MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 120

Cursus

Toegepaste Statistiek

door

Prof. Dr J. Hemelrijk
en
Ph. van Elteren

1953

Hoofdstuk 1

Een pot met erwten

1.1. Inleiding.

De wiskundige statistiek houdt zich bezig met de bestudering van verschijnselen, die, ondanks ongeveer gelijkblijvende omstandigheden, bij herhaling verschillende uitkomsten geven.

Voorbeelden. Het aantal auto's, dat ~~xx~~ een brug passeert tussen 8 en 9 uur op Maandagmorgen is steeds verschillend. Het aantal uren ziekteverzuim in een bedrijf per jaar per bedrijfslid berekend is van jaar tot jaar verschillend. Het aantal ondeugdelijke exemplaren in een partij goederen verschilt van partij tot partij, ook als deze partijen uit eenzelfde serieproductie afkomstig zijn. Enz. enz.

Model. Als model voor een bepaalde klasse van problemen, die zich kunnen voordoen, nemen wij een pot met een onbekend aantal rode en groene erwten. Een eerste probleem, dat wij ons stellen, is: een schatting te geven van het onbekende percentage rode erwten in de pot, zonder daartoe alle erwten te tellen en hun kleur na te gaan. Een tweede probleem is, het totale aantal erwten te schatten, zonder ze te tellen. Wij zullen ons in het begin van deze cursus tot dergelijke eenvoudige problemen bepalen

Toelichting. De erwten in de vaas kunnen allerlei dingen voorstellen, b.v.

	groene erwt	rode erwt
a.	Ambtenaar, die pensioengerechtigde leeftijd bereikt.	Ambtenaar, die voor die tijd overlijdt.
b.	Amerikaans kiezer, die republikeins stemt.	Amerikaans kiezer, die democratisch stemt.
c.	Deugdelijk exemplaar in een partij goederen.	Ondeugdelijk exemplaar in een partij goederen.
d.	Watermonster, dat geen bacteriën bevat.	Watermonster, dat wel bacteriën bevat.

In dergelijke gevallen is het van belang zo veel mogelijk te weten te komen over het onbekende percentage rode erwten in de pot. Bij voorbeeld a kan men gegevens omtrent dit percentage gebruiken om pensioenpremies te berekenen (waarvoor natuurlijk nog meer dan alleen dit gegeven nodig is); in geval b kan men er voorspellingen over de uitslag van verkiezingen op baseren; in geval c is kennis van het percentage nodig, om te beslissen of de partij acceptabel is of niet en geval d heeft b.v. betrekking op de keuring van leidingwater.

e. Men kan zich voorstellen, dat een aantal vissen uit een vijver wordt gevangen en voorzien van een bepaald kenmerk, waarna de vissen weer in de vijver worden geworpen. In dat geval kunnen de erwten in de pot corresponderen met alle vissen in de vijver en de rode erwten met de gemerkte vissen. Men wil dan graag iets weten over het totale aantal vissen in de vijver. Dit probleem is te herleiden tot een bepaling van de **fractie** gemerkte vissen in de vijver. Immers men weet het aantal gemerkte vissen (rode erwten) zodat men, indien tevens de fractie bekend is, gemakkelijk het totale aantal vissen (erwten) kan berekenen.

1.2. Volledige en onvolledige waarneming.

Bij een pot met niet te veel erwten is de eenvoudigste oplossing, de rode en groene erwten te tellen. Dit is het geval van volledige waarneming. Vaak is dit niet mogelijk. Bij voorbeeld a wenst men het percentage niet alleen voor het verleden, maar ook voor de toekomst te kennen. Bij voorbeeld b kan men de verkiezingen afwachten (dan wordt er volledig geteld), maar dan is er niets meer te voorspellen. Bij voorbeeld c is volledige keuring vaak practisch onuitvoerbaar of te kostbaar en bij destructieve keuring (sterkteproeven, ontploffingsproeven, levensduurproeven) zelfs onzinnig. Bij voorbeeld d is volledige waarneming onmogelijk, evenals (in de regel) bij voorbeeld e. Het is dus essentieel methoden te bezitten, die uit onvolledige waarneming tot conclusies omtrent het onbekende percentage leiden.

1.3. Terminologie en notatie.

De verzameling van personen, dieren of voorwerpen, die door de pot met erwten wordt voorgesteld, wordt de populatie (Engels: population) genoemd; wij geven deze aan met de Griekse hoofdletter Π (spreek uit: pi). Het aantal erwten wordt de omvang (Engels: size) genoemd; deze kan ook onbepaald zijn, b.v. wanneer het, zoals bij voorbeeld a, om toekomstvoorspellingen gaat; is de omvang groot, dan gebruikt men als wiskundig model vaak een oneindige populatie. Bij een statistisch onderzoek vat men gewoonlijk bepaalde kenmerken van de elementen van de populatie in het oog: in ons geval de kenmerken "rood" en "groen" of wat daardoor wordt voorgesteld.

De omvang zullen wij, als deze eindig is, aangeven met N, de onbekende fractie rode erwten met θ (spreek uit: thêta). N en θ worden parameters van Π genoemd. Wij zullen steeds, indien niet nadrukkelijk anders vermeld is, onderstellen, dat N een groot getal is.

Hoofdstuk 2

Experimenten met een pot erwten

2.1. Uitvoering van een experiment.

Wij nemen een pot met rode en groene erwten, voorzien van een roerstaafje en een vierkant plankje met 25 holletjes. Na goed roeren vissen wij met behulp van dit plankje 25 erwten uit de pot. Dit noemt men: een steekproef nemen. Het aantal erwten in de steekproef heet de omvang daarvan; deze geven wij aan met n. Dit procédé herhalen wij 30 keer, steeds goed roerende tussen het nemen van twee steekproeven, en wij noteren iedere keer het aantal rode erwten in de steekproef. Dit aantal geven wij aan met x. Het resultaat van dit experiment bestond in een bepaald geval uit de volgende 30 waarden van x (in volgorde van waarneming):

(I) 1,6,3,2,3,1,4,2,1,2,2,1,4,2,2,4,2,3,2,1,1,2,5,4,3,1,3,3,6,4.

2.2. Commentaar.

Wij hebben hier een typisch voorbeeld van een statistisch verschijnsel: herhaling van eenzelfde proef, naar beste weten op dezelfde wijze uitgevoerd geeft verschillende uitkomsten. De grootheid x (het aantal rode erwten in een steekproef) kan verschillende waarden aannemen en het hangt, zoals men vaak zegt, van het toeval af, welke waarde in een bepaald geval aangenomen wordt. Een dergelijke grootheid noemt men een stochastische grootheid (Engels: random variable, of ook: variate). De term "stochastisch" is afgeleid van het Griekse woord στοχαζειν (spreek uit: stochadzein), dat "raden" betekent: men kan als het ware de uitkomst niet voorspellen, maar er hoogstens naar raden. Wij geven in het vervolg dergelijke grootheden aan door onderstreepte symbolen, dus in dit geval:

$\underline{x}$ = het aantal rode erwten in een steekproef met $n = 25$.

De waargenomen waarden van $\underline{x}$ geven wij dan aan met $x_1, x_2, \ldots, x_m$ (in ons geval $m = 30$).

Het is duidelijk, dat de in (I) gegeven waargenomen waarden ieder op zichzelf reeds enige informatie verschaffen over de onbekende fractie θ van rode erwten. Indien deze θ dicht bij 1 lag, dus indien vrijwel alle erwten in de pot rood waren, zouden wij niet zulke lage uitkomsten verkregen hebben. Het is derhalve aannemelijk dat θ klein is, d.w.z. dat er weinig rode erwten in de pot zijn. Het gaat er om, deze vage conclusie te verscherpen.

2.3. Grafische voorstelling van de waarnemingsresultaten.

Om een duidelijk overzicht te verkrijgen van de resultaten, zien wij van de volgorde daarvan af en vatten ze samen in een grafiek (zie fig. 2.1), die een histogram (of frequentiediagram) wordt genoemd. De samenstelling hiervan is uit de figuur duidelijk; op de verticale as staat uitgezet hoe vaak ieder der waarden van x waargenomen is.

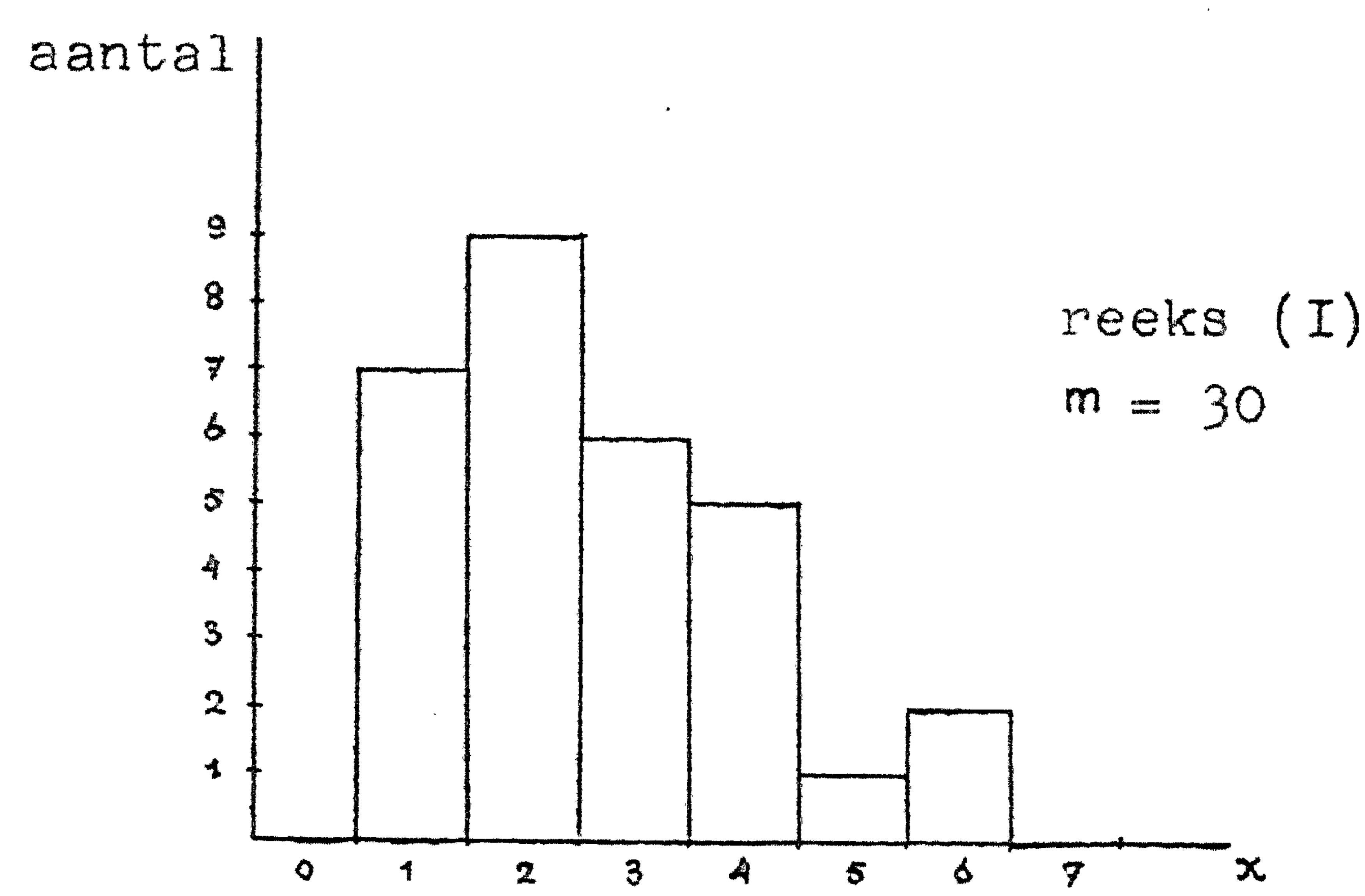


Fig. 2.1. Histogram van 30 waarnemingen van $\underline{x}$ ($n = 25$).

Op deze wijze wordt in één oogopslag een duidelijk overzicht verkregen in gevallen, waarin de volgorde der waarnemingen niet belangrijk is.

Opgave. Maak de histogrammen voor de volgende twee reeksen waarnemingen, die op dezelfde wijze als de bovenstaande verkregen zijn.

(II) 2,0,2,2,4,3,3,6,3,4,3,6,1,3,5,3,3,4,3,3,5,3,3,2,2,2,1,1,2,4.

(III) 4,4,3,1,2,4,4,2,4,1,2,2,1,2,3,2,2,1,1,4,4,1,0,1,2,3,4,3,2,3

Opmerking. Men kan een histogram snel en zonder veel gevaar voor fouten samenstellen door te turven.

2.4. Statistische regelmaat.

Vergelijken wij de drie zo verkregen histogrammen, dan zien wij, dat zij verschillend zijn. Dit was niet anders te verwachten. Niettemin is er een zekere overeenkomst: de waarden 1,2,3 en 4 komen in alle drie vaker voor dan de overige waarden. Hieruit blijkt wel, dat er ook in dit soort onregelmatige verschijnselen toch een zekere regel verborgen zit. Een typisch verschijnsel is daarbij, dat deze regelmaat duidelijker tot uiting komt, indien men de waarnemingsreeksen verlengt (dus als men m groter neemt). Om dit te demonsteren zijn twee reeksen waarnemingen verricht, ieder met m = 100. Zie figuur 2.2, waar deze reeksen in de vorm van turfstaatjes boven elkaar zijn getekend.

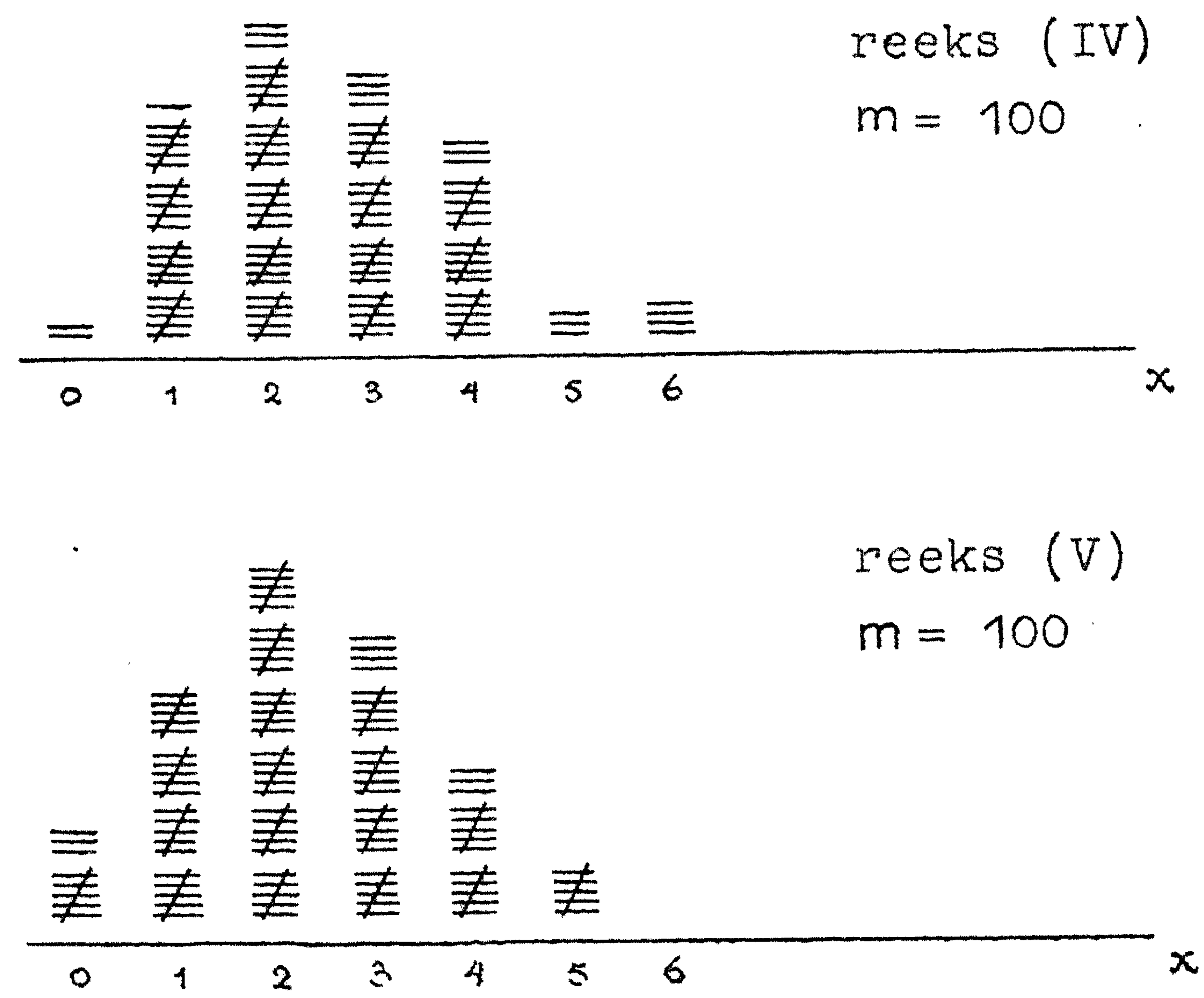


Fig. 2.2. Turfstaatjes van 2 reeksen van 100 waarnemingen van $\underline{x}$ (n = 25).

Het verschil tussen de reeksen (IV) en (V) is reeds minder groot dan b.v. tussen (I) en (III) en de vrij grote onregelmatigheid, die men in reeks (III) aantreft is hier nauwelijks aanwezig.

De "verklaring" van deze regelmaat bewaren wij voor een later hoofdstuk. Wij zullen eerst in het volgende hoofstuk, zonder verklaring van de gevolgde methode, laten zien, welk doel wij nastreven en wat wij met statistische methoden voor de oplossing der in hoofdstuk 1 beschreven problemen kunnen bereiken.

2.5. Invloed van de omvang van de steekproef.

In plaats van een grotere m kunnen wij ook n vergroten. Om de invloed hiervan te demonsteren geven wij in fig. 2.3 een histogram van 100 steekproeven van omvang 100 in plaats van 25. Deze zijn verkregen door telkens 4 opeenvolgende steekproeven van 25 tezamen te nemen, daar er geen steekproefplankje met 100 holletjes beschikbaar is. Het aantal rode erwten noemen wij nu $\underline{y}$.

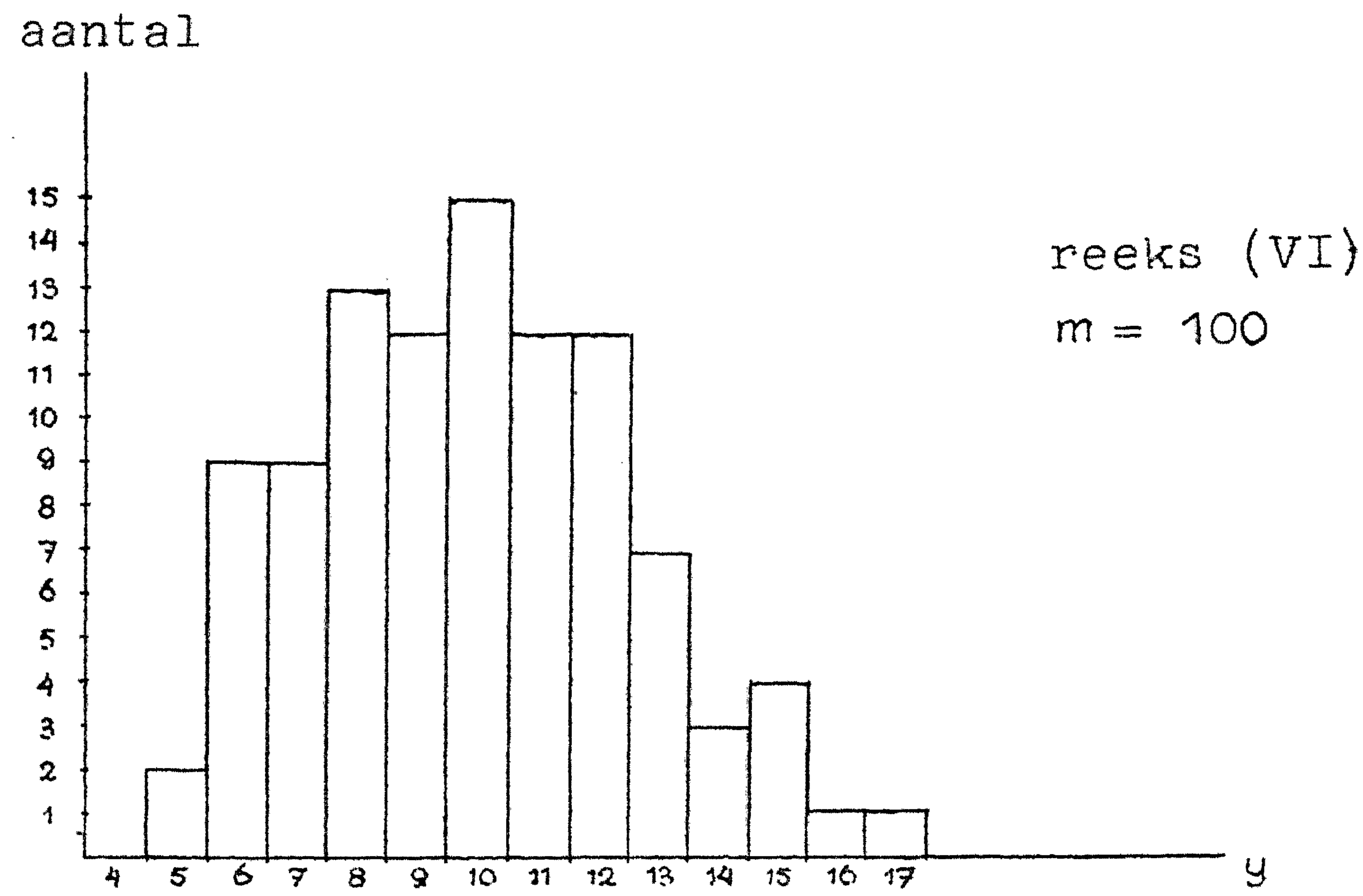


Fig. 2.3 Histogram van 100 waarnemingen van $\underline{y}$ ($n = 100$).

Wij zien uit deze figuur, dat de meest voorkomende waarden van $\underline{y}$ grofweg 4 maal zo groot zijn als de meest voorkomende waarden van $\underline{x}$. Bij $\underline{x}$ waren dit de waarden 2 en 3 en bij $\underline{y}$ de waarden van 8 tot en met 12. Verder zien wij, dat de breedte van het histogram toegenomen is. Bij $\underline{x}$ komen alleen de waarden 1 tot en met 4 vaak voor, bij $\underline{y}$ de waarden 6 tot en met 13. Drukken wij echter de resultaten niet uit in het totale aantal rode erwten, maar in de fractie rode erwten in de steekproef, dan ligt deze fractie voor $\underline{x}$ tussen $\frac{1}{25}$ en $\frac{4}{25}$ en voor $\underline{y}$ tussen $\frac{6}{100}$ en $\frac{13}{100}$. Wat deze fractie betreft — en dit is eigenlijk de grootheid, waar het om gaat — vinden wij dus bij $n = 100$ een kleinere variatie dan bij $n = 25$.

De nauwkeurigheid, waarmede wij de onbekende fractie θ kunnen schatten, neemt blijkbaar toe, als wij n laten toenemen. Dit is een zeer belangrijk feit, dat wij in de statistiek steeds weer zullen ontmoeten.

Hoofdstuk 3

Conclusies op grond van een steekproef.

3.1. Schatting van θ

De eerste vraag, die wij ons kunnen stellen — en die in de practijk vaak van veel belang is — is, hoe wij de onbekende fractie θ aan rode erwten (of wat deze voorstellen) kunnen schatten op grond van het aantal rode exemplaren in de steekproef. Een zeer voor de hand liggende manier, om dit te doen, is de volgende.

Als er in een steekproef van 25 b.v. 3 rode erwten gevonden zijn, nemen wij als schatting voor θ de waarde $\frac{3}{25}$. Dit betekent, dat wij de conclusie trekken, dat θ wel in de buurt van de waarde $\frac{3}{25}$ zal liggen. In het algemeen: als in een steekproef van omvang n het aantal rode exemplaren x is, nemen wij als schatting voor θ de waarde $\frac{x}{n}$.

Het is duidelijk, dat $\frac{x}{n}$ in de regel niet gelijk aan θ zal zijn en dat deze schatting bovendien van steekproef tot steekproef verschilt, m.a.w. aan steekproefvariatie onderhevig is. Het in par. 2.5 vermelde verschijnsel betekent nu, dat deze steekproefvariatie kleiner wordt, naarmate n groter is. Anders uitgedrukt: de schatting is nauwkeuriger bij grote n dan bij kleine n. Dit ligt zeer voor de hand en wij zullen het later quantitatief preciseren.

3.2. Betrouwbaarheidsgrenzen voor θ.

Een onaangename eigenschap van de schatting $\frac{x}{n}$ is, dat deze in de regel fout is: $\frac{x}{n}$ ligt weliswaar gewoonlijk in de buurt van θ, maar meer ook niet. Wensen wij conclusies te trekken, die in de regel goed zijn, dan moeten deze een andere vorm hebben. Dit kunnen wij bereiken door gebruik te maken van tabel 3.I, waarin boven- en benedengrenzen voor θ zijn aangegeven bij alle mogelijke uitkomsten voor x met $n = 25$. De wijze, waarop een dergelijke tabel tot stand komt en de onderstellingen, die vervuld moeten zijn voor het gebruik ervan, zullen wij in een later hoofdstuk bespreken. Hier geven wij slechts aan, hoe de tabel gebruikt kan worden en hoe de interpretatie van het resultaat luidt.

Wij beschouwen eerst het geval, waarbij de bepaling van een bovengrens van belang is. De populatie Π (voorgesteld door de pot met erwten) is bv. een partij goederen, die gekeurd moet worden en x is het aantal ondeugdelijke exemplaren in een steekproef van 25 (voorgesteld door het aantal rode erwten in een steekproef van 25 erwten).

Tabel 3.I

Betrouwbaarheidsgrenzen voor θ

$n = 25$

(onbetrouwbaarheidsdrempel voor beide grenzen apart 0,05).

x	ondergrens voor θ	bovengrens voor θ
0	0,00	0,11
1	0,00	0,18
2	0,01	0,23
3	0,03	0,28
4	0,06	0,33

Tabel 3.I (vervolg)

x	ondergrens voor θ	bovengrens voor θ
5	0,08	0,38
6	0,11	0,42
7	0,14	0,46
8	0,17	0,50
9	0,20	0,54
10	0,24	0,58
11	0,27	0,62
12	0,30	0,66
13	0,34	0,70
14	0,38	0,73
15	0,42	0,76
16	0,46	0,80
17	0,50	0,83
18	0,54	0,86
19	0,58	0,89
20	0,62	0,92
21	0,67	0,94
22	0,72	0,97
23	0,77	0,99
24	0,82	1
25	0,89	1

Als er nu in deze steekproef b.v. 4 ondeugdelijke exemplaren gevonden worden, vindt men in de tabel als bovengrens voor θ de waarde 0,33. De conclusie luidt dan dus: de fractie ondeugdelijke exemplaren in de partij is niet groter dan 0,33. Is deze uitspraak nu zeker juist? Neen, dit is niet zeker, want θ is onbekend en als θ b.v. 0,5 was (dus als de helft der exemplaren in de partij niet deugt), zouden wij toch goed een steekproef met 4 rode erwten kunnen krijgen. In een dergelijk geval zouden wij dus met behulp van tabel 3.I tot een verkeerde conclusie geraken. Dit kunnen wij nooit uitsluiten, daar wij anders slechts triviale conclusies kunnen trekken. Tabel 3.I is echter zo geconstrueerd, dat wij een nauwkeurige uitspraak kunnen doen over de frequentie van het optreden van foute conclusies. Als wij deze tabel gebruiken voor het trekken van een lange reeks van conclusies, dus b.v. als wij de tabel gebruiken bij routinekeuringen van partijen, dan zal de gevonden bovengrens hoogstens gemiddeld in 1 op de 20 gevallen fout zijn. Met andere woorden: in 19 van de 20 gevallen zal de bovengrens zo uitvallen, dat hij ook werkelijk groter dan Θ is. Dit geldt onafhankelijk van de waarde van Θ, die bij ieder der proeven (dus bij ieder der gekeurde partijen) een andere waarde mag bezitten. Hetzelfde geldt voor de ondergrens. Wij zeggen dan, dat zowel de onder- als de bovengrens ieder een onbetrouwbaarheidsdrempel 0,05 bezitten en deze onbetrouwbaarheidsdrempel geven wij in het algemeen aan met de letter α. Het is een vrij algemene gewoonte de waarde $\alpha = 1/20$ als redelijk te aanvaarden, maar men kan ook tabellen construeren, van dezelfde aard als tabel 3.I, met een andere waarde van α. Heeft het maken van een fout ernstige gevolgen, dan zal men α graag klein willen hebben en b.v. $\alpha = 0,01$ of zelfs $\alpha = 0,001$ verkiezen boven $\alpha = 0,05$ zodat de fractie foute conclusies in een lange reeks experimenten tot ongeveer 0,01 resp. 0,001 of minder daalt. Men wenst dan dus meer zekerheid en moet daarvoor dan ook betalen in de vorm van een hogere bovengrens (en een lagere ondergrens), dus een minder scherpe schatting van Θ. Of, indien men dit niet wenst, een grotere n dus een grotere steekproef. Want, zoals wij zullen zien: vergroting van n geeft ook hier verscherping van de schatting van Θ.

Ons weer even tot de bovengrens beperkende, verkrijgen wij dus, bij een bepaalde waarde van x een bovengrens voor Θ, dus wij verkrijgen een uitspraak van de vorm

$$\Theta \leqq \Theta^*(x), \tag{3.2;1}$$

waarin $\Theta^*(x)$ de, van x afhankelijke, bovengrens voorstelt.

Nu is x, het aantal rode erwten in de steekproef, bij herhaling van het experiment het stochastische element, dus de bovengrens $\Theta^*(\underline{x})$ is ook stochastisch. Het is van groot belang dit goed in het oog te houden: Θ bezit bij ieder experiment een onbekende, maar vaste, waarde (die van experiment tot experiment kan wisselen) en de bovengrens $\Theta^*(\underline{x})$ is stochastisch. De ondergrens geven wij aan met $\Theta_*(\underline{x})$. Hiervoor geldt hetzelfde.

Zowel de uitspraak (3.2;1) als

$$\Theta_*(\underline{x}) \leqq \Theta \tag{3.2;2}$$

bezitten dus, als tabel 3.I gebruikt wordt, een onbetrouwbaarheidsdrempel $\alpha = \frac{1}{20}$. De tweezijdige uitspraak

$$\Theta_*(\underline{x}) \leqq \Theta \leqq \Theta^*(\underline{x}) \tag{3.2;3}$$

bezit de onbetrouwbaarheidsdrempel $\frac{1}{10}$, de som van die van $\Theta_*(\underline{x})$ en van $\Theta^*(\underline{x})$. Deze twee grenzen Θ_* en Θ^* worden betrouwbaarheidsgrenzen voor Θ genoemd. Men spreekt ook van intervalsschatting, de schatting van θ door middel van het interval (3.2;1), (3.2;2) of (3.2;3) bedoelende (Engelse terminologie: betrouwbaarheidsgrenzen = confidence limits; onbetrouwbaarheidsdrempel = level of significance; in de Engelse literatuur gebruikt men vaak de term confidence coefficient voor de grootheid $1-\alpha$)

3.3. Invloed van vergroting van n

Wij hebben reeds gezegd, dat vergroting van n, de omvang van de steekproef, verscherping van de schatting van θ ten gevolge heeft. Om dit te laten zien is een gedeelte van de met tabel 3.I corresponderende tabel voor $n = 100$ in tabel 3.II gereproduceerd. Deze tabel wordt dus precies als tabel 3.I gebruikt en heeft ook voor ieder der beide grenzen een onbetrouwbaarheid 0,05, maar geldt voor steekproeven van omvang 100.

Om deze tabel met 3.I te vergelijken, bedenken wij, dat bij de in hoofdstuk 2 beschreven experimenten de waarden $x = 2$ en $x = 3$ (bij $n = 25$) en $y = 8, \ldots, y = 12$ (bij $n = 100$) met elkaar corresponderen en het meeste voorkwamen (dit is natuurlijk een gevolg van de samenstelling van de door ons gebruikte pot erwten; bij een andere samenstelling zouden wij ook andere resultaten gekregen hebben) en ter illustratie van de invloed van de vergroting van n zullen wij daarom de bij deze waarden van x en y behorende grenzen vergelijken. Wij zien dan, dat voor

Tabel 3.II

Betrouwbaarheidsgrenzen voor θ

$n = 100$

(onbetrouwbaarheidsdrempel voor beide grenzen apart 0,05)

y	ondergrens voor θ	bovengrens voor θ
5	0,02	0,10
6	0,03	0,12
7	0,03	0,13
8	0,04	0,14
9	0,05	0,15
10	0,05	0,16
11	0,06	0,18
12	0,07	0,19
13	0,08	0,20
14	0,09	0,21
15	0,09	0,22
16	0,10	0,23
17	0,11	0,24

$x = 2$ en 3 de bovengrens 0,23 en 0,28 is en voor $y = 8, \ldots, 12$ waarden tussen 0,14 en 0,19 aanneemt. Voor de ondergrens zijn deze waarden bij $x = 2$ en 3: 0,01 en 0,03 en bij $y = 8, \ldots, 12$: 0,04, ..,0,07. De winst aan scherpte is dus bij de bovengrens zeer duidelijk en bij de ondergrens weliswaar kleiner, maar niettemin aanwezig.

3.4. Grafische voorstelling van de betrouwbaarheidsgrenzen.

Volledige tabellen voor vele waarden van n van het type van 3.I en 3.II zijn niet beschikbaar. Wel echter grafieken, waaruit zij voor verschillende waarden van n en α afgelezen kunnen worden. Ter illustratie hiervan hebben wij tabel 3.I in figuur 3.1 in beeld gebracht, tezamen met tabel 3.II, waarvan de gegevens slechts een deel der mogelijke uitkomsten van y beslaan.

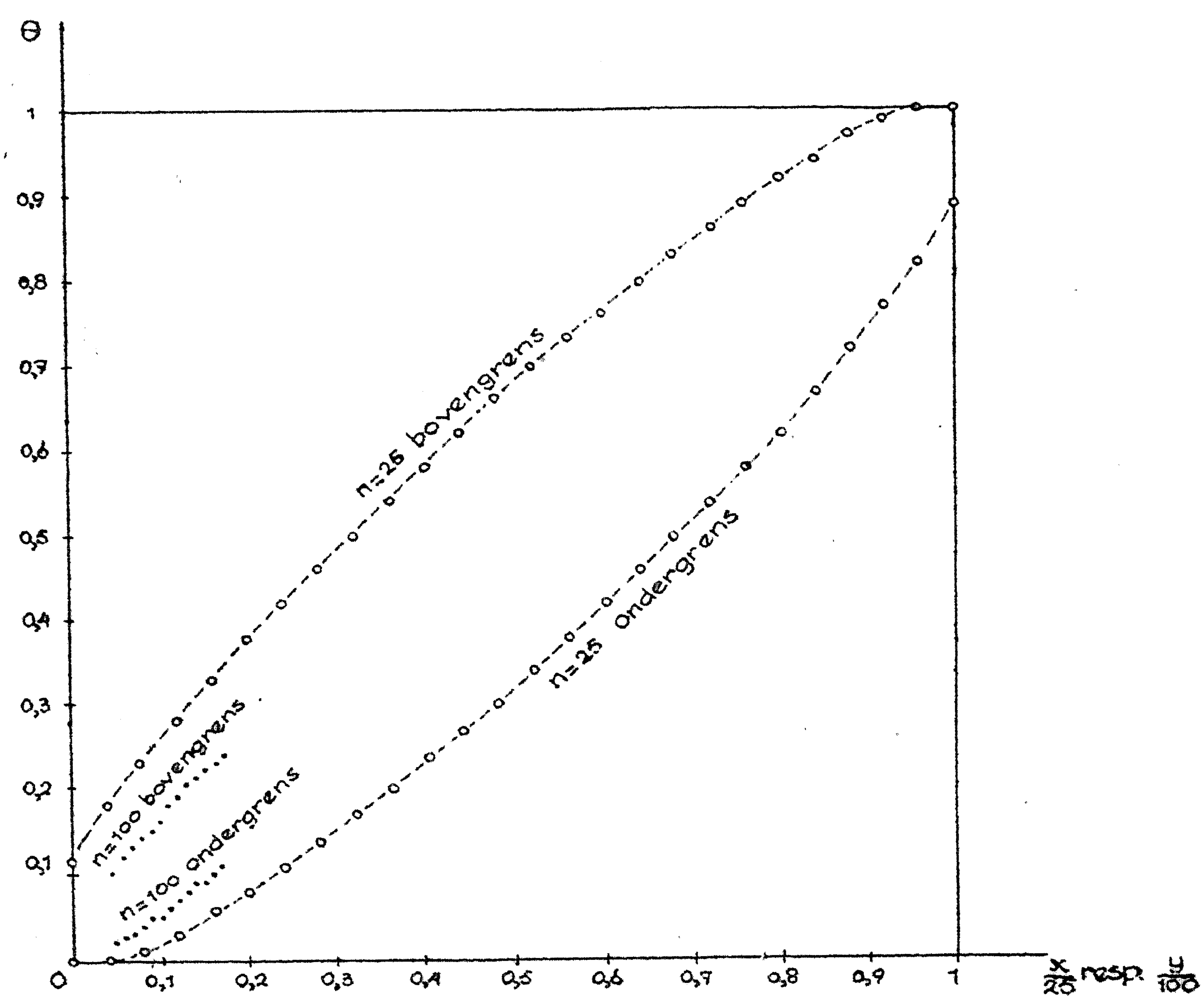


Fig. 3.1. Grafische voorstelling van betrouwbaarheidsgrenzen voor θ, met $\alpha = 0,05$ (éénzijdig) en $n = 25$ resp. 100.

Op de horizontale as zijn $\frac{x}{25}$ en $\frac{y}{100}$ uitgezet, om beide op een schaal van 0 tot 1 te brengen, evenals θ. Vinden wij nu b.v. $x = 3$, dan lezen wij in verticale richting bij $\frac{x}{25} = \frac{3}{25}$ af: ondergrens voor θ is 0,03 en bovengrens 0,28. Het aflezen gaat iets minder vlot dan uit een tabel, maar men kan op deze wijze een groot aantal tabellen (nl. voor een aantal waarden van n) in één grafiek samenbrengen. Bovendien kan men gemakkelijk op het oog interpoleren tussen krommen, die bij verschillende waarden van n behoren, om bij een niet opgenomen waarde van n toch van de grafiek gebruik te maken. In de grafiek zijn eigenlijk alleen de ronde stippen van betekenis, daar alleen de daarbij behorende waarden van de abscis voor kunnen komen. Wij hebben daarom deze punten niet door een lijn vervangen, maar alleen, voor de samenhang van de figuur, door een stippellijn verbonden. Voor $n = 100$ was dit niet nodig. Het feit, dat de lijn niet geheel vloeiend is (vooral voor $n = 100$ is dit duidelijk) wordt grotendeels veroorzaakt door afronding. Deze grafieken vindt men, veel uitgebreider, in het zeer aanbevelenswaardige boek over statistiek van W J.DIXON en F.J.MASSEY, Introduction to statistical analysis, Mac Graw-Hill, New York-Toronto-London, 1951, op blz. 320-323. (Op blz. 320 en 321 staat daar de confidence coefficient verkeerd opgegeven: op blz. 320 moet deze 0,90 en op blz. 321 0,80 zijn; onze grafiek komt overeen met de op blz. 320 gegevene. Dixon en Massey betrekken de confidence coefficient op het tweezijdige geval, d.w.z. op uitspraken van de vorm (3.2;3), terwijl wij onze $\alpha = 0{,}05$ op de boven- en ondergrens apart betrokken hebben.) Het genoemde boek zullen wij voortaan kortweg aangeven als "DIXON and MASSEY" zonder verdere aanduiding. DIXON and MASSEY geven de grafieken als vloeiende lijnen; deze lijnen geven niet de exacte grenzen, die wij in de bovenstaande tabellen en in fig. 3.1 hebben gegeven, maar benaderingen daarvan. Vergelijking van onze tabellen met de grafieken uit DIXON and MASSEY tonen aan dat deze benadering, waarvan wij verderop de details geven, zeer goed is voor niet te kleine n en in het bijzonder voor niet zeer kleine $\frac{x}{n}$ resp. $\frac{y}{n}$. Voor zeer kleine waarden van deze fractie zijn de grafieken trouwens niet goed afleesbaar en moet een andere methode worden gevolgd.

3.5. Berekening van de betrouwbaarheidsgrenzen.

De exacte berekening der betrouwbaarheidsgrenzen zullen wij niet beschrijven, daar hiervoor geen eenvoudig recept te geven is zonder van uitgebreide tabellen gebruik te maken, die gewoonlijk niet voorhanden zijn.
Bovendien heeft men in de practijk gewoonlijk ruim voldoende aan de hieronder beschreven benaderingsmethode.

Als wij in een steekproef van n erwten x rode hebben gevonden vinden wij een bovenste betrouwbaarheidsgrens met $\alpha = 0{,}05$ uit de formule

(3.5;1) $$\Theta^*(x) = \frac{x + \frac{1}{2}\cdot 2{,}71 + \sqrt{2{,}71\left(\frac{x(n-x)}{n} + \frac{1}{4}\cdot 2{,}71\right)}}{n+2{,}71}$$

en een ondergrens met dezelfde onbetrouwbaarheid uit de formule

(3.5;2) $$\Theta_*(x) = \frac{x + \frac{1}{2}\cdot 2{,}71 - \sqrt{2{,}71\left(\frac{x(n-x)}{n} + \frac{1}{4}\cdot 2{,}71\right)}}{n+2{,}71}.$$

Indien wij in plaats van $\alpha = 0{,}05$ een andere waarde wensen te gebruiken, behoeven wij in deze formules slechts het getal 2,71 overal door een ander te vervangen en wel door de in tabel 3.III aangegeven getallen

Tabel 3.III

Waarden voor de constante in de formules (3.5;1) en (3.5;2) voor verschillende waarden van α.

α (éénzijdig)	constante
0,10	1,64
0,05	2,71
0,025	3,84
0,005	6,63

Een tweezijdig begrensd betrouwbaarheidsinterval is te verkrijgen door (3.5;1) en (3.5;2) tegelijk te gebruiken. De onbetrouwbaarheidsdrempel is dan 2α. Deze benaderende formules zijn voor practische doeleinden voldoende betrouwbaar, zolang $n\theta$ niet te klein is, b.v. niet kleiner dan 5. Daar θ niet bekend is, zal men zich uit de waarde van x een oordeel over de bruikbaarheid moeten vormen. Men kan dan als werkregel houden dat zij bruikbaar zijn, als x niet kleiner dan 5 is.

Is x groot (dus n ook), dan zullen bepaalde termen in de genoemde formules zonder bezwaar verwaarloosd kunnen worden, nl. de termen $\frac{1}{2}\cdot 2{,}71$ en $\frac{1}{4}\cdot 2{,}71$ in de teller en de term 2,71 in de noemer (of één der andere constanten in plaats van 2,71). De formules worden dan veel eenvoudiger; zij gaan over in

(3.5;3) $$\Theta^*(x) = \frac{1}{n}\left\{x + 1{,}65\sqrt{\frac{x(n-x)}{n}}\right\}$$

en

(3.5;4). $$\Theta_*(x) = \frac{1}{n}\left\{x - 1{,}65\sqrt{\frac{x(n-x)}{n}}\right\},$$

met in plaats van 1,65 de in tabel 3.IV vermelde, bij verschillende waarden van α behorende constanten.

Tabel 3.IV

Waarden voor de constante in de formules (3.5;3) en (3.5;4) voor verschillende waarden van α.

α (éénzijdig)	constante
0,10	1,28
0,05	1,65
0,025	1,96
0,005	2,58

Men kan zelf, door vergelijking van de uitkomsten van de laatste twee formules met de eerste twee gemakkelijk nagaan of de verwaarlozing van de genoemde termen zonder bezwaar geschieden kan

Opgaven.

3.5.a. Uit een partij goederen wordt een steekproef van 25 exemplaren genomen. Hieronder bevinden zich 2 defecte exemplaren. Bereken een bovengrens voor de fractie defecten in de partij met $\alpha = 0,05$, zowel met formule (3.5;1) als met formule (3.5;3). Vergelijk de uitkomsten onderling en vergelijk ze met de door tabel 3.I aangegeven waarde van $\theta^*(x)$.

3.5.b. Voer dezelfde opgave uit voor een steekproef van 100 exemplaren, waaronder 15 defecte.

3.5.c. Om te onderzoeken of een oesterbank voldoende parels bevat om exploitatie lonend te maken, worden 2000 oesters opgedoken. 6 Oesters bleken bij onderzoek een parel van enige commerciële waarde te bevatten. Welke voor de gestelde vraag belangrijke conclusie kan men hieruit trekken, als men (wegens de grote kosten, die aan exploitatie verbonden zijn) slechts een onbetrouwbaarheid 0,005 wenst toe te laten?

3.6. Betrouwbaarheidsgrenzen voor kleine waarden van $n\theta$.

Voor kleine waarden van $n\theta$, hetgeen tot uiting komt in kleine waarden van x, is de boven beschreven benaderingsmethode niet erg nauwkeurig en de grafieken uit DIXON and MASSEY zijn in dat geval bovendien slecht afleesbaar. Voor dit geval is echter een andere methode beschikbaar, die een goede benadering geeft, als n niet te klein is. Bij deze methode maakt men gebruik van tabel 3.V, waarin grenzen voor $n\theta$ staan aangegeven voor $x = 1,\ldots,10$ en voor verschillende waarden van α. De aangegeven waarde van α is éénzijdig en geldt dus voor ieder van de grenzen afzonderlijk. Ook hier geldt, dat de bij een tweezijdig interval behorende onbetrouwbaarheidsdrempel gelijk aan 2α is.

Tabel 3.V

Betrouwbaarheidsgrenzen voor $n\theta$ voor $x \leqq 10$.

α / x	0,005		0,025		0,05		0,10	
	ondergrens	bovengrens	ondergrens	bovengrens	ondergrens	bovengrens	ondergrens	bovengrens
0	0	5,3	0	3,7	0	3,0	0	2,3
1	0,005	7,4	0,025	5,6	0,05	4,7	0,11	3,9
2	0,1	9,3	0,2	7,2	0,4	6,3	0,5	5,3
3	0,3	11,0	0,6	8,8	0,8	7,8	1,1	6,7
4	0,6	12,6	1,0	10,2	1,4	9,2	1,7	8,0
5	1,0	14,1	1,6	11,7	2,0	10,5	2,4	9,3
6	1,5	15,6	2,2	13,1	2,6	11,8	3,1	10,5
7	2,0	17,1	2,8	14,4	3,3	13,1	3,[illegible]	11,8
8	2,5	18,5	3,4	15,8	4,0	14,4	4,7	13,0
9	3,1	20,0	4,0	17,1	4,7	15,7	5,4	14,2
10	3,7	21,3	4,7	18,4	5,4	17,0	6,2	15,4

Men gebruikt de tabel als volgt. Zijn onder n erwten x rode gevonden, dan leest men de gezochte grens (bij de gekozen α) uit de tabel af en deelt deze door n. Dit geeft de gewenste grens voor θ. Ook in dit geval zouden wij de tabel door een grafiek kunnen vervangen, maar daar wij bij deze benadering alle grenzen (nl. voor iedere n en voor een aantal α's) in één tabel kunnen samenvatten, geeft dat geen ruimtewinst, zodat er niet veel voordeel aan verbonden is. Tabel 3.V is alleen geschikt voor kleine x en grote n, maar niet voor de gevallen: kleine x en kleine n of grote x en grote n. In het laatste geval kan men de in par. 3.5 beschreven formules gebruiken. Het geval van kleine x en kleine n kan eveneens met deze formules vrij goed worden behandeld, als $\frac{x}{n}$ niet te ver van $\frac{1}{2}$ verwijderd is. Men gebruike dan de formules (3.5;1) en (3.5;2). Het geval van kleine x en kleine n met $\frac{x}{n}$ niet ongeveer $\frac{1}{2}$ eist een nauwkeurigere behandeling, die wij hier niet zullen bespreken.

Opgaven.

3.6.a. Los opgave 3.5.a en 3.5.c op met behulp van tabel 3.V en vergelijk de uitkomsten met de eerder verkregene.

3.6.b. De parelvisser, die in opgave 3.5.c 2000 oesters heeft opgedoken, heeft daaraan zoveel werk gehad dat hij zich afvraagt, of hij de volgende keer dat zoiets voorkomt niet met minder toe kan. Hij denkt, dat een schatting, waarbij beneden- en bovengrens ongeveer 0,01 van elkaar verschillen, met tweezijdige onbetrouwbaarheid 0,05, bij een oesterbank, waarin ongeveer 5 per duizend oesters een parel bevatten, ruim goed genoeg is voor een eerste indruk. Hoeveel oesters zal hij daarvoor ongeveer op moeten duiken?

Hoofdstuk 4

Experimentele voorwaarden

4.1. Formulering en toelichting.

De in hoofdstuk 3 beschreven methoden zijn alleen geldig, indien de methode van steekproef nemen aan de volgende voorwaarden voldoet:

a. Eis van gelijkwaardigheid. Indien men een lange reeks van steekproeven van dezelfde omvang n uit dezelfde populatie π van omvang N neemt, met teruglegging van de getrokken steekproef voor de volgende genomen wordt, dan moeten alle mogelijke steekproeven in deze reeks ongeveer even vaak voorkomen.

b. Eis van onafhankelijkheid. Beschouwt men in een dergelijke lange reeks alle steekproeven, die getrokken zijn direct na steekproeven (of na groepen van steekproeven), die een bepaald kenmerk bezitten, dan komen in deze deelreeks, als die lang genoeg is, eveneens alle mogelijke steekproeven ongeveer even vaak voor.

Toelichting. Om deze voorwaarden nader te beschouwen denken wij ons de erwten genummerd van 1 t/m N. Voorwaarde a houdt nu bv. in, dat onder een lange reeks van steekproeven van omvang 1 (dus van één erwt ieder) alle nummers van 1 t/m N ongeveer even vaak voorkomen. Het is duidelijk, dat wij, om dit te controleren, de lengte van de reeks vele malen de lengte N zullen moeten geven. Niettemin is de voorwaarde in principe controleerbaar en de mathematische statistiek beschikt ook over eenvoudiger uitvoerbare methoden om de voorwaarde te controleren. Deze bespreken wij hier niet in extenso, maar één ervan, die voor ons probleem relevant is, geven wij kort aan.

Het gaat er ons nl. in feite niet om, dat alle nummers even vaak uit de pot komen, maar dat de rode en groene erwten in de goede verhouding te voorschijn komen. Zijn er nu k rode en dus $N-k$ groene erwten in de pot, dan houdt voorwaarde a onder andere in, dat in een lange reeks trekkingen van ieder 1 erwt de verhouding van de rode tot de groene erwten ongeveer gelijk moet zijn aan

$$\frac{k}{N-k}.$$

Het is duidelijk, dat dit reeds voor veel kortere reeksen op moet gaan dan voorwaarde a in de uitgebreide vorm. Voorwaarde b betekent bv., dat in de trekkingen direct na het trekken van erwt nr. h ook weer alle nummers (h zelf inbegrepen) ongeveer even vaak voor moeten komen. Om dit te controleren moeten wij dus een nog veel längere reeks hebben, omdat wij nu de deelreeks van de na erwt nr. h getrokken erwten beschouwen, terwijl erwt nr. h in de reeks zelf maar ongeveer 1 op de N maal voorkomt. Ook hier kunnen wij weer vereenvoudigingen toepassen. Zo kunnen wij bv. in plaats van het kenmerk "nr. h " in voorwaarde b substitueren het kenmerk "een even nummer dragend", dat veel vaker dan "nr. h " voor zal komen en alle op even nummers volgende erwten beschouwen. In die deelreeks moeten dan alle nummers weer ongeveer even vaak voorkomen. Ook kunnen wij in deze deelreeks niet op de frequentie van de nummers van de erwten, maar op de frequentie van even en oneven nummers letten en nagaan of deze ongeveer gelijk zijn. En, waar het problemen van rode en groene erwten betreft, kunnen wij de nummering buiten beschouwing laten en alleen op de kenmerken rood en groen letten. Dit maakt de zaak weer veel eenvoudiger; wij krijgen dan een reeks als bv.

$$\begin{array}{ccccccccccccc} & \downarrow & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow & \downarrow & & \downarrow & & \downarrow \\ g & g & g & r & g & g & g & g & r & g & r & g & g \; \ldots\ldots \end{array}$$

waarin volgens voorwaarde a ongeveer $\frac{k}{N-k}$ r's voorkomen, terwijl dit volgens voorwaarde b ook het geval is in de met pijltjes aangegeven deelreeks van waarnemingen direct na een g verricht en eveneens in het complement van deze deelreeks, dat bestaatui uit waarnemingen direct na een r verricht.

Deze eigenschap b moet niet alleen gelden voor deelreeksen direct na g of na r, maar ook bv. voor de deelreeks van waarnemingen, die op tweetallen gg volgen, of op tweetallen gr, etc. In het algemeen voor deelreeksen volgend op groepen waarnemingen, die op de één of andere wijze zijn gedefinieerd.

Wij hebben de voorwaarden a en b nu uitvoerig uiteengezet voor steekproeven van omvang 1. Voor grotere steekproeven gelden dan analoge voorwaarden, die wij, om het eenvoudigste geval te nemen, uiteenzetten voor $n = 2$. Uit de nummers $1, \ldots, N$ kunnen wij

$$\frac{N(N-1)}{2} \quad \left(\text{notatie}: \binom{N}{2}\right)$$

verschillende paren vormen en dit is dus het aantal mogelijke steekproeven van omvang 2. Deze moeten nu alle ongeveer even vaak voorkomen en eis b houdt bv. in dat dit in de deelreeks van steekproeven, direct volgend op steekproeven met de nummers 1 en 2 (of 2 andere willekeurige nummers) ook het geval is, enz. Er is één belangrijk punt, waarop wij hier met nadruk wijzen. Als wij in dit geval speciaal op de kenmerken rood en groen gaan letten, kunnen wij niet zeggen: er zijn 3 mogelijkheden, te weten

A: twee rode
B: één rode en één groene
C: twee groene,

dus komen deze drie combinaties in de steekproevenreeks ongeveer even vaak voor. Het fundamentele feit is nl., dat alle mogelijke steekproeven ongeveer even vaak voorkomen en dat deze uit geheel verschillende aantallen met de kenmerken A, B en C bestaan. Zijn er nl. k rode erwten, dan kunnen wij deze (voor het gemak) de nummers $1, \ldots, k$ geven en de groene erwten $k+1, \ldots, k+n$. Het aantal mogelijke steekproeven met kenmerk A is dan

$$\frac{k(k-1)}{2} = \binom{k}{2}$$

en dat met kenmerk C

$$\frac{(N-k)(N-k-1)}{2} = \binom{N-k}{2}$$

en dat met kenmerk B

$$k(N-k).$$

In deze verhouding (de som der drie aantallen is, hetgeen men ter controle kan uitrekenen, gelijk aan $\frac{N(N-1)}{2}$) komen dan de kenmerken A, B en C in een reeks ongeveer voor.

Opgave.

4.1.a. Aan welke eisen voldoet de volgende rij getallen wèl en aan welke niet, als wij deze rij beschouwen als een reeks steekproeven van omvang 1 uit een populatie met $N = 3$:

123123123 ....

4.1.b. Beantwoord dezelfde vraag voor de rij:

111213212222331323311121321 ....

4.2. Korte samenvatting

Indien men de bovengenoemde voorwaarden kort en wat minder exact wil samenvatten, kan men dit als volgt doen.

a. Alle elementen van de populatie moeten bij de steekproefname gelijkwaardig zijn. Zij komen symmetrisch in het probleem voor en onderlinge verwisseling (of: permutatie) verandert niets aan de situatie.

b. Bij de trekking van een nieuwe steekproef in een reeks is het resultaat van de vorige steekproeven van geen enkel belang. De nieuwe steekproef is geheel onafhankelijk van alle vorige steekproeven.

4.3. Steekproeven met en zonder teruglegging.

Men kan een steekproef van omvang $n > 1$ (b.v. met $n = 25$) in ééns nemen (zoals wij in hoofdstuk 2 hebben gedaan) of hem samenstellen uit n steekproeven van omvang 1 (of in het algemeen uit een aantal steekproeven met omvang $< n$). In het eerste geval spreekt men van steekproeven <u>zonder teruglegging</u>, in het tweede van steekproeven <u>met teruglegging</u> (Engels: without- en with replacement), omdat in het tweede geval de getrokken erwt wordt teruggelegd voor de volgende (na roeren) getrokken wordt. Het is duidelijk, dat deze procedures niet aequivalent zijn, daar in het tweede geval eenzelfde erwt meer dan éénmaal in een steekproef voor kan komen, terwijl dit in het eerste geval niet zo is. In de practijk ontmoet men meestal steekproeven zonder teruglegging, terwijl de theorie gemakkelijker is voor steekproeven met teruglegging. Als $n \ll N$ is, is het verschil tussen de twee methoden gering, daar er dan toch relatief maar weinig steekproeven met meer dan éénmaal dezelfde erwt zullen voorkomen als men met teruglegging trekt. Daar bovendien zeer vaak $n \ll N$ is, wordt het verschil meestal verwaarloosd. Indien niet anders vermeld is, zullen wij steeds onderstellen, dat deze verwaarlozing desgewenst zonder ernstige gevolgen kan worden toegepast.

Deze kwestie hangt samen met het feit, dat de voorwaarden a en b van par. 4.1 in het geval van steekproeven <u>met</u> teruglegging uitsluitend gesteld behoeven te worden voor steekproeven van omvang 1. Zij gelden dan (zoals experimenteel kan worden nagegaan en ook door berekeningen aannemelijk gemaakt kan worden) automatisch voor grotere steekproeven, indien deze met teruglegging genomen worden, zodat zij beschouwd kunnen worden als verkregen door samenvoeging van steekproeven met omvang 1. Bij steekproeven zonder teruglegging kan dit ook, maar dan verandert de populatie bij iedere stap van het procédé, doordat er iedere keer een erwt uit genomen wordt; daardoor worden dan de berekeningen ingewikkelder. Als n niet veel kleiner dan N is, b.v. in de buurt van $\frac{1}{10}N$ komt, dient men hier rekening mee te gaan houden.

4.4. <u>Invloed van de voorwaarden.</u>

De kern van de in hoofdstuk 3 behandelde methoden is, dat wij de onbetrouwbaarheidsdrempel α kunnen interpreteren als een bovengrens voor de fractie van het aantal foute conclusies, dat men in een lange reeks van resultaten ongeveer zal vinden. Deze interpretatie is alleen dan juist, als de steekproeven, waarop deze conclusies gebaseerd zijn, aan de eisen van par. 4.1 voldoen en wel kan men des te meer vertrouwen in deze interpretatie van α stellen naarmate aan de gestelde voorwaarden nauwkeuriger voldaan is.

Ter toelichting hiervan beschouwen wij de in hoofdstuk 1 en 2 gebruikte pot met erwten. Deze pot bevat 1201 erwten, waarvan er 94 rood zijn, zodat voor deze pot de (nu niet meer onbekende) θ gelijk is aan

$$\theta = 0{,}0783.$$

Beschouwen wij nu tabel 3.I, dan zien wij, dat alle in die tabel gegeven bovengrenzen voor θ groter dan de werkelijke waarde zijn, zodat wij voor $n = 25$ bij deze pot erwten steeds tot een juiste conclusie komen. De werkelijke onbetrouwbaarheid is dus voor de bovengrens gelijk aan 0 in plaats van aan $\alpha = 0{,}05$. Dit kan men echter slechts te weten komen, als men θ kent en dan gebruikt men geen steekproeven meer om θ te bepalen. Niettemin blijkt hier uit, dat α inderdaad een bovengrens is voor de fractie fouten, die men kan verwachten, en dat dus de toevoeging "drempel" in de term "onbetrouwbaarheidsdrempel" niet overbodig is.

Wat de ondergrens betreft, is het maken van fouten wel mogelijk. Wij zien uit tabel 3.I, dat hier een foute conclusie getrokken wordt, zodra er in een steekproef van 25 5 of meer rode erwten gevonden worden. Dit is in de twee reeksen (IV) en (V) van 100 steekproeven van fig. 2.2 in totaal 12 maal het geval, zodat de fractie foute conclusies in deze reeks van 200 proeven $\frac{12}{200} = 0{,}06$ bedraagt, hetgeen iets hoger dan 0,05 is, maar toch dicht in de buurt ligt. Met dergelijke afwijkingen van de waarde van α moet men steeds rekening houden, zolang de reeks experimenten niet veel langer dan 200 is.

Voor $n = 100$ (zie tabel 3.II) is bij onze pot erwten de bovengrens nog goed voor $y = 5$ en lagere waarden hebben wij in reeks (VI) (zie fig. 2.3) niet gevonden, zodat hier voor de bovengrens geen foute conclusie gevonden wordt, al is dit voor $n = 100$ wel mogelijk (b.v. als wij $y = 0$, 1 of 2 gevonden hadden; de daarbij behorende bovengrenzen, die niet in tabel 3.II opgenomen zijn, bedragen 0,03, 0,05 en 0,06 en zijn dus $< \theta$, steeds weer: bij de door ons gebruikte pot erwten). De benedengrens is nu fout voor $y \geq 13$ en dit kwam in reeks (VI) 16 maal voor op 100 experimenten in totaal, zodat de fractie fouten in dat geval 0,16 bedroeg, hetgeen aanzienlijk meer dan $\alpha = 0{,}05$ is.

We zien daaruit, dat in dit geval de vroeger gegeven interpretatie van α, nl. als de fractie foute conclusies, niet opgaat.

De oorzaak hiervan is te zoeken in het feit, dat de voorwaarden van par. 4.1 in het geval van onze pot met erwten niet voldoende zijn vervuld. De erwten zijn verschillend van grootte en vermoedelijk zijn de rode erwten gemiddeld groter dan de groene. De grote erwten hebben de neiging meer boven in de pot te blijven hangen, terwijl de kleine gemakkelijk naar beneden zakken. Hierdoor wordt voorwaarde a ernstig verstoord en het resultaat van steekproeven zal sterk afhangen van de wijze, waarop men de steekproef neemt.

Om proefondervindelijk na te gaan of dit inderdaad juist is, hebben wij twee reeksen van 100 steekproeven van 25 laten trekken door twee verschillende proefpersonen, waarvan de eerste de pot op tafel liet staan en, na zorgvuldig roeren, de steekproef nam door het plankje verticaal in de pot te steken en daarna te scheppen terwijl de tweede (onbewust en voor het gemak) de pot in de hand hield, minder zorgvuldig roerde en de steekproef meer uit het bovenste gedeelte van de pot schepte. Het resultaat vindt men in de turfstaatjes van fig. 4.1, waarvan het eerste reeks (IV) van fig. 2.2 voorstelt.

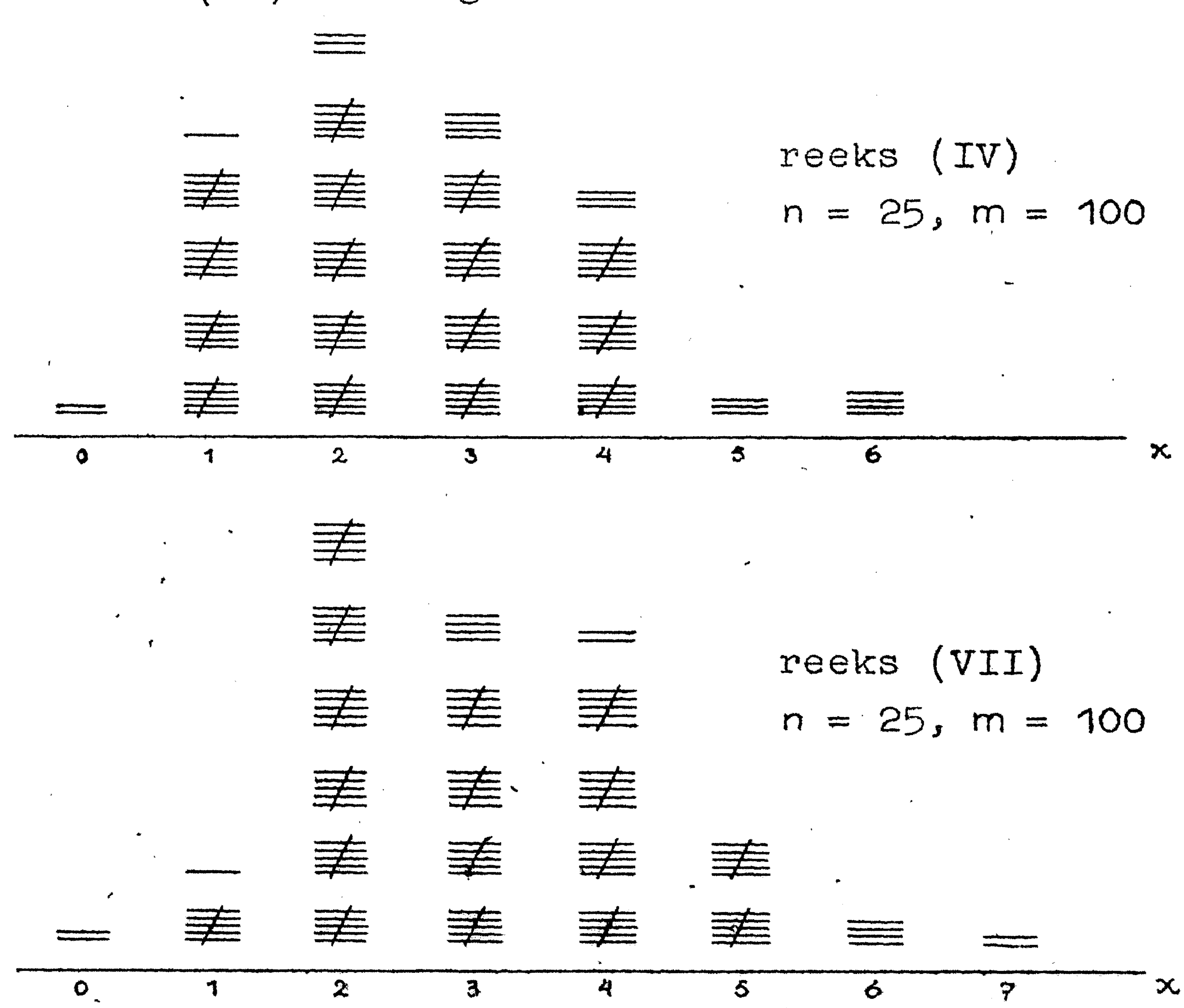


fig. 4.1. Twee reeksen steekproeven, op verschillende wijze genomen.

Het verschil is inderdaad duidelijk te zien. Reeks (VII) bevat aanzienlijk meer steekproeven met veel rode erwten dan reeks (IV) en veel minder met 1 rode erwt. Het aantal foute conclusies in reeks (VII) (in dit geval gelijk aan het aantal malen, dat $x \geqq 5$ gevonden werd) is dan ook zeer groot, nl. 16, dus veel groter dan door $\alpha = 0{,}05$ aangegeven wordt.

In de regel verkeert men niet in de gelukkige positie, waarin men na kan gaan, hoeveel foute conclusies er getrokken worden en derhalve zal men steeds bij het nemen van een steekproef moeten trachten zo zorgvuldig mogelijk aan de in par. 4.1 genoemde voorwaarden te voldoen. De wijze, waarop dit kan geschieden, wordt in de volgende paragraaf besproken.

4.5. Hulpmiddelen voor het nemen van steekproeven.

Uit het voorafgaande is het duidelijk, dat het voor de toepassing van groot belang is, dat aan de voorwaarden a en b zo goed mogelijk voldaan is. Nu blijkt dit bij een pot met erwten reeds op moeilijkheden te stuiten. Men mag dan wel verwachten, dat dit bij andere populaties nog veel moeilijker is. Bij het onderzoek van een menselijke populatie b.v. is "goed roeren" niet letterlijk uitvoerbaar, evenmin trouwens bij de meeste voorwerpen. Men moet dan dus naar hulpmiddelen grijpen.

Zeer vaak bestaat de mogelijkheid om als volgt te werk te gaan. Nummer de elementen van de populatie, waaruit een steekproef genomen moet worden in een willekeurige volgorde. Bij menselijke populaties b.v. alfabetisch-lexicografisch en naar geboortedatum, bij een serieproductie naar de volgorde van fabricage of verpakking, etc. Construeer vervolgens een modelpopulatie met evenveel genummerde elementen (b.v. een pot met allemaal zoveel mogelijk gelijke kogeltjes, die de nummers $1, \ldots, N$ dragen). Controleer of deze model-populatie aan de gestelde eisen voldoet en neem, indien dit het geval is, een steekproef van de gewenste omvang uit deze modelpopulatie. Neem vervolgens uit de te onderzoeken populatie de elementen met dezelfde nummers als in de steekproef uit de modelcollectie en beschouw deze elementen als de gewenste steekproef. Op die wijze is dan toch aan de eisen a en b voldaan en wel in dezelfde mate als de modelcollectie daaraan voldoet. Men noemt dit procédé: trekken door loting of: <u>aselect</u> trekken (aselect = zonder voorkeur). Met enige handigheid kan men volstaan met één modelpopulatie van 10 elementen, genummerd

$$0, 1, 2, \ldots, 9,$$

waaruit men met teruglegging steekproeven van één element trekt. Indien dit procédé aan de eisen a en b voldoet voor $n = 1$ (en dat is zonder al te veel moeite zeer grondig te controleren) en men wenst b.v. een steekproef van 50 elementen uit een populatie met

$N = 1000$ te trekken, dan kan men als volgt te werk gaan. Trek 3 reeksen van 50 cijfers uit de modelcollectie en schrijf deze in 3 kolommen achter elkaar. De 50 op deze wijze verkregen regels zijn getallen ≤ 999, het getal 0 (000) daarbij inbegrepen. Uit de van 1 tot en met 1000 genummerde te onderzoeken populatie nemen wij nu de 50 elementen, die deze 50 nummers dragen, waarbij 000 als 1000 gelezen wordt. Is N niet gelijk aan een macht van 10, dan moet dit procédé enigszins gewijzigd worden.

Opgaven.

4.5.a Gegeven is een modelpopulatie van 10 elementen, die aan eis a en b voldoet. Verder is er een populatie van $N = 900$ elementen, waaruit een steekproef van 20 genomen moet worden, waarbij eveneens aan de eisen a en b voldaan moet worden. Hoe kan men dit bereiken? Beantwoord dezelfde vraag voor $N = 170$.

4.5.b De hierboven beschreven methode leidt tot steekproeven met teruglegging. Welke verandering kan men in de methode aanbrengen om steekproeven zonder teruglegging te verkrijgen?

Wij zien dus, dat men op deze wijze ervoor kan zorgen, dat de voorwaarden a en b vervuld zijn. De modelpopulatie kan bestaan uit 10 fiches in een vaas, genummerd 0,...,9 of uit een tienkantige (cylindervormige) dobbelsteen of tol of iets dergelijks. Deze mechanische methoden hebben echter vaak bezwaren, zoals wij bij de pot met erwten hebben gezien. Er bestaan betere, op wiskundige methoden gebaseerde, wijzen om steekproeven uit een populatie met $N = 10$ te verkrijgen, die zeer goed aan de voorwaarden a en b voldoen. Door toepassing van deze methoden zijn verschillende tabellen van zeer grote steekproeven gemaakt, b.v. M.G.Kendall and B.Babington Smith, "Tables of random sampling numbers", Cambridge University Press 1946. Deze tabel bevat 100.000 trekkingen uit een dergelijke modelpopulatie. Wenst men nu een steekproef van b.v. de omvang 100, dan behoeft men slechts, beginnend op een willekeurig punt in deze tabel, de volgende 100 getallen te nemen. Ook in het boek van Dixon and Massey staan enkele bladzijden van deze "random numbers" (aselecte getallen) opgegeven. Mechanische hulpmiddelen zijn dan overbodig.

Soms is het niet mogelijk de bovenstaande methode toe te passen, eenvoudig omdat het nummeren van de te onderzoeken populatie onmogelijk is (zoals b.v. bij een onderzoek van vissen in een viswater of van de fauna in een bos) of omdat niet alle elementen van de te onderzoeken populatie bereikbaar zijn (indien men b.v. een steekproef uit de lopende productie van een fabriek neemt, kan men geen exemplaren nemen, die reeds verzonden zijn, of die nog niet gefabriceerd zijn). Soms stuit de methode ook op grote practische moeilijkheden, al kan zij in principe wel worden gebruikt. In dergelijke gevallen zoekt men naar een steekproefmethode, waarvan men verwacht, dat toch zo goed mogelijk aan de voorwaarden a en b voldaan is. Wij zullen daar niet verder op in gaan, maar wijzen er slechts op, dat men zich steeds goed rekenschap dient te geven van de noodzaak om, voor zover mogelijk, aan deze voorwaarden te voldoen, indien men tenminste de in de vorige hoofdstukken beschreven eenvoudige statistische technieken toe wil passen. Is aan de voorwaarden slechts gedeeltelijk voldaan, b.v. wel voor aparte delen van de steekproef, maar niet voor het geheel, dan is men bij de verwerking van de gegevens aangewezen op gecompliceerdere statistische methoden dan de hier besprokene.

4.6. Reeksen experimenten.

Wij hebben in de voorafgaande hoofdstukken gezien, dat de statistiek zich bezighoudt met frequentiequotiënten van het optreden van bepaalde gebeurtenissen of kenmerken in reeksen experimenten. In het bijzonder bestonden de experimenten in ons geval uit het trekken van steekproeven (van omvang 1 of groter) uit bepaalde populaties. Daartoe behoeven wij ons echter niet te beperken. Het maakt weinig verschil of wij een lootje trekken uit een pot met 6 lootjes, die de nummers 1,2,...,6 dragen, of dat wij een dobbelsteen op tafel werpen. In beide gevallen is de waarneming, die wij verrichten, één van de getallen 1,2,...,6. Wij kunnen dus, naast het trekken van steekproeven ook andere experimenten beschouwen. Deze kunnen dan trouwens in de regel weer voorgesteld worden als trekkingen uit een populatie.

Voorbeelden.

Een worp met een dobbelsteen kunnen wij beschouwen als een steekproef van de omvang 1 uit een populatie (van willekeurige grootte), bestaande uit in gedachten verrichte worpen met dezelfde dobbelsteen. Een reeks van n worpen met de steen stelt dan een steekproef van omvang n uit die populatie voor.

Een meting van het soortelijk gewicht van een stof met behulp van een bepaalde apparatuur kunnen wij beschouwen als een enkele trekking uit een populatie van vele dergelijke metingen, in gedachten verricht met dezelfde apparatuur.

N.B. In dit soort gevallen zijn steekproeven automatisch steekproeven met teruglegging.

Onder deze voorstellingswijze vallen kortom alle experimenten en waarnemingen, die bij herhaling onder gelijkblijvende omstandigheden verschillende uitkomsten kunnen geven.

In dergelijke gevallen is de populatie veel minder scherp gedefinieerd dan in de vorige hoofdstukken het geval was. De populatie is niet concreet, maar denkbeeldig. Niettemin kunnen wij er steekproeven uit nemen, maar wij kunnen niet meer rechtstreeks controleren of de eis van gelijkwaardigheid (eis a van par. 4.1) vervuld is, daar de populatie niet concreet is en de elementen ervan dus niet meer genummerd kunnen worden. Er is echter wel een indirecte methode. Eis a impliceert nl., dat in twee of meer lange reeksen van trekkingen (in dit geval dus van waarnemingen), de fqn van de verschillende uitkomsten ongeveer met elkaar overeenstemmen. En dit kan men uiteraard experimenteel controleren. Eis b kan dan op soortgelijke wijze gecontroleerd worden: in de deelreeksen, bedoeld onder deze eis, moeten de fqn ongeveer gelijk zijn aan de overeenkomstige fqn in de gehele reeks.

Het enige essentiële verschil met trekkingen uit een concrete (en dus eindige !) populatie is, dat men daaruit zonder teruglegging kan trekken, terwijl dit bij reeksen experimenten niet op natuurlijke wijze mogelijk is. Overigens echter kunnen dit soort experimenten op geheel analoge wijze geanalyseerd worden. De onbekende fractie θ, waarvoor schattingsmethoden besproken zijn, is dan het (onbekende) fq van een bepaald kenmerk (dus: van een bepaalde uitkomst) in de denkbeeldige populatie, dus ook in een zeer lange reeks experimenten. De beschreven schattingsmethoden stellen ons dus in staat om uit een korte reeks waarnemingen een interval af te leiden, waarin dit fq bij een zeer lange reeks zal liggen. Daarbij dient dan het voorbehoud gemaakt te worden, dat de onbetrouwbaarheidsdrempel α alleen dan dezelfde betekenis heeft als vroeger is aangegeven, indien aan de voorwaarden a en b is voldaan.

Wat de voorwaarde b betreft kunnen wij hierbij nog opmerken, dat men vaak reeds op grond van de wijze van uitvoering van het experiment zal kunnen besluiten, dat deze wel vervuld is. Indien men b.v. eën reeks worpen verricht met een dobbelsteen, na schudden in een dobbelbeker, is het niet goed denkbaar, dat de uitkomst van een nieuwe worp door de uitkomsten der vorige worpen beïnvloed zal worden. Dit houdt uiteraard niet in, dat alle uitkomsten 1,...,6 gelijkwaardig moeten zijn, dus dat de dobbelsteen zuiver is. Ook voorwaarde a moet niet als zodanig geïnterpreteerd worden. Een onzuivere dobbelsteen kan precies evengoed aan voorwaarde a voldoen als een zuivere. Of aan a voldaan is hangt nl. af van de wijze van werpen. De precieze formulering van a, waaraan zowel een zuivere als een onzuivere dobbelsteen kan voldoen, is in de vorige alinea te vinden. Bij een <u>concrete</u> populatie, zoals een pot met erwten, spreken wij echter slechts van aselect trekken als alle elementen gelijkwaardig zijn. Is hieraan om de een of andere reden niet voldaan (zoals bij onze pot met erwten het geval bleek te zijn bij een bepaalde techniek van steekproefnemen), dan kan natuurlijk nog best in twee lange reeksen van trekkingen het kenmerk rood ongeveer even vaak voorkomen. In dat geval is aan voorwaarde a voldaan; echter niet voor de concrete populatie, maar voor de denkbeeldige populatie van vele dergelijke trekkingen uit de pot met erwten. Deze gaat dan als het ware over in een ingewikkelde in het algemeen "onzuivere" munt, met een rode en een groene kant. Hetzelfde geldt voor voorwaarde b. Schatten wij nu θ, dan komen wij in zoverre bedrogen uit, dat wij de θ van deze denkbeeldige populatie schatten en niet die van de concrete pot met erwten. En deze twee kunnen sterk verschillen. Dit is de reden, waarom wij bij een concrete populatie de eisen a en b strict formuleren, zoals in het begin van dit hoofdstuk gedaan is.

<u>Hoofdstuk 5</u>.

<u>Waarschijnlijkheidsrekening</u>.

5.1. <u>Het begrip "waarschijnlijkheid"</u>.

De statistiek, als "frequentiequotiënten-rekening", zoals beschreven in de voorafgaande hoofdstukken, is wiskundig moeilijk te hanteren, omdat het woord "ongeveer" er voortdurend in voorkomt. Fundamenteel is nl. steeds, dat fqn in lange reeksen trekkingen "ongeveer" gelijke waarden aannemen. Dit is bijzonder onaangenaam en daarom laten wij deze toevoeging liever tijdelijk weg om hem later weer in te voegen. Het rekenen wordt daardoor veel gemakkelijker.

Indien wij b.v. een pot erwten hebben en indien wij beschikken over een trekkingsmethode van telkens één erwt, die aan de eisen a en b van par. 4.1 voldoet, dan komen dus alle erwten in een lange reeks van trekkingen ongeveer even vaak voor. Wij zeggen dan, dat de <u>kans</u> op het trekken van een bepaalde erwt voor alle erwten even groot is en laten de toevoeging "ongeveer" weg.

Bevat de pot N erwten, dan zal het fq van één bepaalde erwt in een lange reeks trekkingen ongeveer $1/N$ zijn. Wij zeggen dan, dat de <u>kans</u> (of <u>waarschijnlijkheid</u> [1]) van het trekken van die erwt $1/N$ is, met weglating van "ongeveer". Dit geldt dan dus voor iedere erwt.

In het geval van worpen met een dobbelsteen (of van experimenten met een ander toevalsmechanisme, dat ook weer een pot met erwten kan zijn) behoeven niet alle mogelijke uitkomsten gelijkwaardig te zijn. Is echter aan voorwaarden a en b voldaan op de in par. 4.6 beschreven wijze, dan kennen wij weer aan ieder van deze uitkomsten een wh toe, die in de regel onbekend is maar door de fqn van deze uitkomsten in een lange reeks experimenten benaderd wordt. Zijn b.v. bij een dobbelsteen wel alle uitkomsten gelijkwaardig (dus tevens even waarschijnlijk), dan noemen wij deze dobbelsteen <u>zuiver</u>. De wijze van werpen is daarbij inbegrepen.

In het algemeen kennen wij aan het optreden van bepaalde gebeurtenissen of kenmerken bij een gegeven experiment of waarnemingsmethode whn toe, die overeenkomen met de fqn van deze gebeurtenissen (resp. kenmerken) in lange reeksen van deze experimenten, waarbij aan voorwaarden a en b voldaan is, zonder precies gelijk aan die fqn te zijn.

Het is duidelijk, dat whn steeds tussen 0 en 1 liggen.

5.2. <u>Het rekenen met whn.</u>

De wh-rekening houdt zich nu bezig met het berekenen van onbekende whn uit gegeven whn. Daarbij worden enige principes gevolgd, die gewoonlijk in axiomatische vorm gegeven worden, en die wij hier niet uitvoerig zullen bespreken. Deze principes en de daarop gebaseerde rekenregels voor whn komen echter overeen met de rekenwijze, die men met fqn zou volgen en die eenvoudige rekenkundige regels zijn. Wij geven daarom, ter verduidelijking van de methode, enkele voorbeelden, waaruit de algemene methodiek in elementaire vorm duidelijk wordt.

<u>Voorbeelden</u>.

Is een gebeurtenis [2] onmogelijk, dan zal het fq daarvan in iedere reeks experimenten, hoe lang ook, gelijk aan 0 zijn. De bijbehorende wh wordt dan uiteraard gelijk aan 0 gesteld. Is een

---

1) Afkorting: wh, meervoud whn.

2) Wij gebruiken hier steeds het woord "gebeurtenis" tevens in zin van "het optreden van een bepaald kenmerk", en "de uitkomst van een experiment".

gebeurtenis zeker, dan is het fq steeds gelijk aan 1 en de wh wordt dan ook gelijk aan 1 genomen. Beschouwen wij nu worpen met een munt, waarbij twee uitkomsten mogelijk zijn: kruis (K) en munt (M), dan kennen wij dus aan beide uitkomsten een wh toe. Wij geven deze aan met

$$P[K] = p \qquad P[M] = q$$

(P = probability = wh). In iedere reeks worpen is de som der fqn van K en M gelijk aan 1, dus stellen wij ook

(5.2;1) $$p+q=1.$$

Dit geldt eveneens voor experimenten met meer dan twee mogelijke uitkomsten: zijn er h mogelijke uitkomsten $U_1,\dots,U_h$, waarvan er bij iedere uitvoering van het experiment <u>precies</u> 1 optreedt, dan is de som van de fqn van $U_1,\dots,U_h$ gelijk aan 1 en dus wordt dit ook voor de whn geeist.

Geven wij de fqn van U_1 en U_2 (twee uitkomsten uit $U_1,\dots,U_h$) aan met $fq(U_1)$ en $fq(U_2)$, en beschouwen wij nu de samengestelde gebeurtenis

$$V = U_1 \text{ of } U_2,$$

dan geldt in iedere reeks van dergelijke experimenten

$$fq(V) = fq(U_1) + fq(U_2).$$

Derhalve eisen wij ook

(5.2;2) $$P[V] = P[U_1] + P[U_2].$$

Ook deze eigenschap geldt voor meer dan 2 gebeurtenissen, <u>die niet tegelijk op kunnen treden</u>. Neem b.v. een dobbelsteen, met kansen $p_1, p_2, \dots, p_6$ op de uitkomsten 1,2,...,6. Laat W nu zijn: de uitkomst is een even getal (2, 4 of 6 dus), dan is

$$P[W] = p_2+p_4+p_6.$$

In een pot met 1201 erwten, waarvan er 94 rood zijn, en waaruit aselect getrokken wordt, moet dus de kans om bij trekking van één erwt een rode te treffen, gelijk zijn aan $94 \times \frac{1}{1201} = 0{,}0783$. Immers iedere erwt, dus ook iedere rode erwt, heeft wh $\frac{1}{1201}$ om getrokken te worden en er wordt slechts één erwt getrokken. Deze uitkomst kan men ook als volgt beredeneren: het fq van "rood" zal in een lange reeks trekkingen (met teruglegging) van erwten ongeveer gelijk aan $\frac{94}{1201}$ moeten zijn, daar alle erwten ongeveer even vaak getrokken worden (dit zit immers opgesloten in het "aselecte" trekken); derhalve moeten wij de kans op "rood" wel gelijk aan $\frac{94}{1201}$ stellen. Bij aselect trekken komt dus het schatten van de onbekende fractie θ neer op het schatten van de onbekende wh θ van "rood".

Deze eenvoudige eigenschap van optelbaarheid van whn geldt alleen als de beschouwde gebeurtenissen elkaar uitsluiten, d.w.z. niet tegelijk op kunnen treden. In andere gevallen is het iets ingewikkelder. Beschouwen wij b.v. weer een dobbelsteen en nemen wij de volgende twee gebeurtenissen (bij één worp):

$$A \equiv \text{de uitkomst is} \leq 4$$

$$B \equiv \text{de uitkomst is oneven,}$$

dan sluiten deze elkaar niet uit. De whn zijn, volgens het bovenstaande,

$$P[A] = p_1 + p_2 + p_3 + p_4$$

$$P[B] = p_1 + p_3 + p_5 .$$

De kans op de gebeurtenis

$$C \equiv A \quad \text{of/en} \quad B$$

is nu

$$P[C] = p_1 + p_2 + p_3 + p_4 + p_5$$

en dit is kleiner dan $P[A] + P[B]$.

De kans op de gebeurtenis

$$D \equiv A \text{ en } B$$

is

$$P[D] = p_1 + p_3 ,$$

eveneens volgens het bovenstaande. Wij zien dus, dat in dit geval geldt

(5.2;3) $\quad P\,[A \text{ of/en } B] = P[A] + P[B] - P[A \text{ en } B]$.

Deze regel geldt echter geheel algemeen.

Beschouwen wij een experiment, waarbij een gebeurtenis A de kans p bezit om op te treden, dan is dus de kans, dat A niet optreedt gelijk aan $1-p$ (zie (5.2;1)). Voeren wij nu dit experiment tweemaal uit en wel zo, dat de uitkomst van de tweede uitvoering niet van die van de eerste afhangt (<u>onafhankelijke</u> experimenten dus; eis b), dan is de kans op A bij het tweede experiment weer p. De kans, dat wij beide keren A verkrijgen is dan gelijk aan p^2. Immers in een lange reeks experimenten kunnen wij de deelreeks beschouwen van experimenten, uitgevoerd na experimenten met uitkomst A. Deze deelreeks, die zelf ongeveer een fractie p van de gehele reeks beslaat, bevat weer ongeveer een fractie p aan uitkomsten A, volgens eis b. Hieruit valt gemakkelijk in te zien, dat de fractie paren (A,A) in een reeks van paren van onafhankelijke experimenten ongeveer gelijk aan p^2 moet zijn, zodat wij de wh van tweemaal A in twee onafhankelijke experimenten gelijk aan p^2 moeten stellen. Algemener kan men op analoge wijze de volgende eigenschap afleiden. Zijn A en B twee gebeurtenissen, die beide bij een bepaald experiment op kunnen treden, maar die elkaars al of niet optreden niet beïnvloeden, dan geldt

(5.2;4) $$P[A \text{ en } B] = P[A].P[B].$$

In dat geval heten de gebeurtenissen A en B stochastisch onafhankelijk. (5.2;4) geldt analoog voor meer dan 2 gebeurtenissen.

Daar deze kwestie van stochastische onafhankelijkheid, die nauw verwant is aan eis b van hoofdstuk 4, van zeer groot belang is, geven wij enkele voorbeelden van wel en niet onafhankelijke gebeurtenissen.

Voorbeelden.

Bij een reeks worpen met een munt, op de gebruikelijke wijze uitgevoerd, zal in het algemeen aan eis b voldaan zijn. Dit betekent, dat de uitkomst van een volgende worp niet door die van vorige beïnvloed wordt. Noemen wij nu A: de eerste worp geeft kruis, en B: de tweede worp geeft kruis, dan zijn A en B stochastisch onafhankelijk.

Een voorbeeld, waarbij A en B twee uitkomsten zijn, die tegelijk op kunnen treden, zijn de boven voor een dobbelsteen genoemde:

$$A \equiv \text{de uitkomst is } \leq 4$$
$$B \equiv \text{de uitkomst is oneven.}$$

Is de dobbelsteen zuiver, dan is (zie boven)

$$P[A] = \tfrac{2}{3},\ P[B] = \tfrac{1}{2} \text{ en } P[A \text{ en } B] = \tfrac{1}{3},$$

dus aan (5.2;4) is voldaan. A en B zijn dus stochastisch onafhankelijk. Dit kan men, practisch bekeken, ook inzien, door te bedenken, dat kennis over het al of niet optreden van A bij een zuivere dobbelsteen geen kennis verschaft over het al of niet optreden van B. Immers de verhouding van het aantal even tot oneven getallen is onder de getallen 1,...,4 dezelfde als onder 5 en 6 en als onder 1,...,6. Het al of niet optreden van A beïnvloedt dus dat van B niet.

Een voorbeeld van stochastisch afhankelijke gebeurtenissen is het volgende. Laat A het bezit van een auto voorstellen en B het bezit van een rijbewijs. Kiezen wij nu aselect een persoon uit de Nederlandse bevolking, dan is er een zekere kans, dat A vervuld is en eveneens een kans, dat B vervuld is. Is A echter vervuld, dan is B in de meeste gevallen ook vervuld, terwijl B veel minder vaak vervuld zal zijn, indien A niet vervuld is. Deze samenhang tussen A en B heeft ten gevolge, dat (5.2;4) niet opgaat.

Opgaven.

5.2.a. Eén worp met een zuivere dobbelsteen.

$A \equiv$ de uitkomst is ≤ 3,

$B \equiv$ de uitkomst is even.

Zijn A en B stochastisch onafhankelijk?

5.2.b. Als A een gebeurtenis is, die zeker optreedt en B een willekeurige andere, wat kan men dan zeggen over het al of niet onafhankelijk zijn van A en B?

5.2.c. Als A en B elkaar uitsluiten, zijn zij dan stochastisch onafhankelijk of niet?

5.2.d. Beantwoord dezelfde vraag voor het geval, dat B zeker optreedt, indien dit met A het geval is.

5.3. De theoretische wet der grote getallen.

Op grond van de boven fragmentarisch beschreven rekenwijzen kan men nu een zeer belangrijke stelling bewijzen, die wij hier zonder bewijs vermelden. Wij beschouwen een reeks van k onafhankelijke experimenten, die ieder twee mogelijke uitkomsten hebben, die wij met "succes" (S) en "mislukking" (M) aangeven. Daarbij kunnen wij b.v. denken aan statistische voorspellingen of intervalschattingen (zoals de in vroegere hoofdstukken besproken schatting van θ), die goed (S) of fout (M) kunnen zijn. Wij onderstellen verder, dat ieder van deze experimenten zo ingericht is, dat de kans op S gelijk aan p is en de kans op M dus gelijk aan 1-p. Onder een reeks van k experimenten bevinden zich dan een aantal successen, dat wij met $\underline{x}$ aangeven. Dit aantal zal bij een herhaling, d.w.z. bij een nieuwe uitvoering van k experimenten, gewoonlijk niet precies dezelfde waarde aannemen. Op grond van de gemaakte onderstellingen kunnen wij echter voor ieder aantal a tussen 0 en k uitrekenen, hoe groot de kans is, dat $\underline{x}$ deze waarde aanneemt. Wij zeggen dan, dat $\underline{x}$ een wh-verdeling bezit en een dergelijke grootheid noemen wij een stochastische grootheid (Engels: random variable). Dit stochastische karakter geven wij aan door onderstreping van de letter, waarmee wij de grootheid aanduiden. Het fq van S, d.i. de fractie successen

$$\underline{y} = \text{fq}(S) = \frac{\underline{x}}{k} \qquad (5.3;1)$$

is nu ook een stochastische grootheid en wij kunnen dan bewijzen, dat de volgende stelling geldt:

(5.3;2) $\lim_{k\to\infty} P[|y-p|>\varepsilon] = 0$ voor iedere $\varepsilon > 0$.

Dit betekent dus, dat wij een klein (zelfs willekeurig klein) intervalletje $(p-\varepsilon, p+\varepsilon)$ kunnen beschouwen en dat de kans, dat de fractie successen hier buiten zal vallen willekeurig klein gemaakt kan worden door k groot genoeg te nemen.

5.4. Toepassing.

Een kleine kans moet, gezien de wijze van invoering van het begrip wh, wel overeenkomen met een klein fq in een lange reeks waarnemingen en een zeer kleine kans duidt op grote zeldzaamheid. De theoretische wet van de grote getallen stelt ons dus in staat tot voorspellingen, die slechts zeer zelden niet uitkomen. Meer valt met geen enkele wetenschap te bereiken [3]. Wij kunnen nl., als wij voor een bepaalde gebeurtenis (waarvan wij het optreden gemakshalve weer een "succes" noemen) een kans p gevonden hebben, voorspellen, dat in een voldoende lange (en dat kan gespecificeerd worden) reeks experimenten de fractie successen niet meer dan een gegeven klein getal ε van p af zal wijken. De zeldzaamheid van de uitzonderingen wordt daarbij door de lengte der reeks bepaald.

Beschouw nu een reeks van onafhankelijke toepassingen van de in hoofdstuk 1, 2 en 3 beschouwde intervalschattingen. Het "onafhankelijk" betekent in dit geval, dat de toepassingen moeten berusten op stochastisch onafhankelijke groepen van gegevens. De onbetrouwbaarheid van een betrouwbaarheidsinterval is nu de kans op een foute conclusie, dus de kans, dat het interval zo uit zal vallen, dat het de werkelijke waarde van de geschatte parameter niet bevat. Zo worden nl. de formules en de tabellen voor deze intervallen berekend. Wordt nu bij iedere toepassing een onbetrouwbaarheid $\leq \alpha$ (met vaste α) genomen, dan volgt uit de wet der grote getallen, dat de fractie foute conclusies in een reeks onafhankelijke toepassingen slechts zeer zelden boven $\alpha+\varepsilon$ (met bij een lange reeks een zeer kleine ε) zal komen te liggen. Daar wij onbetrouwbaarheden $<\alpha$ wel toelaten ($>\alpha$ niet)

---

3) Zoals bij iedere wetenschap het experiment de proef op de som geeft, moet ook in dit geval experimenteel nagegaan worden, of deze bewering, die uit de theorie voortvloeit, in de practijk ook opgaat. Want in feitehebben wij overal (zie par. 5.1) "ongeveer" verwaarloosd en de gevolgen hiervan zijn theoretisch niet te overzien. Experimenten bevestigen echter het hier gevondene. Zie b.v. J.E.Kerrich, An experimental introduction to the theory of probability, Copenhagen 1950.

kan de fractie mislukkingen wel $< \alpha$ zijn. Dit is dus de betekenis en de achtergrond van het begrip onbetrouwbaarheidsdrempel en het is nu ook duidelijk, waarom aan de voorwaarden van hoofdstuk 4 voldaan moet zijn. Deze liggen nl. ten grondslag aan de whr en de wet der grote getallen zelf, evenals aan de toepassing van de whr bij de afleiding van de formules voor de betrouwbaarheidsintervallen.

Op deze wijze leidt het verwaarlozen van de term "ongeveer" in de experimenteel gevonden wetmatigheden tot een theorie, die een precisering van dit "ongeveer" mogelijk blijkt te maken. Want de wet van de grote getallen geeft juist deze precisering en stelt ons in staat een wh, die wij oorspronkelijk als een theoretische benadering van een fq hebben ingevoerd, in preciezere zin weer als een fq in een lange reeks onafhankelijke waarnemingen te interpreteren.

MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Cursus: "Toegepaste Statistiek".

door

Prof. Dr J. Hemelrijk en Ph. van Elteren.

Hoofdstuk 6 en 7.

S. 120

1954

Hoofdstuk 6

De tekentoets

6.1. Verband met betrouwbaarheidsgrenzen.

In hoofdstuk 3 hebben wij laten zien, hoe men op grond van een steekproef betrouwbaarheidsgrenzen kan construeren voor de fractie elementen van een populatie, die een bepaald kenmerk bezitten. In de practijk komt het vaak voor, dat men een dergelijke intervalschatting feitelijk niet nodig heeft, doch dat men alleen wil weten of genoemde fractie een bepaalde waarde kan hebben of daarvan duidelijk afwijkt. In vele gevallen zal dit de waarde $\frac{1}{2}$ zijn. In het voorbeeld betreffende de Amerikaanse kiezers, genoemd in paragraaf 1.1, is het voor het resultaat van de verkiezingen vooral van belang om te weten of de fractie republikeinse kiezers kleiner dan $\frac{1}{2}$ of groter dan $\frac{1}{2}$ is, dan wel niet duidelijk van $\frac{1}{2}$ verschilt.

In de twee eerste gevallen kan men een verkiezingsoverwinning of nederlaag voorspellen [4]), terwijl men zich in het laatste geval beter niet aan een voorspelling kan wagen. In verband met de terminologie, die wij ingevoerd hebben in hoofdstuk 5, kunnen wij ons probleem voor dit geval ook als volgt formuleren: indien men aselect een Amerikaan trekt, uit de populatie van alle stemgerechtigde Amerikanen, heeft men een kans p, dat deze republikeinsgezind is. Deze kans p is gelijk aan de fractie republikeinen in de populatie. Als de verkiezingen op komst zijn, zal iemand, die wil voorspellen welke partij de overwinning zal behalen, belang hebben bij een methode, waarmee hij kan nagaan of $p \approx \frac{1}{2}$, $p > \frac{1}{2}$ of $p < \frac{1}{2}$ is, terwijl de juiste waarde van p hem minder zal interesseren.

Een ander voorbeeld is het volgende: met behulp van een hier niet verder te beschrijven apparaat wordt de doorslagspanning van een aantal oliemonsters bepaald. Iedere bepaling geschiedt in duplo en het interesseert ons of wellicht door veranderingen in het monster tengevolge van de electrische stroom de tweede bepaling systematisch afwijkt van de eerste. Als er geen verandering optreedt, zal de kans, dat de tweede bepaling groter is dan de eerste gelijk moeten zijn aan $\frac{1}{2}$.

Zowel in dit als in het voorafgaande voorbeeld, laten we de mogelijkheid van onbesliste gevallen (blanco stemmen of gelijke

---

4) Wij maken hierbij enige vereenvoudigende onderstellingen, o.a. dat er maar twee partijen zijn en het systeem van evenredige vertegenwoordiging geldt.

bepalingen) voorlopig buiten beschouwing. Wij zoeken dan naar een methode om aan de hand van een reeks waarnemingen na te gaan of de kans op het optreden van een waarneming, waarbij de tweede bepaling groter is dan de eerste al dan niet gelijk is aan $\frac{1}{2}$. Uiteraard kan de statistiek ons geen methode leveren, waarmee men een zekere uitspraak kan doen.

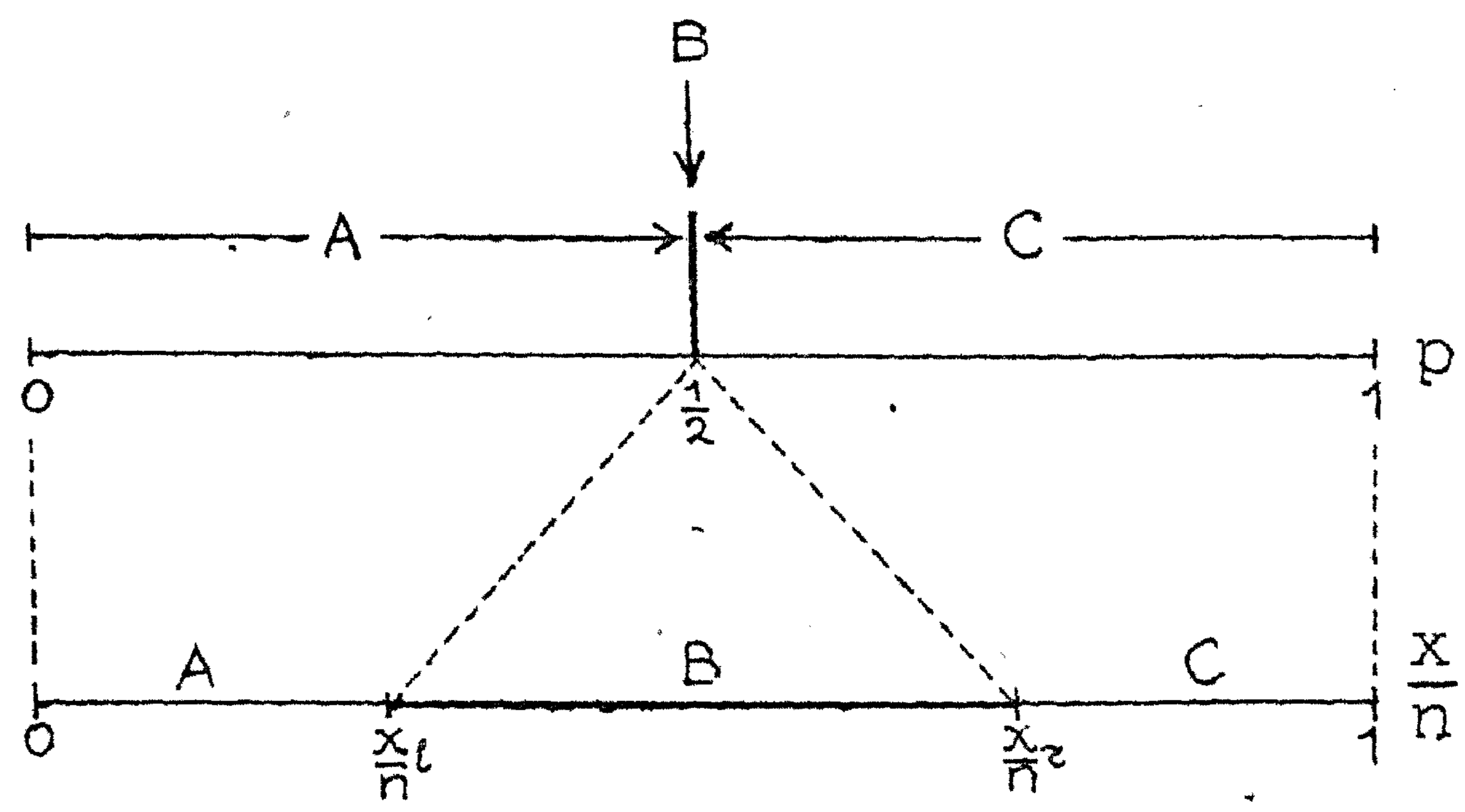
In plaats daarvan stelt men een methode op, waarmee men slechts zelden tot de conclusie komt, dat p niet gelijk is aan $\frac{1}{2}$ indien dit wel het geval is. Een dergelijke methode wordt genoemd een toets voor de hypothese $p = \frac{1}{2}$. De kans, dat men de hypothese $p = \frac{1}{2}$ verwerpt, als zij juist is, wordt de kans op een fout van de eerste soort, of ook wel de onbetrouwbaarheid van de toets genoemd.

Wij kunnen een eenvoudige toets voor de hypothese $p = \frac{1}{2}$ afleiden uit de methode der betrouwbaarheidsgrenzen. Zij n de omvang van de steekproef, en x het aantal malen dat daaronder waarnemingen optreden met het kenmerk K, en wenst men te toetsen $p(K) = \frac{1}{2}$, dan bepalen wij de beide betrouwbaarheidsgrenzen behorende bij een onbetrouwbaarheidsdrempel α (tweezijdig) voor p(K), dus voor de fractie elementen in de populatie met het kenmerk K. Indien nu de waarde $\frac{1}{2}$ buiten deze grenzen ligt, verwerpen wij de hypothese $p = \frac{1}{2}$. Volgens de theorie van de betrouwbaarheidsgrenzen weten wij, dat indien p toch $\frac{1}{2}$ is, men een kans ten hoogste α heeft, dat $p = \frac{1}{2}$ niet tussen de betrouwbaarheidsgrenzen ligt, dus dat de hypothese $p = \frac{1}{2}$ verworpen wordt. De onbetrouwbaarheid van de toets is dus hoogstens α. Evenals bij de betrouwbaarheidsgrenzen kunnen wij hier ook zeggen: de onbetrouwbaarheidsdrempel van de toets is α. De toets voor de hypothese $p = \frac{1}{2}$, die we aldus verkrijgen, wordt de (tweezijdige) tekentoets genoemd. Deze naam is ingevoerd, omdat de toets vaak toegepast wordt op reeksen verschillen van tweetallen waarnemingen, waarbij men de frequentie van de kenmerken positief of negatief onderzoeken wil.

Zoals wij reeds vermeld hebben, wordt de tweezijdige tekentoets toegepast om te beslissen tussen de volgende drie mogelijkheden:

A: $p = \frac{1}{2}$ wordt verworpen ten gunste van $p < \frac{1}{2}$
B: $p = \frac{1}{2}$ wordt niet verworpen
C: $p = \frac{1}{2}$ wordt verworpen ten gunste van $p > \frac{1}{2}$.

We zullen besluiten tot A als $\frac{x}{n}$ zo laag is, dat beide betrouwbaarheidsgrenzen links van $\frac{1}{2}$ liggen, tot B als $\frac{x}{n}$ voldoende dicht bij $\frac{1}{2}$ ligt, zodat de betrouwbaarheidsgrenzen ter weerszijden van $\frac{1}{2}$ liggen, en tot C als $\frac{x}{n}$ zo groot is dat beide grenzen rechts van $\frac{1}{2}$ liggen. Wij kunnen dit schematisch weergeven als in figuur 6.1



mogelijkheden, waartussen beslist moet worden.

criterium voor de beslissing

Fig. 6.1. Beslissingsschema van de tweezijdige tekentoets.

In het onderste gedeelte van deze figuur komt een gebied B voor, dat correspondeert met het punt B in het bovenste gedeelte; men dient hierbij wel te bedenken, dat, indien men een waarde van $\frac{x}{n}$ in het gebied B vindt, de conclusie $p = \frac{1}{2}$ niet gerechtvaardigd is. Immers indien $\frac{x}{n}$ in B ligt, weten we slechts, dat de waarde $\frac{1}{2}$ tussen de bij x behorende betrouwbaarheidsgrenzen ligt, doch ook de andere waarden tussen deze grenzen zijn redelijkerwijze mogelijk. Om die reden hebben wij conclusie B geformuleerd in de vorm: $p = \frac{1}{2}$ wordt niet verworpen.

Indien wij in ons voorbeeld van de oliemonsters bij n = 25, 5 monsters vinden, waarbij de doorslagspanning bij de tweede bepaling lager is dan bij de eerste en 20 oliemonsters, waarbij zij hoger is, dan worden de betrouwbaarheidsgrenzen met onbetrouwbaarheidsdrempel 0,10 (tweezijdig!) volgens tabel 3.I 0,08 en 0,38.

Beide grenzen zijn dan lager dan $\frac{1}{2}$, men zal dus besluiten dat de kans dat de tweede bepaling lager is dan de eerste kleiner is dan $\frac{1}{2}$. Men formuleert dit ook wel door te zeggen: de tweede bepaling is systematisch hoger dan de eerste, hetgeen dus niet betekent, dat dit altijd zo is, maar wèl, dat het, ook in een lange reeks experimenten, vaker wel dan niet het geval zal zijn.

Opgaven.

6.1.a. Welke conclusie zal men bij het bovenbeschreven experiment met de oliemonsters trekken als men n = 25 en x = 10 vindt, wanneer x het aantal monsters is, waarbij de tweede bepaling lager is dan de eerste. Dezelfde vraag wordt gesteld voor n = 25, x = 18, n = 100, x = 17 (tabel 3.II) en n = 100, x = 90 (onbetrouwbaarheidsdrempel 0,10).

6.1.b. In een stad moet een burgemeester gekozen worden, waarvoor twee candidaten gesteld zijn (A en B). Bij een opinieonderzoek vindt men op 1000 ondervraagde kiesgerechtigden, die op aselecte wijze zijn aangewezen, 550 aanhangers van A en 450 van B. Kan men op grond hiervan een voorspelling over de uitslag van de verkiezing wagen als men een onbetrouwbaarheidsdrempel 0,01 kiest?

6.2. Kritieke waarden van de tekentoets, overschrijdingskansen.

De grenzen van het gebied B in het onderste gedeelte van figuur 6.1 worden bepaald door de grootheden, daar aangeduid met x_l en x_r, die de linker resp. rechter kritieke waarde genoemd worden. Deze waarden zijn dus zodanig, dat men besluit tot $p < \frac{1}{2}$ als $x \leqq x_l$ en tot $p > \frac{1}{2}$ als $x \geqq x_r$, terwijl $p = \frac{1}{2}$ niet verworpen kan worden, als $x_l < x < x_r$ is. De beide kritieke waarden liggen symmetrisch t.o.v. het punt $\frac{1}{2}n$; dus als x_l de linker kritieke waarde is, zal $n-x_l$ de rechter kritieke waarde zijn. In tabellen kan men dus met de linker kritieke waarden volstaan. Een dergelijke tabel voor de waarden 1,2,...,100 van n en de onbetrouwbaarheidsdrempels 0,01; 0,02; 0,05; 0,10 zal aan deze syllabus worden toegevoegd. Voor hogere waarden van n kunnen we de kritieke waarden berekenen uit de formules (3.5;1) of (3.5;2). De linker kritieke waarde zal nl. een waarde van x zijn, zodanig dat de bovengrens van het bijbehorende betrouwbaarheidsinterval juist gelijk is aan $\frac{1}{2}$.

Wij stellen dus:

$$\theta^*(x) = \frac{x+\frac{1}{2}c^2+c\sqrt{\frac{x(n-x)}{n}+\frac{1}{4}c^2}}{n+c^2} = \frac{1}{2}$$

en lossen hieruit x op. Men verkrijgt twee oplossingen, nl. $\frac{1}{2}n \pm \frac{1}{2}c\sqrt{n}$. Aangezien men echter weet, dat $x_l < \frac{1}{2}n$ moet zijn, blijft hiervan alleen de oplossing:

$$x_l = \tfrac{1}{2}n - \tfrac{1}{2}c\sqrt{n} \qquad (6.2;1)$$

over.

Evenzo geldt:

$$x_r = \tfrac{1}{2}n + \tfrac{1}{2}c\sqrt{n}. \qquad (6.2;2)$$

De waarden van c, die men bij de onbetrouwbaarheidsdrempels 0,20; 0,10; 0,05 en 0,01 moet gebruiken, vindt men in tabel 3.IV; aangezien daar de éénzijdige onbetrouwbaarheidsdrempels zijn gegeven, zullen we de waarden in de kolom α bij gebruik voor de tweezijdige tekentoets met twee moeten vermenigvuldigen.

Zoals wij boven gezien hebben, is de opgegeven onbetrouwbaarheidsdrempel een bovengrens voor de onbetrouwbaarheid, dus een bovengrens voor de kans dat de hypothese $p = \frac{1}{2}$ verworpen wordt als zij juist is. Nu wordt volgens de tekentoets de hypothese $p = \frac{1}{2}$ verworpen als $\underline{x}$ 5) de waarden x_1 of x_r of verder van $\frac{1}{2}n$ gelegen waarden aanneemt, dus als gezien vanuit het punt $\frac{1}{2}n$, de kritieke waarden bereikt of overschreden worden. De kans dat, indien $p=\frac{1}{2}$ is, $\underline{x}$ een zekere waarde x óf een minstens even ver van $\frac{1}{2}n$ gelegen waarde aanneemt, noemt men de tweezijdige overschrijdingskans van x. Uit deze definitie volgt, dat bij toetsing met een onbetrouwbaarheidsdrempel α, de tweezijdige overschrijdingskans van x_1 of x_r hoogstens gelijk is aan α en voor nog verder van $\frac{1}{2}n$ verwijderde waarden van x nog kleiner. Wij kunnen dus zeggen dat wij de hypothese $p = \frac{1}{2}$ zullen verwerpen, als de bij de gevonden waarde van x behorende overschrijdingskans k niet groter is dan de onbetrouwbaarheidsdrempel α. Wij zullen, indien $k \leq \alpha$ is, met des te meer stelligheid verwerpen, naarmate k kleiner is, immers men kan $p = \frac{1}{2}$ verwerpen bij iedere onbetrouwbaarheidsdrempel, die minstens gelijk is aan k. Anders gezegd: een overschrijdingskans k betekent, dat men, als $p = \frac{1}{2}$ is, slechts in ongeveer een fractie k van een lange reeks toepassingen van de toets een waarde van $\underline{x}$ zal vinden, die nog verder van $\frac{1}{2}n$ verwijderd is dan de gevondene. Naarmate k kleiner is, is de gevonden waarde van x dus een sterkere aanwijzing tegen de getoetste hypothese $p = \frac{1}{2}$, daar men bij $p = \frac{1}{2}$ een waarde van $\underline{x}$ in de buurt van $\frac{1}{2}n$ zou verwachten. Het is om deze redenen dat men bij toepassing van een toets dikwijls niet alleen de conclusie, doch ook de overschrijdingskans opgeeft. Met behulp van de tabel van de kritieke waarden van de tekentoets kunnen wij enige grenzen voor de overschrijdingskans vinden. Indien wij b.v. bij $n = 40$ $x = 13$ vinden, weten wij, dat $0{,}02 < k < 0{,}05$ is, immers de kritieke waarde bij $\alpha = 0{,}02$ is 12, waaruit volgt $k > 0{,}02$ en de kritieke waarde bij $\alpha = 0{,}05$ is $x = 13$, dus $k < 0{,}05$. Wegens de bovengenoemde symmetrie van de kritieke waarden van de tekentoets t.o.v. $\frac{1}{2}n$, geldt hetzelfde voor $x = 40 - 13 = 27$. Verder kan k voor niet te kleine waarden van n benaderd worden met behulp van een tabel van de normale verdeling. Een dergelijke tabel zal eveneens aan deze syllabus worden toegevoegd. Men bepaalt dan de tweezijdige overschrijdingskans van x door

----

5) Aangezien de letter x een stochastische grootheid voorstelt, is zij onderstreept (zie 5.3).

(6.2;3) $$\xi_\alpha = \frac{|x-\frac{1}{2}n|-\frac{1}{2}}{\frac{1}{2}\sqrt{n}}$$ 6)

te berekenen, en de bijbehorende waarde van α in de genoemde tabel op te zoeken. Dan geldt, dat de tweezijdige overschrijdingskans k gelijk is aan 2α. De α in deze tabel heeft in dit geval dus niet de betekenis van een onbetrouwbaarheidsdrempel

Voorbeeld.

n = 60 x = 40

$$\xi_\alpha = \frac{10-\frac{1}{2}}{\frac{1}{2}\sqrt{60}} = 2{,}45.$$

Wij vinden dan in de tabel van de normale verdeling, indien wij zoeken in de rij 2,4 en dan in de kolom waar boven staat 5: α = 0,0071 (alleen de cijfers achter de komma zijn in de tabel weergegeven), dus $k = 2\alpha = 0{,}014$. Deze uitkomst wordt bevestigd door onze tabel van de kritieke waarden van de tekentoets, waar wij vinden dat $0{,}01 < k < 0{,}02$ moet zijn. (Men bedenke daarbij dat de overschrijdingskans van x = 40 dezelfde is als die van x = 20.)

Opgaven.

6.2.a. Hoeveel zal in opgave 6.1.b. bij toetsing met een onbetrouwbaarheidsdrempel 0,01 het aantal aanhangers van één der candidaten minstens van 500 moeten afwijken om tot een voorspelling van de verkiezingsuitslag te kunnen komen. Wat kan men, gegeven de bij deze syllabus gevoegde tabel van de normale verdeling, zeggen over de tweezijdige overschrijdingskans van x = 550.

6.2.b. In 4 reeksen van 25 worpen met een munt worden de volgende resultaten gevonden:

	kruis	munt
1e reeks	16	9
2e reeks	14	11
3e reeks	17	8
4e reeks	15	10

Geef voor iedere reeks afzonderlijk Uw conclusie. Welke conclusie kan men trekken, als men de 4 reeksen tezamen als een reeks van 100 worpen beschouwt? (Onbetrouwbaarheidsdrempel 0,05). Bereken ook voor al deze gevallen een benaderde waarde van de tweezijdige overschrijdingskans met behulp van de normale verdeling.

---

6) $$|x-\tfrac{1}{2}n| = \left\{\begin{matrix} x-\frac{1}{2}n, \text{ als } x > \frac{1}{2}n \\ \frac{1}{2}n-x, \text{ als } x < \frac{1}{2}n \end{matrix}\right\} \text{ is.}$$

De term $-\frac{1}{2}$ in de teller van (6.2;3) wordt de continuïteitscorrectie genoemd, en kan bij grote waarden van n zonder bezwaar weggelaten worden.

6.3. Activeer-stoelen.

Om te kunnen wijzen op een merkwaardige fout, die gemaakt kan worden bij toepassing van de tekentoets (en eveneens bij andere toetsen en statistische analyses in het algemeen), bespreken wij hier een nieuw voorbeeld. Op een groot kantoor werken 100 typisten. Een vertegenwoordiger van een kantoormeubelfabriek doet bij de directie een offerte van zogenaamde "activeer-stoelen". Volgens de vertegenwoordiger zou een op een activeer-stoel gezeten typiste aanzienlijk meer presteren dan een typiste, die op een andere soort kantoorstoel werkt De directie besluit voorlopig 10 activeer-stoelen op proef te nemen. Zij kiest (natuurlijk met behulp van een tabel van aselecte getallen, zie 4.5) een tiental typisten uit haar personeel. De prestaties van deze typisten worden gedurende enige tijd gemeten, b.v. door het aantal per uur getikte regels te tellen. Vervolgens geeft men deze typisten activeer-stoelen en meet wederom hun prestaties. De directie constateert, dat 9 van de 10 typisten op de activeerstoelen meer regels per uur getikt hebben dan op hun vroegere stoelen, terwijl bij één typiste het tegengestelde geconstateerd werd. Met behulp van de tekentoets kan men nu de hypothese toetsen, dat het niets uit maakt of een typiste werkt op een "activeer-stoel" dan wel op een gewone stoel, of wiskundig gepreciseerd: $p = \frac{1}{2}$ als p de kans voorstelt, dat de waargenomen typesnelheid van een typiste werkend op een activeerstoel groter is dan de bij dezelfde typiste op een gewone stoel waargenomen typesnelheid. Deze hypothese moet op grond van het gevonden resultaat verworpen worden, als we werken met een onbetrouwbaarheidsdrempel 0,05. (Zie de tabel van de kritieke waarden der tekentoets.) Men zal hier concluderen dat $p > \frac{1}{2}$ is en dit zal het geval zijn als de activeer-stoelen inderdaad een gunstig effect hebben. Men zou dus op grond hiervan geneigd zijn om de directie te adviseren om voor alle typisten activeer-stoelen te kopen.

Toch is het de vraag of dit advies verantwoord is. Men heeft slechts geconstateerd, dat de typisten beter gingen werken, nadat zij andere stoelen hadden gekregen. De oorzaak van het beter werken behoeft echter niet in de speciale constructie van de stoelen te liggen. Het is mogelijk, dat de prestaties ook verbeterd waren, als men de meisjes stoelen had gegeven, die volgens de vertegenwoordiger van de fabriek wellicht "langzaam-aan"-stoelen zouden moeten worden genoemd. Alléén de verandering zou reeds een tijdelijke verbetering van de prestatie veroorzaakt kunnen hebben. Ook is het mogelijk, dat de dames gevleid zijn door de extra aandacht die aan hen besteed is, zich niet realiserende, dat zij "aselect" gekozen zijn.

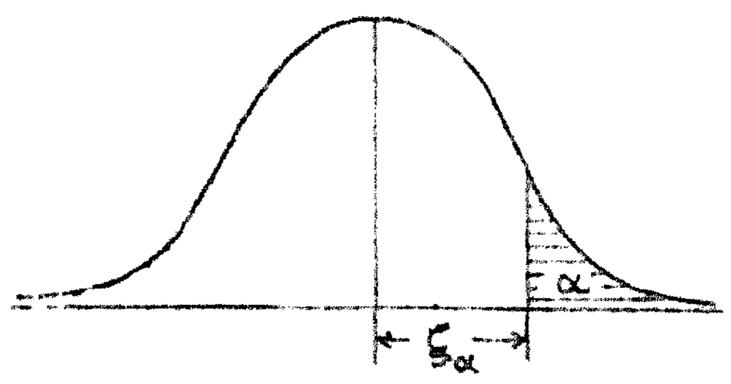
De bedoeling van dit voorbeeld is aan te tonen, dat men met statistische methoden alléén niet de oorzaak van een bepaald verschijnsel kan vaststellen. Men kan hier uit de tekentoets wel concluderen, dat er relatief teveel meisjes beter zijn gaan werken om aan te nemen, dat de verandering van stoelen geen invloed heeft gehad, doch het staat helemaal niet vast dat de verbetering schuilt in het "activeer"-karakter van de nieuwe stoelen. Men kan dan trachten om andere mogelijke oorzaken uit te schakelen, door 10 andere meisjes te kiezen en deze nieuwe stoelen te geven, die in constructie niet afwijken van de oude, maar er een beetje mooier uitzien. Vindt men ook dan een duidelijke tempoverbetering, dan wordt de bijzondere eigenschap van de activeer-stoel in een dubieus licht gesteld.

Bij de hier beschreven proefopzet onderzoekt men alleen het effect van de nieuwe stoelen op korte termijn. Ten aanzien van de te nemen beslissing over het al of niet kopen van activeer-stoelen is het effect op langere termijn van meer betekenis. Men mag aannemen, dat de bovengenoemde psychologische factoren in de regel slechts een tijdelijke verbetering van het tempo kunnen veroorzaken, doch dat als de typisten eenmaal aan de nieuwe situatie gewend zijn, zij weer op hun oude tempo zullen terugvallen, tenzij de "activeer-stoelen" een reële verbetering betekenen. Men kan dus de proefopzet verbeteren door de waarnemingen van het typetempo bij gebruik van "activeer"-stoelen pas te verrichten na een zekere gewenningsperiode. Nog beter is het om de stoelen eerst geruime tijd te laten rouleren, zodat het gehele personeel er mee kennis gemaakt heeft. Pas daarna gaat men metingen verrichten bij een aantal aselect gekozen typisten. Tevens is het beter de typisten niet te laten weten, dat zij waargenomen worden; men vermijdt dan dat zij beter of slechter werken omdat er aandacht aan hen besteed wordt. Men kan bovendien nagaan of er zowel een verbetering optreedt als een typiste eerst op een gewone stoel en daarna op een activeer-stoel werkt, als een vermindering van activiteit in het omgekeerde geval.

Het is ook mogelijk, dat de activeer-stoelen op een deel van het personeel een gunstig effect hebben en op een ander deel een ongunstig effect. Men kan om daarover iets meer te weten te komen, nadat de stoelen enige tijd gerouleerd hebben, b.v. een enquête houden, waarvoor iedere typiste gevraagd wordt of zij

TABEL VAN DE NORMALE VERDELING[1)]

Waarden van $\alpha . 10^4$ voor $\xi_\alpha = 0{,}00(0{,}01)3{,}49$



met $\frac{1}{\sqrt{2\pi}} \int_{\xi_\alpha}^{\infty} e^{-\frac{x^2}{2}} dx = \alpha .$

ξ_α	0	1	2	3	4	5	6	7	8	9
0,0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0,1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0,2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0,3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0,4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0,5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0,6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0,7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0,8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0,9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1,0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1,1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1,2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1,3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1,4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1,5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1,6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1,7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1,8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1,9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
2,0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
2,1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
2,2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
2,3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
2,4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
2,5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
2,6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
2,7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
2,8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
2,9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
3,0	0013	0013	0013	0012	0012	0011	0011	0011	0010	0010
3,1	0010	0009	0009	0009	0008	0008	0008	0008	0007	0007
3,2	0007	0007	0006	0006	0006	0006	0006	0005	0005	0005
3,3	0005	0005	0005	0004	0004	0004	0004	0004	0004	0003
3,4	0003	0003	0003	0003	0003	0003	0003	0003	0003	0002

Een 5 betekent, dat bij afronding naar het dichtsbijgelegen oneven cijfer afgerond moet worden. Een 5 wordt naar het dichtsbijzijnde even cijfer afgerond.

1) Samengesteld met behulp van:
Tabel 1.5 uit: "Techniques of Statistical Analysis" van de Statistical Research Group, Columbia University, 1947.
Tabel I uit: "Tables of the error function and of its first twenty derivatives". The annals of the computation laboratory of Harvard University, Vol. XXIII, Harvard University Press, 1952.

Kritieke waarden van de tweezijdige tekentoets

n	Onbetrouwbaarheidsdrempel 0,01	0,02	0,05	0,10	n	Onbetrouwbaarheidsdrempel 0,01	0,02	0,05	0,10
1	-	-	-	-	51	15	16	18	19
2	-	-	-	-	52	16	17	18	19
3	-	-	-	-	53	16	17	18	20
4	-	-	-	-	54	17	18	19	20
5	-	-	-	0	55	17	18	19	20
6	-	-	0	0	56	17	18	20	21
7	-	0	0	0	57	18	19	20	21
8	0	0	0	1	58	18	19	21	22
9	0	0	1	1	59	19	20	21	22
10	0	0	1	1	60	19	20	21	23
11	0	1	1	2	61	20	20	22	23
12	1	1	2	2	62	20	21	22	24
13	1	1	2	3	63	20	21	23	24
14	1	2	2	3	64	21	22	23	24
15	2	2	3	3	65	21	22	24	25
16	2	2	3	4	66	22	23	24	25
17	2	3	4	4	67	22	23	25	26
18	3	3	4	5	68	22	23	25	26
19	3	4	4	5	69	23	24	25	27
20	3	4	5	5	70	23	24	26	27
21	4	4	5	6	71	24	25	26	28
22	4	5	5	6	72	24	25	27	28
23	4	5	6	7	73	25	26	27	28
24	5	5	6	7	74	25	26	28	29
25	5	6	7	7	75	25	26	28	29
26	6	6	7	8	76	26	27	28	30
27	6	7	7	8	77	26	27	29	30
28	6	7	8	9	78	27	28	29	31
29	7	7	8	9	79	27	28	30	31
30	7	8	9	10	80	28	29	30	32
31	7	8	9	10	81	28	29	31	32
32	8	8	9	10	82	28	30	31	33
33	8	9	10	11	83	29	30	32	33
34	9	9	10	11	84	29	30	32	33
35	9	10	11	12	85	30	31	32	34
36	9	10	11	12	86	30	31	33	34
37	10	10	12	13	87	31	32	33	35
38	10	11	12	13	88	31	32	34	35
39	11	11	12	13	89	31	33	34	36
40	11	12	13	14	90	32	33	35	36
41	11	12	13	14	91	32	33	35	37
42	12	13	14	15	92	33	34	36	37
43	12	13	14	15	93	33	34	36	38
44	13	13	15	16	94	34	35	37	38
45	13	14	15	16	95	34	35	37	38
46	13	14	15	16	96	34	36	37	39
47	14	15	16	17	97	35	36	38	39
48	14	15	16	17	98	35	37	38	40
49	15	15	17	18	99	36	37	39	40
50	15	16	17	18	100	36	37	39	41

een activeerstoel prefereert boven een gewone stoel, een gewone stoel boven een activeer-stoel, of in deze geen mening heeft. Men kan dan voor ieder van deze groepen apart onderzoeken of de activeer-stoel een gunstig effect heeft en de stoel alleen voor die groepen kopen, waarbij men dit gunstige effect ontdekt.

Bij een dergelijk onderzoek bestaan uiteraard nog vele mogelijkheden, die hier niet besproken zijn, zowel wat de methode van waarnemen als van statistische verwerking betreft. Overwegingen als hierboven geschetst zullen daarbij echter steeds van belang zijn.

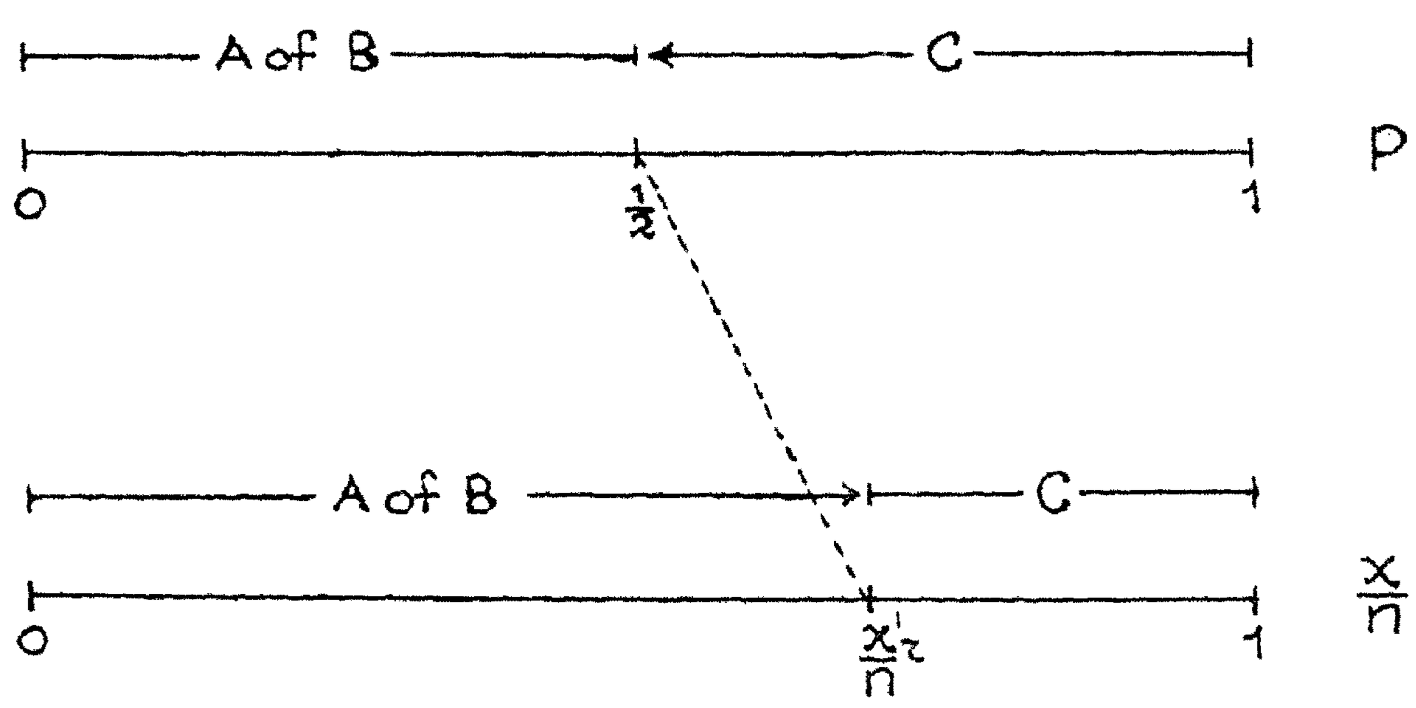
6.4. Eénzijdige tekentoets.

In 6.1 hebben wij uitgelegd dat de tweezijdige tekentoets een statistische methode is om te beslissen tussen drie mogelijkheden:

A: $p < \frac{1}{2}$
B: $p = \frac{1}{2}$ (eigenlijk $p \approx \frac{1}{2}$)
C: $p > \frac{1}{2}$.

Vaak komt het echter voor, dat wij kunnen volstaan met te beslissen tussen de mogelijkheden (A of B) en C of tussen A en (B of C). In ons voorbeeld van de "activeerstoelen" uit 6.3 zal de directie deze stoelen alleen willen kopen als blijkt, dat zij een duidelijk gunstig effect hebben op het werktempo. Dit gunstige effect correspondeert met $p > \frac{1}{2}$, als p wederom de kans voorstelt, dat een typiste op een activeerstoel sneller werkt dan op een gewone stoel. De mogelijkheden $p < \frac{1}{2}$ en $p = \frac{1}{2}$ behoeven eigenlijk niet onderscheiden te worden, want zij corresponderen beide met het advies: "activeerstoelen niet kopen".

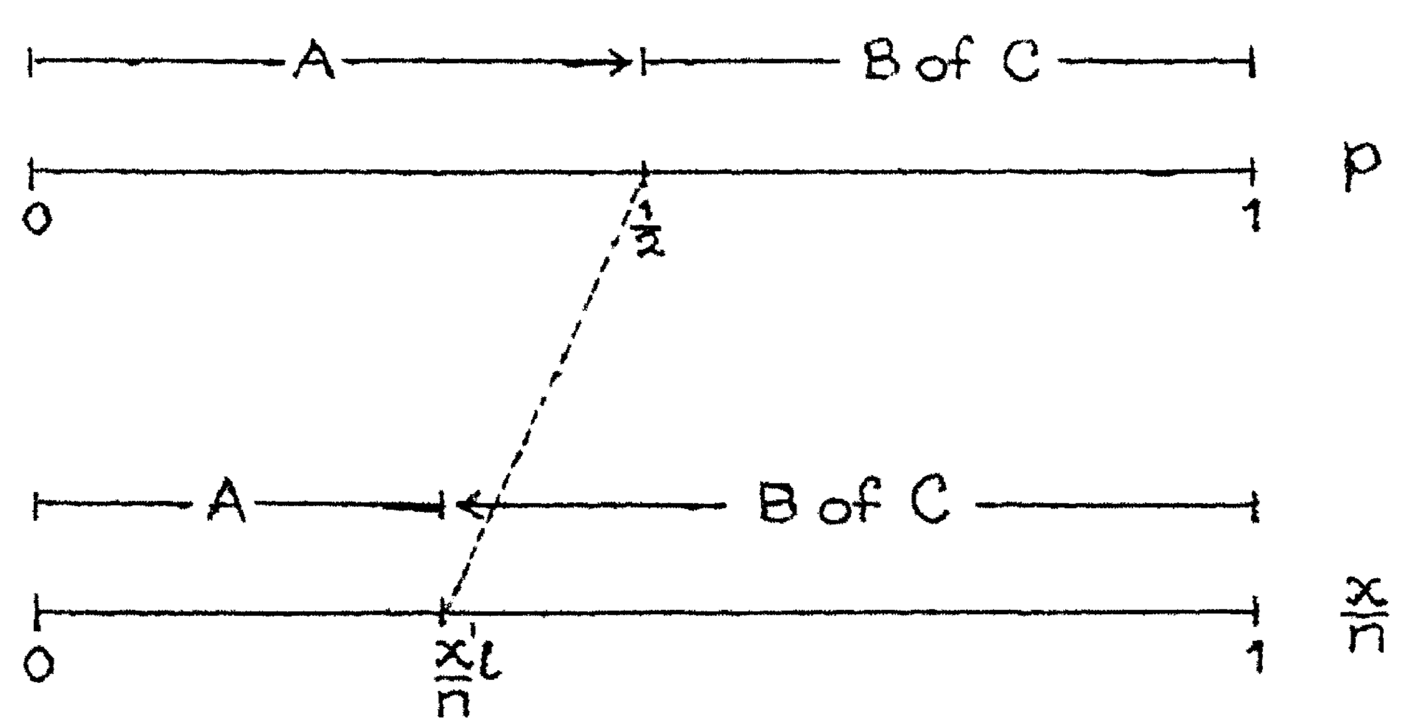
In dergelijke gevallen is het efficiënter om er ook bij het opstellen van de toets rekening mee te houden, dat men slechts twee beslissingsmogelijkheden nodig heeft. Indien wij willen beslissen tussen (A of B) en C, passen wij daarom bij voorkeur de éénzijdige tekentoets toe, waarbij $p \leqq \frac{1}{2}$ getoetst wordt tegen $p > \frac{1}{2}$. Er zijn twee éénzijdige toetsen; de rechtseenzijdige toets, waarbij $p \leqq \frac{1}{2}$ getoetst wordt tegen $p > \frac{1}{2}$ en de linkseenzijdige toets, waarbij $p \geqq \frac{1}{2}$ getoetst wordt tegen $p < \frac{1}{2}$. Het beslissingsschema voor de rechtseenzijdige tekentoets is aangegeven in fig. 6.2 en dat ~~van~~ voor de linkseenzijdige toets in fig. 6.3



Mogelijkheden, waartussen beslist moet worden

Kriterium voor de beslissing

Fig. 6.2. Beslissingsschema voor de rechtseenzijdige tekentoets.



Mogelijkheden, waartussen beslist moet worden

Kriterium voor de beslissing

Fig. 6.3. Beslissingsschema voor de linkseenzijdige tekentoets.

Wij kiezen bij de rechtseenzijdige tekentoets het interval C in het onderste deel van fig. 6.2 (welk interval de rechtse kritieke zone genoemd wordt) zodanig, dat de kans, dat men in dit interval terecht komt, als $p = \frac{1}{2}$ is, hoogstens gelijk is aan de gekozen onbetrouwbaarheidsdrempel α. Men bereikt daardoor, dat de kans, dat men in de rechtse kritieke zone terecht komt, als $p < \frac{1}{2}$ is, zeker kleiner zal zijn dan α, daar de kans op veel "successen" bij $p < \frac{1}{2}$ natuurlijk kleiner is dan bij $p = \frac{1}{2}$. Daaruit volgt dat de kans, dat de conclusie $p > \frac{1}{2}$ getrokken wordt, als $p \leqq \frac{1}{2}$ geldt, hoogstens gelijk is aan de onbetrouwbaarheidsdrempel. Evenzo is in het onderste gedeelte van fig. 6.3 voor de linkseenzijdige toets het interval A (de linkse kritieke zone) zodanig gekozen, dat als $p \geqq \frac{1}{2}$ is de kans, dat men in die zone terecht komt hoogstens gelijk is aan α. De rechtse kritieke zone wordt bepaald door $\frac{x'_r}{n}$; daarom wordt x'_r de rechtse kritieke waarde genoemd. Evenzo wordt de linkse kritieke zone begrensd door een linkse kritieke waarde x'_l.

De beslissingsschema's voor de beide éénzijdige toetsen liggen intuïtief voor de hand. Bij een waarde van $p \leqq \frac{1}{2}$, dus in het interval (A of B) in het bovenste deel van fig. 6.2, zal men geen hoge waarden van $\frac{x}{n}$ verwachten en men vindt waarden van $\frac{x}{n}$ minder in overeenstemming met de hypothese $p \leqq \frac{1}{2}$, naarmate zij meer rechts van $\frac{1}{2}$ liggen. De rechtseenzijdige overschrijdingskans van x wordt in verband hiermee gedefinieerd als de kans dat $\underline{x}$ de waarde x bereikt of overschrijdt, berekend onder de hypothese $p = \frac{1}{2}$. Deze kans wordt in formulevorm aangegeven door

$$k_r(x) = P[\underline{x} \geq x \mid p = \tfrac{1}{2}].$$

Zolang $\frac{x}{n} < \frac{1}{2}$ is zal de rechtseenzijdige overschrijdingskans dus groter dan $\frac{1}{2}$ zijn; naarmate $\frac{x}{n}$ groter wordt, wordt deze overschrijdingskans kleiner. De rechtse kritieke waarde x'_r ligt zodanig, dat $k_r(x'_r)$ niet groter dan de gekozen onbetrouwbaarheidsdrempel α is, terwijl $k_r(x'_r - 1) > \alpha$ is. Men zal op grond van een gevonden waarde $x > x'_r$ de hypothese $p \leq \frac{1}{2}$ met meer stelligheid verwerpen naarmate zij groter is, dus naarmate $k_r(x)$ kleiner is. Met de linkseenzijdige tekentoets, correspondeert een linkseenzijdige overschrijdingskans $k_l(x)$, waarvoor analoog aan het voorafgaande geldt:

$$k_l(x) = P[\underline{x} \leq x \mid p = \tfrac{1}{2}]$$

$$k_l(x) \geq \tfrac{1}{2} \text{ als } \tfrac{x}{n} \geq \tfrac{1}{2}$$

$$k_l(x'_l) \leq \alpha \text{ en } k_l(x'_l + 1) > \alpha.$$

In fig. 6.4 zijn de rechtseenzijdige overschrijdingskans $k_r(6)$, de linkseenzijdige overschrijdingskans $k_l(6)$ en de tweezijdige overschrijdingskans $k(6)$, van $x = 6$ weergegeven, voor de tekentoets met $n = 8$. Daartoe is voor ieder van deze drie gevallen de waarschijnlijkheidsverdeling van $\underline{x}$ onder de hypothese $p = \frac{1}{2}$ getekend in de vorm van een histogram.

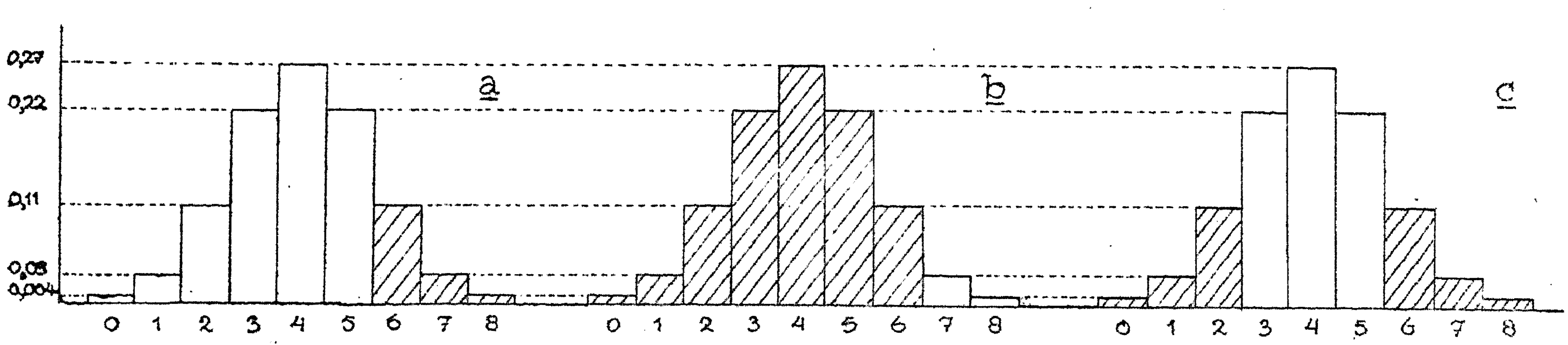


Fig. 6.4. Overschrijdingskansen van $x = 6$ bij de tekentoets ($n = 8$), gegeven door de gearceerde gedeelten der figuren: a rechtseenzijdig, b linkseenzijdig, c tweezijdig.

Opgave 6.4.a. Teken voor de tekentoets met $n = 8$ de overschrijdingskansen van $x = 1$.

Men kan bij figuur 6.4 opmerken, dat de tweezijdige overschrijdingskans van x voor de tekentoets gelijk is aan tweemaal de kleinste der eenzijdige overschrijdingskansen behorende bij dezelfde x.

Hierin schuilt het voordeel van de éénzijdige toetsing, indien men met slechts twee beslissingsmogelijkheden te doen heeft. Indien men bijvoorbeeld $p \leqq \frac{1}{2}$ tegen $p > \frac{1}{2}$ wil toetsen, en men vindt voor x een waarde die groter is dan $\frac{1}{2}n$, dan is de rechtseenzijdige overschrijdingskans daarvan de helft van de tweezijdige overschrijdingskans en men komt dus bij de rechtseenzijdige toets bij lagere waarden van x tot de conclusie $p > \frac{1}{2}$ dan bij de tweezijdige toets het geval is. Men formuleert dit voordeel gewoonlijk als volgt: de rechtseenzijdige tekentoets heeft, voor $p > \frac{1}{2}$, een groter onderscheidingsvermogen dan de tweezijdige; evenzo heeft de linkseenzijdige tekentoets een groter onderscheidingsvermogen voor $p < \frac{1}{2}$.

De kritieke waarden voor de eenzijdige toetsen kunnen zeer gemakkelijk uit de hierbij gevoegde tabel van de tweezijdige toets gevonden worden. In deze tabel zijn uitsluitend linkse kritieke waarden opgegeven. Men vindt bijvoorbeeld bij $n = 30$ dat de linkse kritieke waarde van de tweezijdige toets bij $\alpha = 0{,}10$ gelijk is aan 10. Dat betekent dus dat de tweezijdige overschrijdingskans van $x = 10$ niet groter is dan 0,10; de linkseenzijdige overschrijdingskans is dus niet groter dan 0,05; de linkse kritieke waarde van de eenzijdige toets bij $\alpha = 0{,}05$ is dus 10. De tabel van de kritieke waarden van de tweezijdige toets kan dus gebruikt worden voor de linkseenzijdige toets, mits de onbetrouwbaarheidsdrempels gehalveerd worden. De rechtseenzijdige kritieke waarden vindt men door de linkseenzijdige waarden van n af te trekken. Uit bovengenoemde tabel kan dus voor $n = 30$ worden afgeleid

α =	0,005	0,01	0,02	0,025	0,05	0,10
linkse kritieke waarde, tweezijdig	-	7	8	-	9	10
rechtse " " , "	-	23	22	-	21	20
linkse " " , éénzijdig	7	8	-	9	10	-
rechtse " " , ~~twee~~ één zijdig	23	22	-	21	20	-

Evenzo kan men de benaderingsformule (6.2;2) voor de rechtse kritieke waarde van de tweezijdige tekentoets gebruiken voor de kritieke waarde van de rechtseenzijdige tekentoets, indien men de onbetrouwbaarheidsdrempel halveert, dus juist de in tabel 3.IV opgegeven waarden aanhoudt Dit geldt analoog voor de bepaling van de kritieke waarde der linkseenzijdige toets met behulp van formule (6 2;4)

Indien men eenzijdige overschrijdingskansen wil bepalen met behulp van een tabel van de normale verdeling, gaat men als volgt te werk. Om de rechtseenzijdige overschrijdingskans te vinden berekent men

(6.4;1)
$$\xi'_\alpha = \frac{x - \frac{1}{2}n - \frac{1}{2}}{\frac{1}{2}\sqrt{n}}$$

en voor de linkseenzijdige overschrijdingskans

(6.4;2)
$$\xi''_\alpha = -\frac{x - \frac{1}{2}n + \frac{1}{2}}{\frac{1}{2}\sqrt{n}} .$$

Indien ξ'_α resp. ξ''_α positief is, dan is de rechts- resp. links-eenzijdige overschrijdingskans gelijk aan de bijbehorende waarde van α, die men vindt in de tabel van de normale verdeling; indien daarentegen ξ'_α resp. ξ''_α negatief is, dan bepaalt men de α behorende bij $-\xi'_\alpha$ resp. $-\xi''_\alpha$ en dan is de rechts- resp linkseenzijdige overschrijdingskans gelijk aan $1-\alpha$. Zo vindt men b.v. bij $n = 36$, $x = 26$:

$$\xi'_\alpha = \frac{26 - 18 - \frac{1}{2}}{\frac{1}{2}\sqrt{36}} = 2\tfrac{1}{2} \quad \text{en} \quad \xi''_\alpha = -\frac{26 - 18 + \frac{1}{2}}{\frac{1}{2}\sqrt{36}} = -2\tfrac{5}{6} = -2{,}83$$

ξ'_α is positief, men vindt hierbij $\alpha = 0{,}0062$, dus de rechtseenzijdige overschrijdingskans is 0,0062. ξ''_α is negatief, bij $-\xi''_\alpha$ behoort $\alpha = 0{,}0023$, dus de linkseenzijdige overschrijdingskans is 0,998 Zoals wij reeds opgemerkt hebben bij fig. 6 4 is de tweezijdige overschrijdingskans van x gelijk aan twee maal de kleinste der eenzijdige overschrijdingskansen, dus hier: 2 x 0,0062 = 0,012 Het is gemakkelijk na te gaan, dat we deze tweezijdige overschrijdingskans ook gevonden zouden hebben, zo we te werk gegaan waren als beschreven is bij formule (6.2;3).

<u>Opgaven</u>.

6.4.b. Bepaal bij de tekentoets voor $n = 100$

1) De kritieke waarden voor de tweezijdige toets voor $\alpha = 0{,}01$ en 0,05.
2) De kritieke waarden van de links- en rechtseenzijdige toetsen bij $\alpha = 0{,}01$ en 0,05.
3) De één- en tweezijdige overschrijdingskansen van genoemde kritieke waarden.

Bij vraag 1 en 2 dient men de tabel van de kritieke waarden te gebruiken, bij vraag 3 de tabel van de normale verdeling. Vergelijk de overschrijdingskansen met de bij de kritieke waarden behorende onbetrouwbaarheidsdrempels.

6.4.c. In de gemeente Amsterdam werden in het jaar 1952 14896 kinderen geboren, waaronder 7635 jongens. Toets met behulp van de tekentoets de hypothese, dat een kans op een jongensgeboorte gelijk is aan de kans op een meisjesgeboorte. Verdient één- of tweezijdige toetsing hier de voorkeur?

6.4.d. Een dierenhandelaar krijgt een offerte van een nieuw soort kattenbrood (B), dat iets duurder is dan het kattenbrood, dat hij tevoren steeds verkocht heeft (soort A). Om te beslissen of hij soort B zal kopen neemt hij een proef met 30 katten, die hij tegelijkertijd een bakje met A en een bakje met B voorzet. Hij wil soort B alleen kopen als blijkt, dat B beter is dan A. Bij de proef bleken 10 katten A te prefereren boven B, terwijl 20 katten meer B aten. Wat zal volgens de tekentoets de beslissing van de handelaar zijn, indien hij een kans 0,05 wil riskeren om de offerte te accepteren als A niet slechter is dan B.

6.5. Onbesliste experimenten.

In dit hoofdstuk hebben wij steeds aangenomen, dat er bij de toepassing van de tekentoets geen onbesliste gevallen optraden. Indien het aantal onbesliste gevallen relatief klein is, kan men deze het beste buiten beschouwing laten en op de overige gevallen de tekentoets toepassen. Deze methode is altijd beter dan die, waarbij men de helft van de onbesliste gevallen positief en de helft negatief telt, een methode die ook wel eens voorgesteld wordt. Bij deze laatste methode is de kans dat men de getoetste hypothese verwerpt, als zij niet juist is, kleiner. Volgens de boven reeds vermelde terminologie kunnen wij dus zeggen: de laatste methode heeft een kleiner onderscheidingsvermogen. Wij zullen in het volgende hoofdstuk voor een meer algemeen geval aangeven, welke weg men kan volgen, indien het aantal onbesliste gevallen relatief groot is.

Hoofdstuk 7

Binomiale toetsen

7.1. Inleiding.

In hoofdstuk 6 hebben wij beschreven, hoe men de hypothesen kan toetsen, dat de fractie p der elementen met een bepaald kenmerk K in een populatie gelijk is aan $\frac{1}{2}$, hoogstens gelijk is aan $\frac{1}{2}$ of minstens gelijk is aan $\frac{1}{2}$, respectievelijk met behulp van de tweezijdige, de rechtseenzijdige en de linkseenzijdige tekentoets. Deze methoden kunnen toegepast worden, indien men uit de populatie een steekproef genomen heeft, die voldoet aan de in hoofdstuk 4 genoemde eisen en het aantal x der elementen met kenmerk K daarin bepaald heeft. Men kan desgewenst op grond van dezelfde gegevens ook hypothesen $p = p_0 \neq \frac{1}{2}$, $p \leq p_0$ of $p \geq p_0$ toetsen. Deze toetsen vertonen grote analogie met de tekentoetsen. Zij worden hier binomiale toetsen genoemd, omdat de waarschijnlijkheidsverdeling van de toetsingsgrootheid $\underline{x}$ een zogenaamde binomiale verdeling is. De tekentoets is een bijzonder geval ($p_0 = \frac{1}{2}$) van de binomiale toets. Wij zullen hier volstaan met het opgeven van de nodige formules en verwijzen voor de verdere interpretatie naar het vorige hoofdstuk. In aansluiting hierop zullen wij een methode bespreken voor behandeling van steekproeven met relatief veel onbesliste experimenten, die bij deze toets en bij de tekentoets gebruikt kan worden.

7.2. Tweezijdige binomiale toets.

De methode, die wij in paragraaf 6.1 beschreven hebben om de hypothese $p = \frac{1}{2}$ tweezijdig te toetsen, kan in principe ook gebruikt worden om een hypothese $p = p_0 \neq \frac{1}{2}$ te toetsen. Men zal daartoe slechts betrouwbaarheidsgrenzen voor p behoeven te bepalen op grond van het gegeven waarnemingsmateriaal. Als het interval tussen boven- en ondergrens p_0 bevat, zal men de hypothese $p = p_0$ niet verwerpen. Indien blijkt, dat beide grenzen kleiner zijn dan p_0, zal men $p = p_0$ verwerpen ten gunste van $p < p_0$; indien daarentegen beide grenzen groter zijn dan p_0, zal men $p = p_0$ verwerpen ten gunste van $p > p_0$.

Voor waarden van p_0, die niet te ver van $\frac{1}{2}$ afwijken en voldoende grote waarden van n kan men wederom de kritieke waarden benaderen door x op te lossen uit

$$\frac{x + \frac{1}{2}c^2 \pm c\sqrt{\frac{x(n-x)}{n} + \frac{1}{4}c^2}}{n + c^2} = p_0 .$$

Men vindt dan:

(7.2;1) $$x_l = p_o n - c\sqrt{np_o(1-p_o)}$$

en

(7.2;2) $$x_r = p_o n + c\sqrt{np_o(1-p_o)}.$$

De vereiste waarden van c bij diverse onbetrouwbaarheidsdrempels vindt men dan in tabel 3.IV. Men dient bij tweezijdige toetsing rekening te houden met het feit, dat genoemde tabel de éénzijdige onbetrouwbaarheidsdrempels vermeldt, die dus bij tweezijdige toetsing met 2 vermenigvuldigd moeten worden.

Men kan onder de bij de kritieke waarden genoemde voorwaarden, tweezijdige overschrijdingskansen benaderen met behulp van een tabel van de normale verdeling. Daartoe berekenen wij:

(7.2;3) $$\xi_\alpha = \frac{|x - p_o n| - \frac{1}{2}}{\sqrt{np_o(1-p_o)}}$$

en zoeken de bijbehorende waarde van α op in de (bij deze syllabus gevoegde) tabel van de normale verdeling. De tweezijdige overschrijdingskans k is dan gelijk aan 2α. Wij vestigen er hier nogmaals de aandacht op, dat α in deze tabel geen onbetrouwbaarheidsdrempel voorstelt.

Opgave 7.2.a.

Pas de tweezijdige binomiale toets toe op de volgende gegevens:

$$n = 10.000 \qquad p_o = 0{,}43$$
$$x = 3.578 \qquad \alpha = 0{,}05$$

Bepaal tevens de tweezijdige overschrijdingskans.

7.3. Eénzijdige binomiale toetsen.

Volkomen analoog aan hetgeen wij in paragraaf 6.4 beschreven hebben, zullen wij de hypothese $p = p_o$ tweezijdig toetsen, indien wij willen beslissen tussen de mogelijkheden $p < p_o$, $p = p_o$ en $p > p_o$. Indien wij daarentegen $p \leqq p_o$ willen toetsen tegen $p > p_o$ zullen wij een rechtseenzijdige toets prefereren en voor het geval $p \geqq p_o$ tegen $p < p_o$ een linkseenzijdige toets Kritieke waarden en overschrijdingskansen worden op analoge wijze gevonden. De kritieke waarde x_l' van de linkseenzijdige toets is dus gelijk aan het rechterlid van formule (7.2;1), de onbetrouwbaarheidsdrempels zijn nu echter dezelfde als vermeld in tabel 3.IV; evenzo is de kritieke waarde x_r' van de rechtseenzijdige toets te vinden uit formule (7.2;2) en tabel 3.IV.

De eenzijdige overschrijdingskansen kan men met behulp van een tabel van de normale verdeling benaderen, zoals beschreven is in paragraaf 6.4. Men dient dan echter ξ'_α en ξ''_α niet te berekenen uit de formules (6.4;1) en (6.4;2), doch uit de volgende:

(7.3;1) $$\xi'_\alpha = \frac{x - p_0 n - \frac{1}{2}}{\sqrt{n p_0 (1-p_0)}}$$

en

(7.3;2) $$\xi''_\alpha = - \frac{x - p_0 n + \frac{1}{2}}{\sqrt{n p_0 (1-p_0)}} .$$

Zowel voor de bepaling van de kritieke waarden als van de eenzijdige overschrijdingskansen moeten wederom de voorwaarden gesteld worden dat p_0 niet te ver van $\frac{1}{2}$ afwijkt en n voldoende groot is.

Opgave 7.3.a. Bepaal bij de gegevens uit opgave 7.2.a eveneens de rechts- en linkseenzijdige overschrijdingskansen.

7.4. Onbesliste experimenten bij binomiale toetsen.

Wij willen nu het geval beschouwen, waarbij in principe voor ieder element van de populatie, die wij beschouwen, drie mogelijkheden bestaan, te weten:

1. Het element heeft het kenmerk K .
2. Het element heeft het kenmerk K niet.
3. Wij weten niet, of het element het kenmerk K al dan niet bezit, Het is een grensgeval.

Vaak zullen deze grensgevallen voortspruiten uit de onnauwkeurigheid van de meting of de waarnemingsmethode. In het voorbeeld van de meting van de doorslagspanning van oliemonsters (zie par. 6.1) behoren de monsters, waarbij de tweede bepaling gelijk is aan de eerste tot de grensgevallen. Gelijk zijn houdt in dit geval meestal in: voor zover de meetnauwkeurigheid van het experiment toelaat is er geen verschil te bespeuren tussen de eerste en de tweede bepaling. Een ander voorbeeld van dergelijke grensgevallen hebben we bij een verkiezingsenquête in degenen, die nog niet weten welke partij zij zullen kiezen, of die hun mening niet kenbaar willen maken. Ook deze gevallen kunnen gedeeltelijk voorkomen uit de onnauwkeurigheid van de waarneming; het zal van de bekwaamheid van de enqueteur afhangen in hoeverre de mening van weinig mededeelzame mensen achterhaald wordt.

Wij beschouwen nu een populatie, waarin de fractie elementen met kenmerk K gelijk is aan p_1, de fractie elementen, die kenmerk K niet bezitten gelijk is aan p_2 en de fractie grensgevallen gelijk aan p_3. Er wordt ondersteld dat p_1, p_2 en p_3 onbekende constanten zijn, waarvoor uiteraard moet gelden: $p_1 + p_2 + p_3 = 1$, (zie par. 5.2.). Indien wij de onbesliste experimenten buiten beschouwing laten, zoals wij in par. 6.4 hebben aanbevolen, en vervolgens een der toetsen beschreven in par. 7.2 of 7.3 toepassen, toetsen wij in feite of $\frac{p_1}{p_2}$ een bepaalde waarde kan hebben, of dat $\frac{p_1}{p_2}$ een bepaalde waarde niet overschrijdt.

Wij willen nu echter tevens rekening houden met de grensgevallen, en wel in die zin, dat wij geen conclusie meer trekken betreffende $\frac{p_1}{p_2}$ indien p_3 te groot is. Hoe we dit kunnen bereiken zal in de volgende paragrafen nader beschreven worden.

7.5. Getoetste hypothesen voor binomiale toetsen, waarbij rekening wordt gehouden met onbesliste experimenten.

Indien wij eerst het tweezijdige geval beschouwen, heeft de hypothese die wij wensen te toetsen de gedaante $\frac{p_1}{p_2} = a$, waarin a een bepaalde gegeven waarde voorstelt. In het bijzondere geval van de tekentoets zal a = 1 genomen moeten worden.

Wij willen echter alleen een conclusie trekken betreffende $\frac{p_1}{p_2}$ als p_3 kleiner is dan een eveneens te voren gegeven waarde b. Wij kunnen dit in feit bereiken door de volgende samengestelde hypothese H_0 te toetsen

$$H_0 = \begin{cases} H_0' & p_3 \geqq b \quad \text{of/en} \\ H_0'' & \frac{p_1}{p_2} = a . \end{cases}$$

Als H_0 verworpen wordt kunnen wij besluiten tot de mogelijkheden:

$$H_1 \quad p_3 < b \quad \text{én} \quad \frac{p_1}{p_2} < a$$

of

$$H_2 \quad p_3 < b \quad \text{én} \quad \frac{p_1}{p_2} > a.$$

In fig. 7.1 zijn deze mogelijkheden schematisch weergegeven. Ieder punt in $\triangle OAB$ correspondeert met een der mogelijke combinaties van p_1 en p_3.

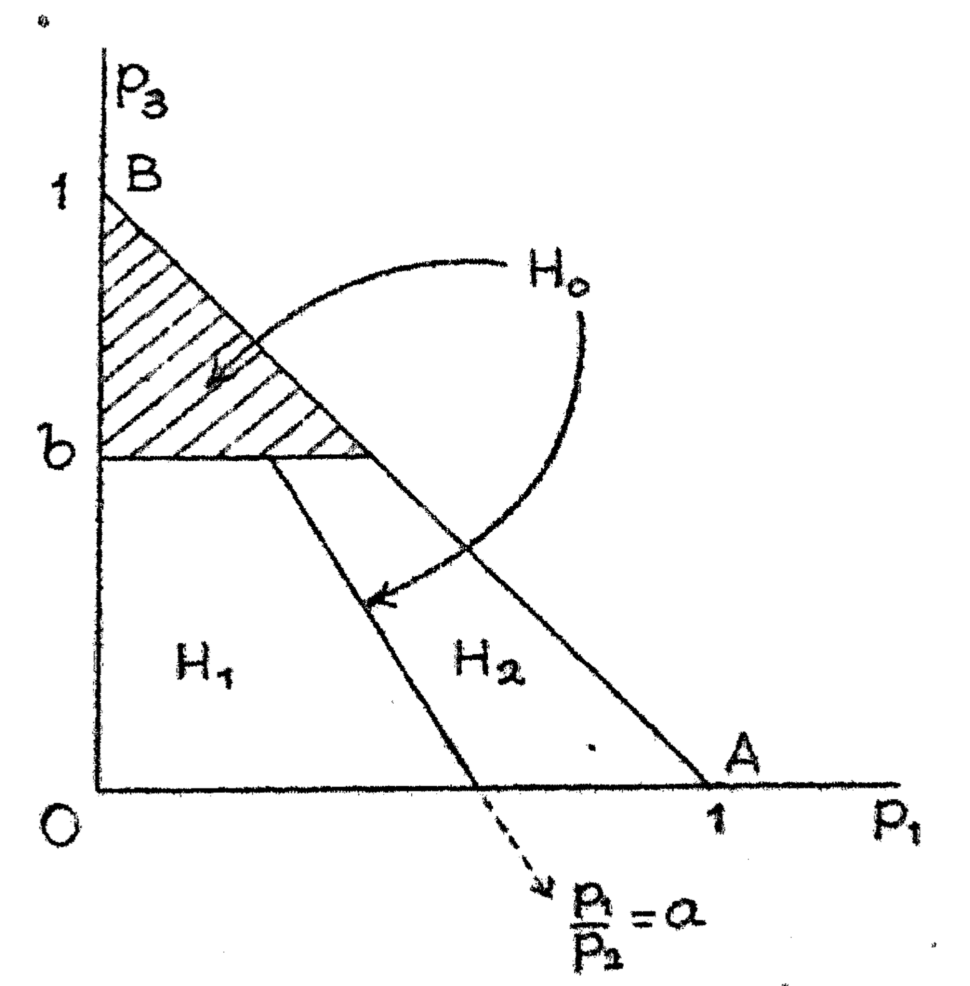


Fig. 7.1. Tweezijdige binomiale toets met inachtneming der twijfelgevallen. (Verdere toelichting zie blz. 52 bovenaan.)

$$H_0: \; p_3 \geqq b \quad \text{of} \quad \frac{p_1}{p_2} = a \; ;$$

alternatieve hypothesen:

$$H_1: \; p_3 < b \quad \text{én} \quad \frac{p_1}{p_2} < a \quad \text{en}$$

$$H_2: \; p_3 < b \quad \text{én} \quad \frac{p_1}{p_2} > a \; .$$

De toets wordt nu zo ingericht, dat H_0' en H_0'' afzonderlijk getoetst worden en alleen als deze beide verworpen worden verwerpen wij H_0. Het voordeel hiervan is, dat, indien H_0' en H_0'' beide met een onbetrouwbaarheidsdrempel α getoetst worden, de onbetrouwbaarheid van de samengestelde toets niet groter zal zijn dan α. Immers als H_0' (dus ook H_0) juist is, dan is de kans dat H_0' verworpen wordt hoogstens gelijk aan α. H_0 wordt echter pas verworpen als H_0' en H_0'' beide verworpen worden. De kans dat dit gebeurt is zeker niet groter dan de kans dat alleen H_0' verworpen wordt en deze kans is al niet groter dan α volgens onze definitie van onbetrouwbaarheidsdrempel van een toets (zie par. 6.1). Eenzelfde redenering kan toegepast worden in het geval, dat H_0'' juist is.

Wij kunnen uiteraard ook een schema opstellen voor het geval dat wij de binomiale toets voor $\frac{p_1}{p_2}$ éénzijdig wensen toe te passen. Wij willen dat hier aangeven voor het rechtseenzijdige geval (het linkseenzijdige geval kan gemakkelijk op analoge wijze behandeld worden). De getoetste hypothese wordt voor dit geval:

$$H_0 = \begin{cases} H_0' & p_3 \geqq b \\ H_0'' & \frac{p_1}{p_2} \leqq a \end{cases}$$

en de alternatieve mogelijkheid:

$$H_1: \; p_3 < b \quad \text{én} \quad \frac{p_1}{p_2} > a .$$

Dit wordt schematisch weergegeven in figuur 7.2.

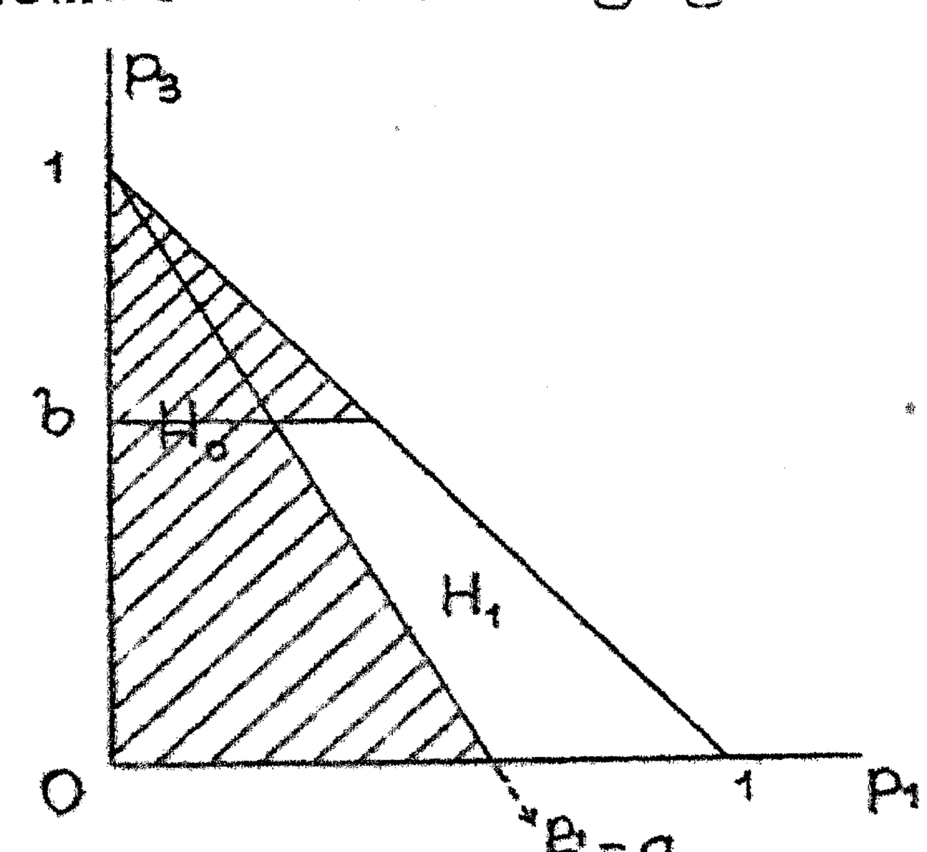


Fig. 7.2. Rechtseenzijdige binomiale toets met inachtneming der twijfelgevallen. (Verdere toelichting op blz. 53.)

$$H_0 : p_3 \geqq b \quad \text{of} \quad \frac{p_1}{p_2} \leqq a \; ;$$

alternatieve hypothese

$$H_1 : p_3 < b \quad \text{en} \quad \frac{p_1}{p_2} > a \; .$$

Bij de hier beschreven samengestelde nulhypothesen wordt vereist dat a en b tevoren gekozen worden. De keuze van a zal in de meeste gevallen geen moeilijkheden opleveren, omdat deze constante direct verband houdt met het doel waarop de toets gericht is. De keuze van b hangt echter af van de invloed, die men aan de fractie grensgevallen op de eindconclusie betreffende $\frac{p_1}{p_2}$ toekent. Er zijn gevallen denkbaar, waar men reeds bij een betrekkelijk kleine fractie grensgevallen wantrouwend zal zijn ten aanzien van een conclusie gebaseerd op $\frac{p_1}{p_2}$ alleen. Dit is bijvoorbeeld denkbaar bij verkiezingsenquêtes, indien deze gehouden worden om de verkiezingsuitslag te voorspellen, aangezien er dan onder de grensgevallen veel kiezers zullen voorkomen, die nog vatbaar zijn voor de propaganda der concurrerende partijen. Daarentegen zijn er ook experimenten denkbaar, die onder zulke omstandigheden moeten worden uitgevoerd, dat men een groot aantal nulwaarnemingen kan verwachten. Men zal dan b zo hoog kunnen stellen, dat men slechts vermijdt om conclusies betreffende $\frac{p_1}{p_2}$ te trekken als p_3 abnormaal hoog zou zijn. Soms kan men b ook kiezen op grond van economische overwegingen.

7.6. Toetsingsschema voor binomiale toetsen met onbesliste experimenten.

Wij willen nu nagaan hoe de bovenbeschreven samengestelde hypothesen H_0 kunnen worden getoetst. Men neemt een steekproef van n waarnemingen uit de populatie en constateert, dat x_1 waarnemingen het kenmerk K hebben, x_2 waarnemingen het kenmerk K niet hebben en x_3 waarnemingen grensgevallen zijn (Dan geldt dus: $x_1 + x_2 + x_3 = n$). De toets van de hypothese H_0': $p_3 \geq b$ is nu de linkseenzijdige binomiale toets, beschreven in par. 7.3. De daar gebruikte letters p_0 en x moeten vervangen worden door b respectievelijk x_3, om met de bovengebruikte notatie in overeenstemming te komen.

Indien H_0' verworpen wordt, gaan we in het tweezijdige geval over tot toetsing van $\frac{p_1}{p_2} = a$. In de populatie van de niet-grensgevallen is de fractie der elementen met kenmerk K gelijk aan $p_1' = \frac{p_1}{p_1 + p_2}$; wij kunnen dus nu de toets, beschreven in par. 7.1, toepassen op de $x_1 + x_2$ waarnemingen, die geen grensgevallen zijn.

Men moet dan de daar voorkomende letters n, p_o en x vervangen door resp.: $x_1 + x_2$, $a' = \frac{a}{1+a}$ en x_1. Indien men een van de éénzijdige toetsen wenst toe te passen, zal men te werk kunnen gaan volgens par. 7.3, waarbij de daar voorkomende letters op dezelfde wijze geïnterpreteerd worden als in het tweezijdige geval.

Voorbeeld.

$n = 600$, $x_1 = 300$, $x_2 = 100$, $x_3 = 200$

$a = 4$, $b = 0{,}4$.

Wij toetsen eerst $p_3 \geq b = 0{,}4$.

Formule (7.3;2) wordt dan

$$\xi''_\alpha = -\frac{x_3 - b n + \frac{1}{2}}{\sqrt{n b (1-b)}} = -\frac{200 - 240 + \frac{1}{2}}{\sqrt{600 \times 0{,}4 \times 0{,}6}} = \frac{39\frac{1}{2}}{12} \approx 3{,}29.$$

De eenzijdige overschrijdingskans hiervan is volgens de tabel van de normale verdeling ongeveer 0,0005. Bij gebruik van een onbetrouwbaarheidsdrempel 0,05 wordt deze hypothese dus zeker verworpen.

Indien wij de hypothese $\frac{p_1}{p_2} = a = 4$ tweezijdig toetsen, passen wij de tweezijdige binomiale toets voor de hypothese $p_1' = a' = \frac{a}{1+a} = 0{,}8$ toe. Wij berekenen dus:

$$\xi_\alpha = \frac{|x_1 - a'(x_1+x_2)| - \frac{1}{2}}{\sqrt{(x_1+x_2)a'(1-a')}} = \frac{|300 - 0{,}8 \times 400| - \frac{1}{2}}{\sqrt{400 \times 0{,}2 \times 0{,}8}} = \frac{19\frac{1}{2}}{8} \approx 2{,}44.$$

Volgens de tabel van de normale verdeling is de tweezijdige overschrijdingskans van dit resultaat 0,015. Bij gebruik van een onbetrouwbaarheidsdrempel 0,05 wordt dus de hypothese H''_o $(\frac{p_1}{p_2} = 4)$ eveneens verworpen en wel ten gunste van $\frac{p_1}{p_2} < 4$ of $p_1' < 0{,}8$.

Opgaven.

7.6.a. Een drogist brengt twee hoestmiddeltjes in de handel (A en B). Om na te gaan of middel A beter of slechter bevalt dan middel B houdt hij een enquête onder 50 klanten, die beide middeltjes gebruiken. Het resultaat is het volgende:

15 klanten beweren dat beide middelen helpen.
10 " " " géén van beide middelen helpt
20 " " " A helpt en B niet.
5 " " " B helpt, doch A niet.

De drogist wil de voorkeur voor beide middeltjes onderzoeken met behulp van de tekentoets (tweezijdig), doch wil daaruit geen verdere conclusies trekken, als hij de hypothese niet kan verwerpen, dat tenminste 2/3 van zijn klanten geen voorkeur heeft voor een van beide middeltjes boven het andere. Welke conclusie zal hij trekken, als de onbetrouwbaarheidsdrempel 0,05 wordt aangehouden?

MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 120

Cursus Toegepaste Statistiek

door Prof. Dr J. Hemelrijk en Ph. van Elteren

Hoofdstuk 8,

De toets van Wilcoxon

benevens oplossingen vraagstukken, inhoud en errata.

Juni 1954

Hoofdstuk 8

De toets van Wilcoxon

8.1. Voorbeeld 1: Echtparen.

Een onderzoeker heeft de lengten bepaald van de mannen en de vrouwen van twintig echtparen, met het doel om na te gaan of gehuwde mannen in het algemeen groter zijn dan hun vrouwen. De (fictieve) resultaten van dit (fictieve) onderzoek zijn weergegeven in tabel 8.I.

Tabel 8.I

Lengten in meters van de mannen en vrouwen van 20 echtparen. (Fictieve cijfers) 7)

nummer van het echtpaar	x_i = lengte van de man	y_i = lengte van de vrouw	teken van $x_i - y_i$
1	1,592	1,633	-
2	1,622	1,663	-
3	1,639	1,531	+
4	1,677	1,612	+
5	1,681	1,557	+
6	1,711	1,718	-
7	1,717	1,645	+
8	1,721	1,741	-
9	1,725	1,675	+
10	1,727	1,631	+
11	1,732	1,647	+
12	1,744	1,615	+
13	1,745	1,657	+
14	1,749	1,649	+
15	1,765	1,674	+
16	1,786	1,692	+
17	1,813	1,742	+
18	1,820	1,828	-
19	1,821	1,743	+
20	1,878	1,890	-

---

7) In de practijk heeft het uiteraard geen zin lichaamslengten nauwkeuriger op te geven dan in hele centimeters. Dit is hier alleen gedaan om te vermijden dat er gelijke waarnemingen optreden, die tot later te behandelen complicaties bij de methode aanleiding geven. (zie verder opgave 8.8a).

Wij beschouwen nu de lengte van de man van het i^e echtpaar als een waarneming van een stochastische grootheid $\underline{x}_i$ en de lengte van zijn vrouw als een waarneming van een stochastische grootheid $\underline{y}_i$. Wij eisen niet, dat alle $\underline{x}_i$ en evenmin dat alle $\underline{y}_i$ dezelfde verdeling hebben; wij laten dus b.v. toe dat de verschillende echtparen bevolkingsgroepen representeren met verschillende lengteverdelingen. De hypothese, dat er geen systematisch verschil bestaat tussen de lengten van mannen en hun vrouwen kunnen we nu als volgt interpreteren:

Voor iedere i zijn $\underline{x}_i$ en $\underline{y}_i$ onafhankelijk en hebben zij dezelfde verdeling. We zullen dit de hypothese H_0 noemen.

Men kan nu gemakkelijk aantonen, dat, indien H_0 juist is, ook geldt:

$$H_0' : \; P\{\underline{x}_i > \underline{y}_i\} = P\{\underline{x}_i < \underline{y}_i\}.$$

De hypothese H_0' kan nu volgens hoofdstuk 5 getoetst worden met behulp van de tekentoets. Men behoeft daartoe slechts te tellen hoeveel echtparen het kenmerk $x_i > y_i$ (de niet onderstreepte letters x_i en y_i stellen de werkelijk gemeten lengten voor) en hoeveel het kenmerk $x_i < y_i$ bezitten. Wij vinden in tabel 8.I, dat bij 14 echtparen de man het grootste is ($x_i > y_i$) en bij 6 echtparen de vrouw ($x_i < y_i$). De tweezijdige overschrijdingskans van de tekentoets is bij dit resultaat ongeveer 0,12; wij kunnen dus de hypothese H_0' en daarom ook H_0 op grond van toepassing van de tekentoets op dit materiaal niet verwerpen.

Wij willen bij de hier gebruikte methode nog opmerken dat de hypothese H_0' veel minder inhoudt dan H_0, zodat men met de tekentoets H_0 op een vrij ruwe wijze toetst. Er bestaan andere methoden, waarbij van meer gegevens gebruik wordt gemaakt en een meer gespecificeerde nulhypothese getoetst wordt 8). Wij willen ons hier echter met deze methoden niet bezig houden, doch een toets behandelen, die voor een principieel ander geval van toepassing is.

8.2. Voorbeeld 2: Mannen en vrouwen.

Wij beschouwen nogmaals tabel 8.I, doch nemen nu aan, dat de 2e kolom de lengten bevat van een twintigtal mannen, die aselect gekozen zijn uit een bepaalde homogene bevolkingsgroep en de derde kolom de lengten van een twintigtal vrouwen, aselect gekozen uit dezelfde bevolkingsgroep. De twee waarnemingen op één

---

8) Zie b.v. J.HEMELRIJK, Symmetrietoetsen en andere toepassingen van de theorie van Neyman en Pearson, Diss. Amsterdam, 1950.

regel hebben dus in de regel niet meer op één echtpaar betrekking.

Wij willen nu nagaan of de mannen van de onderzochte bevolkingsgroep in doorsnee groter zijn dan de vrouwen. Ieder getal in de tweede kolom van tabel 8.I kan beschouwd worden als een waarneming van een stochastische grootheid $\underline{x}$ (de lengte van mannen van de beschouwde bevolkingsgroep) en ieder getal in de derde kolom als een waarneming van een stochastische grootheid $\underline{y}$. Wij zullen deze waarnemingen respectievelijk door $x_1, \ldots, x_{20}$ en $y_1, \ldots, y_{20}$ voorstellen. Op grond van deze waarnemingen gaan wij nu de hypothese H_0 toetsen, dat de waarschijnlijkheidsverdelingen van $\underline{x}$ en $\underline{y}$ hetzelfde zijn.

Indien men ook op dit geval de tekentoets zou willen toepassen, zou men op een of andere manier paren waarnemingen (x_i, y_i) moeten vormen. Dit kan echter op zeer veel verschillende manieren gebeuren, zonder dat er enige reden is hieruit een bepaalde keuze te maken, terwijl het resultaat van de tekentoets van deze keuze afhangt. Indien men de mannen en vrouwen, die in tabel 8.I op dezelfde regel voorkomen (dus de vroegere echtparen) vergelijkt, verkrijgt men uiteraard dezelfde overschrijdingskans als in 8.1, dus ongeveer 0,12. Verwisselt men alleen de waarnemingen $y_8 = 1{,}741$ en $y_{12} = 1{,}615$, dan wordt het aantal paren waarvan de man groter is dan de vrouw vermeerderd met 1 en daalt de tweezijdige overschrijdingskans volgens de tekentoets tot ongeveer 0,04.

Wij kunnen deze moeilijkheid oplossen door de lengte van iedere man te vergelijken met de lengte van iedere vrouw. Wij hebben in totaal 400 mogelijkheden om een man met een vrouw te vergelijken. Uiteraard zijn deze mogelijkheden niet meer onafhankelijk van elkaar, zodat wij op het resultaat niet de tekentoets voor $n = 400$ kunnen toepassen.

Om een methode aan te geven, die we hier wel kunnen toepassen, keren wij terug tot de te toetsen hypothese H_0. Deze hield in, dat $x_1, \ldots, x_{20}$ en $y_1, \ldots, y_{20}$ onderling onafhankelijke waarnemingen waren van eenzelfde stochastische grootheid $\underline{x}$. In onze gedachten zetten wij nu alle waargenomen mannen en vrouwen in een rij, te beginnen met de kleinste, vervolgens de op een na kleinste etc. Indien H_0 juist is, is er geen enkele reden om aan te nemen, dat er voorkeur zou zijn voor een bepaalde volgorde van de mannen en vrouwen in deze rij; er kan inderdaad ook bewezen worden, dat dan alle mogelijke volgorden even waarschijnlijk zijn. Het is nu deze consequentie, verder aan te duiden met H_0', van bovengenoemde hypothese, welke wij in feite gaan toetsen.

Wij tellen nu bij alle 400 mogelijke vergelijkingen tussen een man en een vrouw, het aantal malen, dat daarbij de man groter is dan de vrouw, of, wat op hetzelfde neerkomt, wij tellen, nadat mannen en vrouwen naar opklimmende grootte gerangschikt zijn (als boven is aangegeven), bij iedere man het aantal vrouwen dat vóór hem in de rij staat, en sommeren deze aantallen. Het resultaat, dat gewoonlijk met de letter u wordt aangeduid, is in ons voorbeeld 289; bij de overige 111 vergelijkingen is de man kleiner dan de vrouw, terwijl er hier geen gevallen voorkomen, waarbij beide gelijk zijn. In paragraaf 8.7 zullen wij (voor een iets algemener geval) bespreken hoe wij de overschrijdingskans onder bovengenoemde hypothese H_0' kunnen berekenen, en welke conclusie wij op grond daarvan uit het gevonden resultaat kunnen trekken.

De statistische toetsingsmethode, die hiertoe wordt gebruikt, wordt de toets van WILCOXON genoemd, naar de Amerikaanse statisticus F.WILCOXON [9], die deze methode in 1945 geïntroduceerd heeft. Deze methode is nader uitgewerkt door H.B.MANN en D.R.WHITNEY [10] in 1947.

8.3. Definitie van de toetsingsgrootheid u.

In het algemene geval van de toets van WILCOXON beschouwen wij twee steekproeven:

$$x_1, \dots, x_m \quad \text{en}$$
$$y_1, \dots, y_n$$

waarbij m niet gelijk aan n behoeft te zijn. Wij rangschikken de waarnemingen naar opklimmende grootte en duiden een waarneming uit de eerste steekproef met een x en een waarneming uit de tweede steekproef met een y aan. Laat het resultaat ($m = 4$, $n = 5$) bijvoorbeeld zijn:

$$xxyxyyyxy.$$

Wij tellen nu bij iedere x het aantal letters y, dat daaraan voorafgaat; dit is achtereenvolgens:

---

9) F.WILCOXON, Individual comparison by ranking methods, Biometrics 1 (1945), p. 80 83.

10) H.B.MANN and D.R.WHITNEY, On a test of whether one of two random variables is stochastically larger than the other; Annals of Mathematical Statistics 18 (1947), p. 50-60.

0, 0, 1 en 4 en bepalen de som van deze aantallen, dus hier 5. Deze som wordt aangeduid met $\mathcal{U}$. Er is nog een andere dergelijke grootheid, verkregen door bij iedere x het aantal letters y te bepalen, dat er op volgt en deze aantallen te sommeren (in ons voorbeeld: 5+5+4+1 = 15). Ter onderscheiding zou men de eerste grootheid kunnen aanduiden met $\mathcal{U}(y|x)$ (aantal keren, dat een y voorafgaat aan een x) en de tweede met $\mathcal{U}(x|y)$ (aantal letters y dat achter een x voorkomt), doch het heeft weinig zin dit te doen, aangezien steeds geldt:

$$\mathcal{U}(y|x) + \mathcal{U}(x|y) = mn.$$

Immers, er zijn mn vergelijkingen mogelijk van een x_i met een y_j; als bij een paar x_i, y_j geldt $x_i > y_j$, draagt dit paar een eenheid bij tot $\mathcal{U}(y|x)$ en als $x_i < y_j$ is, een eenheid tot $\mathcal{U}(x|y)$. Wij zullen daarom, na aangegeven te hebben, op welke steekproef de letters x en op welke steekproef de letters y betrekking hebben, steeds $\mathcal{U}(y|x)$ beschouwen en deze grootheid kortweg met $\mathcal{U}$ aanduiden.

In de practijk komt het vaak voor dat enige waarnemingen onderling gelijk zijn. Als een groepje gelijke waarnemingen geheel tot één van beide steekproeven behoort, geeft dit geen moeilijkheden voor de definitie van $\mathcal{U}$. Indien een x-waarneming gelijk is aan een y-waarneming, spreken wij af voor dit paar een bijdrage $\frac{1}{2}$ bij $\mathcal{U}$ te tellen. Indien wij bij $m = 6$, $n = 4$ bij voorbeeld vinden:

$$\begin{array}{ccccccc} & & & & x & & \\ x & & & & x & & \\ x & y & y & y & y & x & x \end{array}$$

waarbij boven elkaar geschreven letters betrekking hebben op gelijke waarnemingen, vinden wij

$$\mathcal{U} = 0 + 0 + 3\tfrac{1}{2} + 3\tfrac{1}{2} + 4 + 4 = 15.$$

8.4. Schema van de tweezijdige toets.

Zoals wij reeds in par. 8.2 gezien hebben, willen wij met de toets van WILCOXON de hypothese H_0 toetsen, dat $x_1, \ldots, x_m$ en $y_1, \ldots, y_n$ onderling onafhankelijke waarnemingen van eenzelfde stochastische variabele zijn en dat in feite daartoe de hypothese H_0' getoetst wordt, inhoudende dat alle mogelijke volgorden der letters x en y bij rangschikking van de waarnemingen naar opklimmende grootte even waarschijnlijk zijn. De waarschijnlijkheidsverdeling van $\underline{\mathcal{U}}$ onder de hypothese H_0 is voor kleine waarden van m en n (en in principe ook voor grote waarden) exact te berekenen. Het blijkt, dat deze verdeling symmetrisch is om het gemiddelde $\frac{1}{2}mn$; dit betekent dus dat de kans op een waarde van $\underline{u}$ v groter dan $\frac{1}{2}mn$ gelijk is aan de kans op een waarde v kleiner dan $\frac{1}{2}mn$, of, wat op hetzelfde neerkomt (in formule):

(8.4.1) $$P\{\underline{u}=u\} = P\{\underline{u}=mn-u\}.$$

Tevens blijken waarden van u minder waarschijnlijk te zijn naarmate zij verder van $\frac{1}{2}mn$ verwijderd zijn. Indien H_0 juist is, zullen waarden, dicht bij 0 of mn dus weinig voorkomen.

Grote waarden van u komen voor als de meerderheid van de letters x achter in de rij van de gerangschikte waarnemingen staat, en de letters y meer voorin. Een dergelijk resultaat zal vaak voorkomen als de stochastische variabelen $\underline{x}$ en $\underline{y}$, corresponderend resp. met $x_1, \ldots, x_m$ en $y_1, \ldots, y_n$, niet dezelfde verdeling hebben, doch $\underline{x}$ in de meerderheid van de gevallen grotere waarden aanneemt dan $\underline{y}$. Precieser gezegd, grote waarden van $\underline{u}$ zullen vaak voorkomen als geldt:

$$p = P\{\underline{x} > \underline{y}\} > \tfrac{1}{2}.$$

Kleine waarden van u zal men daarentegen vaak vinden als

$$p = P\{\underline{x} > \underline{y}\} < \tfrac{1}{2} \text{ is.}$$

Wij kunnen nu een dergelijk schema opstellen, als voor de tekentoets gemaakt is in paragraaf 6.1. Het schema voor de tweezijdige toets van WILCOXON is gegeven in figuur 8.1.

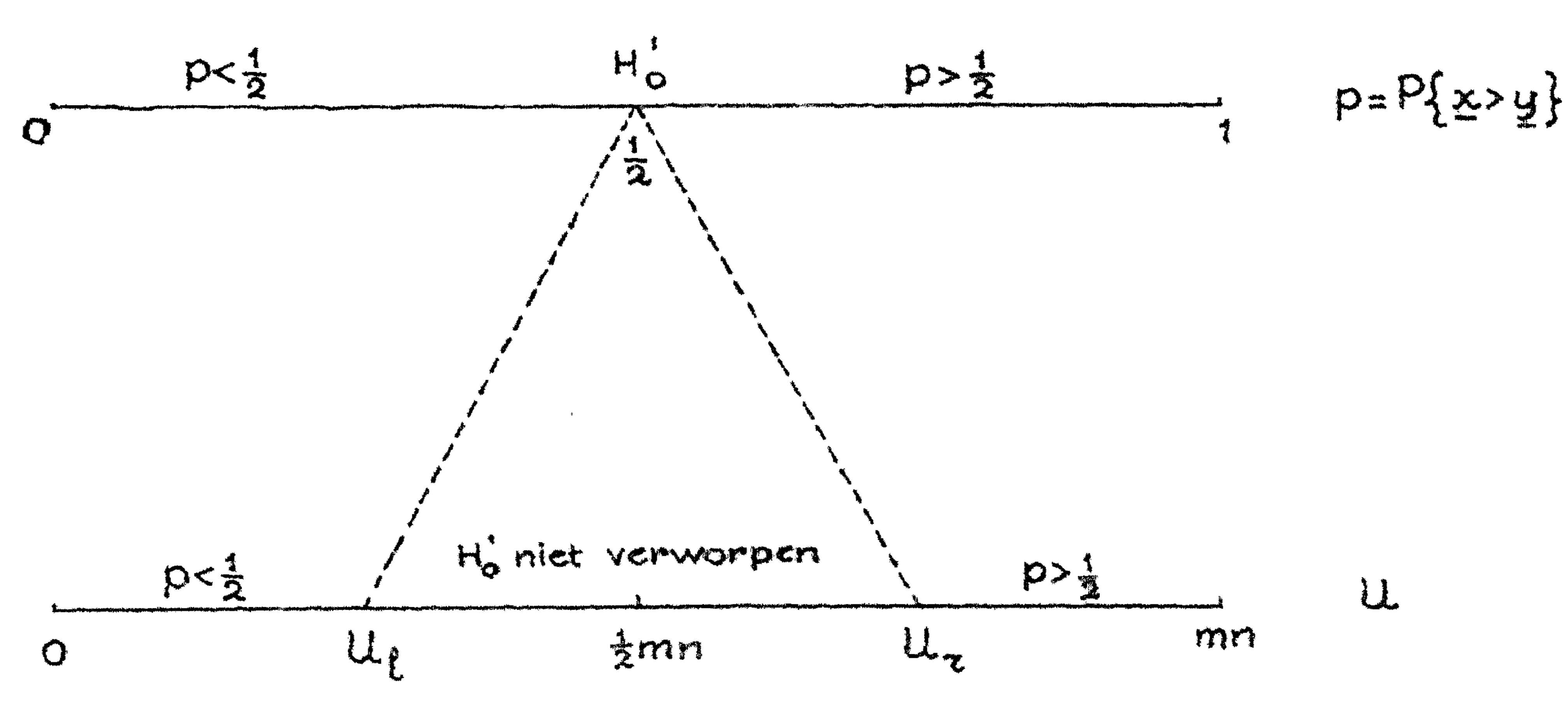


fig. 8.1. Schema voor de tweezijdige toets van WILCOXON.

Op de bovenste lijn zijn wederom de mogelijkheden (uitgedrukt in p) aangegeven, waartussen men wil beslissen, op de onderste lijn is aangegeven, welke conclusie men op grond van een gevonden waarde van u zal trekken. Daarbij treden weer twee kritieke waarden u_ℓ en u_r voor u op, zodanig gekozen, dat de kans dat men rechts van u_r of links van u_l terecht komt, als H_0' geldt, nog juist niet groter is dan α, de gekozen onbetrouwbaarheidsdrempel. Men concludeert $p < \frac{1}{2}$, als $u \leqq u_l$ is, $p > \frac{1}{2}$ als $u \geqq u_r$ is, terwijl H_0 niet verworpen kan worden als $u_l < u < u_r$ is.

Er is een klein verschil met de tekentoets. De hypothese H_0' houdt hier meer in dan alleen $p = \frac{1}{2}$, het kan dus voorkomen dat H_0' niet juist is, en toch $p = \frac{1}{2}$ geldt. Dit treedt b.v. op als $\underline{x}$ en $\underline{y}$ beide symmetrisch verdeeld zijn t.o.v. dezelfde waarde a, maar verschillende spreiding bezitten. Indien H_0' verworpen wordt, kan men dus strikt genomen $p = \frac{1}{2}$ nog niet verwerpen. Bij nader onderzoek van de toets blijkt echter, dat men bij deze toets geen grote noch een met het aantal waarnemingen toenemende kans heeft om in het kritieke gebied (links van u_l of rechts van u_r) terecht te komen als $p = \frac{1}{2}$ is, ook al is H_0 niet vervuld. Hierdoor is het aangegeven beslissingsschema voldoende gerechtvaardigd.

Wij zullen in paragraaf 8.6 aangeven hoe de kritieke waarden bepaald worden.

Bij de tweezijdige toets van WILCOXON kan wederom een tweezijdige overschrijdingskans voor iedere gevonden waarde u van $\underline{u}$ gedefinieerd worden. Dit is de kans, onder de hypothese H_0, dat $\underline{u}$ de waarde u of een minstens evenver van $\frac{1}{2}mn$ verwijderde waarde aanneemt. In formule wordt bijvoorbeeld de tweezijdige overschrijdingskans van $u = 40$ als $m = 10$ en $n = 20$:

$$k(40) = P\{\underline{u} \leqq 40 \text{ of } \underline{u} \geqq 10 \times 20 - 40\} = P\{|\underline{u} - 100| \geqq 60\}.$$

De behandeling van de vraag hoe deze overschrijdingskansen berekend worden, wordt wederom tot paragraaf 8.6 uitgesteld.

8.5. Eénzijdige toetsen van WILCOXON.

Analoog aan de twee éénzijdige tekentoetsen ~~van WILCOXON~~ kunnen wij éénzijdige toetsen van WILCOXON opstellen. Het schema van de rechtséénzijdige toets is gegeven in figuur 8.2. u_r' is hierin de kritieke waarde van de rechtséénzijdige toets, en is dus zo gekozen, dat de kans onder H_0' dat $\underline{u} \geqq u_r'$ is, nog juist kleiner is dan de gekozen onbetrouwbaarheidsdrempel. De analogie met de rechtséénzijdige tekentoets is niet volledig. Wij kunnen hier niet zeggen, dat $p \leqq \frac{1}{2}$ getoetst wordt tegen $p > \frac{1}{2}$, omdat hier niet bewezen kan worden dat uit $P\{\underline{u} \geqq u_r' | H_0'\} \leqq \alpha$ volgt: $P\{\underline{u} \geqq u_r' | p \leqq \frac{1}{2}\} \leqq \alpha$, zodat wij niet weten of de onbetrouwbaarheidsdrempel, die geldt voor H_0', ook geldt voor iedere andere hypothese met $p \leqq \frac{1}{2}$. De situatie is hier zo, dat wij de rechtséénzijdige toets alléén toepassen, als de mogelijkheid $p < \frac{1}{2}$ ons feitelijk niet interesseert. Wij toetsen de hypothese H_0'. Indien u groter uitvalt dan de kritieke waarde u_r', verwerpen wij H_0' ten gunste van $p > \frac{1}{2}$. Strikt genomen weten wij dan nog niet of we iedere hypothese met $p \leqq \frac{1}{2}$ ook zouden kunnen verwerpen. Valt u kleiner uit dan u_r', dan wordt H_0' niet verworpen.

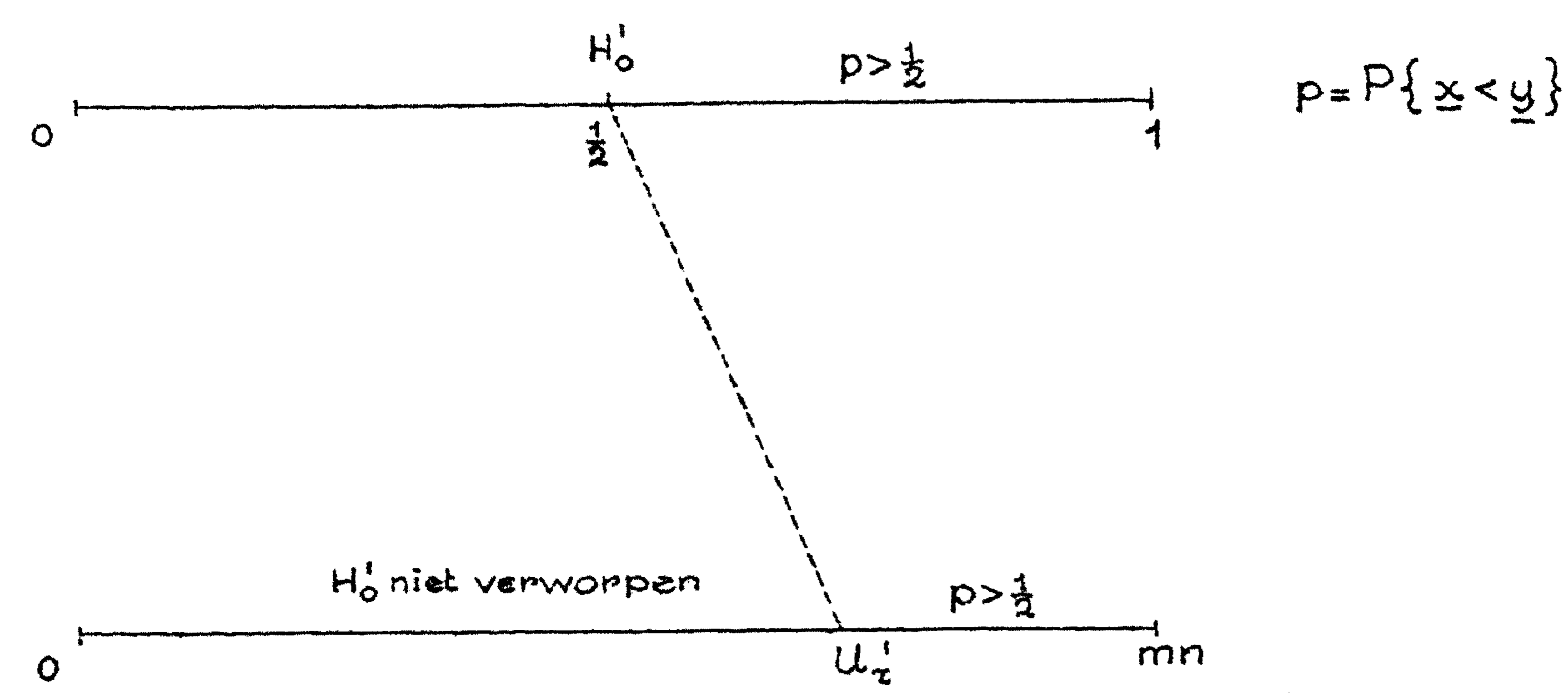


fig. 8.2. Schema voor de rechtseenzijdige toets van WILCOXON.

Bij de rechtséénzijdige toets behoort wederom een rechtséénzijdige overschrijdingskans. Dit is dus de kans, dat $\underline{u}$ onder de hypothese H_0', de gevonden waarde u bereikt of overschrijdt, in formule:

$$k_r(u) = P\{\underline{u} \geqq u \mid H_0'\}.$$

Wederom geldt, dat de rechtséénzijdige overschrijdingskans, van de kritieke waarde u_r' van de rechtséénzijdige toets van WILCOXON hoogstens gelijk is aan de onbetrouwbaarheidsdrempel α.

Bij de linkséénzijdige toets wordt H_0' getoetst tegen $p < \frac{1}{2}$. De kritieke waarde u_l' is kleiner dan $\frac{1}{2}mn$ en ligt zodanig, dat de kans onder de hypothese H_0', dat $\underline{u}$ de waarde u_l' of een lagere waarde aanneemt, hoogstens gelijk is aan de gekozen onbetrouwbaarheidsdrempel α. De linkséénzijdige overschrijdingskans wordt gedefinieerd door:

$$k_l(u) = P\{\underline{u} \leqq u \mid H_0'\}.$$

De tweezijdige overschrijdingskans is gelijk aan twee maal de kleinste der éénzijdige overschrijdingskansen (en gelijk aan 1 als beide $> \frac{1}{2}$ zijn). Hierin schuilt weer het voordeel van het gebruik van een éénzijdige toets als men slechts twee mogelijkheden wil onderscheiden. De rechtséénzijdige toets heeft, voor $p > \frac{1}{2}$, een groter onderscheidingsvermogen (zie par. 6.4) dan de tweezijdige, de linkséénzijdige voor $p < \frac{1}{2}$.

Wij willen er hierbij nog op wijzen, dat men niet op grond van de voor u gevonden waarde mag beslissen of men één- dan wel tweezijdig wil toetsen. Indien men dat wél doet, kan de onbetrouwbaarheidsdrempel niet meer geïnterpreteerd worden als een bovengrens voor de kans dat men de getoetste hypothese zal verwerpen als zij juist is. Indien men bijvoorbeeld steeds rechtséénzijdig toetst als $u > \frac{1}{2}mn$ en linkséénzijdig als $u < \frac{1}{2}mn$ is, dan is de kans, dat men in één van beide kritieke gebieden terecht komt, niet hoogstens gelijk aan α — de onbetrouwbaarheidsdrempel van de éénzijdige toets — doch 2α. Deze methode komt dus in feite neer op tweezijdig toetsen.

8.6. Bepaling van kritieke waarden.

Wij hebben in par. 8.4 reeds opgemerkt, dat de waarschijnlijkheidsverdeling van $\underline{u}$ onder de hypothese H_o symmetrisch is ten opzichte van $\frac{1}{2}mn$, zodat geldt: (zie 8.4;1): $P\{\underline{u} = u | H_o\} = P\{\underline{u} = mn - u | H_o\}$.

Dientengevolge liggen ook de kritieke waarden symmetrisch ten opzichte van $\frac{1}{2}mn$. Als dus de linkse kritieke waarde van de tweezijdige toets gelijk is aan u_ℓ volgt de rechtse kritieke waarde uit:

(8.6;1) $$u_r = mn - u_\ell .$$

Eveneens is de verdeling van $\underline{u}$ symmetrisch in m en n Dat wil zeggen, als we twee steekproeven hebben, één met m en één met n waarnemingen, doet het er niet toe of we de letters x aan de eerste en de letters y aan de tweede steekproef toekennen, of omgekeerd; in beide gevallen verkrijgen we voor $\underline{u} = \underline{u}(y|x)$ (zie 8.3) onder de hypothese H_o', dezelfde verdeling. Dit geldt eveneens voor de kritieke waarden. We kunnen dus bij het tabelleren volstaan met de linkse kritieke waarden voor twee steekproeven met $m \geqq n$.

We hebben in de tabel 8.II de linkse kritieke waarden van de toets van WILCOXON gegeven bij tweezijdige onbetrouwbaarheidsdrempels $\alpha = 0{,}10$; $0{,}05$; $0{,}02$ en $0{,}01$ (éénzijdig: $0{,}05$; $0{,}025$; $0{,}01$ en $0{,}005$) voor waarden van m en n van 2 tot en met 10.

De hier opgegeven kritieke waarden gelden strikt genomen alléén indien er geen gelijke waarnemingen in de steekproeven optreden (verg. par. 8.3). Indien er slechts weinig gelijke waarnemingen vóórkomen in kleine groepjes van 2 of 3, kan men tabel 8.II zonder bezwaar gebruiken. De juiste kritieke waarden kunnen in deze gevallen iets afwijken van de waarden in tabel 8.II opgegeven en zijn dan over het algemeen hoger; bij gebruik van de tabel is het dus mogelijk dat men H_o' niet verwerpt in gevallen, waar dit bij exacte berekening wel mogelijk zou zijn.

Tabel 8.II

Linkse kritieke waarden van de toets van Wilcoxon

m, n : aantallen waarnemingen van de steekproeven

α_1 = onbetrouwbaarheidsdrempel éénzijdige toets

α_2 = onbetrouwbaarheidsdrempel tweezijdige toets

$\alpha_1 = 0{,}05$ $\alpha_2 = 0{,}10$

n \ m	2	3	4	5	6	7	8	9	10
2	-	-	-	0	0	0	1	1	1
3		0	0	1	2	2	3	4	4
4			1	2	3	4	5	6	7
5				4	5	6	8	9	11
6					7	8	10	12	14
7						11	13	15	17
8							15	18	20
9								21	24
10									27

$\alpha_1 = 0{,}025$ $\alpha_2 = 0{,}05$

n \ m	2	3	4	5	6	7	8	9	10
2	-	-	-	-	-	-	0	0	0
3		-	-	0	1	1	2	2	3
4			0	1	2	3	4	4	5
5				2	3	5	6	7	8
6					5	6	8	10	11
7						8	10	12	14
8							13	15	17
9								17	20
10									23

$\alpha_1 = 0{,}01$ $\alpha_2 = 0{,}02$

n \ m	2	3	4	5	6	7	8	9	10
2	-	-	-	-	-	-	-	-	-
3		-	-	-	-	0	0	1	1
4			-	0	1	1	2	3	3
5				1	2	3	4	5	6
6					3	4	6	7	8
7						6	7	9	11
8							9	11	13
9								14	16
10									19

$\alpha_1 = 0{,}005$ $\alpha_2 = 0{,}01$

n \ m	2	3	4	5	6	7	8	9	10
2	-	-	-	-	-	-	-	-	-
3		-	-	-	-	-	-	0	0
4			-	-	0	0	1	1	2
5				0	1	1	2	3	4
6					2	3	4	5	6
7						4	6	7	9
8							7	9	11
9								11	13
10									16

N.B.: Een streepje in de tabel betekent dat de kritieke waarde ontbreekt en dat dus bij de gegeven aantallen waarnemingen en onbetrouwbaarheidsdrempel de hypothese H_0 bij geen enkele waarde van u verworpen kan worden. Dit is steeds het geval als één van beide steekproeven slechts één en de andere niet meer dan tien waarnemingen bevat, bij alle hierboven vermelde onbetrouwbaarheidsdrempels.

Als minstens één van beide steekproeven meer dan 10 waarnemingen bevat en de kleinste steekproef niet minder dan 5, kunnen wij gebruik maken van de volgende benaderingsformules:

(8.6;2) $$u_r = \tfrac{1}{2}mn + c\sqrt{\tfrac{1}{12}mn(m+n+1)} + \tfrac{1}{2}$$

en

(8.6;3) $$u_l = \tfrac{1}{2}mn - c\sqrt{\tfrac{1}{12}mn(m+n+1)} - \tfrac{1}{2}$$

welke zowel voor de tweezijdige als voor de éénzijdige toetsen gebruikt kunnen worden, als voor c de waarden uit tabel 8.III worden ingevuld.

Tabel 8.III

Waarden van c in de formules (8.6;2) en (8.6;3) voor verschillende waarden van α. (Eveneens bruikbaar voor de formules (6.2;1), (6.2;2), (7.2;1) en (7.2;2).)

α		
éénzijdig	tweezijdig	c
0,05	0,10	1,65
0,025	0,05	1,96
0,01	0,02	2,33
0,005	0,01	2,58

Indien er veel gelijke waarnemingen in de steekproeven voorkomen, moeten in de formules (8.6;2) en (8.6;3) correcties worden aangebracht. Deze zullen besproken worden in par. 8.8.

Voorbeelden.

Bij $m = 10$, $n = 10$ zijn de kritieke waarden van de tweezijdige toets van WILCOXON bij $\alpha = 0{,}02$ volgens tabel 8.II en formule (8.6;1)

$$u_l = 19 \quad u_r = 10.10 - 19 = 81,$$

en benaderd volgens formule (8.6;3)

$$u_l = \tfrac{1}{2}.10.10 - 2{,}33\sqrt{\tfrac{1}{12}.10.10.21} - \tfrac{1}{2} =$$

$$= 50 - 11{,}65\sqrt{7} - \tfrac{1}{2} = 18{,}67.$$

Te werk gaande volgens de benaderingsformule vindt men hier dus een te lage waarde. Men zou namelijk volgens de benadering aan $u = 19$ een overschrijdingskans groter dan 0,02 toekennen en $u_l = 18$ kiezen in plaats van de juiste waarde 19. Bij gebruik maken van de benadering komt men dus niet zo snel tot verwerping als bij toepassing van de exacte methode.

Opgave.

8.6.a. Bepaal de kritieke waarden voor de toets van WILCOXON bij:

$m = 3$, $n = 7$, $\alpha = 0{,}01$ (rechtséénzijdige toets),

$m = 3$, $n = 7$, $\alpha = 0{,}01$ (tweezijdige toets) en

$m = 36$, $n = 12$, $\alpha = 0{,}05$ (tweezijdige toets).

8.7. Bepaling van overschrijdingskansen.

De exacte waarden van één- en tweezijdige overschrijdingskansen als beide steekproeven niet meer dan 10 waarnemingen bevatten, kunnen alleen gevonden worden met behulp van volledige tabellen, zoals bijvoorbeeld zijn te vinden in H.R.van der VAART: Gebruiksaanwijzing voor de toets van WILCOXON, Rapport S 32 (M 4) van de Statistische Afdeling van het Mathematisch Centrum. Tabel 8.II is aan de hand van deze tabellen samengesteld; deze tabel kan evenals het geval was bij de tekentoets, gebruikt worden om bepaalde grenzen aan te geven voor de overschrijdingskansen. Zo vindt men voor $m = 9$, $n = 6$ voor de rechtséénzijdige overschrijdingskansen: $0{,}01 > P\{\underline{u} \geq 47\} > 0{,}005$. Immers $u_r (= 54 - u_l)$ ligt voor $\alpha = 0{,}01$ bij 47 en voor $\alpha = 0{,}005$ bij 49.

Voor grotere waarden van m en n kunnen de overschrijdingskansen weer benaderd worden met behulp van de tabel van de normale verdeling. Ter bepaling van de tweezijdige overschrijdingskans berekent men nu:

$$\xi_\alpha = \frac{|u - \frac{1}{2}mn| - \frac{1}{2}}{\sqrt{\frac{1}{12}mn(m+n+1)}} \qquad (8.7;1)$$

en bepaalt de bijbehorende waarde van α met behulp van de tabel van de normale verdeling. De tweezijdige overschrijdingskans is dan gelijk aan 2α.

Ter bepaling van de rechtséénzijdige overschrijdingskans berekent men:

$$\xi'_\alpha = \frac{u - \frac{1}{2}mn - \frac{1}{2}}{\sqrt{\frac{1}{12}mn(m+n+1)}} \qquad (8.7;2)$$

en voor de linkséénzijdige overschrijdingskans:

$$\xi''_\alpha = -\frac{u - \frac{1}{2}mn + \frac{1}{2}}{\sqrt{\frac{1}{12}mn(m+n+1)}} \qquad (8.7;3)$$

en handelt hiermee wederom als is aangegeven op blz. 46.

Voorbeeld: $m = 20 \quad n = 60 \quad U = 300$

$$\xi_\alpha = \frac{|300-600|-\frac{1}{2}}{\sqrt{\frac{1}{12}\times 60\times 20\times 81}} = \frac{299\frac{1}{2}}{90} \approx 3,33$$

Dus de tweezijdige overschrijdingskans is 2 x 0,0004 = 0,0008. De linkséénzijdige overschrijdingskans (ξ''_α is positief) wordt 0,0004; de rechtséénzijdige overschrijdingskans wordt 0,9996, want ξ'_α is negatief.

Wij vonden bij het voorbeeld van de lengten van mannen en vrouwen: $m = n = 20$ en $U = 289$. Volgens formule (8.7;1) geldt:

$$\xi_\alpha = \frac{|289-200|-\frac{1}{2}}{\sqrt{\frac{1}{12}\times 20\times 20\times 41}} = \frac{88\frac{1}{2}}{36,97} \approx 2,39$$

met tweezijdige overschrijdingskans 0,0168.

Wij kunnen dus bij een onbetrouwbaarheidsdrempel 0,05 of 0,02 de hypothese H_0, dus ook de hypothese, dat de lengteverdeling van mannen en vrouwen identiek is verwerpen ten gunste van $p = P\{\underline{x} > \underline{y}\} > \frac{1}{2}$, die dus inhoudt dat in de populatie, waaruit de steekproeven genomen zijn, de mannen over het algemeen groter zijn dan de vrouwen. Men zegt ook wel, de mannen zijn systematisch groter dan de vrouwen, terwijl men de gevonden waarde van U in zo'n geval wel significant noemt, waarmee men wil aangeven, dat het (door de toetsingsgrootheid gemeten) verschil moeilijk aan toevalsvariaties alleen kan worden toegeschreven.

De formules (8.7;2) en (8.7;3) geven minder goede benaderingen als er betrekkelijk veel stellen gelijke waarnemingen in het materiaal voorkomen. In dat geval kan men correcties aanbrengen, die in de volgende paragraaf behandeld worden.

8.8. Correcties voor gelijke waarnemingen.

Wij beschouwen nu nogmaals het stadium van de berekening van U, dat verkregen is nadat de waarnemingen van beide steekproeven naar opklimmende grootte gerangschikt zijn. Laat het resultaat in de notatie van par. 8.3 voor $m = 10 \quad n = 10$ zijn.

$$\begin{array}{cccccccccccc}
 & & & & x & & & & & & & \\
 & & & & x & & x & & x & & & \\
x & & & & x\,x & & x & & y & & & \\
x & y & y & y & y & x & y & y & y & y & y
\end{array}$$

Wij tellen nu in deze rij het aantal waarnemingen, dat niet gelijk is aan een andere; dit stellen we voor door g_1.

Evenzo tellen we de groepen van 2 gelijke waarnemingen. Hun aantal stellen we voor door g_2, het aantal groepen van drie gelijke waarnemingen door g_3 enz. In ons voorbeeld vinden we dus: $g_1 = 6$, $g_2 = 2$, $g_3 = 2$, $g_4 = 1$ en $g_5 = g_6 = \ldots = 0$.

Indien men de aantallen g_1, g_2 etc. bepaald heeft, kan men de formules (8.6;2), (8.6;3), (8.7;1), (8.7;2) en (8.7;3) zodanig corrigeren, dat zij ook een redelijke benadering voor kritieke waarden en overschrijdingskansen geven, indien er veel en betrekkelijk grote groepen van gelijke waarnemingen voorkomen. Men vervangt daartoe de in die formules voorkomende wortelvorm

$$\sqrt{\tfrac{1}{12}mn(m+n+1)}$$

door de meer ingewikkelde vorm:

$$\sqrt{\frac{1}{12}\cdot\frac{mn\{(m+n)^3-(1.g_1+2^3g_2+3^3g_3+4^3g_4+\cdots)\}}{(m+n)(m+n-1)}}$$

In ons voorbeeld zou dit dus worden:

$$\sqrt{\frac{1}{12}\cdot\frac{10.10\{20^3-(6+8.2+27.2+64)\}}{200.199}} =$$

$$= \sqrt{\frac{100}{12}\cdot\frac{8000-140}{380}} \approx 13{,}13.$$

We vinden in ons voorbeeld:

$$U = 3 \times 2\tfrac{1}{2} + 1 \times 3\tfrac{1}{2} + 3 \times 4 + 1 \times 6 = 29.$$

De benaderde tweezijdige overschrijdingskans k kan nu dus worden gevonden uit:

$$\xi_\alpha = \frac{|29-50|-\frac{1}{2}}{13{,}13} = 1{,}56$$

waarbij

$$k = 0{,}12.$$

Indien wij de correctie voor gelijke waarnemingen niet hadden toegepast, hadden wij gevonden

$$\xi_\alpha = \frac{|29-50|-\frac{1}{2}}{\sqrt{\frac{1}{12}\times 10^2\times 21}} \approx 1{,}55,$$

waarbij $k = 0,12$.

Zelfs bij het betrekkelijk grote aantal gelijken in onze waarnemingsreeks heeft de correctie dus nog maar weinig te betekenen. Men kan nog opmerken, dat wij bij gebruik van tabel 8.II tot de conclusie waren gekomen, dat $k > 0{,}10$ moest zijn; daarbij is dan echter geen rekening gehouden met de gelijke waarnemingen.

Opgave.

8.8.a. Verander de gegevens in tabel 8.I door af te ronden op centimeters en pas vervolgens de toets van WILCOXON toe met correctie voor gelijke waarnemingen. Vergelijk het resultaat met dat gevonden in par. 8.7.

8.9. Voorbeeld van toepassing.

Een van de cursisten heeft een psychologisch onderzoek toegepast bij een aantal proefpersonen. Deze kregen daarbij onafhankelijk van elkaar eenzelfde opdracht te vervullen. Bij iedere proefpersoon werd de benodigde tijd en de leeftijd opgegeven. Het resultaat van de waarnemingen is in tabel 8.IV samengevat. Om een eenvoudig overzicht te verkrijgen zijn de personen ingedeeld in 3 leeftijdsklassen ter breedte van 15 jaar en de benodigde tijd in intervallen van 30 seconden.

Wij willen aan de hand van deze waarnemingen nagaan of de per opdracht benodigde tijd t afhangt van de leeftijd der proefpersoon ℓ. Nu is de benodigde tijd per opdracht voor iedere leeftijdsklasse een stochastische grootheid, welke wij voor de klasse 15-29 voorstellen door $\underline{t}_1$ voor de klasse 30-44 door $\underline{t}_2$ en voor de klasse 45-64 door $\underline{t}_3$.

Tabel 8.IV

Resultaat van een psychologisch onderzoek

ℓ = leeftijd proefpersonen in jaren

t = benodigde tijd per opdracht in ~~minuten~~ seconden.

t \ l	15-29	30-44	45-64	totaal
120-149	4	7	1	12
150-179	17	12	1	30
180-209	21	15	7	43
210-239	19	21	8	48
240-269	18	13	4	35
270-299	13	11	11	35
300-329	6	11	6	23
330-359	6	5	6	17
360-389	4	2	2	8
390-419	0	3	5	8
totaal	108	100	51	259

Wij zullen nu nagaan of de waarschijnlijkheidsverdelingen voor t_1, t_2 en t_3 verschillen. Wij beschouwen nu de gevonden tijden als onafhankelijke waarnemingen van t_1, t_2 of t_3 naar gelang de leeftijd van de persoon waarop zij betrekking hebben. In onze tabel 8.III vinden wij niet de waarnemingen zelf, doch alleen het tijdsinterval, waarin zij liggen. Dit is echter geen bezwaar, aangezien wij bij de toets van WILCOXON kunnen volstaan met de <u>grootte-volgorde</u> van de waarnemingen; de juiste grootte is niet vereist. Het enige gevolg van de indeling in tijdsintervallen is, dat wij de waarnemingen op een regel als een groep gelijken moeten beschouwen, waardoor de toets minder gevoelig wordt.

Er bestaan generalisaties van de toets van WILCOXON, waarmee men de drie steekproeven in eens kan vergelijken. Wij zullen hier echter slechts de eerste leeftijdsgroep met de laatste vergelijken (zie ook oefening 8.9.a). Wij stellen dus de waarnemingen van de leeftijdsgroep 15-29 voor door x en de waarnemingen van de groep 45-64 door y. Het rangschikken naar opklimmende waarden van t is in de tabel feitelijk reeds geschied. Het blijkt daaruit dat 4 waarnemingen x en één y in het interval van de laagste waarden van t liggen, 17 x-en en één y in het daarop volgende interval etc. Er zijn géén y's vóór de 4 x-en in interval 120-149, wel één y in dat interval, de bijdrage van deze x-en tot U is dus $4 \times \frac{1}{2} = 2$. Er is één y voor de 17 x-en in het interval 150-179 en één in hetzelfde interval; de bijdrage van deze x-en wordt dus $17 \times 1\frac{1}{2}$; wij vinden zo voortgaande: $U = 4 \times \frac{1}{2} + 17 \times 1\frac{1}{2} + 21 \times 5\frac{1}{2} + 19 \times 13 + 18 \times 19 + 13 \times 26\frac{1}{2} + 6 \times 35 + 6 \times 41 + 4 \times 45 + 0 \times 48\frac{1}{2} = 1712\frac{1}{2}$.

Gezien het feit, dat het materiaal verdeeld is in groepen gelijke waarnemingen, passen wij de correctie toe, beschreven in par. 8.8. Wij vinden 2 leeftijdsklassen met in totaal 5 waarnemingen uit beide steekproeven, 2 met 12, en één met 6, 18, 22, 24, 27 of 28 waarnemingen. Dus geldt:

$g_5 = g_{12} = 2$; $g_6 = g_{18} = g_{22} = g_{24} = g_{27} = g_{28} = 1$. Alle overige g's voorkomende in de wortelvorm van par. 8.8 zijn nul. Dus wordt deze vorm ($m = 108$, $n = 51$)

$$\sqrt{\frac{1}{12}\,\frac{108 \times 51 \times \{159^3 - (2.5^3 + 6^3 + 2.12^3 + 18^3 + 22^3 + 24^3 + 27^3 + 28^3)\}}{159 \times 158}} \approx 268$$

(Bij toepassing van de formule $\sqrt{\frac{1}{12}mn(m+n+1)}$ dus zonder correctie vinden wij voor de wortelvorm 271.)

Wij berekenen vervolgens:

$$\xi_\alpha = \frac{|1712\tfrac{1}{2} - \tfrac{1}{2}\cdot 51.108| - \tfrac{1}{2}}{268} \approx 3{,}88 .$$

De tweezijdige overschrijdingskans van deze waarde van ξ_α is volgens de tabel van de normale verdeling in ieder geval kleiner dan 0,0004 (volgens een meer uitgebreide tabel 10^{-5}).

Wij kunnen dus op grond van dit resultaat met grote stelligheid de hypothese verwerpen dat $\underline{t}_1$ dezelfde verdeling heeft als $\underline{t}_3$, en wel ten gunste van $P\{\underline{t}_1 > \underline{t}_3\} < \frac{1}{2}$. Dat betekent dus, dat personen uit de leeftijdsklasse 15-29 over het algemeen minder tijd nodig hebben voor het uitvoeren van de opdracht dan personen uit de leeftijdsklasse 45-64.

Opgave.

8.9.a. Toets met behulp van de methode van WILCOXON de hypothese dat $\underline{t}_2$ dezelfde waarschijnlijkheidsverdeling heeft als $\underline{t}_1$, en eveneens dat $\underline{t}_2$ dezelfde verdeling heeft als $\underline{t}_3$, bij het materiaal behandeld in bovenstaand voorbeeld en interpreteer de uitkomsten. (α = 0,05).

Opmerking. De drie toetsen waarmee de leeftijdsklassen twee aan twee vergeleken worden, zijn niet onafhankelijk, omdat ieder der beschouwde steekproeven altijd bij twee van de drie toetsen betrokken is. Dit heeft tengevolge, dat als men bijvoorbeeld bij vergelijking van $\underline{t}_1$ en $\underline{t}_3$ ten onrechte de hypothese H_0 verwerpt, men daarna ~~een~~ kans groter dan de onbetrouwbaarheidsdrempel heeft, dat dit ook bij de vergelijking van $\underline{t}_1$ en $\underline{t}_2$ zal gebeuren. Deze kans kan dan zelfs veel groter zijn.

De opgegeven onbetrouwbaarheidsdrempels of overschrijdingskansen gelden alleen als men de toetsen ieder afzonderlijk beschouwt.

Rapport S 120

Cursus Toegepaste Statistiek

Oplossing der opgaven.

3.5.a Formule (3.5;1): 0,22, formule (3.5;3): 0,17.
Tabel 3.I: 0,23. De benadering met (3.5;1) is dus vrij goed, die met (3.5;3) onvoldoende.

3.5.b Formule (3.5;1): 0,22, formule (3.5;3): 0,21.
Tabel 3.II: 0,22. Hier zijn beide benaderingen reeds goed.

3.5.c Formule (3.5;1) geeft.
Bovengrens voor fractie oesters met parel: 0,0082
Benedengrens voor fractie oesters met parel: 0,0011
(Beide afzonderlijk met onbetrouwbaarheidsdrempel 0,005).

3.6.a (3.5.a) 0,25, 3.V geeft dus een vrij slechte benadering.
(3.5.c) Bovengrens: 0,0078, benedengrens 0,00075.

3.6.b Als de fractie oesters met een parel ongeveer 0,005 is, kan men in een steekproef van b.v. 200 oesters 0, 1 of 2 oesters met een parel verwachten. Het betrouwbaarheidsinterval gevonden met tabel 3.V blijkt dan te breed te zijn. Door een aantal verschillende waarden voor de omvang der steekproef te proberen vindt men dat ongeveer 1000 waarnemingen voldoende zijn.

(Bij n = 1000	x = 3	wordt het	interval	0,0011	-	0,009
	x = 4	"	"	0,001	-	0,010
	x = 5	"	"	0,002	-	0,012
	x = 6	"	"	0,002	-	0,013
	x = 7	"	"	0,003	-	0,

4.1.a Wel aan de eis van gelijkwaardigheid, niet aan de eis van onafhankelijkheid.

4.1.b De rij is periodiek (de periode bevat 18 cijfers). Zij voldoet dus niet aan de eis van onafhankelijkheid. Zij voldoet wel aan de eis van gelijkwaardigheid; daaraan voldoen ook de deelrijen van de cijfers voorkomende na een bepaald cijfer.

4.5.a Men nummert de elementen van de gegeven populatie (N = 900) als volgt: 000, 001, 002,...,899. Om nu de cijfers van de eenheden en van de tientallen te loten trekt men uit de gehele modelpopulatie. Voor het loten van het cijfer der honderdtallen laat men daaruit de negens weg. De populatie met N = 170 nummert men 000, 001,...,169. Voor het cijfer der eenheden loot men uit de modelpopulatie. Om de beide andere cijfers te loten kan men b.v. als volgt te werk gaan:

men loot voor het cijfer der tientallen en voor dat der honderdtallen uit de modelpopulatie. Als men voor de honderdtallen een even cijfer (niet = 0) vindt, vervangt men dit door een 0; vindt men een oneven cijfer (niet = 1), dan vervangt men dit door een 1. Alle getallen van 170 t/m 199 die bij dit lotingsproces verkregen worden, worden geschrapt.

4.5.b Om steekproeven zonder teruglegging te verkrijgen dient men de reeds getrokken elementen bij hernieuwde trekking over te slaan.

5.2.a $P[A] = \frac{1}{2}$, $P[B] = \frac{1}{2}$, $P[A \text{ en } B] = \frac{1}{6}$. Dus A en B zijn afhankelijk.

5.2.b $P[A] = 1$; zij $P[B] = p$, dan is $P[A \text{ en } B] = p$. A en B zijn dus onafhankelijk.

5.2.c Zij $P[A] = p_1$ en $P[B] = p_2$. Er geldt $P[A \text{ en } B] = 0$, dus A en B zijn afhankelijk, tenzij $p_1 = 0$ of $p_2 = 0$.

5.2.d Zij $P[A] = p_1$ en $P[B] = p_2$, dan is $P[A \text{ en } B] = p_1$. Dus A en B zijn afhankelijk, tenzij $p_2 = 1$ of $p_1 = 0$.

6.1.a $n = 25$ $x = 10$: H_o kan niet verworpen worden.
$n = 25$ $x = 18$: Tweede bepaling systematisch lager dan de eerste
$n = 100$ $x = 17$: Tweede bepaling systematisch hoger dan de eerste
$n = 100$ $x = 90$: Tweede bepaling systematisch lager dan de eerste.

6.1.b. Men kan voorspellen, dat A zal winnen. Het betrouwbaarheidsinterval voor de fractie aanhangers van A wordt 0,51-0,59.

6.2.a De kritieke waarden worden $500 \pm \frac{1}{2}.2,58\sqrt{1000} \approx 500 \pm 41$. De afwijking moet dus minstens 41 bedragen.
De tweezijdige overschrijdingskans van $x = 550$ ($\xi_\alpha = 3,13$) wordt $k = 0,0018$.

6.2.b 1^o reeks: $k = 0,23$ H_o niet verworpen
2^o reeks: $k = 0,69$ " " "
3^o reeks: $k = 0,11$ " " "
4^o reeks: $k = 0,42$ " " "
Totaal: $k = 0,02$. Munt niet zuiver, kan op kruis $> \frac{1}{2}$.

6.4.b

	$\alpha = 0,01$				$\alpha = 0,05$			
	x_l 36	x_r 64	x'_l 37	x'_r 63	x_l 39	x_r 61	x'_l 41	x'_r 59
$k(x)$	0,007	0,007	0,012	0,012	0,036	0,036	0,089	0,089
$k_l(x)$	0,004	0,998	0,006	0,997	0,018	0,989	0,045	0,971
$k_r(x)$	0,998	0,004	0,997	0,006	0,989	0,018	0,971	0,045

6.4.c Met behulp van de normale benadering vinden wij een tweezijdige overschrijdingskans 0,002. Bij gebruik van een onbetrouwbaarheidsdrempel $> 0,002$ kan men dus concluderen, dat de kans op een jongensgeboorte $> \frac{1}{2}$ is. Bij éénzijdige toetsing sluit men de mogelijkheid van de tegengestelde conclusie uit.

6.4.d Blijkens de laatste zin van de opgave moet men <u>éénzijdig</u> toetsen. De beslissing wordt dan: B kopen.

7.2.a $p < p_0$, $k < 10^{-4}$.

7.3.a $k_r(x) \simeq 1$, $k_l(x) < 10^{-4}$.

7.6.a De hypothese, dat tenminste $\frac{2}{3}$ van de klanten geen voorkeur heeft wordt verworpen met overschrijdingskans 0,0094. De hypothese, dat van de klank met voorkeur, een fractie $p = \frac{1}{2}$ een voorkeur heeft voor A wordt verworpen ten gunste van $p > \frac{1}{2}$ met overschrijdingskans 0,005. De conclusie van de drogist is dus, dat A beter bevalt.

8.6.a $m = 3$, $n = 7$, $\alpha_1 = 0,01$: $U_r' = 21$.
$m = 3$, $n = 7$, $\alpha_2 = 0,01$; kritieke waarden ontbreken.
$m = 36$, $n = 12$, $\alpha_2 = 0,05$, $U_l \approx 133$, $U_r \approx 299$.

8.8.a Men vindt $U = 286$. De gecorrigeerde wortelvorm wordt 36,9. $\xi_\alpha = 2,32$. Dus $k = 0,02$. Het resultaat wijkt dus slechts weinig af van hetgeen wij vonden bij de niet afgeronde waarden van tabel 8.I (zie blz. 67).

8.9.a Men vindt bij vergelijking van $\underline{t}_1$ met $\underline{t}_2$
$U = 5145\frac{1}{2}$, $\xi_\alpha = 0,59$, $k = 0,56$ (tweezijdig)
en bij vergelijking van $\underline{t}_2$ met $\underline{t}_3$
$U = 1741$, $\xi_\alpha = 3,21$, $k = 0,0014$ (tweezijdig).
<u>Conclusie</u>: Personen uit de 3^o leeftijdsklasse hebben systematisch <u>meer</u> tijd nodig voor de opdracht dan personen uit de 2^o leeftijdsklasse. Tussen de 2^o leeftijdsklasse en de eerste wordt geen verschil van betekenis gevonden.

Cursus Toegepaste Statistiek

Inhoud.

Cursus Toegepaste Statistiek

Errata.

pag.	regel	staat	moet staan
1	4 v.b.	dat in een	dat een
2	2 v.b.	bepaalde	bepaald
3	14 v.b.	wij een met x	wij aan met x
5	5 v.o.	in plaats van 2	in plaats van 25
6	2 v.b.	van y	van $\underline{y}$
9	12 v.b.	boven α, zodat	boven $\alpha = 0{,}05$, zodat
11	2 v.b.	en 0,1	en 0,19
12	19 v.b.	321	320
		conficende	confidence
21	12 v.b.	de eers	de eerste
22	21 v.b.	Bij m	Bij men-
	11 v.o.	aan	aan de
23	2 v.b.	5	50
26	20 v.b.	Een deze	En deze
38	5 v.o.	m =	x =
41	4 v.o.	heeft	hebben
tabel vóór p. 42	2 v.b.	3,09	3,49
42	2 v.o.	dat van de	dat voor de
44	17 v.b.	$x \leqq \frac{1}{2}$	$\frac{x}{n} \geqq \frac{1}{2}$
	18 v.b.	$k_1(x_1) \leqq \alpha$ en $k_1(x_1+1) > \alpha$	$k_1(x_1') \geqq \alpha$ en $k_1(x_1'+1) > \alpha$
45	5 v.o.	tweezijdig	eenzijdig
	4 v.o.	(6.2;1)	(6.2;2)
46	3 v.b.	(6.2;2)	(6.2;1)
49	6 v.o.	(7.1;1)	(7.2;1)
	3 v.o.	(7.1;2)	(7.2;2)
54	11 v.b.	p_3 (3 keer)	b (3 keer)
61	20 v.o.	gedefinieer	gedefinieerd
	18 v.o.	a nneemt	aanneemt
	12 v.o.	tekentoetsen van WILCOXON	tekentoetsen
	7 v.o.	analo	analogie
62	18 v.o.	overschrijdingskans,	overschrijdingskans
65	Tabel 8.III	1,64	1,65
	11 v.o.	Tabel 8.I	Tabel 8.II
	2 v.b	200	300
68	15 v.b.	200.199	20.19
	16 v.b.	$\sqrt{\frac{1}{12} \cdot \frac{8000-140}{398}} \approx 12{,}83$	$\sqrt{\frac{100}{12} \cdot \frac{8000-140}{380}} \approx 13{,}13$

pag.	regel	staat	moet staan
68	9 v.o.	12,83	13,13
		1,60	1,56
	7 v.o.	0,11	0,12
69	13 v.o.	minuten	seconden