)^2 \\
&= \frac{M_1(M_1-1)}{\begin{pmatrix} M_1+M_2 \\ n \end{pmatrix}} \begin{pmatrix} M_1+M_2-2 \\ n-2 \end{pmatrix} + n\frac{M_1}{M_1+M_2} - \left(n\frac{M_1}{M_1+M_2}\right)^2 \\
&= \frac{M_1(M_1-1)(M_1+M_2-n)!n!}{(M_1+M_2)!} \frac{(M_1+M_2-2)!}{(M_1+M_2-2-n+2)!(n-2)!} \\
&\quad + n\frac{M_1}{M_1+M_2} - \left(n\frac{M_1}{M_1+M_2}\right)^2 \\
&= \frac{M_1(M_1-1)n(n-1)}{(M_1+M_2)(M_1+M_2-1)} + n\frac{M_1}{M_1+M_2} - \left(n\frac{M_1}{M_1+M_2}\right)^2 \\
&= \frac{n^2M_1^2 + nM_1M_2 - n^2M_1}{(M_1+M_2)(M_1+M_2-1)} - \left(n\frac{M_1}{M_1+M_2}\right)^2 \\
&= \frac{nM_1M_2(M_1+M_2-n)}{(M_1+M_2)^2(M_1+M_2-1)} = n\frac{M_1}{M_1+M_2} \cdot \frac{M_2}{M_1+M_2} \cdot \frac{M-n}{M_1+M_2-1}
\end{aligned}$$

Opmerking: Bij een hypergeometrische verdeling gebeuren de n trekkingen zonder terugleggen. Beschouwen we de n trekkingen maar met terugleggen dan is de verdeling binomiaal. De kans dat we een element trekken uit de populatie A is gelijk aan

$$\mathbf{p} = \frac{M_1}{M_1 + M_2}.$$

Hieruit volgt dat

$$1 - \mathbf{p} = \frac{M_2}{M_1 + M_2}.$$

De gemiddelde waarde bij n trekkingen is hier

$$\mu = n\mathbf{p} = n\frac{M_1}{M_1 + M_2}$$

en de standaardafwijking is

$$\sigma = \sqrt{n\frac{M_1}{M_1 + M_2}\frac{M_2}{M_1 + M_2}}.$$

De gemiddelde waarde is bij een trekking met teruglegging dezelfde als bij een trekking zonder teruglegging. De standaardafwijking bij een trekking zonder teruglegging is kleiner dan bij een trekking met teruglegging aangezien de factor

$$\frac{M_1 + M_2 - n}{M_1 + M_2 - 1} < 1.$$

4.6.3.4 Kleine steekproef

We gaan hier een hypothese toetsen over twee proporties d.m.v. een kleine steekproef. We illustreren met een aantal voorbeelden.

- Men wil nagaan of een nieuw medicament het beter doet dan een klassiek medicament. Hiertoe gaat men als volgt te werk: 12 patiënten krijgen het nieuw medicament en 8 patiënten krijgen het klassiek medicament toegediend. Na verloop van tijd telt men het aantal genezingen. In de groep A van het nieuwe medicament heeft men 7 genezingen en in de andere groep B 2 genezingen.

  De steekproefgegevens kunnen in een vierveldentabel voorgesteld worden (contingentietabel).

	genezen	niet genezen	$\sum$
A	7	5	$M_1 = 12$
B	2	6	$M_2 = 8$
$\sum$	$n = 9$	11	$M = 20$

De alternatieve hypothese is dat het nieuwe medicament beter is dan het klassiek medicament.

Als toevalsveranderlijke X nemen we het aantal personen dat geneest en dat tot de groep A behoort.

$$X \sim \text{HYP}(9; 12; 20)$$

H_0: we verwachten gemiddeld 5,4 genezingen als het nieuwe medicament niet beter is dan het klassiek medicament.
H_1: er zijn gemiddeld meer 5,4 genezingen.
X varieert tussen de waarden 1 en 9. Voor het steekproefresultaat heeft de veranderlijke X de waarde 7. Het toetsen van dit steekproefresultaat noemt men de **exacte toets van Fisher**.

De kans dat er van de 9 mensen die genezen er x zijn die het nieuwe medicament genomen hebben, is

$$\mathbf{P}(X = x) = \frac{\binom{12}{x}\binom{8}{9-x}}{\binom{20}{9}}.$$

We berekenen de p-waarde:

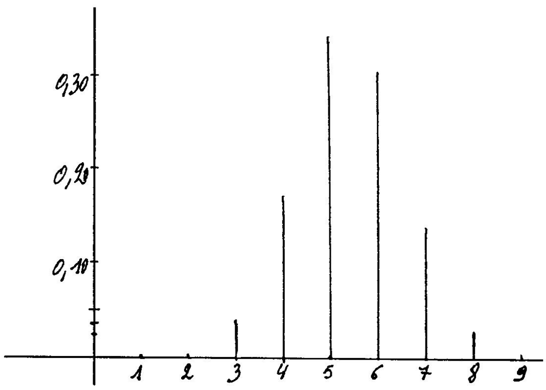
$$\begin{aligned} \mathbf{P}(X \geq 7) &= \mathbf{P}(7) + \mathbf{P}(8) + \mathbf{P}(9) \\ &= \frac{\binom{12}{7}\binom{8}{2} + \binom{12}{8}\binom{8}{1} + \binom{12}{9}\binom{8}{0}}{\binom{20}{9}} \\ &= 0,1569 \end{aligned}$$

De toets is rechts-eenzijdig.
p-waarde $= 0,1569 > 0,05$ betekent dat het steekproefresultaat van 7 genezingen behoort tot het waarschijnlijkheidsgebied van de hypergeometrische verdeling. De nulhypothese mag niet verworpen worden. Het nieuw medicament is niet significant beter. De kansen op genezing zijn voor de beide medicamenten gelijk.
Merk op dat de hypergeometrische verdeling niet symmetrisch is.

Vanaf 8 genezingen met het nieuwe medicament zouden we de nulhypothese verwerpen want dan geldt dat de p-waarde $= \mathbf{P}(X \geq 8) = 0,0249 < 0,05$.

- Een firma wil haar verkopers trainen. Ze verdeelt lukraak haar vertegenwoordigers in twee groepen A en B. De twee groepen worden getraind met twee verschillende trainingsmethodes. Na een bepaalde tijd gaat men de verkoopcijfers na.

  De vierveldentabel is



Figuur 4.2: Hypergeometrische verdeling voor $n = 9$, $M_1 = 12$ en $M_2 = 8$

	boven 10%	beneden 10%	$\sum$
A	15	5	$M_1 = 20$
B	9	10	$M_2 = 19$
$\sum$	$n = 24$	15	$M = 39$

Is er een verschil in de trainingsmethodes?

Als toevalsveranderlijke X nemen we het aantal verkopers van groep A waarvoor het verkoopcijfer stijgt met meer dan 10%.

$$X \sim \mathrm{HYP}(24; 20; 39)$$

H_0: we verwachten gemiddeld 12,3 verkopers waarvoor het verkoopcijfer meer dan 10% verhoogd als de trainingsmethodes niet verschillen van elkaar.
H_1: het gemiddelde is niet gelijk aan 12,3.
Deze toevalsveranderlijke varieert hier tussen de waarden 5 en 20. Voor het steekproefresultaat heeft de veranderlijke X de waarde 15.

Omdat de hypergeometrische verdeling niet symmetrisch is, nemen we als p-waarde bij de tweezijdige toets:

$$2\min\{\mathbf{P}(X \leq 15, \mathbf{P}(X \geq 15)\}$$

$$\mathbf{P}(X \geq 15) = 1 - \mathbf{P}(X \leq 14) = 0,074$$

Uit p-waarde $= 2 \cdot 0,074 = 0,148 > 0,05$ volgt dat de trainingsmethode van groep A niet significant beter is dan die van B.

4.6.3.5 Grote steekproef

De steekproef waarbij men n trekt zonder terugleggen uit $M_1 + M_2$ elementen geeft een resultaat die een waarde is van de stochastische veranderlijke van een hypergeometrische verdeling. De berekeningen worden zeer omslachtig en onuitvoerbaar als we te maken hebben met grote waarden van n, M_1 en M_2.

Het is intuïtief duidelijk dat als M_1 en $M_1 + M_2 = M$ zeer groot zijn het weinig verschil zal uitmaken of we n trekkingen uitvoeren met terugleggen of zonder terugleggen. Hieruit volgt dat we een hypergeometrische verdeling zullen kunnen benaderen door een binomiale verdeling. Deze binomiale verdeling moet dan omwille van de grote waarde van n benaderd worden door een normale verdeling.

STELLING 4.3 *Als $M = M_1 + M_2 \longrightarrow +\infty$ en $M_1 \longrightarrow +\infty$ zodanig dat $\frac{M_1}{M} \longrightarrow \mathbf{p}$, dan is*

$$\mathbf{P}(X = i) = \begin{pmatrix} n \\ i \end{pmatrix} \mathbf{p}^i (1 - \mathbf{p})^{n-i}.$$

Bewijs

$$\mathbf{P}(X = i) = \frac{\begin{pmatrix} M_1 \\ i \end{pmatrix}\begin{pmatrix} M_2 \\ n-i \end{pmatrix}}{\begin{pmatrix} M \\ n \end{pmatrix}} = \frac{\frac{M_1!}{i!(M_1-i)!}\frac{M_2!}{M_2-n+i)!(n-i)!}}{\frac{M!}{n!(M-n)!}} = \frac{n!}{(n-i)!i!}\frac{M_1!}{(M_1-i)!}\frac{M_2!}{M_2-n+i)!}$$

$$= \begin{pmatrix} n \\ i \end{pmatrix} \cdot \left(\frac{M_1}{M} \cdot \frac{M_1 - 1}{M - 1} \cdots \frac{M_1 - i + 1}{M - i + 1}\right)\left(\frac{M_2}{M - i} \cdot \frac{M_2 - 1}{M - i - 1} \cdots \frac{M_2 - n + i + 1}{M - n + 1}\right)$$

$$= \begin{pmatrix} n \\ i \end{pmatrix} \cdot \mathbf{p}^i \cdot (1 - \mathbf{p})^{n-i}.$$

□

$$X \sim \mathrm{HYP}(n; M_1; M)$$

Voor grote waarden van M_1 en M is

$$X \sim \mathrm{B}(n; \frac{M_1}{M})$$

De populatieproportie P is bijgevolg normaal verdeeld

$$P \sim \mathrm{N}(\frac{M_1}{M}, \frac{M_1 M_2}{n M^2})$$

Voorbeelden

- Een farmaceutisch bedrijf brengt een product A tegen haaruitval op de markt. Het beweert dat het beter is dan product B van zijn concurrent. Bij 500 personen met haaruitval wordt een test uitgevoerd waarbij ze lukraak aan product A of aan product B worden toegewezen. De resultaten stellen we voor in een vierveldentabel.

	haaruitval gestopt	haaruitval niet gestopt	$\sum$
product A	180	80	$M_1 = 260$
product B	160	80	$M_2 = 240$
$\sum$	$n = 340$	160	500

  Is het product A beter dan het product B?

  Als toevalsveranderlijke X nemen we het aantal personen waarvan de haaruitval stopt en het product A gebruikte.

  $$X \sim \mathrm{HYP}(340; 260; 500)$$
  $$P \sim \mathrm{N}(\frac{260}{500}; \frac{1}{340}\frac{260}{500}\frac{240}{500})$$
  $$\mathrm{H}_0 : \mathbf{p} = \frac{260}{500}$$
  $$\mathrm{H}_1 : \mathbf{p} > \frac{260}{500}$$

  p-waarde $= \mathbf{P}(P \geq \frac{180}{340}) = 1 - \mathbf{P}(P \leq \frac{180}{340}) = 0,364$.
  De steekproef is rechts-eenzijdig.
  De p-waarde $= 0,364 > 0,05$ betekent dat het steekproefresultaat niet significant is op significantieniveau 0,05. We aanvaarden de nulhypothese dat beide producten dezelfde uitwerking hebben. Het product A is niet significant beter dan product B.

- Een producent van kattevoeding beweert dat zijn product A lever bevat en daardoor zeer geliefd is bij katten. Zijn concurrent beweert dat deze diertjes liever rundsvlees eten en heeft dit in zijn product B gestopt.

  De vierveldentabel is

	lusten het	lusten het niet	$\sum$
product A	120	60	$M_1 = 180$
product B	100	120	$M_2 = 220$
$\sum$	$n = 220$	180	400

We vragen ons af of er een significant verschil bestaat in de voorkeur voor lever of rundsvlees bij katten.

We hebben hier te doen met een tweezijdige toets.

Als toevalsveranderlijke X nemen we het aantal katten die het product met lever lusten.

$$X \sim \mathrm{HYP}(220; 180; 400)$$
$$P \sim \mathrm{N}(\frac{180}{400}; \frac{1}{220}\frac{180}{400}\frac{220}{400})$$
$$\mathrm{H}_0 : \mathbf{p} = \frac{180}{400}$$
$$\mathrm{H}_1 : \mathbf{p} \neq \frac{180}{400}$$

p-waarde $= \mathbf{P}(P \geq \frac{120}{220}) = 1 - \mathbf{P}(P \leq \frac{120}{220}) = 0,0022$

De p-waarde $= 0,0022 < 0,025$ betekent dat het steekproefresultaat significant is op significantieniveau 0,05. We verwerpen de nulhypothese dat katten geen voorkeur hebben voor lever.

OPGAVEN — 167 Voor een bepaalde ziekte wordt een nieuwe therapie uitgetest tegenover de oude therapie. Daartoe worden 30 patiënten lukraak toegewezen aan één van beide behandelingen. Na verloop van 4 weken telt men hoeveel patiënten volledig genezen zijn. De vierveldentabel is

	genezen	niet genezen	
nieuw	12	4	$M_1 = 16$
oud	6	8	$M_2 = 14$
$\sum$	$n = 18$	12	30

Is de nieuwe therapie significant beter dan de oude?

168 Voor het practisch rijexamen zijn er twee examinatoren A en B. Een steekproef van grootte 39 uit de personen die bij A terechtkwamen levert 36 geslaagden. Een steekproef van grootte 16 uit degene die bij B terechtkwamen levert 10 geslaagden. Is de slaagkans bij A groter dan bij B?

169 In een lukraak gekozen steekproef van 100 mannelijke chauffeurs waren er 15 met een ongeval. In een steekproef van grootte 81 bij de vrouwelijke chauffeurs waren er 4 met een ongeval. Is er reden om aan te nemen dat de kans op een ongeval groter is bij de mannen dan bij de vrouwen?

170 Aspirine als preventie tegen hartaanval? (Harward Medical School, 1988-89)
De "Physicians'Health Study" is een beroemde clinische studie, uitgevoerd in de tachtiger jaren bij meer dan 20000 Amerikaanse artsen (vrijwilligers). Een deelaspect van deze studie was nagaan of het regelmatig gebruik van aspirine de kans op hartinfarct vermindert. Daartoe werden de 22071 proefpersonen lukraak toegewezen aan

a. ofwel de aspirine groep (om de andere dag 1 tabletje aspirine);

b. ofwel de placebo groep (om de andere dag 1 neptabletje).

Het experiment was 'blind', d.w.z. de proefpersonen in de studie wisten niet tot welke groep ze behoorden. Na ongeveer 5 jaar follow-up (1982-87) verschenen de eerste resultaten in de *The New England Journal of Medicine*. In het eindrapport vinden we:

	hartaanval	geen hartaanval	
Placebo groep	239	10795	$M_1 = 11034$
Aspirine groep	139	10898	$M_2 = 11037$
$\sum$	$n = 378$	21693	22071

Toon aan dat de kans op hartinfarct significant groter is bij de placebo groep dan bij de aspirine groep. Omwille van het spectaculair gunstig resultaat werd de studie vroegtijdig gestopt om alle proefpersonen van de resultaten op de hoogte te kunnen brengen.

171 Eenzelfde examen werd gegeven in twee klassen A en B van gelijk niveau. In A zijn studenten 40 studenten, de gemiddelde uitslag was 74% en de standaardafwijking 8; in B zijn 50 studenten, de gemiddelde uitslag was 78% en de standaardafwijking 7. Is het verschil in uitslag beduidend?

172 Een groothandelaar koopt een grote partij neonlampen. Hij doet voor de aankoop een eerste steekproef met 30 lampen en vindt een gemiddelde levensduur van $\overline{y} = 1000$ uren en een standaardafwijking van $s_1 = 100$ uren. Bij levering doet hij een steekproef met 40 lampen en vindt een gemiddelde levensduur van $\overline{y} = 950$ uren en een standaardafwijking van $s_2 = 100$ uren. Is het verschil beduidend of niet?

Hoofdstuk 5

Inleiding tot lineaire regressie

5.1 Het principe van de maximum kans

We beschouwen het volgende probleempje. Thomas en Stephanie zijn twee beroemde scheikundigen die proberen erin te slagen om melk artificieel te vervaardigen. Stel dat de ideale natuurlijke melk bevat

88% water,
4,3% vetstoffen en mineralen,
2,9% eiwitten,
4,7% lactose (melksuiker) en
0,1% vitamines en enzymen.

Om één of andere reden blijkt het toch niet zo gemakkelijk te zijn deze percentages perfect te doen kloppen, en na heel wat proeven (in twee betekenissen, welke?) bekomt Thomas de volgende samenstelling:

88,11% water,
4,34% vetstoffen en mineralen,
2,91% eiwitten,
4,59% lactose en
0,05% vitamines en enzymen.

Stephanie van haar kant bekomt met haar methode:

87,89% water,
4,24% vetstoffen en mineralen,

2,85% eiwitten,
4,85% lactose en
0,17% vitamines en enzymen.

De vraag is: wie van onze twee bollebozen, Thomas of Stephanie benadert het best de ideale natuurlijke melk?

Om dit te kunnen antwoorden moeten we een principe vooropstellen, want anders zijn vele antwoorden mogelijk. Bijvoorbeeld het principe dat de som van de relatieve fouten in percentages minimaal moet zijn. Met dit principe is het haast zeker dat Thomas wint, want Stephanie heeft een fout van 70% op de vitamines en enzymen, terwijl Thomas er slechts 50% naast zit. De andere foutprocenten kunnen die 20% niet meer goedmaken. Wij zullen dit principe hier echter niet gebruiken omdat zeldzame stoffen teveel invloed erop hebben.

Het principe dat we zullen gebruiken is het principe van de zogenaamde **maximum kans**. We gebaren dat de twee melkprobeersels stalen zijn uit eenzelfde universum van de ideale melk. Vervolgens beschouwen we de kans dat een steekproef er exact zó uitziet als Thomas respectievelijk Stephanie bekomen. Diegene met de grootste kans zullen we de beste noemen.

Vooralleer die kans te berekenen geven we ander voorbeeld. Onderstel dat Thomas (van de pilletjes) en Kristof (van de whisky), twee ongure elementen uit de maffia wereld, als nieuwe broodwinning valse dobbelstenen maken. De dobbelstenen die Kristof vervaardigt werpt 32 keer op honderd een zes, en deze van Kristof 25 keer. Onderstellen we dat beide dobbelstenen uit hetzelfde — eerlijke — universum komen, dan zien we dat de kans om op 100 worpen 32 respectievelijk 25 keer zes te gooien $C_{100}^{32} \cdot \left(\frac{1}{6}\right)^{32} \cdot \left(\frac{5}{6}\right)^{68} = 0,000074$ respectievelijk $C_{100}^{25} \cdot \left(\frac{1}{6}\right)^{25} \cdot \left(\frac{5}{6}\right)^{75} = 0,009826$ zodat we overduidelijk zien dat Kristof haar teerling de normaalste is. Hier konden we dat natuurlijk al op voorhand voorspellen, maar met ons principe uitgerekend klopt het helemaal. De meest valse dobbelstenen worden dus gemaakt door KristofĖn Kristof ... zij sloot haar kasboek ...

Keren we nu nog even terug naar ons eerste probleem. Hoe gaan we de kans bepalen dat de procenten van Thomas respectievlijk Stephanie voorkomen? Wel uit het hoofdstuk van de normale verdeling weten we dat de fouten een normale verdeling vormen. Dus als X_i de ideale waarde is van de i-de veranderlijke en x_i is de gemeten waarde, dan is de fout $X_i - x_i$ en deze volgt een normale verdeling. Anders gezegd: de kans op een fout $X_i - x_i$ is evenredig met $e^{-\frac{1}{2}(X_i - x_i)^2}$, waar we nu even $\sigma = 1$ onderstellen voor de eenvoud. De kans op het geheel van fouten $X_i - x_i$, $i \in \{1, 2, \ldots, n\}$, is dan evenredig met $e^{-\frac{1}{2}(X_1 - x_1)^2} \cdot e^{-\frac{1}{2}(X_2 - x_2)^2} \cdot \ldots \cdot e^{-\frac{1}{2}(X_n - x_n)^2} =$

$$e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i - x_i)^2}.$$

Deze kans is maximaal als de som

$$\sum_{i=1}^{n}(X_i - x_i)^2$$

minimaal is. We zeggen dat de **kwadratische afwijking** minimaal moet zijn. Thomas haar kwadratische afwijking is nu

$$(88 - 88,11)^2 + (4,3 - 4,34)^2 + (2,9 - 2,91)^2 + (4,7 - 4,59)^2 + (0,1 - 0,05)^2 = 0,0284$$

terwijl Stephanie haar kwadratische afwijking gelijk is aan

$$(88 - 88,11)^2 + (4,3 - 4,34)^2 + (2,9 - 2,91)^2 + (4,7 - 4,59)^2 + (0,1 - 0,05)^2 = 0,0256.$$

Dus Stephanie wint. Maar Thomas is toch tweede.

5.2 Lineaire regressie

Een belangrijke toepassing op de methode van de maximum kans is de *lineaire regressie.* We leggen dit uit aan de hand van een voorbeeld. Het doel van de lineaire regressie is een lineair verband op te sporen tussen twee gekoppelde veranderlijken. Dus nu eens *geen* onafhankelijke. We zullen dat bijvoorbeeld eens doen voor de lengte en het gewicht van Kate in haar eerste levensmaanden. Om de vier weken doen we een meting en ziehier de resultaten:

Leeftijd in weken	Lengte in mm	gewicht in gram
0	493	2871
4	532	3774
8	570	4739
12	589	5585
16	621	6459
20	642	7007
24	662	7586
28	680	8178
32	691	8543
36	702	8989
40	712	9259
44	721	9511
48	731	9810
52	740	9911

Zet even die waarden uit in een orthogonaal assenstelsel. Je ziet dat de punten tamelijk lineair liggen. Ons doel is nu de vergelijking zoeken van de rechte die het best past bij deze puntenrij. De methode om deze vergelijking op te stellen steunt weerom op het principe van de maximum kans (men drukt de kans uit dat voor een rechte $y = mx + q$ men de gegeven waarden krijgt en stelt deze minimaal). Meestal is in dergelijk probleem één van de twee veranderlijken *correct* gemeten, bijvoorbeeld de tijd. Deze kiezen we dan als x-as. In ons voorbeeld is dat niet het geval en is de keuze arbitrair. Bijvoorbeeld de lengtes zijn x-coördinaten en de gewichten y-coördinaten. We berekenen de gemiddelden van de x'en, van de y's en hun respectievelijke standaardafwijking.

$$\mu_x = 649 \qquad \sigma_x = 74,9$$

$$\mu_y = 7301,6 \qquad \sigma_y = 2226.9.$$

We voeren nu een nieuwe maat in, namelijk de gemiddelden van de producten van de afwijkingen van de x'en t.o.v. hun gemiddelde en de overeenkomstige afwijkingen van de y's t.o.v. hun gemiddelde. We noteren dit door

$$\sigma_{x,y} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}.$$

Men rekent gemakkelijk uit dat dit gelijk is aan

$$\sigma_{x,y} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \mu_x \mu_y},$$

wat in ons voorbeeld gelijk is aan

$$\sigma_{x,y} = \sqrt{\frac{68675051}{14} - 649 \cdot 7301,6} = 408,194.$$

Men bewijst nu dat de best passende rechte door het zwaartepunt (μ_x, μ_y) gaat en als richtingscoëfficiënt de uitdrukking

$$\left(\frac{\sigma_{x,y}}{\sigma_x}\right)^2$$

heeft.

De vergelijking wordt dus:

$$y - \mu_y = \left(\frac{\sigma_{x,y}}{\sigma_x}\right)^2 (x - \mu_x),$$

wat hier wordt:

$$y = 29,7x - 11974,28.$$

De coëfficiënt

$$r = \frac{\sigma_{x,y}^2}{\sigma_x \cdot \sigma_y}$$

geeft een getal tussen de -1 en $+1$. Hoe groter de absolute waarde van r, des te beter het lineair verband, d.w.z. des te nauwkeuriger liggen de punten op de gevonden rechte. Hoe dichter tegen nul, des te slechter het lineair verband. Als er geen lineair verband is, kan er eventueel een ander verband zijn. Door een transformatie van één of beide veranderlijken kan men dikwijls terug tot een lineair verband komen. Ziet de grafiek er tamelijk exponentieel uit, dan kunnen we de logaritmen van y uitzetten tegen x en misschien geeft dit een lineair verband.

In ons voorbeeld is $r = 0,998965984$, wat wijst op een zeer sterk lineair verband.

Aangenomen dat het verband tussen lengte en gewicht hetzelfde blijft als men ouder wordt, kan iedere leerling nu eens nagaan hoeveel hij/zij zou moeten wegen (y) voor zijn/haar gegeven eigen lengte (x). Vergeet niet dat de lengte in mm is en het gewicht in gram. Schrijver dezes is 1788 mm groot en zou dus moeten wegen (indien hij een meisje was, want bovenstaande formule is gebaseerd op de gemiddelde gewichten en lengten van pasgeborene meisjes) het engelachtige gewicht van 41, 12932 kilogram. Je kan dus nagaan dat op een bepaald moment in onze jonge jaren het gewicht rapper zal toenemen dan onze lengte. Dit is inderdaad zo. Terwijl de lengte in functie van de leeftijd steeds een bolle kant naar boven heeft, vertoont de gewichtscurve een buigpunt rond de leeftijd van 4 jaar, waarna deze functie een tijdje haar bolle kant naar beneden legt. De formule die we opgesteld hebben zal wèl zeer goed kloppen voor kinderen jonger dan twee jaar. Test ze maar uit op familie!

We zouden ook de x en y kunnen verwisselen en bekomen dan een andere rechte, die echter de eerste rechte zeer goed benadert. Het gemiddelde van de twee rechten is in feite nog de beste benadering van de gezochte lineaire functie.

Je kan nu zelf een aantal verbanden testen op waarnemingen die je in de klas doet. Bijvoorbeeld is er een lineair verband tussen de punten voor wiskunde en deze voor latijn? Laat de extreem goede en extreem slechte leerlingen buiten beschouwing! Of is er een lineair verband tussen voetlengte en handgrootte? Of tussen de frequentie van de polsslag en de tijd die je neerzet om honderd meter te lopen (een tip om de turnles interessanter te maken voor de wiskundefreaks en de wiskundeles interessanter te maken voor de sportfanaten), enz.

SUPPLEMENTAIRE OPGAVEN — 173 Speel een toneelstukje waarin ieder de rol vertolkt die in de voorgaande pagina's op haar lijf geschreven is.

Bibliography

Prof. dr. N. Veravermate. Medische statistiek en basisgebruik Computer.

Inhoudsopgave