

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

Rapport S 386 (C 17)

Syllabus van een cursus Elementaire Statistiek

door

W. Molenaar

en

J. Oosterhoff



augustus 1967

Voorwoord

In een door de Universiteit van Amsterdam georganiseerd interuniversitair seminarium voor sedimentologen werd de behoefte gevoeld aan een inleidende bespreking van de mathematische statistiek. Deze syllabus is geschreven om die bespreking te begeleiden.

Er is niet gestreefd naar volledigheid of naar een streng-mathematische opbouw van de theorie. In een vrij informele inleiding worden de begrippen populatie, steekproef, kans, schatting en toets centraal gesteld. Een aantal vrij klassieke onderwerpen zoals binomiale toets, lineaire regressie en variantie analyse zijn kort geschetst, omdat juist deze technieken nogal eens in geologische onderzoeken gebruikt worden.

In een zo summiere cursus kunnen wij de lezers niet in staat stellen zelfstandig statistische methoden toe te passen. Maar gehoopt wordt dat de cursus tot steun zal zijn bij het lezen van statistiek bevattende publicaties en bij de samenwerking met deskundige statistici ter uitwerking van eigen onderzoeken.

De eerste drie hoofdstukken van deze cursus zijn een bewerking van het overeenkomstige gedeelte van de cursus "Enkele elementaire statistische methoden", rapport S 258 (C 12), door Dr. J. FABIOUS.

Inhoud	blz.
1. Populatie en steekproef	1
2. Enige experimenten	6
3. Kansrekening	10
4. Schattingstheorie	17
5. Toetsingstheorie	22
6. De normale verdeling	29
7. Lineaire regressie	45
8. Variantie analyse	51

Hoofdstuk 1. Populatie en steekproef.

Het prototype, of, zoals men meestal zegt, het model, van een grote groep statistische problemen is het volgende:

Men heeft een vaas, waarin zich een groot aantal rode en witte knikkers bevinden. Men wil een schatting hebben van het percentage rode knikkers in de vaas.

Een iets andere vorm van dit zelfde probleem hebben we, als men om een of andere reden vermoedt, dat de fractie rode knikkers een bepaalde waarde p bezit, en wenst na te gaan of dit vermoeden juist is.

Voorbeelden:

- a. Een partij goederen bevat een onbekende fractie ondeugdelijke exemplaren. Hoe groot is deze fractie?
- b. Een deel der rijksambtenaren bereikt de pensioengerechtigde leeftijd, de anderen overlijden voordien. Hoe groot is de laatste fractie?
- c. Als men ratten een bepaalde dosis van een zekere stof inspuit, ondervinden sommige ratten schadelijke gevolgen, andere niet. Hoe groot is het percentage dat wel schadelijke gevolgen ondervindt van de injectie?
- d. Op grond van de erfelijkheidswetten van Mendel mag men verwachten dat een fractie p van alle planten die door kruising van 2 planten verkregen kunnen worden, een bepaalde eigenschap bezit. Hoe kan men de juistheid van dit vermoeden controleren?

In al deze voorbeelden gaat het erom een uitspraak te doen over een groep van objecten. Deze groep van objecten noemt men de populatie die beschouwd wordt. Zo is de populatie in ons model de vaas met knikkers, in voorbeeld a de partij goederen, in voorbeeld b de verzameling van alle rijksambtenaren, in c de verzameling van alle ratten van een bepaalde soort, in d de verzameling van alle planten welke door kruising uit de twee planten verkregen kunnen worden. Het aantal objecten waaruit de populatie bestaat noemt men de omvang van de populatie.

Het is bij het gebruik van statistische methoden van groot belang, dat men de populatie nauwkeurig omschrijft. In voorbeeld b dient men

vast te stellen wie als rijksambtenaar zal worden beschouwd en wie niet. Over de specifieke moeilijkheden bij het definiëren van populaties van geologische objecten raadplege men pag. 75 e.v. van KRUMBEIN en GRAYBILL, An introduction to statistical models in geology, McGraw Hill, New York 1965.

In het geval van de vaas met knikkers, en ook in voorbeeld a, zou men alle rode en witte knikkers uit de vaas, resp. alle deugdelijke en ondeugdelijke exemplaren uit de partij, kunnen tellen. We spreken dan van een volledige waarneming. Maar bij zeer grote populatie-omvang kan dit bezwaarlijk zijn, en in voorbeeld a kan het gebeuren, dat het middel erger is dan de kwaal, nl. als we een destructieve keuringsmethode moeten gebruiken om uit te maken of een exemplaar al dan niet deugdelijk is. In de andere voorbeelden is een volledige waarneming zelfs in het algemeen onmogelijk!

Het essentiële van de wiskundige statistiek is nu, dat deze ons methoden geeft, die ons in staat stellen uit onvolledige waarnemingen, dat wil zeggen waarnemingen die slechts een (vaak onbekend) deel van de populatie omvatten, conclusies te trekken over de hele populatie. Wij zullen straks de eisen E1 en E2 formuleren waaraan een onvolledige waarnemingsmethode moet voldoen om statistische analyse mogelijk te maken.

Om een onvolledige waarneming te kunnen doen moeten we eerst een aantal objecten uit de populatie afzonderen waaraan we de waarnemingen gaan verrichten. In het geval van de vaas met knikkers moeten we een aantal knikkers eruit pakken, in voorbeeld b moeten we in gedachten een aantal rijksambtenaren afzonderen etc. etc.

We spreken in dit verband van het nemen van een steekproef uit de populatie. Een steekproef kan bestaan uit een willekeurig aantal n objecten. We noemen dit aantal de omvang van de steekproef.

Een steekproef van de omvang n kan men op 2 verschillende manieren verkrijgen:

1. Men trekt telkens één object uit de populatie, doet aan dit ene object de waarneming, en voegt het weer bij de populatie. Eénzelfde object kan op deze wijze méér dan eens in de steekproef voorkomen. We spreken van een steekproef met teruglegging.

2. Men trekt n objecten (al of niet tegelijkertijd) uit de populatie, zonder de getrokken objecten weer aan de populatie toe te voegen.

Nu spreekt men van een steekproef zonder teruglegging.

Voor de theorie maakt het verschil of men steekproeven met, dan wel zonder teruglegging neemt, tenminste als de omvang der populatie eindig is. Als echter de omvang van de populatie zeer groot is ten opzichte van de omvang van de steekproef wordt het verschil verwaarloosbaar. Aangezien de theorie het gemakkelijkst is voor steekproeven met teruglegging, zullen wij ons in het volgende meestal beperken tot steekproeven met teruglegging of - indien geen teruglegging plaatsvindt - tot populaties wier omvang groot is in vergelijking met de daaruit getrokken steekproef. Een dergelijke steekproef ontstaat per definitie door achtereenvolgens n trekkingen uit dezelfde populatie te nemen. Immers na elke trekking en de bijbehorende waarneming aan het getrokken object wordt dit object weer bij de populatie gevoegd, zodat de populatie bij de volgende trekking onveranderd is.

We zullen nu 2 eisen geven waaraan deze trekkingen moeten voldoen:

- E1. Elke trekking moet aselect zijn, d.w.z. alle objecten uit de populatie moeten gelijkwaardig zijn bij de trekking; we mogen niet een bepaald object selecteren; we moeten het te trekken object willekeurig kiezen.
- E2. De trekkingen moeten onderling onafhankelijk zijn, d.w.z. de keuze van een object mag niet afhangen van het resultaat van de vorige trekkingen.

Bij het nemen van een steekproef dient men er steeds voor te waken dat aan deze eisen voldaan is. Bij het trekken van knikkers uit een vaas kan men dit proberen te verwezenlijken door voor elke trekking zo goed mogelijk te schudden en dit te herhalen nadat de getrokken knikker is teruggelegd. Indien echter de rode knikkers bijv. iets groter zijn dan de witte zullen de rode knikkers bij het schudden de neiging hebben zich bovenin de vaas op te hopen. Het gevolg hiervan is dan dat de knikkers niet meer gelijkwaardig zijn bij de trekking; men zal te vaak een rode knikker trekken.

Het volgende hulpmiddel zal in vele gevallen uitkomst brengen. We

nummeren de objecten van de populatie in één of andere volgorde. Een menselijke populatie bijv. alfabetisch, bij goederen uit serie-fabricage naar volgorde van fabricage etc. etc. Nu bepalen we door loting welke nummers we in de steekproef zullen opnemen. Daartoe kunnen we een vaas met tien gelijke knikkers, genummerd 0, 1, ..., 9 nemen en hieruit met teruglegging trekkingen doen.

Wenst men bijv. een steekproef met omvang 50 uit een populatie met omvang 1000, dan nummeren we de objecten van de populatie van 000 t/m 999. Uit onze vaas met 10 knikkers doen we nu 50 maal telkens 3 trekkingen met teruglegging. Zo krijgen we 50 groepjes van 3 cijfers, dus 50 nummers onder de 1000. Onze steekproef stellen we nu samen door die 50 objecten uit de populatie te nemen, wier nummers juist overeenkomen met de 50 groepjes van 3 cijfers die we getrokken hebben. Zo wordt dan een steekproef met teruglegging verkregen, want onder de 50 nummers kunnen gelijke voorkomen. Wenst men een steekproef zonder teruglegging, dan schrappen men nummers, die reeds eerder voorgekomen zijn en zette het lotingsprocédé voort tot 50 verschillende nummers onder 1000 verkregen zijn.

Deze methode heeft nog wel enige bezwaren. Het kost, vooral als het om grote steekproeven gaat, nogal veel tijd om alle nodige lotingen uit te voeren. Er zijn echter tabellen geconstrueerd die de resultaten van grote aantallen trekkingen uit een vaas met tien genummerde knikkers geven. (Men heeft deze getallen niet verkregen door werkelijk knikkers uit een vaas te trekken, doch met behulp van snellere methoden, die daarmee essentieel gelijkwaardig zijn.) Zo bevatten bijv. de "Tables of random sampling numbers" van M.G. KENDALL en B. BABINGTON SMITH (Cambridge University Press 1946) 100.000 trekkingen uit de populatie 0, 1, ..., 9.

Wenst men nu weer een steekproef van 50 elementen uit de populatie met omvang 1000, dan hoeft men slechts, te beginnen op een willekeurig punt in de tabel, de 50 volgende groepjes van 3 cijfers te nemen.

Soms is het niet mogelijk deze methode toe te passen, eenvoudig omdat het niet mogelijk is de te onderzoeken populatie te nummeren, omdat niet alle objecten uit de populatie bereikbaar zijn. Als men een steekproef wil nemen uit de lopende productie van een fabriek, kan men geen exemplaren nemen die reeds verzonden zijn of die nog geproduceerd

moeten worden.

In dergelijke gevallen zoekt men naar een steekproefmethode waarvan men verwacht dat toch zo goed mogelijk aan de eisen E1 en E2 voldaan is. Wij zullen hier niet verder op ingaan, doch wijzen er nogmaals op, dat men zich steeds goed rekenschap dient te geven van de noodzaak aan deze eisen te voldoen. Is aan de eisen slechts gedeeltelijk voldaan, dan zijn de in deze cursus te bespreken eenvoudige statistische technieken niet zonder voorbehoud van toepassing. In sommige gevallen bestaan er echter gecompliceerde methoden die dan nog wel toepasbaar zijn.

Steekproeftrekking uit geologische populaties wordt behandeld in hoofdstuk 7 van het reeds geciteerde boek van KRUMBEIN en GRAYBILL.

Hoofdstuk 2. Enige experimenten.

Keren wij thans terug tot het probleem van de vaas met knikkers, waarvan een onbekende fractie rood gekleurd is. Uit deze vaas nemen we nu een steekproef van n knikkers en we noteren het aantal rode knikkers dat wij daarbij aantreffen. Wij geven dit aantal aan met x . Herhalen we dit experiment, dan zullen we in het algemeen een andere waarde voor x vinden. Dit nu is een typisch statistisch verschijnsel: herhaling van eenzelfde experiment naar beste weten op dezelfde wijze uitgevoerd geeft afwijkende uitkomsten. Het aantal rode knikkers in een steekproef van n stuks varieert van steekproef tot steekproef. Een dergelijke variërende grootte noemen we een stochastische grootte (Engels: random variable). Wij zullen in het volgende stochastische grootheden steeds onderstrepen: $\underline{x}$ is het aantal rode knikkers in een steekproef van n knikkers uit onze vaas.

De waargenomen waarden van $\underline{x}$ in concrete gevallen geven we aan met $x_1, x_2, \dots, x_m$, zonder onderstreping. In een bepaald geval hebben wij 30 keer een steekproef van 25 uit een vaas met witte en rode knikkers getrokken ($n = 25, m = 30$). Wij vonden voor $\underline{x}$ de volgende waarden (in volgorde van waarneming):

(I) 1 3 0 1 0 1 2 4 2 3 1 4 1 1 0 1 2 3 3 2 2 2 4 3 3 3 2 3 4 1.

Het is duidelijk, dat de waargenomen waarden ieder op zichzelf reeds enige informatie verschaffen over de onbekende fractie p van rode knikkers. Als p dicht bij 1 lag, d.w.z. als bijna alle knikkers in de vaas rood waren, zouden we niet zulke lage uitkomsten hebben gekregen. Het is dus aannemelijk dat p klein is, d.w.z. dat er weinig rode knikkers in de vaas zitten.

Om een duidelijk overzicht te krijgen van de resultaten, zien wij van de volgorde daarvan af en vatten we ze samen in een histogram of frequentiediagram (zie fig. 2.1). De samenstelling daarvan is uit de figuur duidelijk: in horizontale richting zijn de waarden uitgezet die $\underline{x}$ heeft aangenomen, en in verticale richting het aantal malen dat $\underline{x}$ deze waarde heeft aangenomen. Een dergelijke figuur geeft in één oog-

aantal reeks I: n = 25
 m = 30

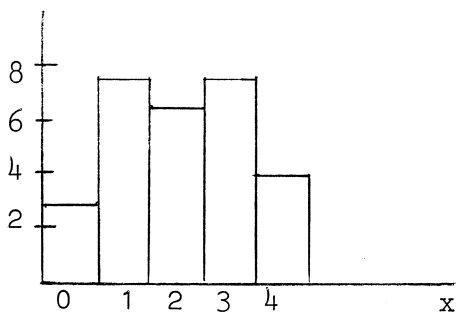


fig. 2.1

aantal reeks II: n = 25
 m = 30

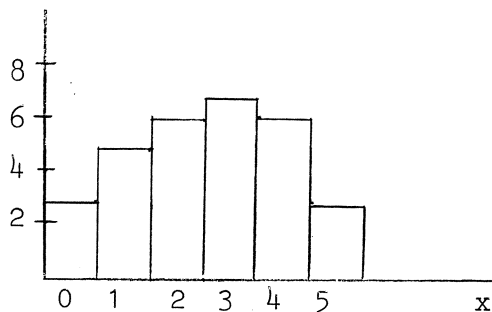


fig. 2.2

opslag een duidelijk overzicht van de waarnemingen. Alleen de volgorde der waarnemingen gaat hierbij verloren.

Herhaling van het experiment gaf de reeks waarnemingen:

(II) 2 2 2 0 3 5 3 5 2 3 4 2 5 3 1 1 4 1 4 0 4 4 1 1 3 3 3 0 4 2.

Ook deze waarnemingen hebben we in de vorm van een histogram samengevat (zie fig. 2.2).

We zien dat er ondanks de verschillen toch wel enige overeenkomst is tussen de waarnemingsreeksen (I) en (II): In beide gevallen varieert x slechts van 0 tot hoogstens 5, en de grote meerderheid der gevonden waarden van x is niet kleiner dan 1 en niet groter dan 4. Bij deze experimenten ligt de fractie rode knikkers in een steekproef van 25 stuks dus voornamelijk tussen $\frac{1}{25} = 0,04$ en $\frac{4}{25} = 0,16$.

Een typisch verschijnsel is dat deze statistische regelmaat duidelijker tot uiting komt indien men de waarnemingsreeksen verlengt, dus als men m groter kiest. Om dit te demonstreren zijn twee series van 100 steekproeven van 25 stuks uit dezelfde vaas genomen. De resultaten zijn samengevat in 2 histogrammen (fig. 2.3 en fig. 2.4).

We zien dat de histogrammen uit fig. 2.3 en fig. 2.4 minder verschillen dan die uit fig. 2.1 en fig. 2.2.

Aangezien het voor de hand ligt om de fractie rode knikkers in de steekproef te gebruiken als schatting voor de onbekende fractie rode knikkers in de vaas, kunnen we concluderen dat de nauwkeurigheid van deze schatting blijkbaar toeneemt naarmate we meer trekkingen doen. Dit is een zeer algemeen ervaringsfeit, dat we overal waar aselechte, onderling onafhankelijke trekkingen uit een populatie genomen worden, kunnen constateren. In de statistiek wordt bijna voortdurend gebruik van deze eigenschap van aselechte en onderling onafhankelijke trekkingen gemaakt. In het volgende zullen we zien op welke wijze dit geschiedt en hoe we het bovengenoemde nader kunnen preciseren.

Hoofdstuk 3. Kansrekening.

Zij π een eindige populatie met omvang N , terwijl $N(A)$ objecten van π een bepaald kenmerk A (bijv. een rode kleur) bezitten. Doen we nu een aselecte trekking uit π dan kan het zijn dat we een object met kenmerk A trekken, doch het is ook mogelijk dat dit niet het geval is. "Het vinden van kenmerk A bij een aselecte trekking" is blijkbaar een mogelijke, doch niet zekere, gebeurtenis. Een dergelijke mogelijke gebeurtenis noemen we een eventualiteit.

De kans of waarschijnlijkheid van deze eventualiteit definiëren we als $\frac{N(A)}{N}$, in formule:

$$(3.1) \quad P(A) \stackrel{\text{def}}{=} \frac{N(A)}{N}$$

Het is duidelijk dat voor elk kenmerk A de kans $P(A)$ tenminste 0 en ten hoogste 1 is. Een onmogelijke eventualiteit, d.w.z. het vinden van een kenmerk B dat geen enkel object uit π bezit, krijgt blijkbaar de waarschijnlijkheid 0 en evenzo heeft de zekere eventualiteit, d.w.z. het vinden van een kenmerk C , dat elk object uit π bezit, de waarschijnlijkheid 1.

Als de omvang van π oneindig is, dan is het quotiënt $\frac{N(A)}{N}$ onbepaald, zodat we het kansbegrip niet op dezelfde wijze als bij eindige populaties kunnen invoeren. Toch kan men in die gevallen nog wel spreken over een fractie p van de populatie van objecten die het kenmerk A bezitten. (Wij zullen hier niet ingaan op de streng-mathematische definitie van het begrip fractie in dit verband, daar dit ons ver buiten het kader van deze cursus zou voeren). We definiëren dan $P(A) \stackrel{\text{def}}{=} p$.

Voorbeelden:

- a. Zij π de vaas met witte en rode knikkers. Als π in totaal N knikkers bevat, waarvan er M rood zijn, dan is de kans dat een aselecte trekking een rode knikker oplevert dus $p \stackrel{\text{def}}{=} \frac{M}{N}$.
- b. Zij π de denkbeeldige verzameling van alle mogelijke worpen met een munt. Als we afzien van de mogelijkheid dat de munt op zijn kant blijft staan, dan is de kans op "kruis" gelijk aan de fractie worpen die de uitslag "kruis" geven; is deze gelijk $\frac{1}{2}$, dan wordt de munt

zuiver genoemd. Dan is dus: $P(\text{kruis}) = P(\text{munt}) = \frac{1}{2}$.

Het komt voor dat bij elk object uit de populatie π één of meer getallen behoren. Bij een aselechte trekking krijgen we één object uit π , en het (de) bijbehorende getal(len) kunnen we dan waarnemen. Aangezien de waargenomen getallen bij andere trekkingen andere waarden kunnen hebben, hebben we hier stochastische grootheden op de populatie π .

Voorbeelden:

- c. Zij π de denkbeeldige verzameling van alle mogelijke worpen met een dobbelsteen. Bij elk dezer worpen hoort een getal, nl. de uitkomst $\underline{x}$. Dit is een stochastische grootheid. Gooien we één keer met een dobbelsteen, dan kunnen we dat opvatten als een aselechte trekking uit π , en de uitkomst x is dan een waarneming van $\underline{x}$. Het is duidelijk dat $\underline{x}$ alleen de waarden 1,2,3,4,5 of 6 kan aannemen. Bij een zuivere dobbelsteen heeft elk van deze waarden een kans van $\frac{1}{6}$.
- d. Zij π de verzameling van alle recruten van de lichterij 1968. Bij elk dezer recruten behoort nu een lengte $\underline{l}$ en een gewicht $\underline{g}$. Dit zijn 2 stochastische grootheden. Trekken we aselekt een rekrut, dan zijn zijn lengte l en gewicht g waarnemingen van $\underline{l}$ en $\underline{g}$. De grootheden $\underline{l}$ en $\underline{g}$ kunnen in theorie elke positieve waarde aannemen. In feite ligt $\underline{l}$ meestal tussen 160 en 190 cm, en vrijwel zonder uitzondering tussen 140 en 210 cm. Waarden buiten deze grenzen zijn extreem onwaarschijnlijk. We kunnen spreken van $P(\underline{l} \leq 160)$, dit is het aantal recruten van 160 cm of minder gedeeld door het totaal aantal recruten. De berekening is ook uitvoerbaar als we 160 vervangen door een willekeurig ander getal l .

Algemeen is bij een stochastische grootheid $\underline{x}$ op een populatie π de mogelijkheid aanwezig om voor elk getal x de kans $P(\underline{x} \leq x)$ te bepalen. Deze kans, opgevat als functie van het getal x , noemt men de (cumulatieve) verdelingsfunctie van de stochastische grootheid $\underline{x}$.

De waarde van een stochastische grootheid $\underline{x}$ bij één experiment, d.w.z. één aselechte trekking, valt niet te voorspellen. Wel kunnen we een soort doorsnee-waarde aangeven. Hiertoe middelen we de bij alle populatie-elementen behorende waarden van $\underline{x}$. Zo ontstaat een getal dat we het populatiegemiddelde of de verwachting van $\underline{x}$ noemen en met $E\underline{x}$ aan-

duiden. In voorbeeld d is $E_{\underline{x}}$ het gemiddelde gewicht van een recruit van de lichte 1968. Het middelen over een oneindige populatie zoals in voorbeeld c geeft mathematische moeilijkheden, maar het is duidelijk dat het gemiddelde aantal ogen bij een worp met een zuivere dobbelsteen $3\frac{1}{2}$ is (d.i. het gemiddelde van de even waarschijnlijke waarden 1 t/m 6). We vestigen er de aandacht op dat $\underline{x}$ een stochastische grootte is, dus aan toevalsfluctuaties onderhevig, terwijl $E_{\underline{x}}$ een vast getal is.

De gemiddelde waarde van $\underline{x}$ in een lange reeks experimenten blijkt doorgaans dicht bij het getal $E_{\underline{x}}$ te liggen, en wel des te dichter naarmate de reeks langer is. Tien wegingen van aselekt gekozen recruten geven meestal een gemiddelde tussen 65 en 80 kg, maar als men toevallig een tiental met veel zware recruten treft kan er bijv. 85 kg uitkomen. Het gemiddelde van duizend wegingen zal echter bij nagenoeg elk duizendtal minder dan 1 kg van het populatiegemiddelde $E_{\underline{x}}$ afwijken. In het volgende zal deze statistische regelmaat van groot belang blijken te zijn.

Vaak wil men ook aangeven in welke mate een willekeurige waarde van $\underline{x}$ van de verwachting $E_{\underline{x}}$ zal afwijken. Voor wie in een vrij beroep 1500 gulden per maand verwacht te verdienen maakt het veel verschil of zijn inkomen zal schommelen tussen 1300 en 1700 gulden dan wel geregeld dalen tot 500 en stijgen tot 2500 gulden. Als maat voor de afwijkingen tussen $\underline{x}$ en $E_{\underline{x}}$ gebruikt men de verwachting van de kwadratische afwijking, variantie genaamd en aangeduid met $\sigma^2(\underline{x})$. Men definieert a.h.w. een nieuwe stochastische grootte $\underline{z} = (\underline{x} - E_{\underline{x}})^2$ op dezelfde populatie als $\underline{x}$ en berekent daarvan het populatiegemiddelde:

$$(3.2) \quad \sigma^2(\underline{x}) \stackrel{\text{def}}{=} E (\underline{x} - E_{\underline{x}})^2.$$

Uiteraard is $\sigma^2(\underline{x}) \geq 0$ voor elke stochastische grootte $\underline{x}$. De positieve vierkantswortel uit de variantie is de standaardafwijking $\sigma(\underline{x})$. Voor verwachting en variantie kan men de volgende eigenschappen bewijzen:

$$(3.3) \quad \left\{ \begin{array}{l} E(c\underline{x}) = c E_{\underline{x}} \quad \text{voor elke constante } c; \\ E(c + \underline{x}) = c + E_{\underline{x}}; \\ E(\underline{x} + \underline{y}) = E_{\underline{x}} + E_{\underline{y}}; \\ \sigma^2(c\underline{x}) = c^2 \sigma^2(\underline{x}); \\ \sigma^2(c + \underline{x}) = \sigma^2(\underline{x}). \end{array} \right.$$

Men noemt twee stochastische grootheden onafhankelijk, als de getalwaarde van de ene grootheid bij een experiment geen invloed heeft op de kansverdeling van de andere grootheid. In voorbeeld d zijn l en g niet onafhankelijk, want lange recruten zullen in het algemeen zwaarder zijn dan kleine. De aantallen ogen bij twee opeenvolgende worpen met een zuivere dobbelsteen zijn wel onafhankelijk: ook bij de tweede worp heeft elk der waarden 1 t/m 6 een kans $\frac{1}{6}$, ongeacht wat de eerste worp heeft opgeleverd. Nu geldt

$$(3.4) \quad \sigma^2(\underline{x} + \underline{y}) = \sigma^2(\underline{x}) + \sigma^2(\underline{y}) \text{ mits } \underline{x} \text{ en } \underline{y} \text{ onafhankelijk zijn.}$$

Bij afhankelijkheid behoeft (3.4) niet te gelden. In het extreme geval $\underline{y} = 1 - \underline{x}$ bijv. heeft de som steeds de waarde 1, dus variantie 0, ook als x en y wel degelijk positieve variantie bezitten.

Als toepassing van (3.3) en (3.4) bepalen wij nu verwachting en variantie van het steekproefgemiddelde. Laten $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ de resultaten voorstellen van de waarneming van x aan n onafhankelijk en aselekt te kiezen populatie-elementen. Dan geldt voor elke index i uit de rij 1 t/m n dat $E\underline{x}_i = E\underline{x}$ en $\sigma^2(\underline{x}_i) = \sigma^2(\underline{x})$, terwijl $\underline{x}_1$ t/m $\underline{x}_n$ bovendien onafhankelijk zijn. Men noteert het steekproefgemiddelde als $\underline{\bar{x}}$:

$$(3.5) \quad \underline{\bar{x}} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \underline{x}_i \quad *)$$

Nu geldt volgens de eerste regel van (3.3)

$$E\underline{\bar{x}} = E\left(\frac{1}{n} \sum_{i=1}^n \underline{x}_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n \underline{x}_i\right).$$

Herhaald toepassen van de derde regel geeft

$$E\left(\sum_{i=1}^n \underline{x}_i\right) = \sum_{i=1}^n E\underline{x}_i = n E\underline{x},$$

immers de laatste som bestaat uit n getallen die elk gelijk zijn aan Ex. Dus geldt

*)

De notatie $\sum_{i=1}^n \underline{x}_i$ betekent $\underline{x}_1 + \underline{x}_2 + \dots + \underline{x}_n$, de som van de $\underline{x}_i$

als de index i de waarden 1 t/m n doorloopt.

$$(3.6) \quad \overline{E\mathbf{x}} = \frac{1}{n} \sum E\mathbf{x} = E\mathbf{x}.$$

Evenzo leidt men uit de vierde regel van (3.3) en uit (3.4) af dat

$$(3.7) \quad \sigma^2(\overline{\mathbf{x}}) = \sigma^2\left(\frac{1}{n} \sum_{i=1}^n \underline{x}_i\right) = \frac{1}{n^2} \sigma^2\left(\sum_{i=1}^n \underline{x}_i\right) =$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sigma^2(\underline{x}_i) = \frac{1}{n^2} \cdot n \sigma^2(\underline{\mathbf{x}}) = \frac{\sigma^2(\underline{\mathbf{x}})}{n}.$$

Wij komen op (3.6) en (3.7) terug in hoofdstuk 6.

De kansrekening houdt zich bezig met het berekenen van onbekende kansen uit bekende kansen. Fundamenteel zijn daarbij een aantal rekenregels die wij hier zonder bewijs geven. In het geval van een eindige populatie met aselechte trekking zijn ze echter gemakkelijk te verifiëren.

Als A en B twee eventualiteiten zijn, dan is

$$(3.8) \quad P(A \text{ of } B) = P(A) + P(B) - P(A \text{ en } B). \quad *)$$

Voorbeeld: Zij bij het gooien met een zuivere dobbelsteen A het gooien van een even aantal ogen en B het gooien van een zes. We hebben dan $P(A \text{ of } B) = \frac{1}{2} + \frac{1}{6} - \frac{1}{6} = \frac{1}{2}$.

Een bijzonder geval krijgen we als A en B elkaar uitsluiten, d.w.z. als A en B niet tegelijkertijd op kunnen treden. Dan is immers $P(A \text{ en } B) = 0$, zodat we hebben:

$$(3.9) \quad P(A \text{ of } B) = P(A) + P(B) \text{ als A en B elkaar uitsluitende eventualiteiten zijn.}$$

*) Met "of" is steeds het niet-exclusieve "of/en" bedoeld. Eventualiteit "A of B" treedt dus op zodra minstens één van beide eventualiteiten A of B optreden.

Voorbeeld: We trekken een speelkaart uit een goed geschud spel kaarten. Als A het trekken van een schoppenkaart is en B het trekken van ruitenboer is, dan is

$$P(A \text{ of } B) = \frac{13}{52} + \frac{1}{52} = \frac{14}{52}.$$

(3.9) kan gegeneraliseerd worden als volgt:

Als $A_1, A_2, \dots, A_k$ elkaar uitsluitende eventualiteiten zijn, d.w.z. als geen tweetal dezer eventualiteiten tegelijkertijd kan optreden, dan is

$$(3.10) \quad P(A_1 \text{ of } A_2 \text{ of } \dots \text{ of } A_k) = P(A_1) + P(A_2) + \dots + P(A_k).$$

Voorbeeld: Zij bij het gooien met een zuivere dobbelsteen A_1 het gooien van een 1, A_2 het gooien van een 2, A_3 het gooien van een 3 en A_4 het gooien van een 4. Als nu B het gooien van hoogstens 4 is, dan is $P(B) = P(A_1) + P(A_2) + P(A_3) + P(A_4) = \frac{2}{3}$.

Als A en B twee eventualiteiten zijn die elkaars al dan niet optreden niet beïnvloeden, dan is

$$(3.11) \quad P(A \text{ en } B) = P(A) \cdot P(B).$$

Als (3.11) geldt, noemt men A en B onafhankelijk.

Voorbeelden:

a. We gooien tweemaal achtereenvolgens met een munt. We noemen A de eventualiteit dat het resultaat van de eerste worp "kruis" is, en B dat het resultaat van de tweede worp "kruis" is. Als men "eerlijk" werpt, dan wordt de uitkomst van de tweede worp niet door die van de eerste worp beïnvloed, zodat A en B onafhankelijk zijn en, als de munt zuiver is,

$$P(A \text{ en } B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

b. Bij het gooien met een zuivere dobbelsteen noemen we A het gooien van hoogstens 4, en B het gooien van een even aantal ogen. Nu is $P(A) = \frac{2}{3}$, $P(B) = \frac{1}{2}$ en $P(A \text{ en } B) = \frac{1}{3}$. Dit laatste omdat (A en B) het gooien van een 2 of een 4 is. Derhalve is aan (3.11) voldaan,

zodat A en B onafhankelijk zijn.

- c. Zij nu, weer bij het gooien met een zuivere dobbelsteen, A het gooien van hoogstens 3 ogen en B als in voorbeeld b. Dan is nu $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$ en $P(A \text{ en } B) = \frac{1}{6}$, want (A en B) is het gooien van een 2, zodat nu niet voldaan is aan (3.11): A en B zijn afhankelijk. Dit kunnen we ook wel anders inzien door te bedenken, dat als A optreedt er nog drie - even waarschijnlijke - mogelijke uitkomsten zijn, te weten 1, 2 en 3, waarbij er slechts één is waarbij B optreedt, nl. 2. Treedt A niet op, dan zijn er ook 3 - even waarschijnlijke - mogelijke uitkomsten, waarbij er twee zijn waarbij B optreedt.

Hoofdstuk 4. Schattingstheorie.

In het vorige hoofdstuk hebben wij de stochastische grootheid g (gewicht) gedefinieerd op de populatie π bestaande uit alle recruten van de lichting 1968. Als er nu een legerorder verschijnt waarbij elke recruit van zwaarder dan 75 kg recht heeft op een bed met versterkte spiraalvering, zal de Chef Intendance graag willen weten hoeveel van de opgeroepen 20.000 recruten daarvoor in aanmerking komen. Dit blijkt uit de 20.000 keuringsrapporten waarop ieders gewicht vermeld is. Maar voor een benadering is het voldoende een aselechte steekproef van bijv. 100 keuringsrapporten te onderzoeken. Als nl. in die steekproef 31 recruten van zwaarder dan 75 kg voorkomen, kan de fractie zware recruten in de populatie op 31% geschat worden. Voor 20.000 man zijn dan circa 31% van 20.000, dus circa 6200, zware bedden nodig.

Het gebruik van "circa" is geen overbodige luxe. Zelfs als de steekproef volkomen aselekt, dus zonder aanzien des gewichts, wordt samengesteld, is er een - kleine - kans dat in de steekproef veel meer of veel minder zwaargewichten zitten dan in de populatie. De fractie zwaargewichten in de steekproef is een getal dat van steekproef tot steekproef kan verschillen: het is zelfs een stochastische variabele f . De zg. wet van de grote aantallen garandeert ons dat de kansverdeling van f voor toenemende steekproefomvang n steeds dichter geconcentreerd ligt om de waarde p van de fractie zwaargewichten in de populatie: waarden van f die veel van p verschillen worden steeds onwaarschijnlijker naarmate n toeneemt.

Het is belangrijk onderscheid te maken tussen de stochastische grootheid en zijn bij een experiment aangenomen waarde. De steekproef-fractie f in een nog te trekken aselechte steekproef is een stochastische grootheid, die een schatter is van de populatiefractie p . De getalwaarde f die ontstaat als inderdaad een steekproef is getrokken en waargenomen wordt schatting genoemd, al is het onderscheid tussen beide begrippen niet bij alle auteurs consequent volgehouden. Men zou kunnen zeggen dat de schatter het berekeningsvoorschrift is en de schatting het resultaat.

Men noemt een schatter van een onbekende grootheid zuiver, als hij in verwachting (dus gemiddeld over alle denkbare steekproeven) de juiste

waarde oplevert. Laat een fractie p van een populatie een eigenschap A hebben, en laat $\underline{f}$ de fractie met A zijn in een aselechte steekproef van omvang n . Dan kan men bewijzen dat $\underline{f}$ een zuivere schatter van p is: er geldt

$$(4.1) \quad E\underline{f} = p.$$

Het bovenstaande geldt voor aselechte steekproeven met én zonder teruglegging. Men kan echter berekenen dat

$$(4.2) \quad \sigma^2(\underline{f}) = \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}$$

voor een aselechte steekproef zonder teruglegging van omvang n uit een populatie van omvang N , terwijl

$$(4.3) \quad \sigma^2(\underline{f}) = \frac{p(1-p)}{n}$$

voor een aselechte steekproef met teruglegging van omvang n . Voor aselechte steekproeven zonder teruglegging geldt (4.3) exact als N oneindig is, en bij benadering als N veel groter is dan n (de laatste factor van (4.2) is dan nagenoeg 1). In beide variantieformules zien we de variantie van $\underline{f}$ afnemen als n toeneemt. De variantie wordt nul als $n = N$ (alles onderzocht!) resp. als $n \rightarrow \infty$.

Wij keren terug naar de Chef Intendance die circa 6200 zware recruten verwacht. Om te voorkomen dat tegen de wens van de generaal in een recruit door zijn bed zakt, wil deze Chef een bovengrens bepalen voor het aantal zware gevallen. Nu is met zekerheid alleen een absurd hoge bovengrens aan te geven: in theorie is het denkbaar dat de 20.000 - 100 niet onderzochte militairen allen meer dan 75 kg wegen, en dan moeten er $19.900 + 31$ dubbele spiraalveren komen. Maar dat is extreem onwaarschijnlijk. Een betere (want lagere) bovengrens is alleen te geven, als men het risico van een fout door een dergelijke extreem toevallige toestand wil lopen. Als de Chef een kans α op een te lage bovengrens acceptabel vindt, kan een statisticus hem een bij de waargenomen steekproeffractie $\underline{f}$ en steekproefomvang n behorend getal $p_b(\underline{f})$ opgeven, zodanig dat de conclusie $p < p_b(\underline{f})$ voor slechts een fractie α van alle

denkbare steekproeven van omvang n onjuist is. Voor α kiest men meestal 0,05. Men noemt $p_b(\underline{f})$ een betrouwbaarheidsbovengrens voor p met onbetrouwbaarheid α , of met betrouwbaarheidscoëfficiënt $1 - \alpha$.

De bepaling van $p_b(\underline{f})$ uit n en $\underline{f}$ en α geschiedt met tabellen, of met benaderingsformules als b groot is. Wij volstaan met het reproduceren van twee figuren waaruit men $p_b(\underline{f})$ ongeveer kan aflezen, voor $\alpha = 5\%$ (fig. 4.1) en $\alpha = 2,5\%$ (fig. 4.2). Men vindt in deze figuren ook betrouwbaarheidsondergrenzen $p_o(\underline{f})$ waarvoor de conclusie $p > p_o(\underline{f})$ in slechts een fractie α van de toepassingen onjuist is. De conclusie $p_o(\underline{f}) < p < p_b(\underline{f})$ is in een fractie 2α van de gevallen verkeerd: men noemt het interval van $p_o(\underline{f})$ tot $p_b(\underline{f})$ een betrouwbaarheidsinterval

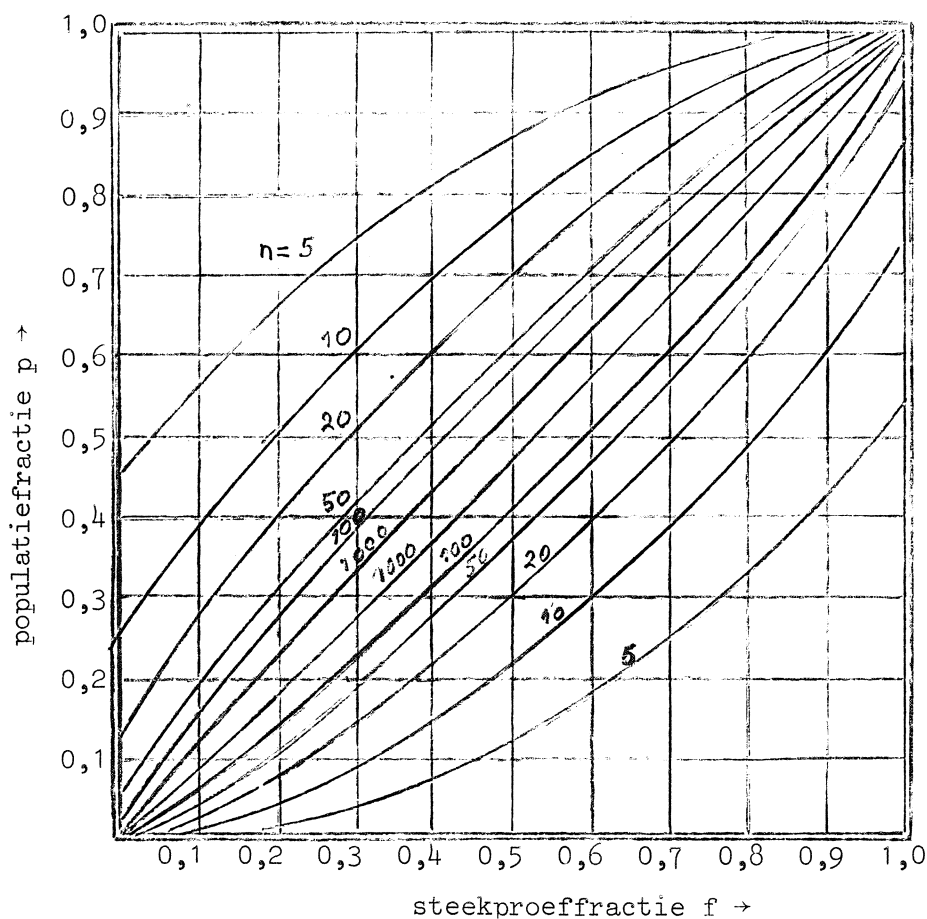


fig. 4.1 Betrouwbaarheidsgrenzen $p_o(f)$ en $p_b(f)$ met onbetrouwbaarheid 0,05 voor onder- of bovengrens apart en 0,10 voor het interval.

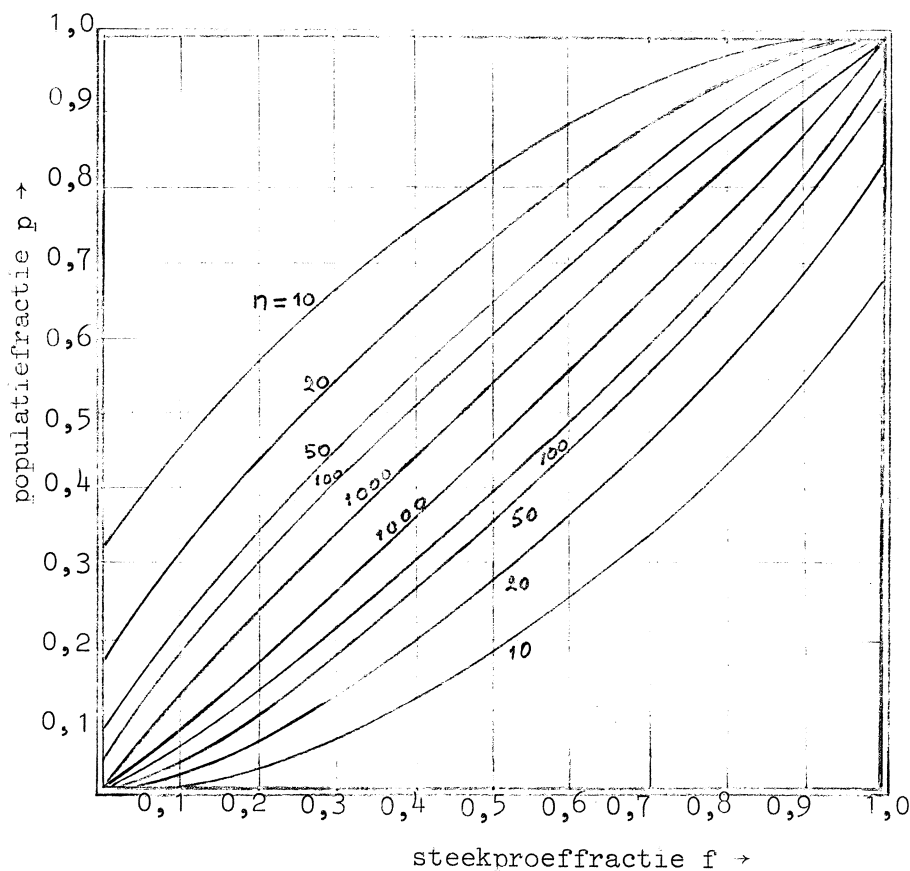


fig. 4.2 Betrouwbaarheidsgrenzen $p_o(f)$ en $p_b(f)$ met onbetrouwbaarheid 0,025 voor onder- of bovengrens apart en 0,05 voor het interval.

met betrouwbaarheidscoëfficiënt $1 - 2\alpha = 90\%$ (fig. 4.1) resp. $1 - 2\alpha = 95\%$ (fig. 4.2).

In ons voorbeeld was $n = 100$ en $f = 0,31$. Bij $\alpha = 0,05$ geeft fig. 4.1 een $p_b(f)$ van 0,40. Aankoop van $0,40 \times 20.000 = 8000$ bedden is dus een politiek die in 95 van de 100 gevallen voldoende zal blijken te zijn. We kunnen ook in fig. 4.2 aflezen dat $p_o(f) = 0,22$ en $p_b(f) = 0,41$: een 95% betrouwbaarheidsinterval voor p loopt dus van 0,22 tot 0,41.

We zien in beide figuren, dat het betrouwbaarheidsinterval steeds de waarde $p = f$ bevat en nauwer wordt naarmate n toeneemt. De intervallen met $\alpha = 0,10$ zijn nauwer dan die met $\alpha = 0,05$ bij dezelfde n en f : wie meer risico op fouten neemt kan nauwkeuriger uitspraken doen.

Samenvattend kunnen we zeggen: een betrouwbaarheidsinterval met onbetrouwbaarheid α is een uit steekproefwaarden berekend (dus stochastisch) interval, dat in een fractie $1 - \alpha$ van de toepassingen de onbekende (maar vaste) populatiegrootte bevat. Het zou onjuist zijn te zeggen dat de fractie zware recruten met een kans van 95% tussen 0,22 en 0,41 ligt, want deze fractie is een vast getal p dat ik kan bepalen door alle recruten te wegen. Dat getal p ligt óf wel tussen 0,22 en 0,41 óf niet, maar niet met een kans van 95%.

Hoofdstuk 5. Toetsingstheorie.

Volgens de zwaartekrachtstheorie van NEWTON vallen in het luchtledige alle voorwerpen even snel, en menige natuurkundeleraar heeft zijn schoolklas tot verbazing gebracht door dit met een veertje en een stukje lood in een leeggezogen buis te demonstreren. Een dergelijke uitkomst bevestigt de theorie. Wie een nieuwe theorie lanceert waarvan de voorspellingen door een experiment gelogenstraft worden, doet er goed aan de theorie nog eens kritisch te bekijken.

Bij de meeste natuurkundige verschijnselen is één experiment voldoende. Immers een onafhankelijke herhaling levert dezelfde uitkomst op, als we afzien van meetfouten, veranderde omstandigheden e.d. Experimentele verificatie van een theorie dat meisjes eerder het melkgebit verliezen dan jongens, of dat roken het ontstaan van longkanker bevordert, is aanzienlijk lastiger. Volmaakt ondubbelzinnig is de uitslag van een experiment, in casu het onderzoek van een aselechte steekproef, in zo'n geval nooit. Wanneer de waarnemingen echter veel beter in overeenstemming zijn met de geponeerde theorie dan met haar tegendeel betekent dit een statistische bevestiging. Deze is minder sterk dan een logisch bewijs, dat ieder overtuigt tenzij er fouten tegen de logica in optreden. Bij een statistische bevestiging van de vroegere gebitswisseling van meisjes bestaat altijd een kleine kans, dat de meisjes in onze steekproef toevallig eerder wisselen dan de jongens, hoewel dat in de populatie niet het geval was. Op dit risico van fouten komen wij nog terug.

Voorbeeld: Wij hebben al enige malen gesproken over een vaas met witte en rode knikkers. Laat er een vermoeden bestaan dat de fractie p van de rode knikkers in de vaas $\frac{1}{4}$ zou zijn. (Als men de knikkers vervangen denkt door potentiële nakomelingen van de kruising van twee planten, zou $p = \frac{1}{4}$ het gevolg kunnen zijn van een erfelijkheidstheorie). Men wenst dit vermoeden te onderzoeken door een aselechte steekproef zonder teruglegging uit de (zeer talrijke) populatie te nemen. Veronderstel dat men hierbij slechts 4 rode knikkers vindt in een steekproef van 50. Als de populatiefractie $\frac{1}{4}$ is, verwacht men $\frac{1}{4} \times 50$, dus ongeveer 12 of

13, rode knikkers te vinden. Het gevonden aantal is dus slecht in overeenstemming met de theorie. Het vermoeden $p = \frac{1}{4}$ is niet tegen statistische toetsing bestand gebleken, en wordt verworpen ten gunste van een nieuwe bewering dat $p < \frac{1}{4}$ is. Zouden we bijv. 10 of 13 rode knikkers onder de 50 hebben aangetroffen dan hadden we gezegd dat waarneming en vermoeden elkaar niet tegenspreken.

Een statistische toets is een procedure om na te gaan in hoeverre waarnemingen aan een aselechte steekproef in overeenstemming zijn met een tevoren opgestelde hypothese omtrent een populatiegrootheid. Deze te toetsen hypothese wordt nulhypothese genoemd en met H_0 aangeduid. Bij een slechte overeenstemming tussen H_0 en waarnemingen verwerpt men H_0 . Bij een goede overeenstemming wordt H_0 niet verworpen. Dit wil niet zeggen dat H_0 nu bewezen is en aanvaard mag worden, maar alleen dat onze waarnemingen H_0 niet tegenspreken.

Omdat een toets dus niet kan bewijzen maar wel ontkennen, formuleert men H_0 meestal als het tegendeel van wat op theoretische gronden wordt verwacht. Wie bijv. uit groeitheorieën of uit vroegere ervaringen gelooft dat meisjes eerder wisselen dan jongens, kan het beste als H_0 formuleren dat meisjes later of even laat wisselen. Als dan in een aselechte steekproef de meisjes veel eerder zijn, wordt H_0 verworpen en mag men zijn mening statistisch bevestigd noemen. Afhankelijkheid van bijv. inkomen en televisiebezit toont men aan door onafhankelijkheid te toetsen.

Hoe meet men nu de verenigbaarheid van H_0 en de waarnemingen? Na de formulering van H_0 kiest men een onbetrouwbaarheidsdrempel α , dit is het toelaatbaar geachte risico dat H_0 ten onrechte wordt verworpen. Een gangbare keuze is $\alpha = 0,05$, terwijl ook $\alpha = 0,01$ geregeld voorkomt. Vervolgens zoekt men een toetsingsgrootheid t ; dit is een functie van de waarnemingen die zich onder H_0 anders zal gedragen dan bij onjuistheid van H_0 . Zolang de steekproef niet getrokken en waargenomen is, zijn de waarnemingen stochastisch en zal ook t stochastisch zijn. In de kansverdeling van t zonderen wij nu een kritieke zone Z af van waarden die onder H_0 extreem onwaarschijnlijk zijn maar bij onjuistheid van H_0 wel kunnen voorkomen. Blijkt nu bij invulling van de waarnemingen de waarde van t in dit kritieke gebied te liggen, dan verwerpen wij H_0 . De keuze van Z wordt zo gemaakt dat

$$(5.1) \quad P(\underline{t} \text{ valt in } Z \text{ als } H_0 \text{ juist}) \leq \alpha$$

is, zodat er een kans van hoogstens α bestaat dat we H_0 verwerpen hoewel H_0 juist is.

In ons voorbeeld van de witte en rode knikkers toetsen wij H_0 : $p = \frac{1}{4}$ met $\alpha = 0,05$. Als toetsingsgrootte $\underline{t}$ kiezen we het aantal rode knikkers in een aselechte steekproef van 50 stuks; dit leidt tot de zg. binomiale toets. De stochastische grootte $\underline{t}$ kan theoretisch alle waarden 0 t/m 50 aannemen; als H_0 juist is zijn de waarden rondom 12 en 13 het meest waarschijnlijk. Voor de kritieke zone Z kiezen we de waarden dichtbij 0 en 50, die immers H_0 het sterkste tegenspreken. Men kan nu in een tabel opzoeken ^{*)} dat onder H_0 geldt $P(\underline{t} \leq 6) = 1,94\%$ en $P(\underline{t} \geq 19) = 2,87\%$. Als Z uit de punten 0 t/m 6 en 19 t/m 50 bestaat, is $P(\underline{t} \text{ in } Z \text{ als } H_0 \text{ juist}) = 1,94\% + 2,87\% = 4,81\% < \alpha$, zoals vereist. Omdat onze steekproef 4 rode knikkers telde en 4 tot Z behoort, verwerpen wij H_0 met een onbetrouwbaarheidsdrempel van 5%. De grenspunten 6 en 19 waar Z begint en ophoudt noemt men de linker en rechter kritieke waarde van de toets: het zijn de uitkomsten die nog juist tot verwerping van H_0 leiden.

De kritieke zone was hier tweezijdig, omdat zowel zeer grote als zeer kleine waarden van $\underline{t}$ in strijd met H_0 zijn. Wie als H_0^* stelt dat de fractie rode knikkers in de populatie minstens $\frac{1}{4}$ is, zal H_0^* alleen verwerpen bij weinig rode exemplaren in de steekproef. Men kan dan bij $\alpha = 0,05$ voor Z de waarden 0 t/m 7 gebruiken, omdat onder H_0 geldt $P(\underline{t} \leq 7) = 0,0453$. Nu bevat Z tevens de waarde 7, wat maakt dat H_0^* eerder verworpen wordt dan H_0 . Een dergelijke eenzijdige toetsing is echter alleen geoorloofd als er vóór de bestudering van de steekproef

*) Op de wiskundige formule van deze kans gaan wij hier niet in. Het aantal successen in n onafhankelijke experimenten die elk met kans p een succes opleveren is een stochastische grootte met een zg. binomiale verdeling (in ons geval $n = 50$ en $p = \frac{1}{4}$). Voor $n \leq 150$ kan men de bijbehorende kansen in tabellen opzoeken, terwijl voor grote waarden van n benadering mogelijk is met de normale verdeling (echter voor p zeer dicht bij 0 of 1 met de Poissonverdeling).

redenen waren om een systematische afwijking in één richting te veronderstellen, of als alleen een afwijking in één richting van belang is. De beslissing over één- of tweezijdig toetsen en de keuze van α mogen nimmer op grond van de uitkomsten worden gedaan, omdat dit tot bevooroordeelde conclusies leidt.

Voorbeeld: Een bedrijf overweegt de aanschaf van een nieuwe machine, maar wil hier alleen toe overgaan als de nieuwe machine beter is dan de bestaande. Deze bestaande geeft een fractie 1% aan uitval. Voor de uitvalfractie p van de nieuwe machine toetst men dan $p \geq 0,01$. Alleen bij verwerping gaat men tot aanschaf over. Het zou geen zin hebben tweezijdig te toetsen, want bij verwerping van $p = 0,01$ ten gunste van $p > 0,01$ wordt evenmin een nieuwe machine gekocht als bij niet-verwerpen van H_0 .

Voorbeeld: Een reclamebureau meent, dat de televisiereclame in het doorsnee Nederlandse gezin even intensief gevolgd wordt door man en vrouw. Een enquêteur bezoekt 50 aselekt gekozen gezinnen en onderzoekt bij beide echtgenoten apart hoeveel van tien recente reclamespots zij zich weten te herinneren. Wij gaan niet in op de techniek van de ondervraging, hoewel die van groot belang is om een vertekend beeld van de toestand te voorkomen. De fictieve uitslag van de fictieve enquête vindt men in tabel 5.1.

Volgens de bewering zouden man en vrouw evenveel aandacht aan reclame schenken. Een eventueel verschil in aantal herkende spots zou dan toevallig zijn, en dus even vaak voorkomen ten gunste van de vrouw (plusteken in de tabel) als ten gunste van de man (minteken). Wij passen daarom de tekentoets toe, d.w.z. wij gaan na of de fractie plustekens onder de 28 gezinnen met een verschil in score systematisch afwijkt van $\frac{1}{2}$. Dit komt neer op een binomiale toets van $p = \frac{1}{2}$ met $n = 28$. Kiezen wij $\alpha = 0,05$ dan voldoet de keuze 0 t/m 8 en 20 t/m 28 voor Z , want onder $p = \frac{1}{2}$ is $P(\underline{t} \leq 8) = P(\underline{t} \geq 20) = 1,8\%$. In de tabel staan 7 plustekens, zodat H_0 verworpen wordt ten gunste van de conclusie dat mannen beter op reclamespots letten.

Tabel 5.1. Fictieve resultaten van een fictief onderzoek naar de herkenning van reclamespots

gezin nr	aantal herkende spots			gezin nr	aantal herkende spots		
	vrouw	man	teken		vrouw	man	teken
1	2	5	-	17	0	5	-
2	0	0	=	18	0	7	-
3	8	10	-	19	2	0	+
4	4	0	+	20	8	5	+
5	6	6	=	21	10	2	+
6	2	5	-	22	0	9	-
7	1	3	-	23	0	5	-
8	0	8	-	24	2	4	-
9	4	8	-	25	0	0	=
10	2	0	+	26	4	6	-
11	4	7	-	27	8	1	+
12	0	7	-	28	6	2	+
13	4	8	-	29	6	9	-
14	1	4	-	30	0	2	-
15	5	9	-	31	0	5	-
16	6	8	-				

Geen televisie in 3 gezinnen; man afwezig in 12 gezinnen; vrouw afwezig in 2 gezinnen; beide afwezig in 2 gezinnen.

Alvorens nu te adviseren de reclame meer op mannelijke kijkers te richten, willen we opmerken dat bovenstaande klakkeloze toepassing van de tekentoets niet gerechtvaardigd lijkt. Omdat er 12 gezinnen zijn waar de man niet thuis was en slechts 2 met afwezigheid van de vrouw, moeten we onze conclusie op zijn minst wat afzwakken. Immers het is plausibel dat een aantal van deze afwezigen vaak s'avonds uit is en dus weinig T.V.-spots zal herkennen. Bij het radi-cale standpunt dat degene die niet thuis is steeds minder spots her-kent dan zijn of haar huwelijkspartner, krijgen we $28 + 12 + 2 = 42$ gezinnen met verschil in score, waaronder $7 + 12 = 19$ plustekens. Voor $n = 42$ is de kritieke zone bij toetsing van $p = \frac{1}{2}$ echter 0 t/m 14 en

28 t/m 42, zodat de waarde 19 niet in Z ligt en heel wel met $p = \frac{1}{2}$ verenigbaar is. Het lijkt dus wenselijk de enquête te herhalen in 50 nieuwe gezinnen en er door een afsprakensysteem zoveel mogelijk voor te zorgen dat beide huwelijkspartners de vragen beantwoorden.

Dit laatste voorbeeld is helaas kenmerkend voor de moeilijkheden die men bij het gebruik van statistische methoden kan ontmoeten. Als we de gesignaleerde moeilijkheid nu verder terzijde laten, kunnen we nog opmerken dat een enquêteresultaat van 4 plustekens bij 28 scoreverschillen natuurlijk nog veel meer in strijd zou zijn met H_0 dan de in tabel 5.1 gevonden 7 stuks. Wie in beide gevallen alleen rapporteert dat "de toets met onbetrouwbaarheid 0,05 leidde tot verwerping van H_0 ", of dat "een op niveau $\alpha = 0,05$ significant verschil tussen mannen en vrouwen werd gevonden", vergeet niet alleen te vermelden op hoeveel waarnemingen dit gebaseerd is, welke toets werd gebruikt en in welke richting het verschil wees, maar bovendien komt niet tot uiting hoe sterk de tegenspraak tussen H_0 en de waarnemingen was. Een maat voor deze discrepantie is de overschrijdingskans, die de kans onder H_0 aangeeft op de gevonden waarde van $\underline{t}$ of een nog extremere waarde. Bij 7 successen in 28 experimenten met $p = \frac{1}{2}$ geldt $P(\underline{t} \leq 7) = 0,63\%$, maar de overschrijdingskans is bij deze tweezijdige toetsing $2 \times 0,63\% = 1,26\%$ omdat 21 of meer successen onder H_0 even onverwacht is als 7 of minder, en $P(\underline{t} \geq 21)$ is hier eveneens $0,63\%$ uit symmetrie-overwegingen. Bij slechts 4 plustekens was de overschrijdingskans $2P(\underline{t} \leq 4)$ en dit is kleiner dan $0,02\%$, zodat vermelding van de overschrijdingskans in dat geval duidelijk zou maken dat het resultaat volmaakt onverenigbaar is met H_0 . Men kan stellen dat overschrijdingskansen van minder dan α behoren bij t-waarden in de kritieke zone en omgekeerd.

Wie toetst maakt onvermijdelijk fouten. We hebben al gezien dat met een kans ten hoogste α de nulhypothese ten onrechte wordt verworpen, d.w.z. een t-waarde in Z wordt gevonden hoewel H_0 juist is. Dit heet een fout van de eerste soort; deze kan alleen optreden als H_0 correct is. Daarentegen wordt een fout van de tweede soort, het ten onrechte niet verwerpen van H_0 , gemaakt als H_0 in feite onjuist is maar toch een t-waarde buiten Z wordt gevonden. Voor deze fout is geen eenvoudige

grens te geven. Wij volstaan met de opmerking dat men streeft naar toetsen met een zo groot mogelijk onderscheidingsvermogen (Engels: power), d.i. een zo klein mogelijke kans op een fout van de tweede soort. In het algemeen wordt het onderscheidingsvermogen van een toets beter naarmate het aantal waarnemingen toeneemt. Verlaging van α bij gelijkblijvende steekproefomvang doet het onderscheidingsvermogen juist dalen: om de ene fout te vermijden vergroot men de kans op een andere.

Wij merken nog op dat de kritieke zone van een binomiale toets voor $p = p_0$ (p_0 een vast getal) juist bestaat uit die waarden van t waarvoor de waarde p_0 buiten het betrouwbaarheidsinterval van de fractie $\frac{t}{n}$ ligt. Bij 4 rode erwten uit 50 en toetsing van $p = \frac{1}{4}$ met $\alpha = 0,05$ ligt de waarde $\frac{1}{4}$ buiten het interval van 0,02 tot 0,20 (fig. 4.2 voor $n = 50$ en $f = \frac{4}{50}$). De kritieke zone loopt voor deze toets tot aan $0,12 \times 50 = 6$ en vanaf $0,38 \times 50 = 19$, omdat de lijn $p = 0,25$ de curven $n = 50$ snijdt bij $f = 0,12$ en $f = 0,38$. Voor eenzijdige toetsen bestaat een soortgelijk verband met de betrouwbaarheidsboven- of ondergrens.

Hoofdstuk 6. De normale verdeling.

In de voorgaande hoofdstukken hebben we gesproken over stochastische grootheden, die slechts een beperkt aantal verschillende waarden kunnen aannemen, zoals het aantal rode ballen in een steekproef van n ballen, het verkregen aantal ogen bij een worp met een dobbelsteen en de fractie personen zwaarder dan een bepaald gewicht in een steekproef van omvang 100. Indien men echter een lengte, gewicht, temperatuur of tijdsverloop tussen twee gebeurtenissen meet, dan is het aantal mogelijke verschillende uitkomsten in principe oneindig groot, nl. alle reële getallen in een zeker interval (dit interval kan zich ook van $-\infty$ tot $+\infty$ uitstrekken). Ook dergelijke waarnemingen, waarvan de uitkomsten een schijn van grote precisie geven, kunnen een stochastisch karakter hebben, d.w.z. bij herhaling van hetzelfde experiment tot verschillende resultaten leiden. Deze variabiliteit kan bijv. veroorzaakt worden door een bij de meting optredende meetfout, of door een oncontroleerbare verandering van de omstandigheden waaronder het experiment wordt uitgevoerd. De populatie bestaat nu uit de collectie van alle mogelijke waarnemingsuitkomsten, die men met dezelfde meetapparatuur onder zo goed mogelijk gelijk gehouden omstandigheden zou kunnen verkrijgen. Voert men in de praktijk een dergelijke meting enige malen uit, er daarbij zorg voor dragend dat de waarnemingen elkaar niet beïnvloeden, dan kan men de waarnemingen veelal beschouwen als een aselechte steekproef uit de genoemde fictieve populatie.

Uiteraard kan men in de praktijk nooit metingen verrichten met een oneindig grote precisie, maar in vele gevallen vormt dit "model" toch een zo goede beschrijving (benadering) van de werkelijke situatie, dat men hier van uit kan gaan bij de statistische analyse.

Metten we van een zeer grote populatie nauwkeurig een aantal objecten, die wel van dezelfde soort doch niet geheel identiek zijn, dan is het aantal mogelijke verschillende meetresultaten eveneens zeer groot, ook al meten we zonder meetfout. Ook in dit geval zal men het besproken model als beschrijving van de werkelijke situatie vaak toepassen. Het is duidelijk, dat thans een aselechte steekproef van waarnemingen wordt verkregen, indien men de te onderzoeken objecten aselekt uit de populatie trekt.

Een stochastische grootheid van het besproken type noemt men gewoonlijk een continue stochastische grootheid (de grootheid kan een "continuum" van waarden aannemen), dit in tegenstelling tot discrete stochastische grootheden, die slechts bepaalde discrete waarden kunnen aannemen, zoals in de vorige hoofdstukken het geval was.

Zij $\underline{x}$ een continue stochastische grootheid. We hebben in hoofdstuk 3 reeds opgemerkt, dat men de kans dat $\underline{x}$ een waarde kleiner dan of gelijk aan x aanneemt, $P(\underline{x} \leq x)$, de verdelingsfunctie van $\underline{x}$ noemen. We kunnen de grootte van deze kansen eenvoudig grafisch weergeven, zie figuren 6.1 en 6.2.

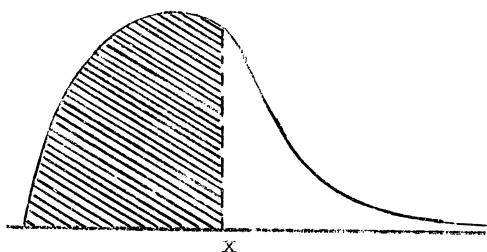


fig. 6.1 Grafiek van een scheve kansverdeling. Het gearceerde oppervlak is gelijk aan $P(\underline{x} \leq x)$.

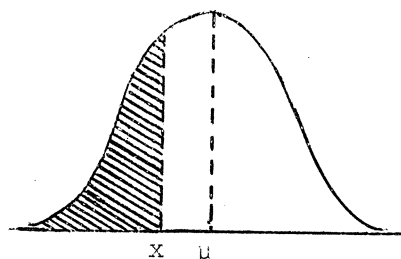


fig. 6.2 Grafiek van een symmetrische kansverdeling met symmetriepunt μ . Het gearceerde oppervlak is gelijk aan $P(\underline{x} \leq x)$.

De kansverdeling van een continue stochastische grootheid wordt nl. gekarakteriseerd door een grafiek, met oppervlak ter grootte 1 onder de grafiek en boven de horizontale as, waarop de waarden van $\underline{x}$ worden uitgezet. De kans $P(\underline{x} \leq x)$ is nu gelijk aan het oppervlak links van het punt x onder de kromme en is in beide figuren voor een zekere x gearceerd. Evenzo is de kans, dat $\underline{x}$ een waarde groter dan x aanneemt, $P(\underline{x} > x)$ gelijk aan het oppervlak rechts van het punt x onder de kromme.*)

*) Uit de interpretatie van een kans als een oppervlak volgt direct, dat bij continue stochastische grootheden de kansen $P(\underline{x} \leq x)$ en $P(\underline{x} < x)$ gelijk zijn; evenzo is $P(\underline{x} \geq x) = P(\underline{x} > x)$, etc.

Het zal nu duidelijk zijn, dat de kans, dat $\underline{x}$ een waarde tussen x_1 en x_2 aanneemt, $P(x_1 < \underline{x} < x_2)$, gelijk is aan het oppervlak tussen x_1 en x_2 onder de kromme. Al deze kansen zijn inderdaad steeds kleiner dan of gelijk aan 1; immers het totale oppervlak onder de grafiek is gelijk aan 1. Figuur 6.1 suggereert, dat de grafiek naar rechts tot hoogste nul nadert zonder ooit exact nul te worden. Dit betekent, dat $\underline{x}$ willekeurig grote waarden nog met een (zeer) kleine kans aanneemt. De grootheid $\underline{x}$ met de kansverdeling uit fig. 6.2 kan echter alleen waarden in een bepaald interval aannemen.

De grafiek van de kansverdeling in fig. 6.1 is niet symmetrisch; de kansverdeling wordt daarom scheef genoemd. De grafiek in fig. 6.2 is symmetrisch om het symmetriepunt μ ; men noemt de kansverdeling dan symmetrisch (om μ). Merk op, dat in dit geval $P(\underline{x} \leq \mu) = P(\underline{x} \geq \mu)$. Een getal μ met deze eigenschap noemt men (ook bij scheve verdelingen) de mediaan van de kansverdeling. Bovendien is bij een symmetrische verdeling $E\underline{x} = \mu$, zodat bij symmetrische verdelingen mediaan en verwachting samenvallen in het symmetriepunt. Bij een scheve verdeling zijn $E\underline{x}$ en de mediaan gewoonlijk verschillend.

Een zeer belangrijke continue kansverdeling is de standaard-normale kansverdeling. Als de stochastische grootheid $\underline{u}$ standaard-normaal verdeeld is, kunnen we de kansen $P(\underline{u} \leq u)$ voor een groot aantal waarden van u aflezen in tabel 6.1. Dergelijke kansen kunnen we tevens interpreteren als het oppervlak links van het punt u onder de standaard-normale grafiek, zie figuur 6.3.

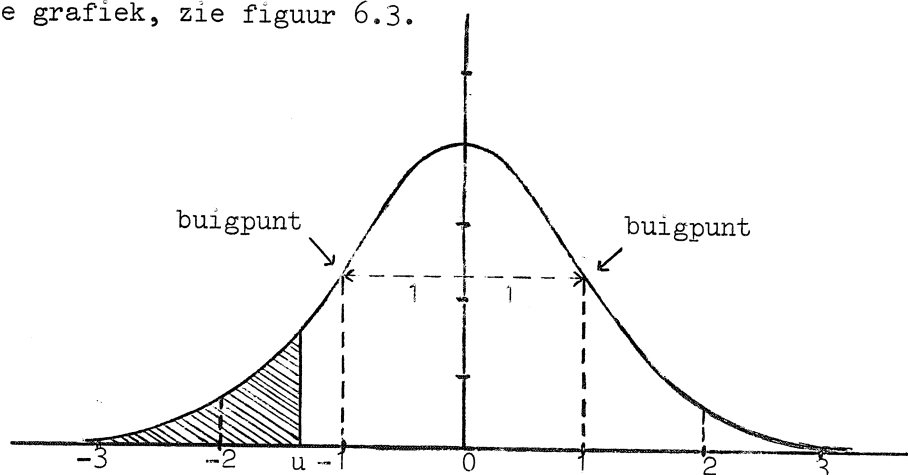
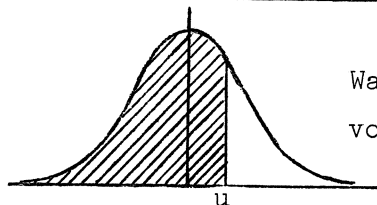


fig. 6.3 De standaard-normale verdeling. Het gearceerde oppervlak is gelijk aan de kans $P(\underline{u} \leq u)$.

Tabel 6.1 Verdelingsfunctie $P(\underline{u} \leq u)$ voor de standaard-normale verdeling



Waarden van $10^4 \cdot P(\underline{u} \leq u)$
voor $u = 0,00 (0,01) 3,49$

u	0	1	2	3	4	5	6	7	8	9
0,0	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0,6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1,9	9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
3,0	9987	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,1	9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,2	9993	9993	9994	9994	9994	9994	9994	9995	9995	9995
3,3	9995	9995	9995	9996	9996	9996	9996	9996	9996	9997
3,4	9997	9997	9997	9997	9997	9997	9997	9997	9997	9998

Een 5 betekent, dat bij afronding naar omlaag moet worden afgerond.

Een 5 wordt naar boven afgerond.

Deze grafiek is symmetrisch om 0 en strekt zich uit van $-\infty$ tot $+\infty$. De symmetrie impliceert, dat $P(\underline{u} < u) = P(\underline{u} > -u)$ voor elke u . Daar het totale oppervlak onder de grafiek 1 is, geldt bovendien $P(\underline{u} > u) = 1 - P(\underline{u} < u)$ voor elke u . Deze eigenschappen stellen ons in staat kansen te berekenen, die niet in tabel 6.1 vermeld staan; deze tabel geeft immers alleen linkse kansen $P(\underline{u} \leq u)$ voor positieve waarden van u . Zo is

$$P(\underline{u} > 1) = 1 - P(\underline{u} < 1) = 0,1587,$$

$$P(\underline{u} > 3) = 1 - P(\underline{u} < 3) = 0,0013,$$

$$P(\underline{u} < -2) = P(\underline{u} > 2) = 1 - P(\underline{u} < 2) = 0,0228,$$

$$P(-2 < \underline{u} < 2) = P(\underline{u} < 2) - P(\underline{u} < -2) = 0,9772 - 0,0228 = 0,9544.$$

De grootheid $\underline{u}$ neemt dus met een kans van ruim 0,95 waarden aan tussen -2 en +2, terwijl waarden groter dan 3(en kleiner dan -3) zeer onwaarschijnlijk zijn.

In tabel 6.1 staan bij verschillende waarden van u de kansen $P(\underline{u} \leq u)$ vermeld. Omgekeerd is het vaak belangrijk bij een gegeven kans ter grootte α de waarde van u te bepalen, zodat $P(\underline{u} \geq u) = \alpha$. We noemen dergelijke waarden van u rechtse α -punten van de standaard-normale verdeling en geven ze aan met u_α , dus

$$P(\underline{u} \geq u_\alpha) = \alpha.$$

In tabel 6.1 kunnen we, door te zoeken naar de kansen $1 - \alpha$, deze rechtse α -punten ook wel vinden; bijv. $u_{0,5} = 0$, $u_{0,05} = 1,645$, $u_{0,025} = 1,96$.

De grootheid $\underline{u}$ heeft verwachting 0 en variantie 1; de standaard-normale verdeling wordt daarom vaak aangeduid met $N(0,1)$. Er bestaat echter een hele familie van normale verdelingen, waarvan de standaard-normale er één is. De grafiek van een normaal verdeelde grootheid $\underline{x}$ met verwachting μ en variantie σ^2 ontstaat uit de standaard-normale grafiek door verschuiving over een afstand μ , vermenigvuldiging van de horizontale schaal met σ en van de verticale schaal met $1/\sigma$. Bij elke μ en σ^2 behoort dus een normale verdeling, die door deze twee grootheden wordt vastgelegd. Men noemt μ en σ^2 de parameters van de normale verdeling. Deze normale verdelingen worden vaak aangeduid met $N(\mu, \sigma^2)$ en zijn symmetrisch om het punt μ . De mediaan van een normale verdeling is dan ook gelijk aan de verwachting μ .

Het is zeer eenvoudig kansen met betrekking tot een normaal verdeelde grootheid $\underline{x}$ uit te rekenen door reductie op de $N(0,1)$ verdeling. Immers als $\underline{x} \sim N(\mu, \sigma^2)$ verdeeld is, dan is de grootheid

$$\frac{\underline{x} - \mu}{\sigma}$$

standaard-normaal verdeeld. Daarom is

$$P(\underline{x} \leq x) = P\left(\frac{\underline{x} - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(u \leq \frac{x - \mu}{\sigma}\right),$$

en deze laatste kans kan men, bij gegeven x , μ en σ , bepalen met behulp van de tabel van de $N(0,1)$ verdeling (zie tabel 6.1).

We noemen nog de volgende belangrijke eigenschappen van normaal verdeelde grootheden:

- (i) als $\underline{x}$ normaal verdeeld is en c een constante, dan zijn $c + \underline{x}$ en $c\underline{x}$ ook normaal verdeeld;
- (ii) als $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ alle normaal verdeeld zijn, dan is hun som ook normaal verdeeld.

In de statistiek spelen de normale verdelingen een centrale rol. Hiervoor zijn twee redenen aan te voeren. In de eerste plaats is de meetfout bij fysische metingen vaak normaal verdeeld. De ware (onbekende) waarde van de te meten grootheid is μ , doch door de meetfout vindt men uitkomsten, die om μ variëren. Men gaat er hierbij van uit, dat de meetfout een $N(0, \sigma^2)$ verdeling heeft, dus niet tot systematisch te hoge of te lage waarden leidt. De onderzoeker zal nu uit een aantal uitkomsten deze onbekende waarde μ willen schatten en misschien ook hypothesen omtrent de waarde van μ willen toetsen. Ook vertonen kenmerken van objecten van grote biologische populaties vaak een normale verdeling, bijv. lengten van bruine bonen of van personen van een bepaald ras. In dit geval zal de onderzoeker door steekproeftrekking informatie trachten te verkrijgen over het populatie-gemiddelde μ . Soms is ook de variantie σ^2 van belang.

De tweede reden is van andere, meer wiskundige aard. Vele stochastische grootheden, die bij de statistische analyse van belang zijn (zoals bijv. toetsingsgrootheden bij het toetsen van hypothesen), blijken nl. verdelingen te bezitten, die bij grote steekproefomvang veel gelijk zijn op normale verdelingen. Bij het berekenen van kansen maakt men

hier dan dankbaar gebruik van.

Zij $x_1, x_2, \dots, x_n$ een aselechte steekproef van onafhankelijke waarnemingen uit een $N(\mu, \sigma^2)$ verdeling met onbekende parameters μ en σ^2 . Hoe zullen we μ en σ^2 thans schatten? Het ligt voor de hand de verwachting (populatie-gemiddelde!) μ van $\underline{x}$ te schatten met het steekproef-gemiddelde

$$(6.1) \quad \bar{\underline{x}} = \frac{1}{n} \sum_{i=1}^n x_i.$$

In hoofdstuk 3 is reeds bewezen (zie formule (3.6)), dat

$$(6.2) \quad E\bar{\underline{x}} = \mu,$$

d.w.z. $\bar{\underline{x}}$ is een zuivere schatter van μ . De variantie van deze schatter is gelijk aan (zie formule (3.7))

$$(6.3) \quad \sigma^2(\bar{\underline{x}}) = \frac{\sigma^2}{n}.$$

Men zou μ ook wel op andere wijze kunnen schatten, bijv. met de middelste van de naar opklimmende grootte geordende x_i 's (bij even n is er geen middelste; men kan dan het gemiddelde van de twee middelste waarnemingen nemen). Men kan laten zien, dat dit ook een zuivere schatter van μ is. Als men tussen een aantal zuivere schatters moet kiezen, dan zal men de voorkeur geven aan die schatter, die de kleinste variantie heeft. Men kan aantonen, dat $\bar{\underline{x}}$ de zuivere schatter van μ is met de kleinste variantie, zodat $\bar{\underline{x}}$ de best mogelijke schatter van μ is. Uit (6.3) blijkt, dat de variantie van $\bar{\underline{x}}$ omgekeerd evenredig is met het aantal waarnemingen, m.a.w. de nauwkeurigheid waarmee μ wordt geschat door $\bar{\underline{x}}$ neemt toe met toenemend aantal waarnemingen. De verdeling van $\bar{\underline{x}}$ is op grond van de eigenschappen (i) en (ii) ook weer normaal en volgens bovenstaande formules dus $N(\mu, \frac{\sigma^2}{n})$. Voor een tweetal waarden van n zijn de kansverdelingen van $\bar{\underline{x}}$ geschetst in fig. 6.4, tezamen met de kansverdeling van de x_i zelf.

Als bijv. $\mu = 5$ en $\sigma^2 = 9$, dan is bij $n = 4$ de kans op een waarde van $\bar{\underline{x}}$ kleiner dan 4 gelijk aan (zie tabel 6.1)

$$P(\bar{\underline{x}} < 4) = P\left(\frac{\bar{\underline{x}} - \mu}{\sqrt{\sigma^2/n}} < \frac{4-5}{\sqrt{9/4}}\right) = P(\underline{u} < -\frac{2}{3}) = 0,252;$$

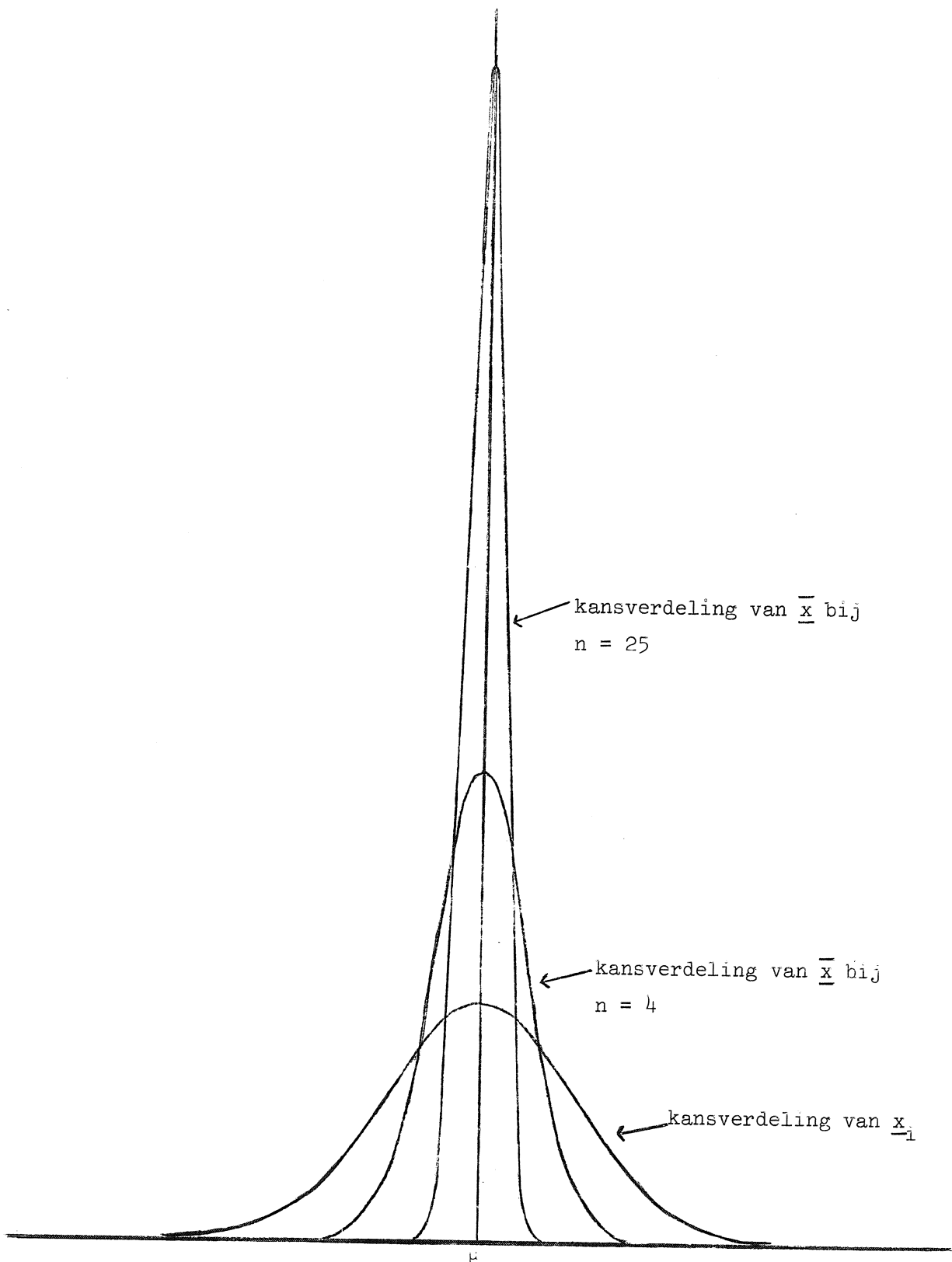


fig. 6.4. Kansverdeling van x_i , van $\bar{x}$ bij $n = 4$ en van $\bar{x}$ bij $n = 25$.

bij $n = 25$ wordt deze kans

$$P(\bar{x} < 4) = P\left(\frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} < \frac{4-5}{\sqrt{9/25}}\right) = P(u < -\frac{5}{3}) = 0,048.$$

De variantie σ^2 van $\underline{x}$ is de verwachte waarde van $(\underline{x} - \mu)^2$. Bij bekende μ is het daarom intuïtief aantrekkelijk om σ^2 te schatten met het corresponderende steekproefgemiddelde

$$\frac{1}{n} \sum_{i=1}^n (\underline{x}_i - \mu)^2.$$

Daar μ onbekend is ondersteld, zou men in deze uitdrukking μ kunnen schatten met $\bar{x}$, hetgeen leidt tot de schatter

$$(6.4) \quad \frac{1}{n} \sum_{i=1}^n (\underline{x}_i - \bar{x})^2.$$

Men kan echter laten zien, dat de kwadratische vorm

$$(6.5) \quad \underline{S}^2 = \sum_{i=1}^n (\underline{x}_i - \bar{x})^2$$

verwachting

$$(6.6) \quad E\underline{S}^2 = (n-1)\sigma^2$$

heeft, zodat (6.4) verwachting $\frac{n-1}{n} \sigma^2$ heeft en dus géén zuivere schatter van σ^2 is. Het is nu echter duidelijk, dat

$$(6.7) \quad \underline{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (\underline{x}_i - \bar{x})^2$$

wel een zuivere schatter van σ^2 is; $\underline{s}^2$ heet de steekproefvariantie. Men zegt wel, dat $\underline{S}^2$ een kwadratische vorm is met $n-1$ vrijheidsgraden. Een zuivere schatter van de variantie van de $\underline{x}_i$ ontstaat dus, als men de kwadratische vorm deelt door zijn vrijheidsgraden. De naam vrijheidsgraden wordt duidelijker, wanneer men bedenkt dat van de grootheden $x_1 - \bar{x}$, $x_2 - \bar{x}$, ..., $x_n - \bar{x}$ er slechts $n-1$ vrij gekozen kunnen worden, daar hun som nul is. Het aantal vrijheidsgraden van $\underline{S}^2$ is één kleiner dan het aantal waarnemingen, omdat in $\underline{S}^2$ de verwachtingen van de $\underline{x}_i$ (die hier alle gelijk zijn) geschat zijn door één parameter μ te schatten.

De verdeling van $\underline{s}^2$ is niet normaal, doch hangt samen met de zo-

genaamde chikwadraat-verdeling. We gaan hier verder niet op in en merken slechts op, dat de verdeling van $\underline{s}^2$ alleen van σ^2 en van het aantal vrijheidsgraden $n-1$ afhangt, doch niet van μ .

Uiteraard is $\underline{s}$, de positieve wortel van $\underline{s}^2$, een schatter van de standaardafwijking σ van x .

We merken nog op, dat steekproefgemiddelde $\bar{x}$ en steekproefvariantie $\underline{s}^2$ onafhankelijk zijn, d.w.z. de waarde van $\bar{x}$ beïnvloedt de waarde, die $\underline{s}^2$ aanneemt, niet. Deze eigenschap, die alleen geldt voor normaal verdeelde waarnemingen $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ met dezelfde verwachting en variantie, speelt in de statistiek een grote rol.

De formules (6.2) en (6.6) zijn afgeleid zonder gebruik te maken van de normaliteit van de waarnemingen. Dit betekent, dat $\bar{x}$ en $\underline{s}^2$ ook zuivere schatters zijn van μ resp. σ^2 bij niet-normaal verdeelde waarnemingen!

Schat men μ met $\bar{x}$, dan kan men zich afvragen hoe nauwkeurig deze schatting is. Teneinde hiervan een indruk te geven zou men bijv. de standaardafwijking van $\bar{x}$ kunnen vermelden, doch deze is gelijk aan $\sigma/\sqrt{n}$ en hangt dus af van de onbekende σ . Schat men σ met s , dan geeft $s/\sqrt{n}$ wel een idee van de grootte van de standaardafwijking van $\bar{x}$ en dus van de nauwkeurigheid waarmee men μ schat door $\bar{x}$. Men noemt $s/\sqrt{n}$ de standaardfout van $\bar{x}$. In publicaties geeft men deze standaardfout vaak als volgt aan:

$$\bar{x} \pm s/\sqrt{n},$$

bijv. $2,53 \pm 0,08$. Hiermee wordt dus géén betrouwbaarheidsinterval bedoeld!

Weliswaar geeft de standaardfout een zekere indruk van de nauwkeurigheid waarmee men μ schat, maar we kunnen er geen precieze uitspraken aan verbinden. Bij normaal verdeelde waarnemingen verdient het daarom de voorkeur betrouwbaarheidsgrenzen voor μ te bepalen. Hiertoe introduceren we een nieuwe familie van (continue) kansverdelingen, de t-verdelingen van STUDENT. Deze verdelingen hangen nog af van een parameter v , het aantal vrijheidsgraden van de t-verdeling ($v = 1, 2, 3, \dots$). De grafiek van een t-verdeling is symmetrisch om 0 en lijkt bij toenemend aantal vrijheidsgraden v steeds meer op de grafiek van de $N(0, 1)$

verdeling, doch de "staarten" van deze verdelingen zijn iets groter dan van de standaard-normale verdeling, zie fig. 6.5

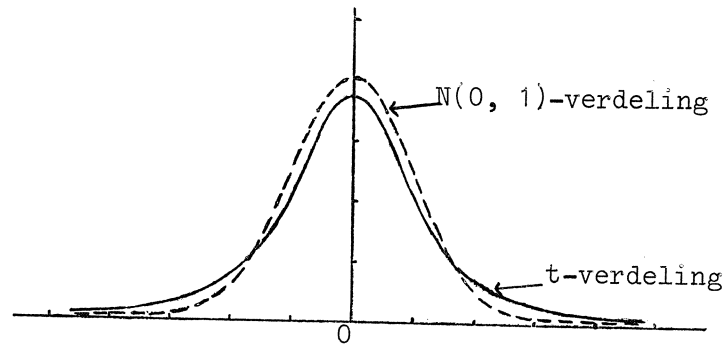


fig. 6.5 Grafieken van de $N(0, 1)$ verdeling en een t -verdeling.

Een betrouwbaarheidsbovengrens voor μ met onbetrouwbaarheid α wordt nu gegeven door

$$(6.8) \quad \bar{x} + t_{n-1;\alpha} \frac{s}{\sqrt{n}},$$

waarin $t_{n-1;\alpha}$ het rechtse α -punt is van de t -verdeling met $n-1$ vrijheidsgraden d.w.z.

$$(6.9) \quad P(t_{n-1} > t_{n-1;\alpha}) = \alpha$$

(t_{n-1} is een stochastische grootheid met een t -verdeling met $n-1$ vrijheidsgraden). Analooq is een betrouwbaarheidsondergrens voor μ met onbetrouwbaarheid α gegeven door

$$(6.10) \quad \bar{x} - t_{n-1;\alpha} \frac{s}{\sqrt{n}},$$

terwijl een tweezijdig betrouwbaarheidsinterval met onbetrouwbaarheid α wordt gegeven door

$$(6.11) \quad \bar{x} - t_{n-1;\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1;\frac{\alpha}{2}} \frac{s}{\sqrt{n}},$$

waarin $t_{n-1;\frac{\alpha}{2}}$ het rechtse $\frac{\alpha}{2}$ -punt van de t -verdeling met $n-1$ vrijheidsgraden voorstelt. Heeft men uit de gegeven steekproef $\bar{x}$ en s berekend, dan kan men het betrouwbaarheidsinterval voor μ als volgt interpreteren:

in een fractie $1 - \alpha$ van de gevallen zal de uitspraak (6.11) juist zijn (zal μ inderdaad binnen de berekende grenzen liggen), in een fractie α van de gevallen is ze onjuist. Hetzelfde geldt voor de éézijdige uitspraken $\mu < \bar{x} + t_{n-1;\alpha} s/\sqrt{n}$ resp. $\mu > \bar{x} - t_{n-1;\alpha} s/\sqrt{n}$.

Daar voor grote aantallen vrijheidsgraden de t -verdeling sterk gelijkert op de $N(0, 1)$ verdeling, kan men bij grote n ($n > 40$) $t_{n-1;\alpha}$ resp. $t_{n-1;\frac{\alpha}{2}}$ goed benaderen met de rechtse α -punten u_α resp. $u_{\frac{\alpha}{2}}$ van de standaard-normale verdeling.

Onderstel thans, dat de onderzoeker erin is geïnteresseerd de hypothese

$$H_0 : \mu \leq \mu_0$$

te toetsen met als mogelijk alternatief

$$H_1 : \mu > \mu_0,$$

waarin μ_0 een gegeven getal is. Dit kan bijv. het geval zijn, als een automobilist wil nagaan of het (verwachte) benzinegebruik μ van zijn auto inderdaad zo laag is als de fabrikant opgeeft (μ_0). We gaan er wederom van uit, dat onafhankelijke waarnemingen $x_1, x_2, \dots, x_n$ uit een $N(\mu, \sigma^2)$ verdeling ter beschikking staan. We zullen eerst de weinig realistische onderstelling maken, dat σ^2 bekend is. Het steekproefgemiddelde $\bar{x}$ is $N(\mu, \frac{\sigma^2}{n})$ verdeeld, zodat

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sqrt{n}$$

standaard-normaal verdeeld is. Als toetsingsgrootte gebruiken we nu de grootte

$$(6.12) \quad \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sqrt{n},$$

die $N(0, 1)$ verdeeld is als $\mu = \mu_0$. Toetsen we met onbetrouwbaarheid α , dan zullen we H_0 verwerpen indien (6.12) groter dan u_α is. Dit is dus een rechts-éézijdige toets met als kritieke waarde het rechtse α -punt van de $N(0, 1)$ verdeling. Als $\mu = \mu_0$, dan is de kans op een fout van de eerste soort (de kans dat we H_0 verwerpen terwijl H_0 waar is) inderdaad gelijk aan α . Als $\mu = \mu_1 < \mu_0$, dan is $E(\bar{x} - \mu_0) = \mu_1 - \mu_0 < 0$, zodat (6.12) systematisch kleinere waarden aanneemt dan

wanneer $\mu = \mu_0$. De verdeling van (6.12) is dan naar links verschoven over een afstand $\mu_0 - \mu_1$ en de kans dat we H_0 verwerpen is (gelukkig) kleiner dan α , zie fig. 6.6. Als echter $\mu = \mu_2 > \mu_0$, dan is $E(\bar{x} - \mu_0) = \mu_2 - \mu_0 > 0$, zodat (6.12) juist systematisch grotere waarden aanneemt dan wanneer $\mu = \mu_0$. De verdeling van (6.12) is dan naar rechts verschoven over een afstand $\mu_2 - \mu_0$ en de kans dat we H_0 verwerpen is groter dan α en neemt toe naarmate $\mu_2 - \mu_0$ toeneemt. Daar in dit geval H_1 waar is, komt dit juist goed uit.

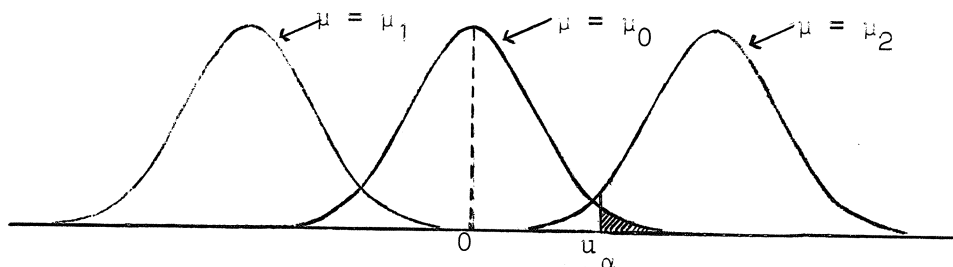


fig. 6.6 Kansverdeling van $\frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$ voor $\mu = \mu_0$,

$\mu = \mu_1 < \mu_0$ en $\mu = \mu_2 > \mu_0$. Het gearceerde oppervlak is gelijk aan α .

Het onderscheidingsvermogen van de toets, d.i. de kans waarmee we H_0 verwerpen als H_1 waar is, kunnen we in fig. 6.6 aflezen als het oppervlak rechts van u_α onder de meest rechtse grafiek. Naarmate μ_2 toeneemt, verschuift de rechtse grafiek verder naar rechts en neemt het onderscheidingsvermogen dus toe. Dit is ook het geval voor toenemende steekproefomvang n bij vaste $\mu_2 > \mu_0$.

Laten we de onderstelling, dat σ^2 bekend is, vallen, dan kunnen we de toetsingsgrootheid (6.12) niet meer toepassen, omdat σ dan onbekend is. Het ligt nu voor de hand σ te schatten met $\underline{s}$ en

$$(6.13) \quad \frac{\bar{x} - \mu_0}{\underline{s}} \sqrt{n}$$

als toetsingsgrootheid te gebruiken. Deze toetsingsgrootheid is nu echter niet meer normaal verdeeld, maar heeft, als $\mu = \mu_0$, een t-verdeling van STUDENT met $n-1$ vrijheidsgraden. We verwerpen H_0 ten gunste van H_1

als

$$(6.14) \quad \frac{\bar{x} - \mu_0}{\underline{s}} \sqrt{n} > t_{n-1; \alpha}.$$

Op dezelfde wijze als bij de toetsingsgrootheid (6.12) kan men weer laten zien, dat voor $\mu = \mu_1 < \mu_0$ de hypothese H_0 met een kans kleiner dan α wordt verworpen en voor $\mu = \mu_2 > \mu_0$ juist met een kans groter dan α . Het onderscheidingsvermogen neemt weer toe als $\mu_2 - \mu_0$ toeneemt en als, voor vaste $\mu_2 > \mu_0$, n toeneemt.

Wensen we de hypothese

$$H_0 : \mu \geq \mu_0$$

te toetsen met als alternatief

$$H_1 : \mu < \mu_0,$$

dan verwerpen we de hypothese H_0 ten gunste van H_1 indien

$$(6.15) \quad \frac{\bar{x} - \mu_0}{\underline{s}} \sqrt{n} < -t_{n-1; \alpha};$$

dit is een links-éénzijdige toets met onbetrouwbaarheid α , die overigens dezelfde gedaante heeft als de rechts-éénzijdige toets.

Wenst men tenslotte de hypothese

$$H_0 : \mu = \mu_0$$

te toetsen tegen het alternatief

$$H_1 : \mu \neq \mu_0,$$

met tweezijdige onbetrouwbaarheid α , dan toetst men tweezijdig en verworpt H_0 ten gunste van H_1 indien

$$(6.16) \quad \left| \frac{\bar{x} - \mu_0}{\underline{s}} \sqrt{n} \right| > t_{n-1; \frac{\alpha}{2}}.$$

De kritieke zone bestaat nu uit twee gedeelten: waarden van de toetsingsgrootheid (6.13) kleiner dan $-t_{n-1; \frac{\alpha}{2}}$ en groter dan $t_{n-1; \frac{\alpha}{2}}$. Indien de toetsingsgrootheid kleiner dan $-t_{n-1; \frac{\alpha}{2}}$ uitvalt zal men concluderen tot $\mu < \mu_0$; indien deze groter dan $t_{n-1; \frac{\alpha}{2}}$ uitvalt zal men de conclusie $\mu > \mu_0$ trekken.

De boven behandelde toetsen heten toetsen van STUDENT. Deze worden dikwijls toegepast in een iets andere situatie. Onderstel bijv. dat men

de neerslaghoeveelheden in Groningen en Maastricht wil vergelijken. Over een n-tal opeenvolgende maanden bepaalt men de neerslagcijfers in beide plaatsen en verkrijgt zo een n-tal gepaarde waarnemingen

$$(\underline{y}_1, \underline{z}_1), (\underline{y}_2, \underline{z}_2), \dots, (\underline{y}_n, \underline{z}_n).$$

De $\underline{y}_i$ en $\underline{z}_i$ stellen hier de maandcijfers van de i-de onderzochte maand in Groningen resp. Maastricht voor. Men zou deze paren kunnen beschouwen als een steekproef uit de populatie van alle mogelijke maandcijfers in de loop der jaren. Nemen we aan, dat deze waarnemingen alle onafhankelijk en normaal verdeeld zijn met verwachtingen

$$E\underline{y}_i = \mu_i, \quad E\underline{z}_i = \nu_i,$$

dan zijn de verschillen

$$\underline{x}_1 = \underline{y}_1 - \underline{z}_1, \underline{x}_2 = \underline{y}_2 - \underline{z}_2, \dots, \underline{x}_n = \underline{y}_n - \underline{z}_n$$

ook onafhankelijk en normaal verdeeld met verwachtingen

$$E\underline{x}_i = \mu_i - \nu_i.$$

Indien de $\underline{x}_i$ alle dezelfde variantie hebben (hetgeen bijv. het geval is als alle $\underline{y}_i$ dezelfde variantie hebben en alle $\underline{z}_i$ ook), dan kan men de hypothese

$$H_0 : \mu_i - \nu_i = 0 \text{ (resp. } \leq 0 \text{ resp. } \geq 0) \text{ voor alle } i$$

toetsen met de boven beschreven toetsen van STUDENT. In dit geval is $\mu_0 = 0$ en de toetsingsgrootte is dus

$$\frac{\bar{\underline{x}}}{\underline{s}} \sqrt{n}.$$

Het is van grote betekenis, dat de μ_i , evenals de ν_i , onderling best mogen verschillen; niveauverschillen tussen de paren verstoren de toets niet (we letten immers alleen op de verschillen $\mu_i - \nu_i$). In het genoemde voorbeeld treden dergelijke niveauverschillen inderdaad op, daar zich in de loop van een jaar systematische neerslagschommelingen voordoen.

In vele praktische situaties zullen de waarnemingen niet normaal verdeeld zijn, zodat de toetsen van STUDENT niet mogen worden toegepast. Men kan dan echter gebruik maken van verdelingsvrije toetsen, waarbij slechts zeer weinig onderstellingen omtrent de verdeling van de waarnemingen worden gemaakt. In hoofdstuk 5 bespraken we reeds de tekentoets.

Veelal verdient het echter de voorkeur de symmetrietoets van WILCOXON (Engels: signed rank test) toe te passen bij het toetsen van hypothesen omtrent verwachtingen, omdat deze toets een groter onderscheidingsvermogen heeft dan de tekentoets. Door hun algemene toepasbaarheid spelen verdelingsvrije toetsen een grote rol bij het statistisch onderzoek van experimenten.

Hoofdstuk 7. Lineaire Regressie.

In vele praktische situaties is men geïnteresseerd in de invloed van een variabele op een andere variabele. We beschouwen hier het geval dat y een stochastische grootheid is, waarvan de verdeling afhangt van een andere, niet-stochastische grootheid x . Zo kan y de lengte van de remweg van een auto voorstellen bij een bepaalde beginsnelheid x , of y is de levensduur van een gloeilamp bij een aangebrachte spanning x . In deze gevallen kan men de waarde van x zelf kiezen, terwijl y bij deze waarde van x vervolgens wordt waargenomen. Bij vele ijkingsproblemen is dit model eveneens van betekenis; x stelt dan de meting van een grootheid voor volgens een zeer nauwkeurige (maar tijdrovende of kostbare) methode, terwijl y een meting van dezelfde grootheid is met een grove methode, waarbij een meetfout optreedt. In al deze gevallen noemt men x de onafhankelijke variabele en y de (van x) afhankelijke variabele.

De vraag rijst nu, hoe de kansverdeling van y afhangt van x . We zullen alleen het eenvoudigste geval bespreken, nl. dat de verwachting van y lineair afhangt van x :

$$(7.1) \quad E(y|x) = \alpha + \beta x, \quad *)$$

waarin α en β onbekende grootheden zijn, en de variantie σ^2 van y niet van x afhangt (dus steeds even groot is). We spreken van lineaire regressie van y op x . We hebben hier dus te maken met drie onbekenden, de parameters α , β (die men de regressiecoëfficiënten noemt) en σ^2 . Op grond van verkregen waarnemingen kan men deze parameters schatten of hypothesen omtrent deze parameters toetsen.

De waarnemingen bestaan thans uit paren:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n);$$

bij iedere beschouwde waarde x_i van x behoort een waarneming y_i . We

*) Met $E(y|x)$ bedoelen we de verwachting van y bij gegeven waarde x .

zullen steeds onderstellen, dat de y_i onafhankelijk zijn. De waarnemings-uitkomsten kan men grafisch uitzetten in een strooiingsdiagram, zoals bijv. in fig. 7.1 is geschied. In het geval van lineaire regressie zal de zo verkregen puntenwolk een duidelijk lineair verloop vertonen.

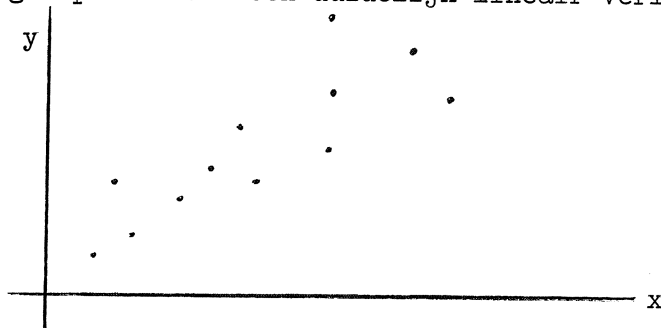


fig. 7.1 Strooiingsdiagram.

Teneinde de onbekende parameters te schatten, zoeken we de rechte lijn (de regressielijn van y op x), die het beste bij de puntenparen in de grafiek past. De aangewezen methode om dit te bereiken is de methode der kleinste kwadraten, die we thans zullen bespreken. Zij

$$y = a + bx$$

de vergelijking van een rechte lijn in het strooiingsdiagram. Men noemt de verticale afstand van het i -de waargenomen punt tot deze rechte lijn,

$$(7.2) \quad r_i = y_i - a - bx_i,$$

een residu. We zullen nu a en b zo trachten te bepalen, dat de kwadraatsom van de residuen

$$(7.3) \quad Q = \sum_{i=1}^n r_i^2$$

minimaal is (zie fig. 7.2).

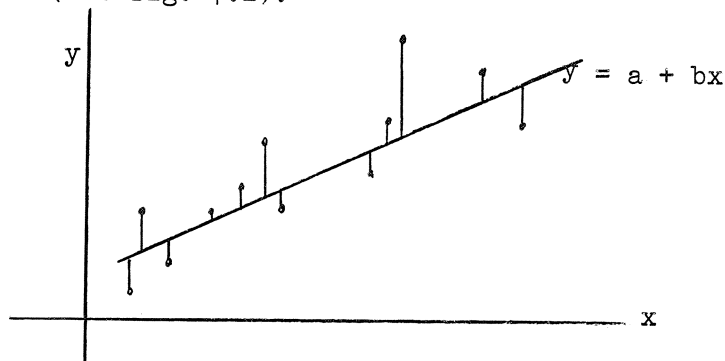


fig. 7.2 Strooiingsdiagram met aan puntenwolk aangepaste rechte lijn en residuen.

Zij $\bar{x}$ het gemiddelde der x_i en $a' = a + b\bar{x}$ (zodat $a = a' - b\bar{x}$). Dan kunnen we Q ook als volgt schrijven:

$$\begin{aligned} Q &= \sum_{i=1}^n \{y_i - a' - b(x_i - \bar{x})\}^2 = \\ (7.4) \quad &= \sum_{i=1}^n (y_i - a')^2 + b^2 \sum_{i=1}^n (x_i - \bar{x})^2 - 2b \sum_{i=1}^n y_i (x_i - \bar{x}), \end{aligned}$$

waarbij we gebruik hebben gemaakt van de relatie $\sum_{i=1}^n (x_i - \bar{x}) = 0$, die we nog vaak zullen toepassen. Q bestaat dus uit twee stukken, waarvan het ene alleen a' en het andere alleen b bevat, zodat deze afzonderlijk geminimaliseerd kunnen worden. De eerste term is minimaal, als men voor a' het gemiddelde der y_i kiest, d.w.z.

$$(7.5) \quad a = \bar{y} - b\bar{x}.$$

De laatste twee termen van Q zijn minimaal, als men voor b kiest

$$(7.6) \quad b = \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Men kan deze resultaten eenvoudig verifiëren door naar a' resp. b te differentiëren en de afgeleiden nul te stellen.

Kunnen de aldus gevonden waarden a en b nu ook dienst doen als schattingen van de parameters α en β uit (7.1)? We merken allereerst op, dat a en b stochastische grootheden zijn, immers beiden hangen af van de stochastische waarnemingen y_i (bij bovenstaande afleiding zijn we van gegeven waarnemingsuitkomsten uitgegaan en hebben daarom niet onderstreept.). Vroeger zijn zuiverheid en een kleine variantie reeds als gunstige eigenschappen van schatters ter sprake gekomen. We zullen thans laten zien, dat $\underline{a}$ en $\underline{b}$ zuivere schatters zijn van α resp. β .

Immers

$$\begin{aligned} E\left\{\sum_{i=1}^n y_i (x_i - \bar{x})\right\} &= \sum_{i=1}^n (x_i - \bar{x}) E y_i = \sum_{i=1}^n (x_i - \bar{x}) (\alpha + \beta x_i) = \\ &= \beta \sum_{i=1}^n (x_i - \bar{x}) x_i = \beta \sum_{i=1}^n (x_i - \bar{x})^2, \end{aligned}$$

zodat

$$E\underline{b} = \beta,$$

en anderzijds

$$\begin{aligned} E\underline{a} &= E(\underline{\bar{y}} - \underline{b}\bar{x}) = E\left(\frac{1}{n} \sum_{i=1}^n y_i\right) - \beta\bar{x} = \frac{1}{n} \sum_{i=1}^n E y_i - \beta\bar{x} = \\ &= \frac{1}{n} \sum_{i=1}^n (\alpha + \beta x_i) - \beta\bar{x} = \alpha + \frac{\beta}{n} \sum_{i=1}^n x_i - \beta\bar{x} = \alpha. \end{aligned}$$

Zonder bewijs vermelden we, dat a en b bovendien onder alle zuivere schatters, die lineair zijn in de y_i , de kleinste variantie hebben.

De kleinste kwadraten methode leidt dus tot schatters van α en β met zeer gunstige eigenschappen. We zoeken nu nog een schatter van de derde onbekende parameter, σ^2 . Hiertoe substitueren we in de definitie van de kwadraatsom Q van residuen, zie (7.2) en (7.3), de kleinste kwadraten schatters a en b. De aldus ontstane grootte noemt men gewoonlijk de residuele kwadraatsom of "error sum of squares" en wordt vaak aangeduid met $\underline{SS}_e$:

$$(7.7) \quad \underline{SS}_e = \sum_{i=1}^n (y_i - \underline{a} - \underline{b}x_i)^2.$$

Men kan laten zien, dat

$$(7.8) \quad E(\underline{SS}_e) = (n - 2)\sigma^2,$$

zodat

$$\underline{SS}_e / (n - 2)$$

een zuivere schatter is van σ^2 . Hier is $n - 2$ het aantal vrijheidsgraden van de kwadraatsom $\underline{SS}_e$. Intuïtief is dit resultaat wel te begrijpen. Bij bekende verwachtingen μ_i van de y_i zou men de variantie σ^2 immers kunnen schatten met het gemiddelde

$$\frac{1}{n} \sum_{i=1}^n (y_i - \mu_i)^2.$$

Deze μ_i zijn in het regressie-model gelijk aan $\alpha + \beta x_i$ en kunnen we schatten met a + b x_i . Doen we dit, dan ontstaat juist $\underline{SS}_e/n$. Daar we hier voor het schatten van de μ_i twee parameters α en β hebben geschat, lijkt het plausibel, dat het aantal vrijheidsgraden twee kleiner is dan het aantal waarnemingen. Dit is inderdaad het geval en we moeten $\underline{SS}_e$ dus door $n - 2$ delen om een zuivere schatter te verkrijgen van σ^2 (vergelijk de opmerking na formule (6.7) in hoofdstuk 6).

De geschatte regressielijn

$$(7.9) \quad Y = \underline{a} + \underline{b}x$$

is een schatting van de ware regressielijn (7.1).*) Deze regressielijn stelt ons in staat bij gegeven waarde van x de bijbehorende waarde van y te "voorspellen". Eigenlijk voorspellen we aldus de verwachte waarde van y . De waarden, die y van geval tot geval aanneemt, zullen om de verwachte waarde (dus om de ware regressielijn) heen schommelen met variantie σ^2 . Naarmate σ^2 kleiner is (hetgeen we kunnen beoordelen aan de hand van $\frac{SS_e}{(n-2)}$ of zonder formules ook aan de mate van concentratie van de puntenwolk om de geschatte lijn), zullen we uitkomsten y dus beter kunnen voorspellen. Dit voorspellen van y bij gegeven x is een van de belangrijkste toepassingen van regressie. Bij ijkingsproblemen doet zich echter de omgekeerde situatie voor; uit een waargenomen y , de uitkomst van een grove meting met een meetfout, zou men graag de juiste waarde x willen afleiden. De geschatte regressielijn stelt ons in staat bij elke waarde van y de bijbehorende x te schatten.

Tot nu toe hebben we over de precieze verdeling van de y_i geen onderstellingen gemaakt. Bij het schatten van de parameters is dit ook overbodig. Zijn de y_i normaal verdeeld, dan kan men ook betrouwbaarheidsintervallen voor α en β , voorspellingsintervallen voor een toekomstige waarde van y bij gegeven x (waarbinnen de waarde van y met een bepaalde kans zal liggen) en betrouwbaarheidsintervallen voor x bij een uitkomst y construeren. Tevens kunnen dan hypothesen omtrent de waarden van α en β worden getoetst. We gaan hier verder niet op in.

In het begin van dit hoofdstuk hebben we ondersteld, dat de regressie van y op x lineair is. De schatters van α en β zijn hier ook op gebaseerd. In vele situaties is het a priori echter geenszins zeker, dat de regressie lineair is. Aan de hand van een strooiingsdiagram kan men met enige ervaring vaak wel beoordelen of dit een redelijke onderstelling is, maar men dient dan wel over een goed "timmermansoog" te be-

*) We merken nog op, dat het punt $(\bar{x}, \bar{y})$ steeds op de geschatte regressielijn ligt.

schikken. Als men bij elke x_i niet één, maar meerdere waarnemingen y_i verricht en deze waarnemingen normaal verdeeld zijn, dan kan men de hypothese van lineaire regressie bovendien ook toetsen.

We besluiten deze korte beschouwing over lineaire regressie met een waarschuwing. Ook al blijkt een rechte lijn heel aardig te passen bij de waargenomen puntenparen, dan nog heeft men geen enkele garantie, dat deze regressielijn ook dienst kan doen voor waarden van x , die ver buiten het gebied liggen waarbinnen men bij de uitgevoerde experimenten x heeft gekozen. Men denke in dit verband aan regressielijnen van geboortecijfers of verkoopcijfers van auto's op de tijd, die soms gehanteerd worden om prognoses te maken voor tijdstippen in een ver verwijderde toekomst. De enorme fouten, die met deze extrapolatie gemaakt zijn, spreken wel voor zichzelf.

Hoofdstuk 8. Variantie analyse.

Onder de naam variantie analyse vat men een groot aantal statistische methoden samen, die alle gemeen hebben, dat men bij de statistische analyse gebruik maakt van kwadratische vormen in de waarnemingen. We bespreken eerst een eenvoudig voorbeeld.

Onderstel dat men de korreldiameter van dekzanden in een I-tal plaatsen in Nederland wil vergelijken. Hiertoe kan men uit elk der dekzanden een aselechte steekproef van korrels trachten te verzamelen en hun diameter bepalen. Voor elk dekzand vormen de korreldiameters een populatie met een kansverdeling. Noemen we de verwachte korreldiameter van de i-de populatie (het i-de dekzand) μ_i , dan kunnen we de waargenomen diameter y_{ij} van de j-de aselekt getrokken korrel uit deze populatie schematisch weergeven met

$$(8.1) \quad y_{ij} = \mu_i + e_{ij}.$$

Hierin stelt e_{ij} dus de afwijking van de j-de waarneming van het i-de populatiegemiddelde μ_i voor. In deze e_{ij} zijn ook mogelijke meetfouten bij de bepaling van de diameter opgenomen. Indien deze meetfouten geen systematisch effect hebben (hetgeen we steeds zullen onderstellen), is

$$(8.2) \quad Ee_{ij} = 0 \quad \text{voor alle } i \text{ en } j.$$

We zullen bovendien steeds aannemen dat alle e_{ij} onafhankelijk zijn. Zijn we geïnteresseerd in mogelijke verschillen tussen de populatiegemiddelden, dan zal onze aandacht zich richten op de parameters $\mu_1, \mu_2, \dots, \mu_I$. Het is niet moeilijk om de verschillende μ_i te schatten. Indien we de korreldiameter bepalen van n_i korrels in het i-de dekzand ($i = 1, 2, \dots, I$), beschikken we over de waarnemingen

$$\begin{array}{l} y_{11}, y_{12}, \dots, y_{1n_1}, \\ y_{21}, y_{22}, \dots, y_{2n_2}, \\ \vdots \\ y_{I1}, y_{I2}, \dots, y_{In_I}. \end{array}$$

Zoals gebruikelijk, kan men elk der μ_i dan schatten met het corresponderende steekproefgemiddelde

$$(8.3) \quad \bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad *)$$

We zijn hier echter niet zozeer geïnteresseerd in schatters van de μ_i , als wel in toetsing van de hypothese

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_I$$

met als mogelijk alternatief

$$H_1 : \text{niet alle } \mu_i \text{ zijn gelijk.}$$

Teneinde de hypothese H_0 te toetsen zullen we aannemen, dat alle y_{ij} normaal verdeeld zijn met dezelfde (onbekende) variantie σ^2 . In dat geval zijn alle e_{ij} normaal $N(0, \sigma^2)$ verdeeld. In praktische situaties zal men steeds moeten nagaan of deze onderstelling gemaakt mag worden!

Indien we slechts met twee dekzanden te maken hebben ($I = 2$), kan men een variant van de in hoofdstuk 7 behandelde toets van STUDENT toepassen (de zg. 2-steekproeven toets van STUDENT) om H_0 te toetsen. Als echter $I > 2$ is, kunnen we deze toets niet gebruiken en zoeken we naar een andere toetsingsmethode (die overigens ook bij $I = 2$ toegepast kan worden). Hiertoe beschouwen we de kwadraatsom

$$(8.4) \quad \sum_{i=1}^I \sum_{j=1}^{n_i} (\bar{y}_{ij} - \bar{y}_{..})^2,$$

waarin $\bar{y}_{..}$ het algemeen gemiddelde van alle waarnemingen voorstelt:

$$(8.5) \quad \bar{y}_{..} = \frac{1}{N} \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} \quad \text{met } N = \sum_{i=1}^I n_i.$$

De vorm (8.4) is een kwadraatsom met $N-1$ vrijheidsgraden (één minder dan het aantal der $\bar{y}_{ij} - \bar{y}_{..}$, omdat hun som nul is) en beschrijft de variatie van de waarnemingen om hun algemeen gemiddelde. We noemen (8.4) de "totale kwadraatsom" en duiden deze aan met SS_{totaal} ^{***)}. Deze kwa-

*) Met een punt op de plaats van een index geven we het gemiddelde over de betreffende index aan.

***) SS is de afkorting van het Engelse "sum of squares".

draatsom kan men als volgt uitsplitsen:

$$\begin{aligned}
 \sum_{i=1}^I \sum_{j=1}^{n_i} (\underline{y}_{ij} - \underline{y}_{..})^2 &= \sum_{i=1}^I \sum_{j=1}^{n_i} (\underline{y}_{ij} - \underline{y}_{i.} + \underline{y}_{i.} - \underline{y}_{..})^2 = \\
 &= \sum_{i=1}^I \sum_{j=1}^{n_i} \left[(\underline{y}_{ij} - \underline{y}_{i.})^2 + (\underline{y}_{i.} - \underline{y}_{..})^2 + \right. \\
 &\quad \left. + 2(\underline{y}_{ij} - \underline{y}_{i.})(\underline{y}_{i.} - \underline{y}_{..}) \right] = \\
 (8.6) \quad &= \sum_{i=1}^I \sum_{j=1}^{n_i} (\underline{y}_{ij} - \underline{y}_{i.})^2 + \sum_{i=1}^I n_i (\underline{y}_{i.} - \underline{y}_{..})^2.
 \end{aligned}$$

De eerste van de beide kwadratische vormen in (8.6) heet de "binnen-steekproef kwadraatsom", aangeduid met $\underline{SS}_{\text{binnen}}$; de tweede heet de "tussen-steekproef kwadraatsom", aangeduid met $\underline{SS}_{\text{tussen}}$. Uit de onafhankelijkheid van steekproefgemiddelde en steekproefvariantie (zie hoofdstuk 6) kan men afleiden, dat de beide grootheden onafhankelijk zijn; we zullen ze thans wat nader bekijken.

Beschouw eerst $\underline{SS}_{\text{binnen}}$. Deze grootheid is zelf weer een som van een aantal onafhankelijke kwadraatsommen.

$$\underline{S}_i^2 = \sum_{j=1}^{n_i} (\underline{y}_{ij} - \underline{y}_{i.})^2 \quad (\text{vaste } i)$$

met $n_i - 1$ vrijheidsgraden (de verwachting μ_i geschat met steekproefgemiddelde $\underline{y}_{i.}$ van de steekproef uit de i -de populatie). Deze sommen $\underline{S}_i^2$ zijn vergelijkbaar met de grootheid $\underline{S}^2$ uit formule (6.5) in hoofdstuk 6, die we gebruikten bij een steekproef uit één populatie om de variantie te schatten. In ons geval staan I steekproeven ter beschikking met waarnemingen, die alle dezelfde variantie σ^2 hebben. Het ligt daarom voor de hand thans σ^2 te schatten op grond van informatie uit alle I steekproeven gezamenlijk. Daar $E\underline{S}_i^2 = (n_i - 1)\sigma^2$ (zie formule 6.6), is

$$(8.7) \quad E(\underline{SS}_{\text{binnen}}) = E\left(\sum_{i=1}^I \underline{S}_i^2\right) = \sum_{i=1}^I (n_i - 1)\sigma^2 = (N - I)\sigma^2,$$

zodat $\underline{SS}_{\text{binnen}}/(N - I)$ een zuivere schatter is van σ^2 , gebaseerd op alle I steekproeven. Men kan $N - I$ weer het aantal vrijheidsgraden van

$\underline{SS}_{\text{binnen}}$ noemen. Daar de kansverdeling van S_1^2 niet afhangt van μ_1 , hangt de verdeling van $\underline{SS}_{\text{binnen}}$ niet af van $\mu_1, \mu_2, \dots, \mu_I$, d.w.z. $\underline{SS}_{\text{binnen}}$ heeft onder de hypothesen H_0 en H_1 dezelfde verdeling.

De grootte $\underline{SS}_{\text{tussen}}$ is een kwadratische vorm met $I - 1$ vrijheidsgraden, die aangeeft hoe de steekproefgemiddelden $\underline{y}_i$ om het algemeen gemiddelde $\underline{y}_{..}$ variëren. De verdeling van deze grootte hangt wél af van de hypothese. Dit is intuïtief ook wel in te zien; immers als de μ_i 's onderling veel verschillen, zal dit ook wel het geval zijn met de $\underline{y}_i$ (schatters van de μ_i) en zal $\underline{SS}_{\text{tussen}}$ relatief grote waarden aannemen. Dit blijkt ook uit de verwachte waarde van $\underline{SS}_{\text{tussen}}$:

$$(8.8) \quad E(\underline{SS}_{\text{tussen}}) = (I - 1)\sigma^2 + \sum_{i=1}^I n_i(\mu_i - \mu)^2,$$

waarin μ het gemiddelde der μ_i voorstelt. Onder H_0 is de tweede term in het rechterlid van (8.8) nul en is $\underline{SS}_{\text{tussen}}/(I - 1)$ dus ook een zuivere schatter van σ^2 . Als echter H_1 waar is, dan is deze term positief en wel groter naarmate de μ_i verder uit elkaar liggen en het aantal waarnemingen n_i toeneemt.

Het ligt nu voor de hand de hypothese H_0 te toetsen met behulp van de toetsingsgrootte

$$(8.9) \quad \frac{\underline{SS}_{\text{tussen}}/(I - 1)}{\underline{SS}_{\text{binnen}}/(N - I)};$$

teller en noemer zijn onder H_0 beide zuivere en onafhankelijke schatters van σ^2 , terwijl onder H_1 de teller en dus ook de toetsingsgrootte systematisch grotere waarden zal aannemen dan onder H_0 . Men zal dus een rechts-éénzijdige toets moeten toepassen, d.w.z. H_0 verwerpen bij grote waarden van de toetsingsgrootte (8.9). Om de kritieke waarde te bepalen bij gegeven onbetrouwbaarheid α , dienen we de verdeling van (8.9) onder de nulhypothese te kennen. Hiertoe introduceren we een nieuwe familie van verdelingen, de zg. F-verdelingen. Deze hangen af van twee vrijheidsgraden, v_1 en v_2 . Rechtse α -punten van deze verdelingen zijn voor vele waarden van v_1 en v_2 en diverse waarden van α getabelleerd. Men kan nu laten zien, dat de toetsingsgrootte (8.9) onder H_0 een F-verdeling heeft met $v_1 = I - 1$ en $v_2 = N - I$ vrijheidsgraden. Stelt $F_{I-1, N-I; \alpha}$ het rechtse α -punt voor van de betreffende F-verdeling, dan

wordt H_0 dus verworpen als

$$(8.10) \quad \frac{SS_{\text{tussen}}/(I - 1)}{SS_{\text{binnen}}/(N - I)} \geq F_{I-1, N-I; \alpha}.$$

Men trekt dan de conclusie, dat niet alle μ_i gelijk zijn. Deze conclusie is nog betrekkelijk onbevredigend, omdat ze ons niet in staat stelt te zeggen welke μ_i verschillend zijn en ons evenmin een indicatie geeft of een μ_i nu groter of kleiner dan een andere μ_i is. Er bestaan echter statistische methoden, die ons in staat stellen de μ_i aan een nader onderzoek te onderwerpen. Deze zal men gewoonlijk toepassen, als men met bovenstaande "F-toets" de hypothese H_0 heeft verworpen.

De hier aan de hand van een voorbeeld beschreven statistische analyse noemt men een één-weg-klassificatie of één-factor variantie analyse. Indien men niet overtuigd is van de normaliteit van de waarnemingen, kan men beter een verdelingsvrije toets toepassen. In het geval $I = 2$ komt hiervoor de 2-steekproeven toets van WILCOXON (verdelingsvrije variant van de eerder genoemde 2-steekproeven toets van STUDENT) in aanmerking; als $I > 2$ is kan men de toets van KRUSKAL-WALLIS toepassen. We volstaan hier met het noemen van deze veel gebruikte verdelingsvrije toetsen.

Het model (8.1), dat als uitgangspunt diende van de statistische analyse, kan men ook op enigszins andere wijze formuleren:

$$(8.11) \quad y_{ij} = \mu + \alpha_i + e_{ij}.$$

Hierin stelt μ het gemiddelde van de μ_i voor en $\alpha_i = \mu_i - \mu$ de afwijking van μ_i van dit gemiddelde. De α_i voldoen dan aan de "randvoorwaarde"

$$(8.12) \quad \sum_{i=1}^I \alpha_i = 0.$$

Deze notatie is in de variantie analyse gebruikelijk en komt op hetzelfde neer als (8.1), immers $\mu + \alpha_i = \mu_i$. De hypothese H_0 houdt nu in, dat alle α_i nul zijn. De statistische analyse blijft overigens ongewijzigd.

We hebben deze notatie hier genoemd met het oog op de bespreking van een iets gecompliceerdere situatie, waar niet één factor (de verschillende dekzanden in ons voorbeeld), maar twee factoren een syste-

matisch effect op de waarnemingen kunnen hebben. De invloed van beide factoren willen we hierbij tegelijkertijd onderzoeken. We spreken thans van een twee-weg-klassificatie of een twee-factor variantie analyse. Deze situatie doet zich bijv. voor, als de opbrengst van verschillende graanrassen wordt vergeleken bij verschillende bemestingen of als de poreusiteit van een aantal zandstenen met verschillende methoden wordt geanalyseerd. Elke beschouwde factor wordt gevarieerd op een aantal "niveau's" (de verschillende graanrassen en bemestingen in ons eerste voorbeeld). Een combinatie van een niveau van de eerste factor met een niveau van de tweede factor noemt men een cel. Deze cellen kan men in een rechthoekig schema weergeven, zie fig. 8.1.

factor A factor B		1	2	...	I
1					
2					
.					
.					
.					
J					

fig. 8.1 Schema van een twee-weg-klassificatie;
factor A op I niveau's, factor B op J niveau's.

Indien we het effect van beide factoren op de waar te nemen grootheid willen onderzoeken, is het niet noodzakelijk in elke cel een gelijk aantal waarnemingen te verrichten (dit was in de één-weg-klassificatie evenmin het geval), noch is het altijd wenselijk in elke cel tenminste éénmaal een waarneming te doen. Dit laatste is ook niet altijd mogelijk, bijv. als we van een viertal bandenmerken de bandenslijtage willen onderzoeken met een aantal motorfietsen. Op elke motorfiets kunnen immers slechts twee banden worden gemonteerd, zodat niet elke combinatie van motorfiets en bandenmerk in het proefschema kan worden opgenomen. Men spreekt dan van onvolledige proefschema's. De theorie van de proefopzetten houdt zich bezig met de optimale verdeling van de waarnemingen

over de cellen in de gegeven omstandigheden en vormt een essentieel onderdeel van de variantie analyse.

We zullen hier slechts een volledig proefschema bespreken met een gelijk aantal onafhankelijke waarnemingen K per cel. We onderstellen dat de factor A over I en de factor B over J niveau's wordt gevarieerd. We nemen tevens aan, dat alle waarnemingen dezelfde variantie σ^2 bezitten. Zij y_{ijk} de k -de waarneming bij factor A op niveau i en factor B op niveau j . Als y_{ijk} verwachting μ_{ij} heeft, kunnen we deze waarneming schrijven als

$$(8.13) \quad y_{ijk} = \mu_{ij} + e_{ijk} \quad (\text{alle } i, j \text{ en } k)$$

met

$$(8.14) \quad E e_{ijk} = 0, \quad \sigma^2(e_{ijk}) = \sigma^2 \quad (\text{alle } i, j \text{ en } k).$$

We willen nu meer expliciet tot uitdrukking brengen, dat de verwachting μ_{ij} van de niveau's van de factoren A en B afhangt. Definieren we μ als de gemiddelde verwachting

$$\mu = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \mu_{ij},$$

dan zullen we willen aangeven in hoeverre μ_{ij} van μ afwijkt. Het volgende model kan hiervoor dienst doen:

$$(8.15) \quad \mu_{ij} = \mu + \alpha_i + \beta_j,$$

waarin de α_i en de β_j voldoen aan de randvoorwaarden

$$(8.16) \quad \sum_{i=1}^I \alpha_i = 0, \quad \sum_{j=1}^J \beta_j = 0.$$

Dan stelt α_i het effect voor van het i -de niveau van de factor A in afwijking van de gemiddelde verwachting en β_j het effect van het j -de niveau van de factor B in afwijking van de gemiddelde verwachting. De parameters α_i en β_j heten de hoofdeffecten van de factoren A en B . Men noemt dit een additief model, omdat we het effect van een niveau van A en het effect van een niveau van B optellen om het effect te verkrijgen van het samengaan van beide niveau's. Dit is niet altijd realistisch; soms geeft het samengaan van twee bepaalde omstandigheden een extra effect in positieve of negatieve zin. Men spreekt dan van inter-

actie van de factoren A en B en kan het model hiervoor aanpassen.

We bespreken het additieve model thans wat nader. De waarnemingen kunnen geschreven worden als

$$(8.17) \quad y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk},$$

waarin e_{ijk} voldoet aan (8.14) en de parameters α_i en β_j aan (8.16). Het schatten van de onbekende parameters is eenvoudig. Het ligt al zeer voor de hand om μ te schatten met het algemeen gemiddelde van alle waarnemingen

$$(8.18) \quad \bar{y}_{...} = \frac{1}{IJK} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K y_{ijk}.$$

Wederom de punt-notatie gebruikend definiëren we

$$\bar{y}_{ij.} = \frac{1}{K} \sum_{k=1}^K y_{ijk}, \quad \bar{y}_{i..} = \frac{1}{J} \sum_{j=1}^J \bar{y}_{ij.}, \quad \bar{y}_{.j.} = \frac{1}{I} \sum_{i=1}^I \bar{y}_{ij.}.$$

De parameters α_i schatten we met

$$(8.19) \quad \bar{y}_{i..} - \bar{y}_{...},$$

de parameters β_j met

$$(8.20) \quad \bar{y}_{.j.} - \bar{y}_{...}.$$

Ook deze schatters spreken aan, indien men aan de betekenis van de hoofdeffecten α_i en β_j denkt.

Het hoofddoel van een variantie analyse is echter gewoonlijk het toetsen van hypothesen omtrent de hoofdeffecten. De belangrijkste hypothesen zijn:

$$H_A : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0,$$

m.a.w. het variëren van de niveau's van de factor A heeft geen effect op de waarnemingen, en

$$H_B : \beta_1 = \beta_2 = \dots = \beta_J = 0,$$

m.a.w. het variëren van de niveau's van de factor B heeft geen effect. In ons eerste voorbeeld zou de hypothese H_A betekenen, dat tussen de opbrengsten van de verschillende graanrassen geen systematische verschillen bestaan, en de hypothese H_B , dat de toegepaste bemestingen hetzelfde effect op de opbrengsten hebben.

Teneinde deze hypothesen te toetsen maken we de extra onderstelling, dat de y_{ijk} alle normaal verdeeld zijn. We gaan op dezelfde wijze te werk als bij de één-weg-klassificatie en splitsen de totale kwadraat-som

$$(8.21) \quad \underline{SS}_{\text{totaal}} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - \bar{y}_{...})^2$$

wederom uit. Gemakshalve laten we daarbij de sommatiegrenzen weg.

$$\begin{aligned} \underline{SS}_{\text{totaal}} &= \sum_i \sum_j \sum_k \left[(y_{ijk} - \bar{y}_{ij.}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) \right]^2 = \\ &= \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2 + K \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \\ (8.22) \quad &+ KJ \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + KI \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2. \end{aligned}$$

Men kan weer laten zien, dat de vier kwadraatsommen in (8.22) onafhankelijk zijn. De "binnen-cel kwadraatsom"

$$(8.23) \quad \underline{SS}_{\text{binnen}} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - \bar{y}_{ij.})^2$$

is zelf weer de som van IJ kwadraatsommen van de vorm

$$S_{ij}^2 = \sum_{k=1}^K (y_{ijk} - \bar{y}_{ij.})^2 \quad (\text{vaste } i \text{ en } j)$$

elk met $K-1$ vrijheidsgraden (de verwachting μ_{ij} van waarnemingen in de (i, j) -de cel geschat met $\bar{y}_{ij.}$). Net als bij de één-weg-klassificatie kunnen we uit de som der S_{ij}^2 weer een schatter afleiden van de variantie σ^2 , gebaseerd op de steekproefjes in alle IJ cellen gezamenlijk. Immers

$$(8.24) \quad E(\underline{SS}_{\text{binnen}}) = E\left(\sum_{i=1}^I \sum_{j=1}^J S_{ij}^2\right) = \sum_{i=1}^I \sum_{j=1}^J (K-1)\sigma^2 = IJ(K-1)\sigma^2.$$

Delen we $\underline{SS}_{\text{binnen}}$ dus door $IJ(K-1)$, het aantal vrijheidsgraden van $\underline{SS}_{\text{binnen}}$, dan ontstaat een zuivere schatter van σ^2 . De verdeling van $\underline{SS}_{\text{binnen}}$ hangt niet af van de hypothesen H_A of H_B , noch van de model-onderstelling (8.15), daar we elke μ_{ij} afzonderlijk hebben geschat met $\bar{y}_{ij.}$.

De grootheid

$$(8.25) \quad \underline{SS}_A = JK \sum_{i=1}^I (\underline{y}_{i..} - \underline{y}_{...})^2$$

is een kwadraatsom met $I-1$ vrijheidsgraden, waarvan de verdeling afhangt van de hypothese H_A . Dit blijkt ook uit het feit, dat

$$(8.26) \quad E(\underline{SS}_A) = (I - 1)\sigma^2 + JK \sum_{i=1}^I \alpha_i^2.$$

Als H_A waar is, dan is $\underline{SS}_A/(I-1)$ dus een zuivere schatter van σ^2 , maar als H_A niet waar is, dan neemt deze grootheid systematisch grotere waarden aan dan onder H_A en dit te sterker naarmate de α_i meer van nul afwijken.

De grootheid

$$(8.27) \quad \underline{SS}_B = IK \sum_{j=1}^J (\underline{y}_{.j.} - \underline{y}_{...})^2$$

is een kwadraatsom met $J-1$ vrijheidsgraden, waarvan de verdeling afhangt van de hypothese H_B . In dit geval is

$$(8.28) \quad E(\underline{SS}_B) = (J - 1)\sigma^2 + IK \sum_{j=1}^J \beta_j^2.$$

Als H_B waar is, dan is $\underline{SS}_B/(J-1)$ een zuivere schatter van σ^2 ; als H_B niet waar is, dan neemt deze grootheid systematisch grotere waarden aan dan onder H_B en wel te sterker naarmate de β_j meer van nul afwijken.

De zg. "interactie kwadraatsom"

$$(8.29) \quad \underline{SS}_{AB} = \sum_{i=1}^I \sum_{j=1}^J (\underline{y}_{ij.} - \underline{y}_{i..} - \underline{y}_{.j.} + \underline{y}_{...})^2$$

heeft $(I-1)(J-1)$ vrijheidsgraden en een verdeling, die afhangt van de additiviteits-onderstelling van het model (8.15). Als er géén interactie is, dan is $\underline{SS}_{AB}/(I-1)(J-1)$ een zuivere schatter van σ^2 , terwijl deze grootheid systematisch grotere waarden aanneemt bij interactie tussen de factoren A en B.

De hypothese H_A kan men nu toetsen met de toetsingsgrootheid

$$(8.30) \quad \frac{\underline{SS}_A/(I - 1)}{\underline{SS}_{\text{binnen}}/IJ(K - 1)},$$

die onder de hypothese H_A een F-verdeling heeft met $v_1 = I-1$ en $v_2 = IJ(K-1)$ vrijheidsgraden. Daar de teller en dus de toetsingsgrootheid systematisch grotere waarden aanneemt als H_A niet waar is, passen we een rechts-éénzijdige toets toe. Bij een gegeven onbetrouwbaarheid α verwerpen we H_A dus als

$$\frac{\underline{SS}_A/(I-1)}{\underline{SS}_{\text{binnen}}/IJ(K-1)} \geq F_{I-1, IJ(K-1); \alpha}.$$

Geheel analoog kan men de hypothese H_B toetsen met behulp van de toetsingsgrootheid

$$(8.31) \quad \frac{\underline{SS}_B/(J-1)}{\underline{SS}_{\text{binnen}}/IJ(K-1)},$$

die onder H_B een F-verdeling met $v_1 = J-1$ en $v_2 = IJ(K-1)$ vrijheidsgraden heeft. Ook in dit geval passen we een rechts-éénzijdige toets toe.

Tenslotte kan men de additiviteits-onderstelling zelf ook toetsen met de toetsingsgrootheid

$$(8.32) \quad \frac{\underline{SS}_{AB}/(I-1)(J-1)}{\underline{SS}_{\text{binnen}}/IJ(K-1)},$$

die onder het model (8.15) een F-verdeling met $v_1 = (I-1)(J-1)$ en $v_2 = IJ(K-1)$ vrijheidsgraden heeft. Een rechts-éénzijdige toets is ook hier aangewezen.

Indien men de hypothese H_A verwerpt, luidt de conclusie, dat niet alle α_i gelijk zijn. Men zal dan, net als bij de één-weg-klassificatie, de parameters α_i aan een nader statistisch onderzoek onderwerpen. Dit is ook het geval met de parameters β_j , indien men de hypothese H_B verwerpt.

In deze beschouwing over de twee-weg-klassificatie hebben we steeds (stilzwijgend) ondersteld, dat het aantal waarnemingen K per cel groter is dan één. Als $K = 1$, dan is $\underline{SS}_{\text{binnen}} = 0$, omdat dan $y_{ijk} = y_{ij}$; ook het aantal vrijheidsgraden van $\underline{SS}_{\text{binnen}}$ is dan nul. We kunnen de hypothesen H_A en H_B en de additiviteits-onderstelling dan niet meer op de aangegeven wijze toetsen. Indien men er echter van overtuigd is, dat het model (8.15) juist is, kan men in de noemer van de toetsingsgrootheden (8.30) en (8.31) $\underline{SS}_{\text{binnen}}/IJ(K-1)$ vervangen door $\underline{SS}_{AB}/(I-1)(J-1)$ en aldus de hypothesen H_A resp. H_B toch nog toetsen.