

## 1 Basisbegrippen van de Statistiek

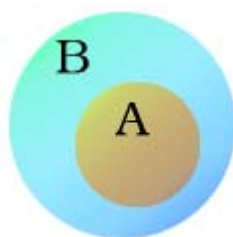
### 1.1 Introductie

Statistiek is een manier om informatie te verkrijgen van data. Niets meer, niets minder. Veel studenten vragen zich af waarom statistiek boeken zo schrikbarend dik zijn als statistiek zich louter bezighoudt met het verkrijgen van informatie over data. Dit komt doordat studenten vandaag de dag geconfronteerd worden met verschillende soorten data die onderzocht kunnen worden met verscheidene methoden.

Om de beginselen van de statistiek te begrijpen is kennis van een aantal basisbegrippen en symbolen een noodzaak. In de tabel hieronder is een aantal symbolen met hun betekenis te vinden.

$A \cup B$	De vereniging van gebeurtenissen $A$ en $B$ die plaatsvindt als $A$ of $B$ of beiden plaatsvinden
$A \cap B$	De doorsnede van de gebeurtenissen $A$ en $B$ die plaatsvindt als $A$ en $B$ beiden plaatsvinden
$A^c$	Het complement van gebeurtenis $A$ vindt plaats als $A$ niet plaatsvindt
$A - B$	Het verschil tussen $A$ en $B$
$A \subset B$	$A$ is een deelverzameling van $B$

Een Venndiagram is een diagram dat gebruikt wordt om relaties tussen verzamelingen aan te geven. Het Venndiagram hieronder geeft een versimpelde illustratie van  $A$  als deelverzameling van  $B$ .



Figuur 1<sup>a</sup>.  $A \subset B$  (bron: [www.wikipedia.nl](http://www.wikipedia.nl))

### 1.2 Kansrekening

Als er een willekeurig experiment (ook wel random experiment = een actie of proces dat leidt tot een of meerdere mogelijke uitkomsten) wordt uitgevoerd zijn een paar begrippen van belang.

Alle mogelijke uitkomsten van zo'n experiment bij elkaar noemt men de uitkomstenruimte ( $\Omega$ ). Het complement van  $\Omega$  noteert men als  $\emptyset (= \Omega^c)$ . Elk element van  $\Omega$  heeft een uitkomst  $\omega$ . De notatie  $\omega \in \Omega$  staat voor "  $\omega$  is element van  $\Omega$  " en deelverzamelingen van  $\Omega$  heten gebeurtenissen.

Een paar  $(P, \Omega)$  dat bestaat uit verzameling  $\Omega$  en kansfunctie  $P$  noemt men een kansruimte.

Deze kansruimte voegt aan elke deelverzameling  $A \subset \Omega$  een reëel getal op het interval  $[0;1]$  toe en wel op zo'n manier dat er sprake is van de twee volgende axioma's (axioma = een niet bewezen, maar als grondslag aanvaarde stelling):

1.  $P(\Omega) = 1$

2.  $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$  als de deelverzamelingen  $A_1, A_2, \dots$  van  $\Omega$  paarsgewijs disjunct zijn

De kans op gebeurtenis  $A$  noteert men als  $P(A)$ . De term disjunct vereist wat toelichting. Twee gebeurtenissen  $A$  en  $B$  zijn disjunct als geldt  $A \cap B = \emptyset$ . Met andere woorden,  $A$  en  $B$  sluiten elkaar uit.

Uit de 2 bovenstaande axioma's kan het volgende geconcludeerd worden:

Als $A \subset B$ dan $P(B - A) = P(B) - P(A) = P(A^c \cap B)$
$P(A^c) = 1 - P(A)$ ; in geval van $A = \Omega : P(\emptyset) = 0$
Als $A \subset B$ dan $P(A) \leq P(B)$
$P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$
$P(\bigcap_{n=1}^k A_n) \geq 1 - \sum_{n=1}^k P(A_n)$ ; dit is ook wel de ongelijkheid van Bonferroni
Als $A_1, A_2, \dots, A_n$ paarsgewijs disjunct zijn dan geldt $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$

Er is een uitzondering op al het bovenstaande: Heeft  $\Omega$  een eindig aantal elementen ( $N$ ) met allen gelijke kansen dan geldt  $P(\omega) = 1/N$  voor iedere  $\omega \in \Omega$ . Tevens geldt dat  $P(A) = \frac{|A|}{N}$ .  $|A|$  is het aantal elementen van  $A$ .

Voorbeeld 1<sup>a</sup>

Beschouw de verzameling  $\Omega = \{1,2,3,4\}$ . Definieer voor iedere deelverzameling  $A \subset \Omega$ .

$$P(A) = \frac{|A|}{4}$$

Dit voorbeeld kan model staan voor een worp met (let op!) een 4-zijdige dobbelsteen.



Figuur 1<sup>b</sup>: 4-zijdige dobbelsteen (bron: www.wikipedia.nl)

#### Voorbeeld 1<sup>b</sup>

Beschouw de verzameling  $\Omega = \{1,2,3,4\} \times \{1,2,3,4\}$ . Definieer voor iedere deelverzameling  $A \subset \Omega$ .

$$P(A) = \frac{|A|}{16}$$

Dit voorbeeld kan model staan voor twee worpen met een 4-zijdige dobbelsteen. De gebeurtenis "twee gelijke ogen" zou omschreven worden als  $\Omega = \{(1,1), (2,2), (3,3), (4,4)\}$ .

### 1.3 Voorwaardelijke kans

Vaak wil men weten hoe twee gebeurtenissen gerelateerd zijn. Sterker nog, men wil de kans weten van een gebeurtenis gegeven dat een andere, gerelateerde gebeurtenis plaatsvindt. Deze kans noemt men de voorwaardelijke kans wordt als volgt genoteerd:

De kans op  $A$ , gegeven  $B$  is

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

De kans op  $B$ , gegeven  $A$  is

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

Voortbordurend op deze definities kan men de productregel toepassen:

$$P(A \cap B) = P(B) \times P(A | B) = P(A) \times P(B | A)$$

#### Voorbeeld 1<sup>c</sup>

Toptennisser Nedal speelt in de finale van het tennistoernooi Ronald Barros tegen de winnaar van de halve finale tussen Foderer en Kraaijenbek. Aangezien Kraaijenbek zijn beste jaren tennis al gehad heeft, schat men de kans dat Foderer de finale haalt op 86%. De kans dat Nedal Foderer in de finale verslaat is 67%, daar Nedal heer en meester is op gravel. De kans dat Nedal Kraaijenbek in de finale verslaat is nog groter, te weten 93%. De kans dat Nedal het toernooi wint, wordt dan als volgt berekend:

$$P(\text{Nedal wint Ronald Barros}) = P(\text{Foderer wint halve finale}) \times P(\text{Nedal wint Ronald Barros} | \text{Foderer wint halve finale}) + P(\text{Kraaijenbek wint halve finale}) \times P(\text{Nedal wint Ronald Barros} | \text{Kraaijenbek wint halve finale}) = 0,86 \times 0,67 + 0,14 \times 0,93 = 70,6 \%$$

#### 1.4 Onafhankelijke gebeurtenissen

Een van de doelen van het berekenen van de voorwaardelijke kans is het vaststellen of twee gebeurtenissen gerelateerd aan elkaar zijn of niet. Anders gezegd, men wil er achter komen of gebeurtenissen onafhankelijk zijn van elkaar.

Twee gebeurtenissen  $A$  en  $B$  zijn onafhankelijk van elkaar als

$$P(A | B) = P(A)$$

of

$$P(B | A) = P(B)$$

Met andere woorden, twee gebeurtenissen zijn onafhankelijk van elkaar als de kans van een gebeurtenis niet beïnvloed wordt door het plaatsvinden van de andere gebeurtenis.

*Voorbeeld 1<sup>d</sup>*

Daan gooit 2 euromunten op. Beschouw de gebeurtenissen

*A*: de eerste euromunt is kop

*B*: de tweede euromunt is munt

De kans hierop is

$$P(A \cap B) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(A) \times P(B)$$

Gegeven de bovenstaande kans kan men concluderen dat *A* en *B* onafhankelijk zijn. Immers zal de uitkomst van de worp van de eerste euromunt geen invloed hebben op de uitkomst van de worp met de tweede euromunt (en andersom is uiteraard ook het geval).

### **1.5 De somregel**

Met de somregel kan men de kans van de vereniging van twee gebeurtenissen berekenen. De kans dat gebeurtenis *A*, of gebeurtenis *B*, of beiden plaatsvinden is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

De meeste studenten vragen zich nu per direct af waarom  $P(A \cap B)$  van  $P(A)$  en  $P(B)$  wordt afgetrokken. Dit komt doordat bij de optelling van  $P(A)$  en  $P(B)$  de kans  $P(A \cap B)$  dubbel wordt geteld en deze dient er weer vanaf getrokken te worden.

*Voorbeeld 1<sup>e</sup>*

Arnold is bezig met zijn derde jaar Bachelor van Bedrijfskunde en heeft nog twee vakken openstaan die hij moet halen om in september aan zijn Master te beginnen. Aangezien Arnold erg druk is met zijn studie en daarnaast ook nog werkt, kan zijn werk invloed hebben op zijn studieresultaten. Zie

onderstaande tabel met de kansen dat hij wel of niet aan zijn Master begint gecombineerd met 1 of 2 dagen in de week werken.

	2 dagen werken	1 dag werken
Master in september	0,11	0,29
Geen Master in september	0,06	0,54

Beschouw de gebeurtenissen

$A_1$ : Master in september

$A_2$ : Geen Master in September

$B_1$ : 2 dagen werken

$B_2$ : 1 dag werken

Nu kan men het volgende berekenen:

$$P(A_1) = P(A_1 \cap B_1) + P(A_1 \cap B_2) = 0,11 + 0,29 = 0,40$$

$$P(B_1) = P(A_1 \cap B_1) + P(A_2 \cap B_1) = 0,11 + 0,06 = 0,17$$

Als men nu probeert de kans van de vereniging van  $A_1$  en  $B_1$  te berekenen door middel van optelling van de kans op  $A_1$  en  $B_1$ , vindt men

$$P(A_1) + P(B_1) = 0,11 + 0,29 + 0,11 + 0,06$$

Let op dat hier twee keer  $P(A_1 \cap B_1)$  optellen (0,11). Om deze dubbele optelling te corrigeren haalt men een keer  $P(A_1 \cap B_1)$  van de som  $P(A_1) + P(B_1)$  af, oftewel:

$$P(A_1 \cup B_1) = P(A_1) + (P(B_1) - P(A_1 \cap B_1))$$

$$P(A_1 \cup B_1) = [0,11 + 0,29] + [0,11 + 0,06] - 0,11$$

$$P(A_1 \cup B_1) = 0,40 + 0,17 - 0,11 = 0,46$$

In geval van 3 willekeurige gebeurtenissen wordt automatisch de somregel ook langer:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Hierop voortbordurend kan men voor  $n$  gebeurtenissen de volgende somregel afleiden:

$$P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(\bigcap_{i=1}^n A_i)$$

Let op! Voor de laatste term gebruikt men een plusteken voor een oneven  $n$  en een minteken voor een even  $n$ .

### 1.6 Theorema van Bayes

Voorwaardelijke kansen worden vaak gebruikt om relaties tussen twee gebeurtenissen te peilen. In voorbeeld 1<sup>c</sup> werd al aangegeven dat een voorwaardelijke kans de kans meet dat een gebeurtenis plaatsvindt, gegeven dat een mogelijke oorzaak van deze gebeurtenis al heeft plaatsgevonden. Als de kans op een gebeurtenis bekend is, maar de kans op een van de mogelijke oorzaken berekend dient te worden, kunnen problemen ontstaan. Het theorema van Bayes biedt hiervoor de oplossing. Hieronder zal alleen de formule van het theorema van Bayes gegeven worden, daar er in de studieboeken van de Bedrijfsstatistiek hier maar kort op ingegaan wordt.

Theorema van Bayes:

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

Voor meer achtergrondinformatie over het theorema van Bayes dient de lezer er geavanceerde statistiekboeken op na te slaan.

### 1.7 Combinatoriek of combinatieleer

In de combinatoriek of combinatieleer bestudeert men eindige [verzamelingen](#) van objecten die aan gespecificeerde eigenschappen voldoen. In het bijzonder houdt men zich bezig met het "tellen"

van objecten in deze verzamelingen en het bepalen of er zekere "optimale" objecten in een verzameling aanwezig zijn<sup>1</sup>.

Binnen de combinatoriek kent men 4 "hoofdcategorieën". Hieronder worden deze categorieën zo overzichtelijk mogelijk uiteengezet met de corresponderende formules.

### 1.7.1 Herhalingsvariatie

Men kiest  $k$  elementen uit  $n$ ; hier is de volgorde van belang en er sprake van trekken met terugleggen. De formule die men hiervoor hanteert is

$$n^k$$

#### Voorbeeld 1<sup>f</sup>

Oscar gooit tien keer met een euromunt. Elke worp heeft twee mogelijkheden: óf kop, óf munt. Er zijn dan  $2^{10} = 1024$  uitkomsten voor deze tien worpen met de euromunt.

### 1.7.2 Variatie (permutatie)

Men trekt  $k$  elementen uit een verzameling van  $n$ ; hier is de volgorde van belang en er is sprake van trekken zonder terugleggen. De formule die men hiervoor hanteert is

$$\frac{n!}{(n-k)!} = P(n, k)$$

#### Voorbeeld 1<sup>g</sup>

Maurits, Karel en Paul besluiten op een zonnige middag naar de bioscoop te gaan. Na hun kaartjes gekocht te hebben lopen ze de desbetreffende zaal in en komen er achter dat de gehele zaal, bestaande uit 350 zitplaatsen, leeg is. Bereken op hoeveel verschillende manieren ze in de zaal kunnen gaan zitten.

$$\frac{350!}{(350-3)!} = 350 \times 349 \times 348$$

<sup>1</sup> <http://nl.wikipedia.org/wiki/Combinatoriek>

### 1.7.3 Combinatie

Men trekt  $k$  elementen uit een verzameling  $n$ ; hier is de volgorde niet van belang en er is sprake van trekken zonder terugleggen. De formule die men hiervoor hanteert is

$$\frac{n!}{k!(n-k)!}$$

#### Voorbeeld 1<sup>h</sup>

Pieter heeft lang gespaard en kan 3 nieuwe videospellen kopen voor zijn Woo-spelcomputer. In de winkel valt zijn oog op 5 spellen die allemaal even duur zijn. Hij kan er echter maar 3 kiezen. Bereken hoeveel keuzemogelijkheden Pieter heeft.

$$\frac{5!}{3!(5-3)!} = 10$$

### 1.7.4 Herhalingscombinatie

Men trekt  $k$  elementen uit een verzameling  $n$ ; hier is de volgorde niet van belang en er is sprake van trekken met terugleggen. De formule die men hiervoor hanteert is

$$\frac{(n+k-1)!}{k!(n-1)!}$$

#### Voorbeeld 1<sup>i</sup>

Eelco besluit voor het Paasontbijt zijn ouders te verrassen met geschilderde eieren. Hij beschikt over twee dozen van zes eieren en haalt drie potten verf uit de kelder, elk met een andere kleur verf. Bereken op hoeveel verschillende manieren hij de eieren kan beschilderen.

$$\frac{(12+3-1)!}{3!(12-1)!} = \frac{14!}{3! \times 9!}$$

Samengevat:

	Volgorde van belang	Volgorde niet van belang
Trekken zonder terugleggen	$\frac{n!}{(n-k)!} = P(n, k)$	$\frac{n!}{k!(n-k)!}$
Trekken met terugleggen	$n^k$	$\frac{(n+k-1)!}{k!(n-1)!}$