

TECHNISCHE HOGESCHOOL EINDHOVEN

Afdeling Algemene Wetenschappen

Onderafdeling der Wiskunde

TOEGEPASTE STATISTIEK

Syllabus van het College van

Prof. Dr. H.C.Hamaker

1961

Inhoudsbeschrijving Toegepaste Statistiek(1961) :

(De indeling in hoofdstukken en paragrafen vertoont enkele inconsistenties!
Bladzijde 60 komt ook voor als variant: 60-60a)

Hoofdstuk I	KANSREKENING	
§1.1	Kansrekening	1
§1.2	Splitsing in deelpopulaties	2
§1.3	Grondformules	3
	Vraagstukken	5
§1.4	Uitbreiding van de som- en productregel	6
	Vraagstukken	8
Hoofdstuk II	PERMUTATIES EN COMBINATIES	
§2.1	n-faculteit	8
§2.2	Permutaties, Variaties, Combinaties	9
	Vraagstukken	11
§2.3	Eigenschappen van binomiaalcoëfficiënten	11
	Vraagstukken	12
§2.4	De hypergeometrische verdeling	13
§2.5	Algemeneren gevallen	16
§2.6	De binomiale verdeling	17
§2.7	Algemeneren problemen	18
§2.10	Vraagstukken	20
§2.11	Een aanvulling	21
§2.12	Toevalscijfers	21

Zie ommezijde

Hoofdstuk VII	VARIANTIEANALYSE	
§7.1	Een eenvoudige situatie	87
§7.2	Variantie-analyse (vervolg)	91
§7.3	Toepassing van de F -toets	93
	Vraagstukken	96
§7.4	Een betrouwbaarheidsinterval voor het gemiddelde	98
§7.5	Een variantie-analyse in twee richtingen	101
§7.6	Variantie-analyse	106
§7.7	Enkele opmerkingen over de variantie-analyse	109

STATISTISCHE TABELLEN

(30 Mei 2005, JdG)

TECHNISCHE HOGESCHOOL EINDHOVEN

Afdeling Algemene Wetenschappen

Onderafdeling Wiskunde

TOEGEPASTE STATISTIEK

Syllabus 1961. 1

Aanbevolen boeken : M.J.Moroney, Facts from figures,
Pelican Serie , f 3,50.
Wijvekate , Verklarende Statistiek,
Aulareeks, f 1,95.

Een goed, pittig leerboek over kansrekening is :
W.Feller, Probability theory and its applications.
Wiley and Sons, New York.

1.1. Kansrekening

Zij ontstond uit dobbelspelen, kaartspelen, loterijen.
Gegeven een loterij met N loten waarvan N_A met kenmerk A.
Kans op trekken van een lot A is dan

$$p_A = \frac{N_A}{N} = \frac{\text{het aantal gunstige loten}}{\text{totaal aantal loten}} .$$

Aangenomen wordt dat alle loten een gelijke kans bezitten te worden
getrokken.

Voor zekerheid is $p = 1$, voor onmogelijkheid is $p = 0$.

Enige termen die veel zullen worden gebruikt :

Populatie met omvang N,
Elementen,
Kenmerken : A,B,C,
Fracties of percentage : $p_A = \frac{N_A}{N}$ of $100 \frac{N_A}{N}$,
Populatie parameters,
Experiment of trekking,
Aselecte trekking,
Waarneming of uitkomst,
Toevalsvariabele X ,
Kans = $\Pr(\underline{X} = A) = \Pr(A) = p_A$.

De kansrekening heeft steeds betrekking op experimenten die we in
principe onbeperkt kunnen herhalen.

Een serie van n experimenten geeft een waarnemingsreeks met aantallen

$$n, n_A, n_B, n_C,$$
$$\hat{p}_A = \frac{n_A}{n}, \hat{p}_B = \frac{n_B}{n} = \text{waargenomen fracties.}$$

We zullen verwachten dat $\hat{p}_A \rightarrow p_A$ als $n \rightarrow \infty$. Tabel 1.1 geeft een voorbeeld.

TABEL 1.1

Resultaat van 5 series van 10, 100, en 1000 tossen met een munt ;
A \equiv "munt"

Seriegrootte	waargenomen fractie \hat{p}_A				
10	0.6	0.1	0.4	0.6	0.6
100	0.45	0.57	0.51	0.44	0.55
1000	0.505	0.490	0.514	0.516	0.531

Voorbeeld, waarbij aan de verwachting niet wordt voldaan : een tollende munt. 50 x tollen met een gulden gaf

$$n = 50, \quad n_A = 12, \quad \hat{p}_A = 0.24.$$

Vraag : Bewijst dit dat $p_A \neq 0.50$? Wat kunnen we omtrent p_A concluderen ?

1.2. Splitsing in deelpopulaties

\bar{A} = niet -A = ontkenning van A = complement van A.

Een element kan ook meer dan één kenmerk dragen : notatie $AB \equiv A$ en B.
Twee kenmerken in één populatie geven een splitsing in 4 deelpopulaties:

$$\bar{A}\bar{B}, \quad A\bar{B}, \quad \bar{A}B, \quad AB.$$

Dan geldt

$$A = A\bar{B} + AB, \quad (1.2.1)$$

$$\bar{A}B = \bar{A}\bar{B} + A\bar{B} + \bar{A}B, \quad (1.2.2)$$

$$(A \text{ of } B) = A\bar{B} + \bar{A}B + AB, \quad (1.2.3)$$

$$(I) = \bar{A}\bar{B} + A\bar{B} + \bar{A}B + AB, \quad (1.2.4)$$

$$(I) = A + \bar{A} = AB + \bar{A}\bar{B} \quad (1.2.5)$$

waarin (I) de gehele populatie aanduidt.

Bij drie kenmerken zijn er 8 deelpopulaties.

$$\bar{A}\bar{B}\bar{C}, \quad A\bar{B}\bar{C}, \quad \bar{A}B\bar{C}, \quad \bar{A}\bar{B}C, \quad A\bar{B}C, \quad \bar{A}BC, \quad ABC,$$

Dan geldt bijv.

$$AB = A\bar{B}C + ABC \quad (1.2.6)$$

$$\bar{A}\bar{B}C = \bar{A}\bar{B}\bar{C} + A\bar{B}\bar{C} + \bar{A}B\bar{C} \quad (1.2.7)$$

$$(A \text{ of } B \text{ of } C) = A\bar{B}\bar{C} + \bar{A}B\bar{C} + \bar{A}\bar{B}C + A\bar{B}C + \bar{A}BC + ABC = (I) - \bar{A}\bar{B}\bar{C} \quad (1.2.8)$$

of

$$\overline{(A \text{ of } B \text{ of } C)} = \bar{A}\bar{B}\bar{C} \quad (1.2.9)$$

Bovenstaande formules gelden eveneens voor de aantallen elementen in de deelpopulaties en de kansen. Bijv. (1.2.1) geeft

$$N_A = N_{A\bar{B}} + N_{AB}, \quad (1.2.10)$$

en
$$\Pr(A) = P(\bar{A}\bar{B}) + \Pr(AB). \quad (1.2.11)$$

enz.

1.3. Grondformules

a) Somregel voor elkaar uitsluitende waarnemingen

$$\boxed{\Pr(A \text{ of } B \text{ of } C) = \Pr(A) + \Pr(B) + \Pr(C)} \quad (1.3.1)$$

Vormen A, B en C een volledig stelsel elkaar uitsluitende waarnemingen, dan geldt

$$P(A) + P(B) + P(C) = 1. \quad (1.3.2)$$

Gevolgen :

$$\Pr(A) + \Pr(\bar{A}) = 1, \quad (1.3.3)$$

$$\Pr(AB) + \Pr(\bar{A}\bar{B}) = 1, \quad (1.3.4)$$

en uit (1.2.8)

$$\Pr(A \text{ of } B \text{ of } C) = 1 - \Pr(\bar{A}\bar{B}\bar{C}) \quad (1.3.5)$$

een veel toegepaste formule.

b) Productregel voor onafhankelijke waarnemingen

$$\boxed{\Pr(ABC) = \Pr(A) \times \Pr(B) \times \Pr(C)} \quad (1.3.6)$$

Komen A, B en C in één populatie voor, dan noemen we A, B en C onderling onafhankelijk indien (1.3.6) geldt voor iedere combinatie van kenmerken die we uit A, B en C kunnen vormen ; dus wanneer (1.3.6) geldt voor A en B, A en C, B en C, en A, B en C.

Dan moet

$$N_{AB} = \frac{N_A \cdot N_B}{N},$$

$$N_{AC} = \frac{N_A \cdot N_C}{N},$$

$$N_{BC} = \frac{N_B \cdot N_C}{N},$$

$$N_{ABC} = \frac{N_A \cdot N_B \cdot N_C}{N^2}.$$

(1.3.7)

Soms weten we a priori dat waarnemingen onafhankelijk zijn ; (werpen met twee dobbelstenen) in andere gevallen moeten we dit uit waarnemingsseries nagaan (blond haar, blauwe ogen).

Voorbeelden

- (a) De kans met twee dobbelstenen een dubbel 6 te gooien bedraagt $1/36$.
- (b) De kans met twee dobbelstenen een som 8 te gooien bedraagt $5/36$.
- (c) De kans met 3 dobbelstenen de combinatie 3,4,5 te gooien bedraagt $1/36$.
- (d) De kans met 3 dobbelstenen 3 opeenvolgende punten tallen te gooien bedraagt $1/9$.
- (e) De kans met 3 dobbelstenen minstens één zes te gooien bedraagt $1-125/216 = 91/216$

Vraagstukken

1.1. In welke deelpopulatie kunnen

$$\overline{\overline{AB}} \quad \text{en} \quad \overline{A \overline{BC}}$$

verder worden gesplitst.

1.2. Bewijs dat

$$(\overline{A} \text{ of } \overline{B}) = \overline{AB}$$

$$(A \text{ of } B)C = (AC \text{ of } BC).$$

1.3. Indien A, B en C onderling onafhankelijk zijn, zijn ook

$$\overline{A}, \overline{B}, \overline{C} \quad \text{of} \quad \overline{A}, B, C$$

onafhankelijk. Bewijs dat.

1.4. Hoe groot zijn de kansen met 3 dobbelstenen te werpen

drie gelijke puntentallen	:	aaa	$\frac{1}{36}$
twee gelijke en één verschillende	:	aab	$\frac{1}{36}$
drie verschillende	:	abc	$\frac{25}{36}$

De som van deze drie kansen moet = 1 zijn.

Verifieer de gevonden uitkomsten door worpen met 3 dobbelstenen.

1.5. Hoe groot is de kans bij vier maal werpen met één dobbelsteen ten minste één 6 te gooien ?

En hoe groot de kans bij 24 worpen en met twee dobbelstenen tenminste éénmaal dubbel 6 te gooien ?

1.6. Hoe groot zijn de kansen dat een gezin van 4 kinderen bestaat uit

- (a) 4 jongens of 4 meisjes,
- (b) 3 jongens, 1 meisje, of 1 jongen, 3 meisjes,
- (c) 2 jongens en 2 meisjes.

Welke onderstelling ligt aan deze berekening ten grondslag ?

1.7. Twee tennisspelers spelen een partij van maximaal 3 games. Wie twee games gewonnen heeft, wint de partij

Wanneer A een kans p en B een kans q = 1-p bezit een game te winnen, hoe groot is dan de kans dat A, resp. B de partij wint ?

Hoe groot moet p ongeveer zijn, opdat A's kans de partij te winnen 0.80 bedraagt ?

1.8. In een telefonische verbindinglijn worden twee versterkers parallel geschakeld. Op een gegeven moment worden beide ingeschakeld, en later beide uitgeschakeld. Voor iedere versterker bestaat een kans p_1 dat het inschakelen niet functioneert, en een kans p_2 dat het uitschakelen niet functioneert. Het is voldoende indien tussen in- en uitschakelen slechts één versterker werkt.

Hoe groot is de kans dat het samenstele goed functioneert ; dus goed in- en uitschakelt ?

HOOFDSTUK I. Kansrekening

1.4. Uitbreiding van de som- en productregel

a) De algemene somregel

$$\begin{aligned}\bar{A} &= I - A \\ \bar{AB} &= B - AB\end{aligned}$$

$$\begin{aligned}\text{Verschil} &\equiv \bar{A}\bar{B} = I - A - B + AB \\ \bar{A}\bar{B}\bar{C} &= C - AC - BC + ABC\end{aligned}$$

$$\text{Verschil} \equiv \bar{A}\bar{B}\bar{C} = I - A - B - C + AB + AC + BC - ABC$$

of in kansen

$$\begin{aligned}\Pr(\bar{A}\bar{B}\bar{C}) &= 1 - \Pr(A \text{ of } B \text{ of } C) = \\ &= 1 - \Pr(A) - \Pr(B) - \Pr(C) + \Pr(AB) + \Pr(AC) + \Pr(BC) - \Pr(ABC)\end{aligned}\tag{1.4.1}$$

Zo kunnen we doorgaan. Algemeen

$$\begin{aligned}\Pr(\bar{A}\bar{B}\dots\bar{K}) &= 1 - \Pr(A \text{ of } B \text{ of } \dots \text{ of } K) = \\ &= \Pr\{(1 - A)(1 - B)\dots(1 - K)\}\end{aligned}\tag{1.4.2}$$

waarbij we het gedurig product uitschrijven en alle termen door de overeenkomstige kansen vervangen.

Voorbeeld : 1.4.a. Hoe groot is de kans met 3 dobbelstenen ten minste één zes te gooien.

We noemen A, B, C het gooien van een zes met de eerste, tweede, en derde steen. Dan is

$$\begin{aligned}\Pr(A) &= \Pr(B) = \Pr(C) = \frac{1}{6} \\ \Pr(AB) &= \Pr(AC) = \Pr(BC) = \frac{1}{36} \\ \Pr(ABC) &= \frac{1}{216},\end{aligned}$$

en dus

$$\Pr(A \text{ of } B \text{ of } C) = 3 \times \frac{1}{6} - 3 \times \frac{1}{36} + \frac{1}{216} = \frac{91}{216}.$$

Zelfde antwoord als in voorbeeld 1.3.c. De berekening volgens

$$\Pr(A \text{ of } B \text{ of } C) = 1 - \Pr(\bar{A}\bar{B}\bar{C}) = 1 - \left(\frac{5}{6}\right)^3 = \frac{91}{216}$$

is in dit geval eenvoudiger, omdat \bar{A} , \bar{B} en \bar{C} onderling onafhankelijk zijn. Is dit niet het geval dan moet (1.4.2) worden toegepast.

b) De algemene productregel

Zijn A en B niet onafhankelijk dan onderscheiden we zgn. voorwaardelijke kansen

$$\Pr(A|B) = \frac{N_{AB}}{N_B}, \quad \Pr(B|A) = \frac{N_{AB}}{N_A}. \quad (1.4.3)$$

$\Pr(A|B)$ = de kans dat een element het kenmerk A bezit, wanneer reeds bekend is dat het het kenmerk B draagt. Evenzo is $\Pr(B|A)$ de kans op B als A gegeven is, en

$$\Pr(A|BC) = \frac{N_{ABC}}{N_{BC}} \quad (1.4.4)$$

de kans op A wanneer gegeven is dat een element de kenmerken B en C bezit.

De algemene productregel luidt

$\Pr(ABC) = \Pr(A) \times \Pr(B|A) \times \Pr(C|AB)$

 (1.4.5)

Zijn A, B, en C onderling onafhankelijk dan geldt

$$N_{AB} = \frac{N_A N_B}{N} \quad \text{of} \quad \frac{N_A}{N} = \frac{N_{AB}}{N_B}$$

of $\Pr(A|B) = \Pr(A)$
Evenzo $\Pr(A|BC) = \Pr(A)$ } (1.4.6)

In woorden : de kans op A wordt niet beïnvloed door kennis omtrent de aan- of afwezigheid van B of C ; B en C leveren geen informatie over het al of niet optreden van A.

Voorbeeld : 1.4.b. Vier loten genummerd 1 tot 4 worden na elkaar getrokken. Wat is de kans dat ze in natuurlijke volgorde worden getrokken. Deze bedraagt

$$\frac{1}{4} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{1} = \frac{1}{24}.$$

Voorbeeld : 1.4.c. Een worp met twee dobbelstenen I en II.

A = een oneven aantal punten met steen I.

B = een som 8 met beide stenen samen.

Dit geeft

$$\Pr(A) = \frac{1}{2}, \quad \Pr(B) = \frac{5}{36}$$

$$\Pr(A|B) = \frac{2}{5}, \quad \Pr(B|A) = \frac{1}{9}$$

en dus

$$\Pr(AB) = \frac{1}{2} \times \frac{1}{9} = \frac{5}{36} \times \frac{2}{5} = \frac{1}{18},$$

en

$$\Pr(A \text{ of } B) = \frac{1}{2} + \frac{5}{36} - \frac{1}{18} = \frac{21}{36}.$$

Voorbeeld : 1.4.d : Een vaas bevat knikkers, 3 rode en 2 blauwe. Eerst wordt een knikker aselekt getrokken zonder naar de kleur te kijken, en daarna een tweede. Hoe groot is de kans dat die tweede knikker blauw is ?
Deze kans bedraagt :

$$\frac{2}{5} \times \frac{1}{4} + \frac{3}{5} \times \frac{1}{2} = \frac{8}{20} = \frac{2}{5} .$$

Vraagstukken

- 1.4.1. Twee munten worden getost. Hoe groot is de voorwaardelijke kans dat twee maal "munt" wordt waargenomen, wanneer gegeven is dat "munt" tenminste éénmaal is waargenomen.
- 1.4.2. Uit een pak kaarten wordt aselekt één kaart getrokken. Zijn de uitkomsten "een schoppen" en "een aas" afhankelijk of onafhankelijk ?
- 1.4.3. Bij werpen met twee dobbelstenen definiëren we :
A = de eerste steen geeft een even aantal punten,
B = de tweede steen geeft een oneven aantal punten,
C = de som is oneven.
Ga na, in hoeverre A, B, en C onderling onafhankelijk zijn.
- 1.4.4. Bij een worp met 3 dobbelstenen zijn de puntentallen n_1 , n_2 , en n_3 . We definiëren $A \equiv n_1 + n_2 = 8$, $B \equiv n_2 + n_3 = 8$, $C \equiv n_3 + n_1 = 8$. Hoe groot zijn $\Pr(A|B)$, $\Pr(C|A)$ en $\Pr(C|AB)$?
Hoe groot is de kans dat bij een worp met 3 dobbelstenen ten minste twee van de drie samen 8 punten geven ?
- 1.4.5. Een vaas bevat 4 witte en 6 zwarte knikkers. Hieruit worden eerst aselekt 2 knikkers getrokken en op zij gelegd zonder naar de kleur te kijken en daarna wordt een derde knikker getrokken. Hoe groot zijn de kansen dat deze derde knikker wit of zwart is ? Hoe groot is de kans dat de 3de knikker wit is, wanneer gegeven is dat onder de eerste twee knikkers tenminste één witte aanwezig was ?

HOOFDSTUK II. Permutaties en combinaties

2.1. n-faculteit $\equiv n!$

$$\left. \begin{array}{l} n! = n(n-1)(n-2)\dots 1 \quad \text{voor } n \text{ geheel positief} \\ n! = 1 \quad \text{voor } n = 0 \\ n! = \infty \quad \text{voor } n \text{ geheel negatief} \end{array} \right\} \quad (2.1.1)$$
$$n! = \int_0^{\infty} x^n e^{-x} dx.$$

In de kansrekening speelt de functie $n!$ een belangrijke rol. Numerieke berekeningen waarin $n!$ voorkomt geschieden het eenvoudigst door gebruik te maken van tabellen van $\log n!$

Tabel 23 geeft waarden tot $n = 100$.

Tot $n = 1000$ vindt men $\log n!$ in Tabel 51 van de Biometrische Tables for Statisticians, Volume I, Cambridge University Press, 1954.

Wanneer men geen tabel van $\log n!$ bij de hand heeft kan men gebruik maken van de benaderingsformule van Stirling ; geldig voor $n > 0$:

$$n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \quad (2.1.2)$$

hetgeen ook reeds voor vrij kleine waarden van n een goede benadering is :

$n=$	$n!$	$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$
5	120	118.02
10	3.6288×10^5	3.5988×10^5
20	2.4329×10^8	2.4228×10^8

2.2. Permutaties, Variaties, Combinaties

Permutaties

n verschillende objecten kunnen op $n!$ manieren in een rij worden gerangschikt. Dit heet het aantal permutaties van n objecten.

Variaties

Uit n verschillende objecten worden r aangewezen en op een rij gezet ; dit kan op

$$(n)_r = n(n-1)\dots(n-r+1) = \frac{n!}{(n-r)!} \quad (2.2.1)$$

manieren geschieden. Dit heet het aantal variaties van r uit n objecten.

Combinaties

Uit n objecten kan op

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \quad (2.2.2)$$

verschillende manieren een groep van r objecten worden aangewezen. Dit is het aantal combinaties van r uit n objecten. De grootheden

$\binom{n}{r}$ heten ook de binomiaal coëfficiënten. Uit (2.2.2) volgt onmiddellijk dat

$$\binom{n}{r} = \binom{n}{n-r} , \quad (2.2.3)$$

een belangrijke eigenschap.

Voorbeelden

De letters ABC kunnen op $3! = 6$ manieren worden gepermuteerd :
ABC, ACB, BAC, BCA, CAB, CBA.

Uit deze letters kan men op $\frac{3!}{1!} = 6$ manieren een variatie van 2 letters vormen :

AB, BA, AC, CA, BC, CB ;

en op $\frac{3!}{2!1!} = 3$ manieren een combinatie van twee letters : AB, AC, BC.

Een verzameling van n objecten kan op

$$\frac{n!}{n_1! n_2! \dots n_k!}, \quad n_1 + n_2 + \dots + n_k = n \quad (2.2.4)$$

manieren in k groepen worden verdeeld waarbij de eerste groep n_1 , de tweede n_2, \dots , en de laatste groep n_k objecten omvat, en de volgorde van de objecten binnen de groepen er niet toe doet.

Men kan deze formule zien als het product van een aantal combinaties

$$\frac{n!}{n_1! n_2! \dots n_k!} = \binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \dots \binom{n-n_1-n_2-\dots-n_{k-1}}{n_k} \quad (2.2.5)$$

We kiezen eerst n_1 uit n , dan n_2 uit de resterende $(n - n_1)$ enz. Men moet er echter vooral op letten dat indien onder de aantallen n_1, n_2, \dots, n_k gelijken voorkomen de permutaties van deze groepen in (2.2.3) mee in rekening zijn gebracht.

Voorbeeld : Men kan de 4 letters ABCD volgens (2.2.3) op $\frac{4!}{2!2!} = 6$ manieren in twee groepen van twee splitsen :

AB - CD, AC - BD, AD - BC, BC - AD, BD - AC, CD - AB.

Hierbij zijn dan echter AB - CD en CD - AB als twee verschillende splitsingen gerekend. Wenst men deze permutatie der groepen buiten beschouwing te laten, dan moet men nog eenmaal door $2!$ delen. De 4 letters ABCD kunnen dus op

$$\frac{4!}{2!2!} \times \frac{1}{2!} = 3$$

manieren in twee groepen van twee worden verdeeld. Evenzo kunnen 12 objecten op

$$\frac{12!}{3!3!2!2!} \times \frac{1}{2!} \times \frac{1}{3!} = 138600$$

manieren in twee groepen van 3 en 3 groepen van 2 worden gesplitst. Een kaartspel kan op

$$\frac{52!}{(13!)^4} \times \frac{1}{4!}$$

manieren in 4 groepen van 13 kaarten worden verdeeld. Deze kunnen op

hun beurt op $4!$ manieren onder vier spelers worden gepermuterd. Een bridgespeler kan dus op

$$\frac{52!}{(13!)^4}$$

manieren onder 4 spelers worden verdeeld.

Vraagstukken

- 2.2.1. Op hoeveel manieren kunnen de letters van het woord "paard" worden gerangschikt? Hetzelfde voor de woorden "aantal", "telefoon", "kenmerken".
- 2.2.2. Op hoeveel manieren kunnen n personen rond een tafel worden gerangschikt?
- 2.2.3. Op hoeveel manieren kunnen we uit een kaartspel van 52 kaarten een groep van 13 kaarten vormen bestaande uit : 5 schoppen, 4 harten, 3 ruiten en 1 klaver ; of uit 4 schoppen, 4 harten, 1 ruiten, en 4 klaver.

2.3. Eigenschappen van binomiaal coëfficiënten

De binomiaal coëfficiënten

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

treden op in de ontwikkeling van het gedurig product :

$$(1+t)^n = \binom{n}{0} + \binom{n}{1}t + \binom{n}{2}t^2 + \dots + \binom{n}{n}t^n. \quad (2.3.1)$$

Hieruit zijn belangrijke eigenschappen van de binomiaalcoëfficiënten af te leiden. Bijv. :

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \dots + (-1)^n \binom{n}{n} = 0, \quad (2.3.2)$$

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n. \quad (2.3.3)$$

Door uitschrijven van

$$(1+t)^n = (1+t)(1+t)^{n-1}$$

vinden we verder

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1} \quad (2.3.4)$$

hetgeen ook gemakkelijk rechtstreeks te bewijzen is. Hierop berust de bekende driehoek van Pascal :

	1	t	t ²	t ³	t ⁴
1	1				
(1+t)	1	1			
(1+t) ²	1	2	1		
(1+t) ³	1	3	3	1	
(1+t) ⁴	1	4	6	4	1
		enz.			

Vraagstukken

2.3.1. Bewijs :

$$\binom{n}{r} = \binom{n-2}{r} + 2 \binom{n-2}{r-1} + \binom{n-2}{r-2}.$$

2.3.2. Bewijs :

$$\begin{aligned} \binom{2n}{n} &= \binom{n}{n} \binom{n}{0} + \binom{n}{n-1} \binom{n}{1} + \binom{n}{n-2} \binom{n}{2} + \dots + \binom{n}{0} \binom{n}{n} \\ &= \binom{n}{0}^2 + \binom{n}{1}^2 + \binom{n}{2}^2 + \dots + \binom{n}{n}^2 \end{aligned}$$

2.4. De hypergeometrische verdeling

Gegeven een populatie uit N elementen, waarvan M het kenmerk A dragen. We nemen hieruit een aselechte steekproef van n elementen ; hoe groot is de kans dat we hierin m elementen A aantreffen ?

Een aselechte steekproef is een steekproef getrokken op zodanige wijze dat alle $\binom{N}{n}$ combinaties van n elementen een gelijke kans bezitten te worden getrokken. We nemen aan dat aan deze voorwaarde wordt voldaan indien de populatie voor het trekken van de steekproef voldoende wordt "geschud".

Dan geldt de zg. hypergeometrische verdeling

$$\begin{aligned} \Pr(m) = \Pr(m | N, M, n) &= \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}} = \frac{\binom{n}{m} \binom{N-n}{M-m}}{\binom{N}{M}} = \\ &= \frac{M!(N-M)! N!(N-n)!}{N! m!(M-m)!(n-m)!(N-M-n+m)!} \end{aligned} \tag{2.4.1}$$

waarbij m aan de volgende restricties is gebonden

$$\begin{aligned} m &\geq 0 & \text{en} & & m &\geq M + n - N, \\ m &\leq n & \text{en} & & m &\leq M. \end{aligned}$$

Voldoet m niet aan één van deze voorwaarden, dan wordt een van de faculteiten in de noemer van (2.4.1) de faculteit van een negatief getal en dus $\pm \infty$. De betreffende kans wordt nul. Met de in (2.1.1) gegeven definitie van n! kunnen we (2.4.1) dus toepassen voor $0 \leq m \leq n$. We kunnen dit opvatten als een splitsing van de populatie naar twee kenmerken A, \bar{A} en wel of niet tot de steekproef behorend. Dit geeft het volgende eenvoudige schema :

$N - M - n + m$	$n - m$	$N - M$
$M - m$	m	M
$N - n$	n	N

Tabel 2.4.1. geeft een voorbeeld van een berekening op (2.4.1) gebaseerd ; de kans op 0,1,2,3, of 4 azen in een bridge spel.

Tabel 2.4.2 geeft de kansen op 0,1,...,13 kaarten van één soort, met een experimentele verificatie.

Het ligt voor de hand te vragen een gemiddelde te definiëren als volgt.

$$E(\underline{m}) = \sum_0^n m \Pr(m).$$

Een dergelijk gemiddelde noemt men ook wel de verwachtingswaarde van de toevalsvariabele \underline{m} ; de E (van het Engelse "Expectation") is het meest gangbare symbool. Voor de verdeling van de azen vinden we uit Tabel

$$E(\underline{m}) = 1.0001$$

en voor het aantal kaarten van één soort uit Tabel

$$E(\underline{m}) = 3.2504$$

hetgeen te verwachten.

Tabel 2.4.1

Berekening van de kans op 0,1,2,3 of 4 azen in een groep van 13 kaarten aselect getrokken uit een spel van 52 kaarten

$N = 52, M = 4, n = 13, m = 0, 1, 2, 3, 4.$ $N - M = 48, (N - n) = 39.$						
$\log \frac{M!(N-M)!n!(N-n)!}{N!} = 50.6714$						
m	log m!	log(M-m)!	log(n-m)!	log(N-n-M+m)!	log p _m	p _m
0	.0000	1.3802	9.7943	40.0142	.4827-1	.3039
1	.0000	.7782	8.6803	41.5705	.6424-1	.4389
2	.3010	.3010	7.6012	43.1387	.3295-1	.2136
3	.7782	.0000	6.5598	44.7185	.6149-2	.0412
4	1.3802	.0000	5.5598	46.3096	.4218-3	.0026
					Som =	1.0002
<p>Gemiddelde aantal azen $E(\underline{m}) = \sum_0^4 mp_m.$</p> <p>$E(\underline{m}) = 0.3039 \times 0 + 0.4389 \times 1 + 0.2136 \times 2 + 0.0412 \times 3 +$ $+ 0.0026 \times 4 = 1.0001.$</p>						

Tabel 2.4.2

Berekende kansen en waargenomen fracties voor het optreden van 0,1,...,13 kaarten van één soort in 13 kaarten aselect getrokken uit een spel van 52.

De waarneming geschiedde door een spel van 52 kaarten 15 keer in 4 groepen van 13 te verdelen, en daarna in iedere groep het aantal schoppen, harten, ruiten, en klaveren te tellen.

m	Pr(m) ber.	\hat{p}_m waargen.	freq.	Turfstand
0	0.0128	0.0125	3	///
1	0.0801	0.0667	16	/// // // /
2	0.2059	0.2042	49	/// // // // // // // // // // //
3	0.2863	0.3167	76	/// //
4	0.2386	0.2458	59	/// //
5	0.1247	0.1000	24	/// //
6	0.0416	0.0416	10	/// //
7	0.0088	0.0083	2	//
8	0.0012	0.0042	1	/
9	0.0001	0.0000	0	
10 -14	< 0.0001			
	1.0001	1.0000	240	

Voor de hypergeometrische verdeling moet

$$\sum_0^n \text{Pr}(m) = 1, \quad (2.4.2)$$

want de serie $m = 0, 1, \dots, n$ vormt een volledig stelsel elkaar uitsluitende waarnemingsuitkomsten. () kan worden bewezen door de coëfficiënt van t^n in

$$(1+t)^M (1+t)^{N-M} = (1+t)^N \quad (2.4.3)$$

of in $(1+t)^n (1+t)^{N-n} = (1+t)^N$

aan elkaar gelijk te stellen.

Verder is

$$E(\underline{m}) = \sum m \text{Pr}(m) = \sum m \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}} = \quad (2.4.4)$$

$$\frac{Mn}{N} \sum \frac{\binom{M-1}{m-1} \binom{N-M}{n-m}}{\binom{N-1}{n-1}} = \frac{Mn}{N},$$

een te verwachten resultaat. De fractie met kenmerk A in de steekproef, $\frac{m}{n}$, heeft gemiddeld dezelfde waarde als de fractie in de populatie, $\frac{M}{N}$.

2.5. Algemeneren gevallen

a) Hoe groot is de kans dat 13 kaarten aselekt getrokken bestaat uit 3 schoppen, 3 harten, 2 ruiten en 5 klaveren ?

$$P = \binom{13}{3} \binom{13}{3} \binom{13}{2} \binom{13}{5} / \binom{52}{13} = 0.0129 \quad (2.5.1)$$

b) Hoe groot is de kans dat 13 kaarten aselekt getrokken over de 4 soorten zijn verdeeld in de aantallen 2,3,3,5 ?

Om deze kans te vinden moeten we de voorgaande kans vermenigvuldigen met het aantal manieren waarop we de getallen 2,3,3,5 over de 4 soorten kunnen verdelen. Dat is met

$$\frac{4!}{2!} \cdot$$

De gevraagde kans is

$$\frac{4!}{2!} \binom{13}{3}^2 \binom{13}{2} \binom{13}{5} / \binom{52}{13} = 0.1548. \quad (2.5.2)$$

c) Hoe groot is de kans dat de schoppen over de 4 spelers zijn verdeeld in de aantallen 2,3,3,5.

We moeten nu eerst de schoppen in deze aantallen verdelen : dit kan op

$$\frac{13!}{2!3!3!5!} \times \frac{1}{2!} \quad (2.5.3)$$

manieren.

Daarna de resterende 39 kaarten in groepen van 11,10,10,8 ; dit kan op

$$\frac{39!}{11!10!10!8!} \times \frac{1}{2!} \quad (2.5.4)$$

manieren. Het aantal manieren waarop een kaartspel in 4 groepen van 13 kaarten kan worden gesplitst, zodanig dat daarin resp. 2,3,3, en 5 schoppen voorkomen, is $2!^4$ het product van (2.5.3) en (2.5.4) want we kunnen de twee groepen met 10 niet-schoppen op $2!$ manieren combineren met de twee groepen van 3 schoppenkaarten. Tenslotte kunnen we de 4 aldus gevormde groepen van 13 kaarten nog op $4!$ manieren over de 4 spelers distribueren.

Het totaal aantal mogelijkheden is $\frac{52!}{13!^4}$.

De gevraagde kans wordt dus

$$\frac{13! 39! 2! 4!}{2! 3! 5! 11! 10! 8! 2! 2! 52!} = 0.1548.$$

De kans is hetzelfde als in vraagstuk 2.5.2.

2.6. De binomiale verdeling

Bij het trekken van n loten uit een loterij kunnen we de trekking uitvoeren zonder of met teruglegging.

Zonder terugleggen wil zeggen dat men n loten na elkaar trekt en de getrokken loten steeds apart houdt ; of dat men ineens n loten tegelijk trekt. Het terugleggen wil zeggen dat men de loten één voor één trekt en na de waarneming het lot weer aan de loterij toevoegt voor men het volgende lot trekt, zodat hetzelfde lot bij volgende trekkingen opnieuw meespeelt. In 2.4 en 2.5 beschouwden we trekking zonder teruglegging. Hieronder wordt de trekking met teruglegging besproken.

We denken ons weer een loterij met N elementen waarvan M met kenmerk A . We trekken hieruit $n \times$ één element met teruglegging. Hoe groot is de kans dat hieronder m elementen A worden waargenomen ?

Bij iedere trekking zijn de kansen of A en \bar{A} steeds hetzelfde

$$P_A = p = \frac{M}{N}, \quad P_{\bar{A}} = q = 1 - p = 1 - \frac{M}{N}. \quad (2.6.1)$$

De kans dat onder n trekkingen er m met A en $(n - m)$ met \bar{A} in een bepaalde rangschikking worden waargenomen is

$$p^m q^{n-m}.$$

De m elementen A en $(n - m)$ elementen \bar{A} kunnen in $\binom{n}{m}$ verschillende rangschikkingen optreden. De gevraagde kans is dus

$$\Pr(m) = \Pr(m|n, p) = \binom{n}{m} p^m (1 - p)^{n-m} = \binom{n}{m} p^m q^{n-m}. \quad (2.6.2)$$

Dit is de formule voor de binomiale verdeling.

We hebben

$$\sum_0^n \binom{n}{m} p^m q^{n-m} = (p + q)^n. \quad (2.6.3)$$

De binomiale verdeling wordt gegeven door de termen in de ontwikkeling van het gedurig product $(p + q)^n$. Daar $(p + q) = 1$ volgt uit (2.6.3) direkt dat

$$\sum_0^n \Pr(m) = 1. \quad (2.6.4)$$

Na differentiëren naar p en daarna vermenigvuldigen met p geeft (2.6.3)

$$\sum m \Pr(m) = np(p + q)^{n-1} = np. \quad (2.6.5)$$

Het gemiddelde van de binomiale verdeling is np .

Trekken we 13 kaarten met teruglegging of zonder uit een spel van 52 dan zijn de kansen op 0,1,2,... azen als in Tabel 2.6.1.

Tabel 2.6.1

Kans op 0,1,2,... azen in 13 kaarten aselekt getrokken uit een spel van 52 met en zonder teruglegging.

Tabel 2.6.1.

Aantal azen m	Kans met teruglegging	Kans zonder teruglegging
0	0.3533	0.3039
1	0.3826	0.4389
2	0.1914	0.2135
3	0.0585	0.0412
4	0.0122	0.0026
5	0.0018	-
6	0.0002	-
Som =	1.0000	1.0001

Met teruglegging kunnen meer dan 4 azen voorkomen, doch de kans hierop is gering. Het verloop van de kansen vertoont overigens veel overeenkomst.

Het verschil wordt daardoor veroorzaakt, dat bij trekking zonder terugleggen de kansen veranderen naarmate men elementen aan de populatie onttrekt. Dit effect wordt echter geringer naarmate de populatie groter is. Voor $N \rightarrow \infty$ gaat de hypergeometrische verdeling in de binomiale over. Trekken met terugleggen kan men daarom ook beschrijven als trekken uit een oneindige populatie.

In de praktijk maken we een steekproef steeds zonder terugleggen ; doch het aantal elementen in de populaties is in vele gevallen zó groot, dat de binomiale formule als een goede eerste benadering voor het berekenen van kansen dienst kan doen. Dit heeft grote voordelen, daar deze formule wiskundig veel eenvoudiger is dan de hypergeometrische. Zij hangt alleen van n en p af en is uitvoerig getabelleerd.

Daarnaast kennen we processen als het werpen met een dobbelsteen waar bij herhaling steeds weer dezelfde kansen optreden. Hier is de binomiale verdeling exact van toepassing.

We zullen later nog uitvoerig op toepassingen van de binomiale verdeling terugkomen.

2.7. Algemene problemen

Tot algemene problemen leidt bijv. het pokerspel met dobbelstenen. Hoe groot is de kans met 5 dobbelstenen 2 x een paar gelijke puntentallen te gooien ; dus een worp van het type aa bb c.

Eerst kiezen we 3 uit de 6 mogelijke puntentallen ; dit kan op

$$\binom{6}{3}$$

manieren. Dan splitsen we de 5 worpen in 2 + 2 + 1 ; dit kan op

$$\frac{5!}{2!2!1!} \times \frac{1}{2!}$$

manieren. We kunnen tenslotte de 3 aangewezen puntentallen op 3! manieren over de 3 groepen worpen permuteren. Er zijn dus

$$\binom{6}{3} \frac{5!}{2!2!1!} \frac{3!}{2!}$$

gunstige combinaties, tegen 6 mogelijke combinaties in totaal. De gezochte kans is

$$\binom{6}{3} \frac{5!}{2!2!1!} \frac{3!}{2!} / 6^5 = 0.2314. \quad (2.7.1)$$

Het pendant van dit probleem zonder teruglegging zou zijn het aselekt trekken van kaarten uit een spel van 24 verdeeld in 6 groepen van 4. De kans op een samenstel 2 + 2 + 1 is dan

$$\binom{6}{3} \binom{3}{1} \binom{4}{2}^2 \binom{4}{1} \binom{24}{5} = 0.2033. \quad (2.7.2)$$

Op soortgelijke wijze werden de pokerkansen voor andere pokercombinaties berekend ; zij zijn weergegeven in Tabel 2.7.1.

Tabel 2.7.1.

Pokerkans bij pokeren met 5 stenen en met 14 kaarten zonder terugleggen.

Combinatie	Kans met 5 stenen	Kans met 24 kaarten
5 gelijken	.0008	-
4 + 1	.0193	.0028
3 + 2	.0386	.0169
3 + 1 + 1	.1543	.0903
2 + 2 + 1	.2314	.2033
2 + 1 + 1 + 1	.4630	.5421
1 + 1 + 1 + 1 + 1	.0926	.1445

2.10 Vraagstukken

- 2.10.1. Wanneer speler A in een bridgespel 2 azen in handen heeft. hoe groot zijn dan de kansen dat zijn partner 0,1, of 2 azen heeft ?
- 2.10.2. 6 Partijen van 100 producten bevatten resp. 1,2,5,10,15 of 20 ondeugdelijke exemplaren. Uit deze partijen worden aselekt steekproeven van 10 stuks genomen. Hoe groot zijn de kansen dat in deze steekproeven geen enkel ondeugdelijk exemplaar wordt aangetroffen ? Bereken deze kansen met de hypergeometrische en met de binomiale verdeling, en zet de kans uit tegen het percentage ondeugdelijke exemplaren in de partij.
- 2.10.3. Hoe groot zijn bij een worp met 4 dobbelstenen de kansen dat 0,1,2,3, of 4 zessen worden gegooid ?
- 2.10.4. Hoe groot zijn de kansen dat bij $5 \times$ tossen met een munt 0,1,2,3,4, of $5 \times$ "munt" wordt waargenomen ?
- 2.10.5. Bij een bridgespel kunnen de azen als volgt zijn verdeeld :

4	0,		
3	1,		
2	2,		
2	1	1,	
1	1	1	1.

Wat zijn de kansen van deze 5 mogelijke verdelingen.

- 2.10.6. Twee tennisspelers zijn van gelijke kracht. Hoe groot is de kans dat een spel als "long set" eindigt ? Beantwoord dezelfde vraag wanneer A een kans p heeft een game te winnen en B een kans $q = 1 - p$.
- 2.10.7. Toevalscijfers zijn cijfers van 0 tot 9 in aselechte volgorde (geloot met teruglegging). Hoe groot zijn de kansen dat een groep van 4 toevalscijfers bestaat uit

4 gelijken,	,aaaa
3 gelijken, 1 verschillend	,aaab, abaa enz.
2 paren,	,aabb, baba, enz.
2 gelijke, 2 verschillend	,aabc, aaba, enz.
4 verschillende	,abcd, dacb, enz.

Verifieer dit met behulp van Tabel 21 van de bij dit college behorende Statistische Tabellen.

2.11. Een aanvulling

Het volgende vraagstuk leidt tot een toepassing van de algemene somregel.

De gehele getallen van 1 tot n worden in aselechte volgorde geplaatst. Hoe groot is de kans dat geen enkel getal op zijn natuurlijke plaats terecht komt?

De kans dat het getal 1 op de eerste plaats komt, ongeacht de plaats van de andere getallen, is $\frac{1}{n}$; hetzelfde geldt voor alle $\binom{n}{1}$ andere getallen.

De kans dat de 1 en de 2 op hun juiste plaats komen is $\frac{1}{n(n-1)}$ en hetzelfde geldt voor alle $\binom{n}{2}$ paren die we uit de n getallen kunnen vormen.

Voor drie getallen op hun plaats is de kans $\frac{1}{n(n-1)(n-2)}$ en er zijn $\binom{n}{3}$ groepen van 3.

Toepassing van de algemene somregel (1.4.2) geeft nu

Kans dat geen getal op zijn plaats staat = $p =$

$$\begin{aligned} &= 1 - \binom{n}{1} \times \frac{1}{n} + \binom{n}{2} \frac{1}{n(n-1)} - \binom{n}{3} \frac{1}{n(n-1)(n-2)} \dots + (-1)^n \binom{n}{n} \frac{1}{n!} = \\ &= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + (-1)^n \frac{1}{n!}. \end{aligned} \quad (2.11.1)$$

De limiet voor $n \rightarrow \infty$ is

$$p = 1 - e^{-1} = 0.6321.$$

Vraagstuk:

2.11.1. Hoe groot is de kans dat bij n worpen met een dobbelsteen de 6 zijden allen tenminste één maal zijn waargenomen?

2.12. Toevalscijfers

Een belangrijk hulpmiddel in de toegepaste statistiek vormen de zg. toevalscijfers; dat zijn reeksen cijfers van 0 tot 9 in aselechte volgorde, dus verkregen door voortdurend trekken met teruglegging. Deze kunnen op verschillende manieren worden gegenereerd, bijv. (demonstratie)

- werpen met een symmetrische 10-zijdige cylinder (Hamaker)
- met een dobbelsteen volgens een regelmatig twintig vlak. (Japanse Normalisatie).
- met behulp van een bol met holle glazen stengel en 10 kralen; 9 witte en een zwarte.
- langs electronische weg door een snel electronisch telwerk te laten reageren op kosmische stralen.

In de moderne leerboeken over statistiek is bijna altijd aan het eind een tabel met toevalscijfers opgenomen. Tabel 21 bij dit college geeft een verzameling van 2500 toevalscijfers, gemakshalve gerangschikt in paren.

Uitgebreidere tabellen zijn die van :

a) M.G.Kendall en B.Babington-Smith, Tracts for computers XXIV, Tables of random sampling numbers, Cambridge University Press, (100 000), \$ 5 /.

b) The RAND Corporation, A million random digits, The Free Press, Glencoe, Illinois 1955, \$ 10,-

Voorbeelden

- 2.12. a) In Tabel 2.6.1. vermeldden we de kansen op 0,1,2,3, of 4 azen in een hand van 13 kaarten aselekt getrokken uit een spel van 52. Door afrondingsfouten was de som 1.0001. Bij de navolgende toepassing is het gewenst, dat de som exact 1.0000 is, hetgeen we bereiken door de grootste kans, die op 1 aas, met 0.0001 te verminderen. Dit geeft de volgende waarde.

TABEL 2.12

Toepassing van toevalscijfers : Aantal azen in 13 bridgekaarten.

Aantal azen	Kans	Cumulatieve kansen	Interval	Score
0	0.3039	0.3039	0000 - 3038	0
1	0.4388	0.7427	3039 - 7426	1
2	0.2135	0.9562	7427 - 9561	2
3	0.0412	0.9974	9562 - 9973	3
4	0.0026	1.0000	9974 - 9999	4
	1.0000			

We tellen nu deze kans successievelijk bij elkaar op ; dit geeft de cumulatieve kansen in de 3de kolom, en construeren hieruit de intervallen in kolom 4. Tenslotte nemen getallen samengesteld uit 4 toevalscijfers toe ; wanneer dit getal valt in het interval 0000 - 3038 kennen we de score 0 toe, in het interval 3039 - 7426 de score 1 etc.

Tabel 21 de eerste kolom van boven naar beneden geeft bijv.

Toevalsgetal	3606	0725	0233	5187	8972	9467	4020	9452
Score	1	0	0	1	2	2	1	2

De serie scores aldus verkregen gedraagt zich statistisch gezien geheel als het aantal azen bij successieve verdeling van een bridgespel. Natuurlijk is het daarbij ongewenst steeds dezelfde toevalscijfers te gebruiken. Normaal begint men daarom op een willekeurig punt in de tabel en schrijft vandaar in een willekeurige richting voort. Beginnen we bv. in Tabel 21 in de 16e kolom en de 35e regel en gaan van beneden naar boven dan vinden we.

Toevalsgetal	9478	0131	2917	5515	2822	1685	1041	6808
Score	2	0	0	1	0	0	0	1

2.12.b) De kans dat in een groep van 25 mensen tenminste twee op eenzelfde dag jarig zijn bedraagt

$$p = 0,57.$$

Dit is ook de kans dat in een groep van 25 getallen aselekt met teruglegging getrokken uit de getallen 1 tot 365 er tenminste 2 gelijken voorkomen. Dergelijke series vinden we door toevalsgetallen van 3 cijfers te nemen en hieruit het getal 000 en alle getallen groter dan 365 te schrappen ; of beter door alleen de getallen groter dan 730 te schrappen en die van 366-730 met 365 te verminderen.

De laatste 3 toevalscijfers uit iedere regel in Tabel 21 geven

Toevalsgetal	412	154	337	809	853	277	853	556
Verjaardag	47	154	337	-	-	277	-	181

In 8 series van 25 verjaardagen, aldus geconstrueerd, vond ik respectievelijk

paren.

2.12.c) Verlotingsseries. Tabel 22

Algemeen gebruikt men toevalscijfers om uit een populatie waarvan de elementen genummerd zijn, (bijv. uit een bevolkingsregister) een aselechte steekproef te trekken. In principe is dit steeds trekking met teruglegging. We maken hiervan trekking zonder teruglegging door getallen die meer dan eenmaal voorkomen, alleen de eerste maal mee te tellen en daarna te laten vervallen. Bij het trekken van een relatief kleine steekproef uit een grote populatie is dit geen groot bezwaar, want dan is de kans op duplicatie gering.

Bij kleine populatie kan men met voordeel gebruik maken van de verlotingsseries in Tabel 22. Dit zijn eindige series getallen in aselechte volgorde zonder teruglegging.

Wenst men bijv. uit een populatie van 35 elementen een steekproef van 10 stuks te trekken, dan neemt men één van de series van 00 tot 49, bijv. de 3e en begint op een willekeurig punt, bijv. met het 8e getal ; dan zijn

01 04 20 24 45 00 11 27 47 32 16 36 25 07

de nummers van de elementen in de steekproef. Het getal 00 en getallen boven de 35 worden weer geschrapd.

Verlotingsseries kunnen worden verkregen door een loting uit te voeren, doch ook met behulp van toevalscijfers. Wensen we bijv. de cijfers 1,2,...,6 in aselechte volgorde te zetten dan kunnen we aan ieder een toevalsgetal van bijv. 4 cijfers toe, aldus

1	2	3	4	5	6
6784	1526	3787	8324	4001	4787

en hergroeperen de cijfers naar de grootte van deze toevalsgetallen ; dit geeft

2 3 5 6 1 4.

Bij grotere series kan dit hergroeperen machinaal gebeuren door gebruik van zg. ponskaarten.

Vraagstuk

- 2.12.1. Verifieer de uitkomst van het in 2.11 behandelde vraagstuk aan enige verlotingsseries.
- 2.12.2. Construeer met behulp van toevalsgetallen een serie van 100 getallen die de verdeling van het aantal schoppen kaarten in een bridgehand weergeven, gerangschikt in 20 regels van 5 getallen.
(Deze zullen later nog verder als oefenmateriaal worden gebruikt).

HOOFDSTUK III. Continue toevalsvariabelen en continue verdelingen

- 3.1. We maken een grote gedachtensprong. Als kenmerken A, B, en C waarover tot nu toe werd gesproken beschouwen we in dit hoofdstuk metingen, bijv. van de dikte of het gewicht. Tabel 3.1.1. bevat het gewicht in milligrammen gemeten aan 90 hardpapieren schijfjes. In Tabel 3.1.2. zijn deze weergegeven in een turfstaatje.

TABEL 3.1.1.

Gewichten in mgr. aan 90 hardpapieren schijfjes

792	799	804	780	776	806	799	774	779
807	802	792	807	761	781	787	805	803
760	809	804	799	793	801	803	810	807
793	752	791	805	807	793	796	803	791
802	793	807	784	802	793	772	791	804
787	787	781	807	792	814	796	782	793
803	789	811	791	806	786	799	813	801
783	787	805	812	789	776	802	805	786
787	787	799	812	795	800	806	796	821
787	801	759	799	758	803	810	806	790

TABEL 3.1.2.

Turfstaat geconstrueerd uit Tabel 3.1.1.

Interval in mgr.	Turfstaat	Frequentie
750-756	/	1
757-763	////	4
764-770		0
771-777	////	4
778-784	//// //	7
785-791	//// // // //	15
792-798	//// // // // //	14
799-805	//// // // // // // /	26
806-812	//// // // // /	16
813-819	//	2
820-826	/	1
Som =		90

In principe zouden we de intervallen kleiner kunnen kiezen en veel meer metingen kunnen doen. De ervaring leert, dat men dan in vele gevallen een zeer regelmatig verloop van de frequentie verkrijgt. Dit suggereert dat we hier te maken hebben met een toevalsvariabele, het gewicht, waarvan de verdeling door een continue functie $f(x)$ van de vorm in fig. 3.1.1. getekend kan worden beschreven.

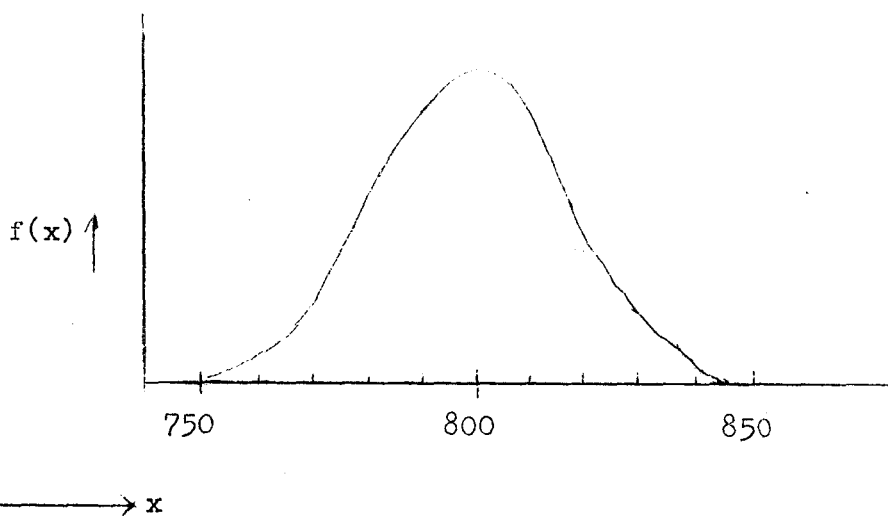


Fig.3.1.1. Een frequentieverdeling die de resultaten van Tabel 3.1.2. schematisch weergeeft.

Deze principiële gedachtengang wordt hieronder verder uitgewerkt.

3.2. De kansdichtheid of frequentieverdeling $f(x)$

De functie $f(x)$ moet zó worden geïnterpreteerd dat de kans dat een meting \underline{x} zal vallen in het interval (x, dx) gegeven wordt door

$$\Pr(x < \underline{x} < x + dx) = f(x) dx. \quad (3.2.1)$$

De functie $f(x)$ heet de kansdichtheid of de verdelingsfunctie van \underline{x} ; deze functie wordt steeds continu ondersteld.

De kans $\Pr(\underline{x} = x)$ dat \underline{x} een voorgeschreven waarde x aanneemt is nul; of we in (3.2.1) $<$ of \leq x schrijven maakt daarom geen verschil.

De Somregel voor elkaar uitsluitende kenmerken wordt nu

$$\Pr(x_1 < \underline{x} < x_2) = \int_{x_1}^{x_2} f(\xi) d\xi, \quad (3.2.2)$$

en voor een volledig stelsel

$$\int_{-\infty}^{+\infty} f(\xi) d\xi = 1. \quad (3.2.3)$$

waarbij we de integratiegrenzen gemakshalve van $-\infty$ tot $+\infty$ laten lopen. Zij kunnen soms ook tot een eindig interval beperkt zijn.

Men noemt

$$\Pr(\underline{x} < x) = F(x) = \int_{-\infty}^x f(\xi) d\xi \quad (3.2.4)$$

de cumulatieve kansdichtheid of de cumulatieve verdeling van x.

Voorts is

$$\text{het populatiegemiddelde} = \mu = E(\underline{x}) = \int_{-\infty}^{+\infty} \xi f(\xi) d\xi. \quad (3.2.5)$$

μ en $E(\underline{x})$ zijn hiervoor gangbare problemen.

Zij

$$y = g(x) \quad (3.2.6)$$

een functie van x , dan zal met iedere waarneming x een waarde van y corresponderen. Met de toevalsvariabele \underline{x} correspondeert dan een toevalsvariabele \underline{y} :

$$\underline{y} = g(\underline{x}) \quad (3.2.7)$$

Het populatiegemiddelde van \underline{y} wordt dus gegeven door

$$E(\underline{y}) = E\{g(\underline{x})\} = \int_{-\infty}^{+\infty} g(\xi) f(\xi) d\xi. \quad (3.2.8)$$

Het symbool $E(\)$ betekent de operator:

$$\int_{-\infty}^{+\infty} \dots f(\xi) d\xi. \quad (3.2.9)$$

Is bijv.

$$\underline{y} = a\underline{x} + b$$

dan is

$$E(\underline{y}) = aE(\underline{x}) + b. \quad (3.2.10)$$

De meeste frequentieverdelingen die we in de praktijk tegenkomen, hebben de eenvoudige klokvormige gedaante in Fig. 3.1.1. geschetst. Voor een volledige beschrijving van een dergelijke kromme zouden we de functie $f(x)$ exact moeten kennen, doch een zo gedetailleerde kennis heeft praktisch geen waarde. Men kan er voor de meeste problemen mee volstaan de verdeling door een paar karakteristieke parameters te beschrijven. Het gemiddelde μ of $E(\underline{x})$ is een van deze. Een tweede, de variantie, levert ons een maat voor de spreiding ; zij is gedefinieerd door

$$\begin{aligned} \text{variantie van } \underline{x} = \text{var}(\underline{x}) &= c^2 = \int_{-\infty}^{+\infty} (\xi - \mu)^2 f(\xi) d\xi = \\ &= \int_{-\infty}^{+\infty} \{\xi - E(\underline{x})\}^2 f(\xi) d\xi. \end{aligned} \quad (3.2.11)$$

De wortel uit de variantie σ heet de standaard afwijking. De betekenis van deze grootheden zal in het vervolg wel blijken. Is $\underline{y} = g(\underline{x})$ dan is de variantie van \underline{y} in overeenstemming met (3.2.11) gegeven door

$$\text{var}(\underline{y}) = \int_{-\infty}^{+\infty} [g(\xi) - E\{g(\underline{x})\}]^2 f(\xi) d\xi. \quad (3.2.12)$$

Is $\underline{y} = a\underline{x} + b$ dan geeft toepassing van (3.2.12)

$$\text{var}(\underline{y}) = \text{var}(a\underline{x} + b) = a^2 \text{var } \underline{x}. \quad (3.2.13)$$

3.2. Aanvulling

Voor de variantie kunnen we ook schrijven

$$\text{var}(\underline{x}) = \sigma^2 = E(\underline{x}^2) - \{E(\underline{x})\}^2 \quad (3.2.14)$$

en evenzo wanneer $\underline{y} = g(\underline{x})$

$$\text{var}(\underline{y}) = E[\{g(\underline{x})\}^2] - [E\{g(\underline{x})\}]^2 \quad (3.2.15)$$

Deze formules worden meestal gebruikt om varianties te berekenen. Is verder

$$y = g(x)$$

een monotone differentieerbare functie van x en

$$x = h(y)$$

de inverse functie dan correspondeert met de toevalsvariabele \underline{x} een toevalsvariabele $\underline{y} = g(\underline{x})$ met een verdeling

$$\varphi(y) = \pm f(h(y)) \frac{dh(y)}{dy} . \quad (3.2.16)$$

Het + teken geldt als $y = g(x)$ een monotoon stijgende, het - teken wanneer het een monotoon dalende functie is. Is y geen monotone functie van x dan is $\underline{y} = g(\underline{x})$ nog steeds een toevalsvariabele doch is die verdeling niet op zo eenvoudige wijze uit te drukken. Formules (3.2.8), (3.2.12) en (3.2.15) blijven steeds geldig.

3.3. Enige voorbeelden van continue verdelingen

3.3.1. De rechthoekige verdeling

Definities van de standaard rechthoekige verdeling :

$$\begin{aligned} f(x) &= 1 & \text{voor} & & 0 \leq x \leq 1, \\ f(x) &= 0 & \text{voor} & & x < 0 \text{ of } x > 1. \end{aligned} \quad (3.3.1)$$

Hieruit volgt

$$\left. \begin{aligned} F(x) &= x, \\ E(\underline{x}) &= 1/2 \\ E(\underline{x}^2) &= 1/3 \\ \text{var}(\underline{x}) &= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}, \\ \sigma(x) &= \sqrt{\frac{1}{12}} = 0.289. \end{aligned} \right\} \quad (3.3.2)$$

Getallen verkregen door toevalsgetallen te schrijven als decimale breuken vormen een benadering van waarnemingen met een standaard rechthoekige verdeling.

Heeft \underline{x} een standaard rechthoekige verdeling dan heeft \underline{y}

$$\underline{y} = a + (b - a)\underline{x}$$

een rechthoekige verdeling in het interval a, b , met

$$\left. \begin{aligned} f(y) &= \frac{1}{b - a} && \text{voor } a < y < b, \\ f(y) &= 0 && \text{voor } y < a \text{ of } y > b, \\ E(\underline{y}) &= a + \frac{1}{2}(b - a) = \frac{1}{2}(a + b), \\ \text{var}(\underline{y}) &= \frac{1}{12} (b - a)^2. \end{aligned} \right\} \quad (3.3.3)$$

3.3.2. De exponentiële verdeling

Definitie :

$$\left. \begin{aligned} f(x) &= \lambda e^{-\lambda x}, && \text{voor } x \geq 0, \\ f(x) &= 0, && \text{voor } x < 0. \end{aligned} \right\} \quad (3.3.4)$$

Hieruit volgt

$$\left. \begin{aligned} F(x) &= 1 - e^{-\lambda x}, \\ E(\underline{x}) &= \frac{1}{\lambda} \\ E(\underline{x}^2) &= \frac{2}{\lambda^2} \\ \text{var}(\underline{x}) &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \\ \sigma(\underline{x}) &= \frac{1}{\lambda} \end{aligned} \right\} \quad (3.3.5)$$

Heeft \underline{x} een standaard rechthoekige verdeling dan heeft

$$\underline{y} = -\frac{1}{\lambda} e_{\log \underline{x}}$$

een exponentiële verdeling met parameter λ . Uit de 5e en 6e kolom van tabel 21 vinden we de toevalsgetallen

$$x = 0.0948 \quad 0.9384 \quad 0.5202 \quad 0.2022 \quad 0.0475 \quad 0.7149 \quad \text{enz.}$$

en hieruit voor $\lambda = 0.1$ de y -waarden

$$y = 23.56 \quad 0.64 \quad 6.54 \quad 15.98 \quad 44.29 \quad 3.36 \quad \text{enz.}$$

Zo kunnen we steekproeven uit een willekeurige exponentiële verdeling construeren.

De exponentiële verdeling speelt in de praktijk een belangrijke rol bij de wachttijd theorie, bij levensduurproblemen en in de operationsresearch. Wanneer in een continu proces op volmaakt willekeurige tijdstippen storingen kunnen optreden is het redelijk in eerste benadering te onderstellen dat in ieder tijdsinterval dT een kans

$$p = \lambda dT$$

op een storing bestaat. De kans dat in een tijds interval $0 - T$ geen storing optreedt is dan

$$P = (1 - \lambda dT)^n \quad \text{waarbij } n = \frac{T}{dT} .$$

In de limiet $dT \rightarrow 0$, $n \rightarrow \infty$ wordt dit

$$P = e^{-\lambda T} .$$

De kans dat in het interval $0, T$ geen en tussen T en $T + dT$ wèl een storing optreedt bedraagt dan

$$f(T)dT = e^{-\lambda T} \lambda dT .$$

De verdeling van de tijden tussen opeenvolgende storingen is dus een exponentiële verdeling. $1/\lambda =$ de gemiddelde tijd tussen storingen, of de gemiddelde looptijd; $\lambda =$ het gemiddelde aantal storingen per tijds-eenheid of de storingsdichtheid.

Men onderstelt bv. een exponentiële verdeling bij

- a) de binnenkomst van gesprekken bij een telefooncentrale,
- b) de binnenkomst van orders in een werkplaats, of patiënten op een spreekuur,
- c) de duur van een telefoongesprek, een werkplaatsopdracht, een doktersconsult,
- d) de levensduur van artikelen die door plotseling onvoorspelbare oorzaken stuk gaan; bv. vaatwerk in de huishouding, matrijzen waaruit gramfoonplaten worden geperst, enz.

Een voorbeeld levert weer Tabel 21. De afstand tussen opeenvolgende nullen moet in eerste benadering door een exponentiële verdeling met $1/\lambda = 10$ worden beschreven. Tabel 3.3.1 geeft een frequentietabel gebaseerd op 200 waarnemingen aan Tabel 21 verricht; de waargenomen frequenties zijn vergeleken met de gemiddelde frequenties berekend uit de exponentiële verdeling.

TABEL 3.3.1.

De frequenties van de afstand tussen opeenvolgende nullen in Tabel 21 vergeleken met de gemiddelde frequentie berekend uit de exponentiële verdeling met $\lambda = 0.1$.

Afstand interval	Waargenomen frequenties	Freq. berekend uit de exp. verd.
1 - 5	80	84.6
6 - 10	51	45.4
11 - 15	34	27.6
16 - 20	9	16.7
21 - 25	10	10.1
26 - 30	3	6.1
31 - 35	5	3.7
36 - 40	4	2.3
41 - 45	0	1.4
46 - 50	2	0.8
51 - 55	2	0.5
> 56	0	0.8
	200	200

De wijze waarop de frequenties in de laatste kolom werden berekend is in principe weergegeven in Tabel 3.3.2. De kans dat een waarneming in een interval x_1, x_2 valt is p .

TABEL 3.3.2

Principe van de berekening van de frequenties in Tabel 3.3.1 uit de exponentiële verdeling met $\lambda = 0.1$.

Interval	Interval-grenzen = x	$e^{-\lambda x}$	p	freq. = 200p
1 - 5	0.0	1.000	0.4230	84.6
6 - 10	5.5	0.5770	0.1272	45.4
11 - 15	10.5	0.3499	0.1377	27.5
16 - 20	15.5	0.2122	0.0835	16.7
	20.5	0.1287		
	enz			

$$p = \int_{x_1}^{x_2} f(\xi) d\xi = F(x_2) - F(x_1) = e^{-\lambda x_2} - e^{-\lambda x_1}.$$

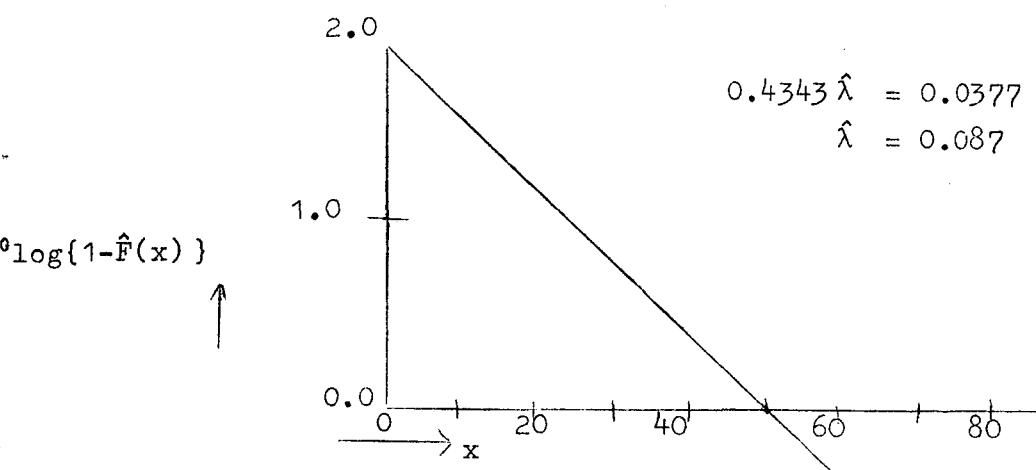
Hiervan is gebruik gemaakt ; de kans dat we een afstand 1 - 5 zullen waarnemen is gelijkgesteld aan de kans dat een waarneming zal vallen in het interval 0 - 5.5 ; de kans op een waarneming 6 - 10 gelijk aan de kans op een waarneming in het interval 5.5 - 10.5, enz. Men lette vooral op de betekenis van de klasse grenzen bij deze berekening. In Tabel 3.3.1 en 3.3.2 verkeerden we in de gunstige omstandigheid dat we te voren de waarde $\lambda = 0.1$ konden voorspellen. In de praktijk komt het vaker voor dat we vermoeden dat een verschijnsel door een exponentiële verdeling kan worden beschreven

TABEL 3.3.3

Onderzoek naar de toepasbaarheid van de exponentiële verdeling op de waarnemingen uit Tabel 3.3.1.

Interval	Waargen. frequentie	Rel. freq. in %	Interval grenzen	Cumulat. frq. = \hat{F} in %	$1 - \hat{F}$ in %	$-\log(1 - \hat{F})$
1 - 5	80	40	5.5	40.0	60.0	1.78
6 - 10	51	25.5	10.5	65.5	34.5	1.54
11 - 15	34	17.0	15.5	82.5	17.5	1.24
16 - 20	9	4.5	20.5	87.0	13.0	1.11
21 - 25	10	5.0	25.5	92.0	8.0	0.90
26 - 30	3	1.5	30.5	93.5	6.5	0.81
31 - 35	5	2.5	35.5	96.0	4.0	0.60
36 - 40	4	2.0	40.5	98.0	2.0	0.30
41 - 45	0	0.0	45.4	98.0	2.0	0.30
46 - 50	2	1.0	50.5	99.0	1.0	0.00
51 - 55	2	1.0	55.5	100.0	0.0	- ∞
> 56	0	0.0				
	200					

Fig.3.3.1. $\log\{1 - \hat{F}(x)\}$ uitgezet tegen x.



We zien uit Fig. 3.3.1 dat de punten inderdaad redelijk op een rechte lijn liggen. Uit de uit de hand getrokken rechte vinden we

$$\hat{\lambda} = 0.086 \quad \text{i.p.v.} \quad 0.10.$$

Het probleem in hoeverre het verschil van toevallige aard kan zijn of niet zal later worden behandeld.

3.4. De normale verdeling of verdeling volgens Gauss.

Definitie :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ - \frac{(x - \mu)^2}{2\sigma^2} \right\}, \quad -\infty < x < +\infty. \quad (3.4.1)$$

Hieruit volgt :

$$\begin{aligned} E(\underline{x}) &= \mu, \\ \text{var}(\underline{x}) &= \sigma^2, \\ \sigma(\underline{x}) &= \sigma. \end{aligned} \quad (3.4.2)$$

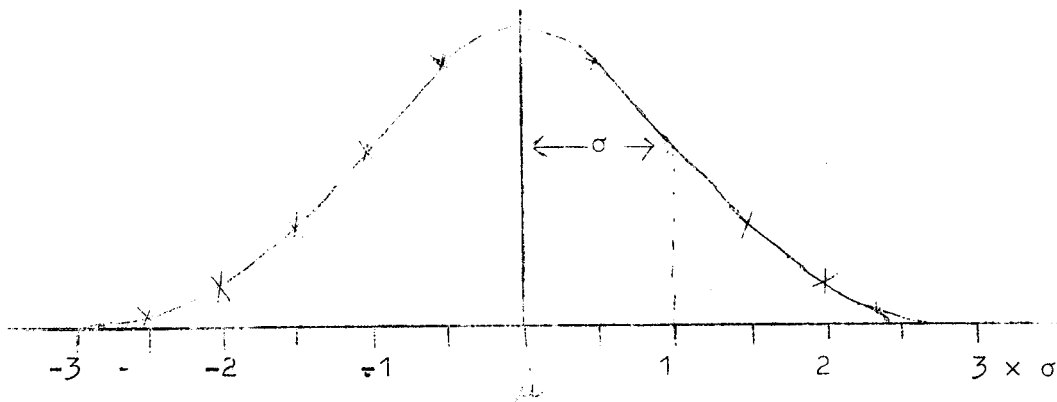


Fig.3.4.1. De normale verdeling

De normale verdeling is een klokvormige verdeling symmetrisch t.o.v. $x = \mu$. In theorie strekt zij zich uit van $-\infty$ tot $+\infty$ doch praktisch is alleen het interval van $\mu - 3\sigma$ tot $\mu + 3\sigma$ van belang. De normale verdeling wordt door twee parameters, het gemiddelde μ en de standaardafwijking σ , volledig bepaald.

Is

$$\underline{u} = \frac{\underline{x} - \mu}{\sigma} \quad \text{of} \quad \underline{x} = \mu + \underline{u} \sigma \quad (3.4.3)$$

dan is de verdeling van \underline{u}

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp. \left(- \frac{u^2}{2} \right). \quad (3.4.4)$$

Dit is de standaard normale verdeling, d.w.z. een normale verdeling met $\mu = 0$, $\sigma = 1$. Waarden van $f(u)$ zijn gegeven in de laatste kolom van Tabel 1. Deze kunnen dienst doen bij het tekenen van een normale verdeling als in fig. 3.4.1. Het symbool \underline{u} zal steeds worden gebruikt voor een toevalsvariabele met deze verdeling. De notatie (3.4.3), $\underline{x} = \mu + \underline{u} \sigma$ betekent dan \underline{x} heeft een normale verdeling met gemiddelde μ en standaardafwijking σ .

Voor de cumulatieve verdeling geldt

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ - \frac{(\xi - \mu)^2}{2\sigma^2} \right\} d\xi =$$

$$= \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} \exp \left(- \frac{t^2}{2} \right) dt = F(u),$$
(3.4.5)

met $u = \frac{x - \mu}{\sigma}$.

(3.4.5) houdt in dat moet gelden

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp \left(- \frac{t^2}{2} \right) dt = 1$$
(3.4.6)

hetgeen we niet zullen bewijzen.

Daar $f(u)$ symmetrisch is t.o.v. $u = 0$ geldt

$$F(u) = 1 - F(-u)$$
(3.4.7)

Voor positieve waarden van u is de cumulatieve normale verdeling $F(u)$ getabelleerd in Tabel 1. De linker pagina geeft een algemene tabel de rechter pagina speciale waarden die veel worden gebruikt.

Voorbeelden :

$$\begin{aligned} \Pr(\underline{u} < 1.14) &= F(1.14) = 0.8729, \\ \Pr(\underline{u} < -1.14) &= F(-1.14) = 1 - F(1.14) = 0.1271, \\ \Pr(\underline{u} > -1.14) &= 1 - F(-1.14) = F(1.14) = 0.8729, \\ \Pr(0.55 < \underline{u} < 2.05) &= F(2.05) - F(0.55) = \\ &= 0.9798 - 0.7088 = 0.2710, \\ \Pr(|\underline{u}| > 0.75) &= 2\{1 - F(0.75)\} = 0.4532. \end{aligned}$$

Is $f(x)$ een bekende verdeling met cumulatieve verdeling $F(x)$ dan kan men door de transformatie

$$\underline{y} = F(\underline{x}),$$

een toevalsvariabele y met standaardrechthoekige verdeling omzetten in een toevalsvariabele x met de verdeling $f(x)$. (zie 3.2.16)
Zo vinden we uit de toevalsgetallen (Tabel 21, rij 20).

0.4161 0.5584 0.9475 0.1963 0.2360 0.0644 enz. ,
door terugzoeken in Tabel 1. de u-waarden

-0.21 0.15 1.62 -0.86 -0.72 -1.52 enz.

Dit is een serie aselechte trekkingen uit een standaard normale verdeling. Tabel 20 geeft 780 waarnemingen die op deze wijze zijn verkregen. Een tabel met 25000 trekkingen is in de handel verkrijgbaar:

H.Wold, Random Normal Deviates,
Tracts for Computers XXV, Cambridge University Press 1954, 5 s/

Is $x = \mu + u\sigma$ en x een gegeven waarde dan noemt men

$$a_r = \Pr\left(\frac{x - \mu}{\sigma} > \frac{x - \mu}{\sigma}\right) = 1 - F\left(\frac{x - \mu}{\sigma}\right) = \quad (3.4.8)$$

de rechteroverschrijdingskant van x ,

$$a_l = \Pr\left(\frac{x - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right) = F\left(\frac{x - \mu}{\sigma}\right) = \quad (3.4.9)$$

de linkeroverschrijdingskans van x , en

$$a = \Pr\left(\left|\frac{x - \mu}{\sigma}\right| > \left|\frac{x - \mu}{\sigma}\right|\right) = 2(1 - F(|\frac{x - \mu}{\sigma}|)) = \quad (3.4.10)$$

de tweezijdige overschrijdingskans van x .

Tabel 1, rechter pagina, geeft waarden voor overschrijdingskansen bij gegeven $u = \frac{x - \mu}{\sigma}$, en waarden van u bij gegeven overschrijdingskansen.

Voorbeelden :

voor $u = 2.20$ is $a_r = 0.0139$, $a = 0.0278$

voor $a_r = 0.05$ is $u = 1.645$,

voor $a = 0.001$ is $u = 3.290$.

De linkeroverschrijdingskans is niet apart getabelleerd daar a dezelfde waarde heeft als a_r wanneer we u van teken verwisselen ; bijv.

$a_l = 0.05$ voor $u = -1.645$.

Bij de kritische beoordeling van de waarde van statistische conclusies spelen overschrijdingskansen een belangrijke rol zoals naderhand nog zal blijken.

Vraagstukken

3.5.1. Gegeven : $f(x) = A x e^{-\lambda x}$, $0 < x < \infty$.

Gevraagd :

- a) Wat is het verband tussen A en λ ?
- b) Bereken $E(\underline{x})$,
- c) Bereken $\text{var}(\underline{x})$.

Hulpformule : $\int_0^{\infty} \zeta^n e^{-\zeta} d\zeta = n!$

3.5.2. Wat is $\text{var}(\underline{x}^2)$ wanneer \underline{x}

- a) een standaard rechthoekige verdeling bezit.
- b) een exponentiële verdeling bezit.

3.5.3. Gegeven een driehoekige verdeling :

$$f(x) = x \quad \text{voor } 0 < x < 1,$$

$$f(x) = 1 - x \quad \text{voor } 1 < x < 2,$$

$$f(x) = 0 \quad \text{voor } x < 0 \text{ en } x > 2.$$

Hoe groot zijn $\text{var}(\underline{x})$ en $\sigma(\underline{x})$.

3.5.4. Wanneer \underline{x} een exponentiële verdeling heeft met $\lambda = 0.20$, wat is dan

- a) $\text{Pr}(\underline{x} > 10)$,
- b) $\text{Pr}(\underline{x} < 5)$,
- c) $\text{Pr}(7 < \underline{x} < 15)$.

3.5.5. In een werkplaats heeft men de bewerkingsduur van ontvangen opdrachten nagegaan met de volgende uitkomst.

Bewerkingsduur	Aantal
0 - 1 uur	78
1 - 2 uur	37
2 - 3 uur	22
3 - 4 uur	15
> 4 uur	15

Kunnen deze gegevens door een exponentiële verdeling worden weergegeven
Zo ja, met welke waarde van λ ? Wat is de gemiddelde bewerkingsduur?

3.5.6. Wanneer $\underline{x} = \mu + \underline{u} \sigma$, wat is dan de verdeling van $\underline{y} = (\underline{x} - \mu)^2$?

3.5.7. Teken met behulp van de laatste kolom van Tabel 1 de normale verdelingen voor

$$\mu = 10, \quad \sigma = 1,$$

$$\mu = 10, \quad \sigma = 2,$$

$$\mu = 10, \quad \sigma = 3.$$

3.5.8. Gegeven $\underline{x} = 50 + 7\underline{u}$.

Gevraagd : a) $\Pr(\underline{x} > 60)$,

b) $\Pr(42 < \underline{x} < 63)$,

c) de rechteroverschrijdingskans van $x = 65$,

d) de tweezijdige overschrijdingskans van $x = 68$,

e) voor welke waarden van x is resp.

$$a_r = 0.02,$$

$$a_l = 0.20$$

$$a = 0.01$$

3.6. De normale verdeling (vervolg)

Voorkomen van de normale verdeling

Zeer vele in de praktijk voorkomende verdelingen zijn ten naaste bij normaal zodat zij bij theoretische beschouwingen zonder bezwaar door een normale verdeling kunnen worden weergegeven.

Dit is begrijpelijk op grond van de zg. centrale limietstelling, die zegt, dat wanneer een variabele y tot stand komt als som van n onafhankelijke toevalsvariabelen x_i , dus

$$y = x_1 + x_2 + \dots + x_n \quad (3.6.1)$$

en de variabelen x_1, \dots, x_n allen in gelijke mate tot de variabiliteit van y bijdragen, dus allen ongeveer dezelfde standaardafwijking bezitten, dan zal y ten naaste bij een normale verdeling bezitten, ook al geldt dit niet voor x_1, \dots, x_n . Strikt genomen convergeert de verdeling van y naar één normale wanneer $n \rightarrow \infty$, doch de convergentie is een zeer snelle zodat ook reeds voor kleine waarden van n , y praktisch een normale verdeling bezit.

We kunnen deze stelling hier niet bewijzen maar haar wel aan een eenvoudig voorbeeld demonstreren. Tabel 3.6.1 bevat een frequentietabel van sommen van 4 toevalscijfers. Terwijl deze toevalscijfers zelf ongeveer een rechthoekige verdeling bezitten zien we uit Tabel 3.6.1 dat de som van 4 toevalscijfers reeds duidelijk een nagenoeg normale verdeling vertoont.

TABEL 3.6.1

Verdeling van de som van 4 toevalscijfers.
Tabel 21. Kolommen 16 + 17, 18 + 19.

Interval voor de som	Turfstaatje	Freq. f	Cum.Freq. $\hat{F}(x)$ %	u	Intervalgr. x
1 - 5		1	1	-2.33	5.5
6 - 10		4	5	-1.65	10.5
11 - 15		19	24	-0.71	15.5
16 - 20		34	58	0.20	20.5
21 - 25		29	87	1.13	25.5
26 - 30		12	99	2.33	30.5
31 - 35		1	1		
		100			

Een ander voorbeeld is dat van de verdeling van het aantal schoppen in 13 bridgekaarten van Tabel 2.4.2. Beschouwen we een schoppenkaart als een waarneming = 1 en de andere kaarten als een waarneming = 0, dan is het aantal schoppenkaarten gelijk aan de som van 13 waarnemingen, die met een kans $\frac{1}{4}$ de waarde 1 aannemen en met een kans $\frac{3}{4}$ de waarde 0. De resulterende verdeling van deze som vertoont weer duidelijk de kenmerken van een normale verdeling.

Bij vele productieprocessen treden variaties in de kwaliteit van het product op als resultante van een aantal storingsbronnen. Bij een automatische draaibank bv. zullen variaties in de afmetingen optreden door speling in het support, speling in de kop, variaties in afmeting en hardheid van het uitgangsmateriaal, temperatuurschommelingen, enz. Wanneer een van deze storingsbronnen overheerst, zal men dit vroeg of laat bemerken en maatregelen nemen om deze buiten werking te stellen. Bij een goed functionerend productieproces blijven we dan tenslotte zitten met een aantal storingsbronnen die allen in ongeveer gelijke mate aan de variabiliteit van het product bijdragen en deze geven samen ten naaste bij een normale verdeling. Zeer algemeen neemt men aan, dat de kwaliteits-eigenschappen van een partij producten die allen onder gelijke omstandigheden vervaardigd zijn, een normale verdeling bezitten. De praktijk wijst uit, dat dit in het algemeen een aanvaardbare onderstelling is. Ook de meetproducten die fluctuaties geven tussen herhaalde metingen zijn meestal te wijten aan een samenspel van verschillende oorzaken, en ook hier wordt daarom doorgaans een normale verdeling ondersteld.

Een controle of men met een normale verdeling te maken heeft kan langs grafische weg geschieden. De frequentietabel 3.6.1. levert ons een serie experimentele waarden voor de cumulatieve verdelingsfunctie $\hat{F}(x)$ met de bijbehorende intervalgrenzen x . Bij deze waarden van $\hat{F}(x)$ zoeken we nu in Tabel 1 de bijbehorende waarden van u op. Hebben we met een normale verdeling te doen dan moet gelden

$$x = \mu + u \sigma$$

zodat u en x tegen elkaar uitgezet een rechte lijn moeten geven. In Fig.3.6.1. is dit uitgevoerd. Men gebruikt hiertoe zg. normaal waarschijnlijkheidspapier, waarvan één van de schalen in wezen een lineaire schaal in u is, doch ingedeeld naar waarden van F in %. Het opzoeken van u in Tabel 1 wordt daardoor overbodig; men kan \hat{F} uit Tabel 3.6.1. direct tegen de intervalgrenzen x uitzetten. Door de punten in Fig.3.6.1. kan uit de hand redelijk een rechte lijn worden getrokken. Bij $u = 0$, of $F = 50\%$ lezen we van deze lijn af

$$x = \bar{x} = 18.9 ; \quad (3.6.2)$$

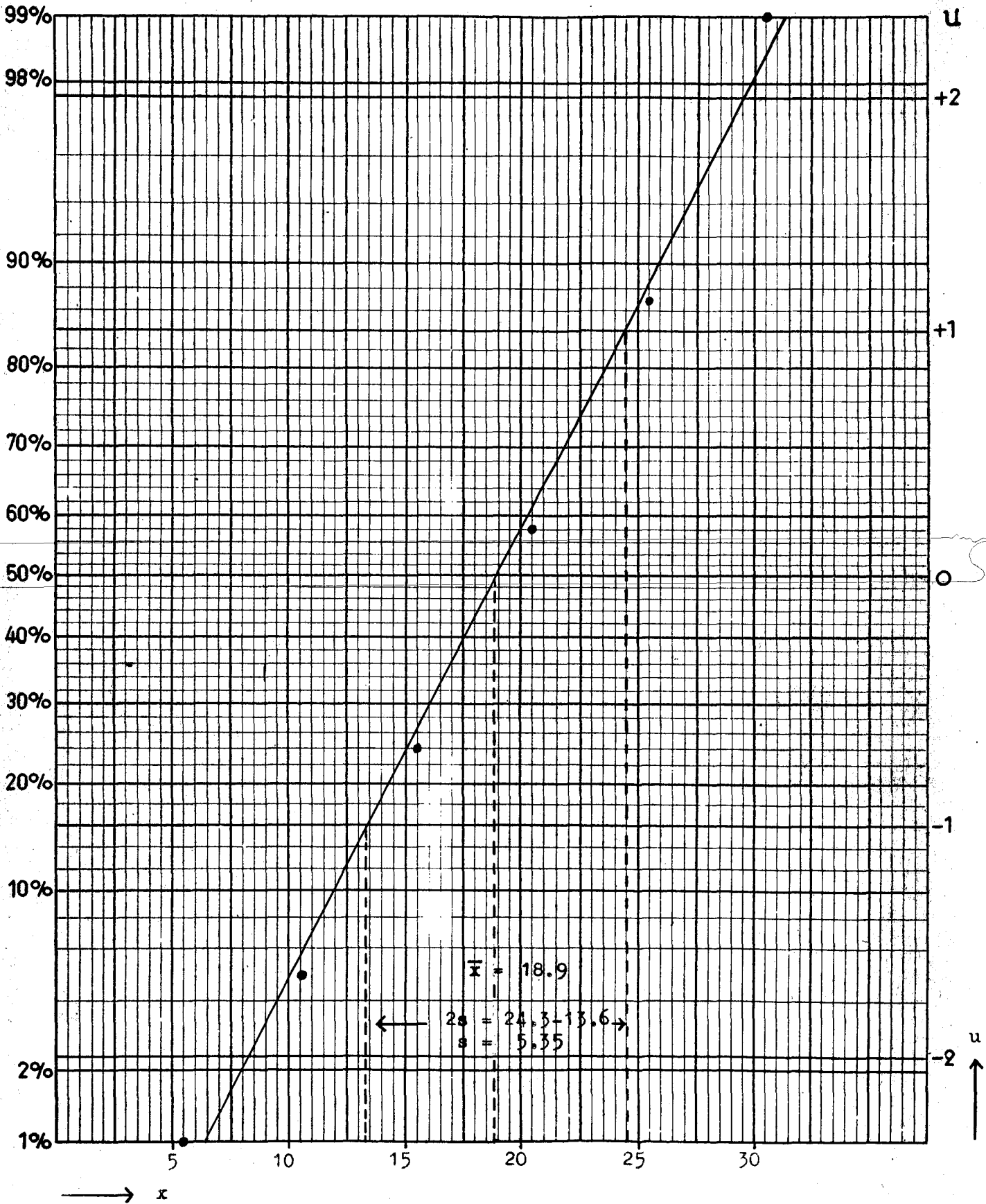
dit is een benaderde waarde voor μ . Bij $u = \pm 1$ of bij $F = 84\%$ en $F = 16\%$ vinden we evenzo $x = 24.3$ en $x = 13.6$. Het verschil tussen deze twee waarden is 2σ de standaardafwijking; dus

$$2s = 24.3 - 13.6 ; \quad s = 5.35. \quad (3.6.3)$$

\bar{x} en s zijn niet exact gelijk aan μ en σ , doch zijn schattingen van μ en σ uit experimentele gegevens gevonden. In het kader van de tot dusver gebruikte symbolen zouden we deze consequent met $\hat{\mu}$ en $\hat{\sigma}$ moeten aangeven; \bar{x} en s zijn echter de algemeen gangbare symbolen. Normaal waarschijnlijkheidspapier wordt in Nederland uitgegeven door de N.V. Drukkerij Mercurius te Wormerveer.

Fig. 3.6.1

De gegevens uit Tabel 3.6.1, uitgezet op normaal waarschijnlijkheidspapier



3.7. De logaritmisch normale verdeling

Definitie :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{0.4343}{x} \exp \left\{ - \frac{(\log x - \mu)^2}{2\sigma^2} \right\}, \quad 0 < x < \infty. \quad (3.7.1)$$

Heeft \underline{x} een log.normale verdeling dan heeft $^{10}\log \underline{x}$ een normale verdeling, De log. normale verdeling voor $\mu = 1.0$, $\sigma = 0.5$ is weergegeven in Fig.3.7.1.

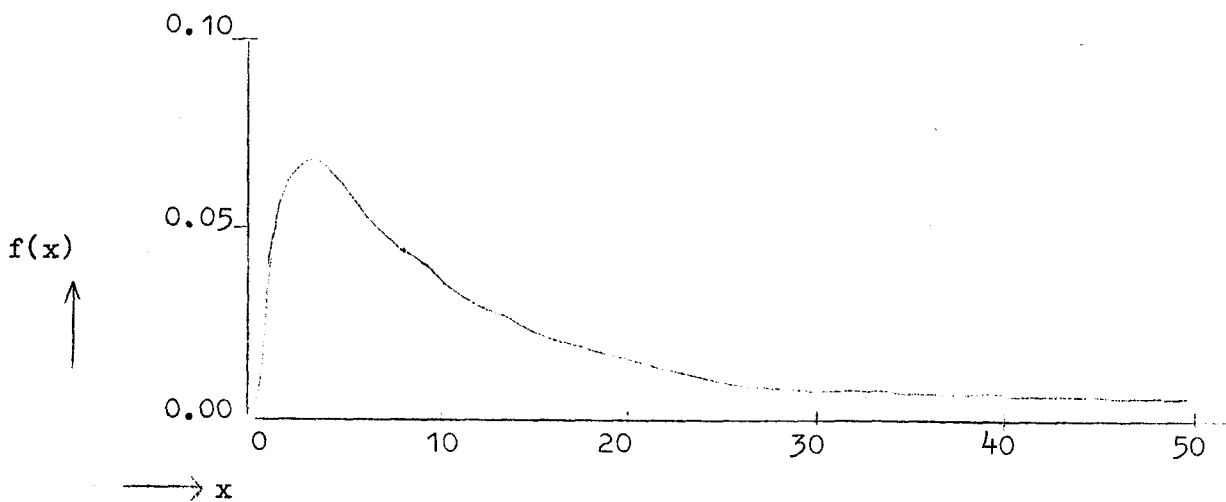


Fig.3.7.1. De log.normale verdeling voor $\mu = 1.0$, $\sigma = 0.5$.

Voor deze waarde van σ is het een sterk asymmetrische verdeling. Waarden van $x = 2.0$ of 3.0 komen het meest frequent voor, maar de verdeling heeft een lange uitloper naar rechts die bij $x = 50$ en hoger nog een duidelijk van nul verschillende waarde bezit.

Op verdere details gaan we hier niet in ; de log.normale verdeling is uitvoerig behandeld in :

J.Aitchison en J.A.C.Brown

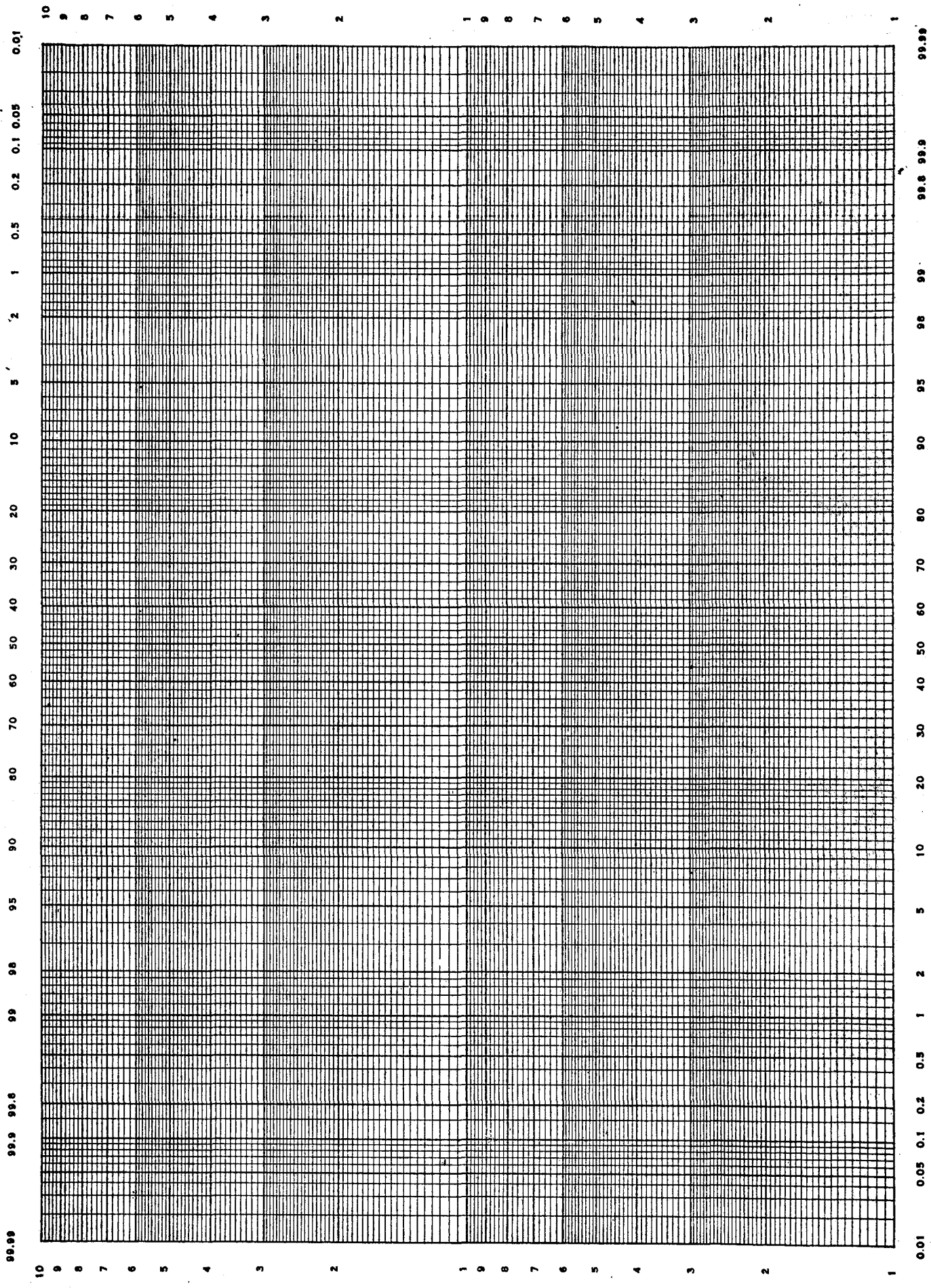
The lognormal distribution.

Cambridge University Press 1957.

De lognormale verdeling speelt onder meer in de economie een rol, Salarissen, en het electriciteitsverbruik van de aangeslotenen op een electriciteitsnet kunnen bevredigend door een lognormale verdeling worden beschreven.

Ook in de techniek komt deze verdeling soms voor ; voorbeelden zijn : de restdruk in een televisie- of radiobuis, lekstromen van gloeidraad naar kathode in een radiobuis, de tegenstroom bij transistoren. Algemeen kan men een verdeling van het log normale type verwachten bij eigenschappen van producten die worden bepaald door een chemische reactie die exponentieel van de temperatuur afhangt. Treden tijdens het productieproces normaal verdeelde temperatuurvariaties op, dan

Fig.3.7.2. Een voorbeeld van logaritmisch waarschijnlijkheidspapier



zullen deze log normaal verdeelde eigenschappen in het eindproduct geven.

Of waarnemingen door een log normale verdeling kunnen worden weergegeven ken eveneens met normaal waarschijnlijkheidspapier worden vastgesteld door \hat{F} tegen $\log x$ uit te zetten. Hiervoor kan ook zg. logaritmisch waarschijnlijkheidspapier worden gebruikt, dat voor x een logaritmische schaal bezit. Zie fig. 3.7.2.

3.8. De betekenis van verdelingen

Algemeen gaat de statistische interpretatie van waarnemingsreeksen uit van de onderstelling dat deze waarnemingen mogen worden opgevat als een serie aselechte trekkingen uit een populatie met een continue kansdichtheidsverdeling.

Meestal weet men dan uit de aard van de waarnemingen ook welke vorm van verdeling men kan verwachten. Zo nodig kan men dit op de boven aangegeven wijze nader onderzoeken. In verreweg de meeste praktijkgevallen neemt men een normale verdeling aan. Aan deze verdeling zullen we dan ook in de volgende lessen voornamelijk aandacht besteden.

Naast de exponentiële, de normale, en de log normale verdeling komen nog andere verdelingen voor, die we echter niet nader zullen bespreken. We noemen slechts het systeem Pearson'sche krommen gebaseerd op de algemene differentiaalvergelijking:

$$\frac{dy}{dx} = \frac{y(x-a)}{b_0 + b_1 x + b_2 x^2}, \quad (3.8.1)$$

waarin $y = f(x)$ en dat al naar de aard van de noemer tot een 7-tal verdelingen van verschillend type leidt. In de jaren 1900-1920 is dit systeem veel toegepast voor de beschrijving van antropologische metingen e.d.

Ook is soms de zg. Weibull verdeling, gegeven door

$$1 - F(x) = \exp \left\{ - \left(\frac{x}{a} \right)^b \right\}, \quad (3.8.2)$$

of

$$f(x) = \frac{b x^{b-1}}{a^b} \exp \left\{ - \left(\frac{x}{a} \right)^b \right\},$$

van nut. Dit is een variant op de exponentiële verdeling die men soms kan toepassen wanneer de exponentiële verdeling zelf niet voldoet. Of men met een Weibull verdeling te doen heeft kan men grafisch onderzoeken door

$$\log \{ - \log (1 - \hat{F}) \} \quad \text{tegen} \quad \log x \quad (3.8.3)$$

uit te zetten. Bij een Weibull verdeling geeft dit een rechte lijn.

3.9. Het praktisch nut van een frequentietabel of histogram

Een eerste stap bij een onderzoek naar de aard en vorm van een verdeling is de constructie van een turfstaatje of frequentietabel als in Tabel 3.1.2. (pag.24). Soms is een dergelijke samenvatting van de gegevens van onmiddellijk nut. Een paar praktijk gevallen mogen dit toelichten.

Voorbeeld 3.9.1. (Ontleend aan B.G.Wiggers, Sigma 1 No.3. 1955, p.50).

De A.K.U. levert garen af gewikkeld op metalen spoelen de zg. copshulsen. Van 100 exemplaren uit een ontvangen partij copshulzen werd het gewicht bepaald en in een frequentietabel vastgelegd. Het resultaat is in fig. 3.9.1. weergegeven in de vorm van een zg. histogram ; de intervallen zijn op de horizontale as uitgezet en op ieder interval als basis is een verticale kolom gezet waarvan het oppervlak evenredig is met de frequentie f . Dit geeft een illustratieve figuur ; in rapporten en publicaties verdient het aanbeveling een frequentietabel in deze vorm

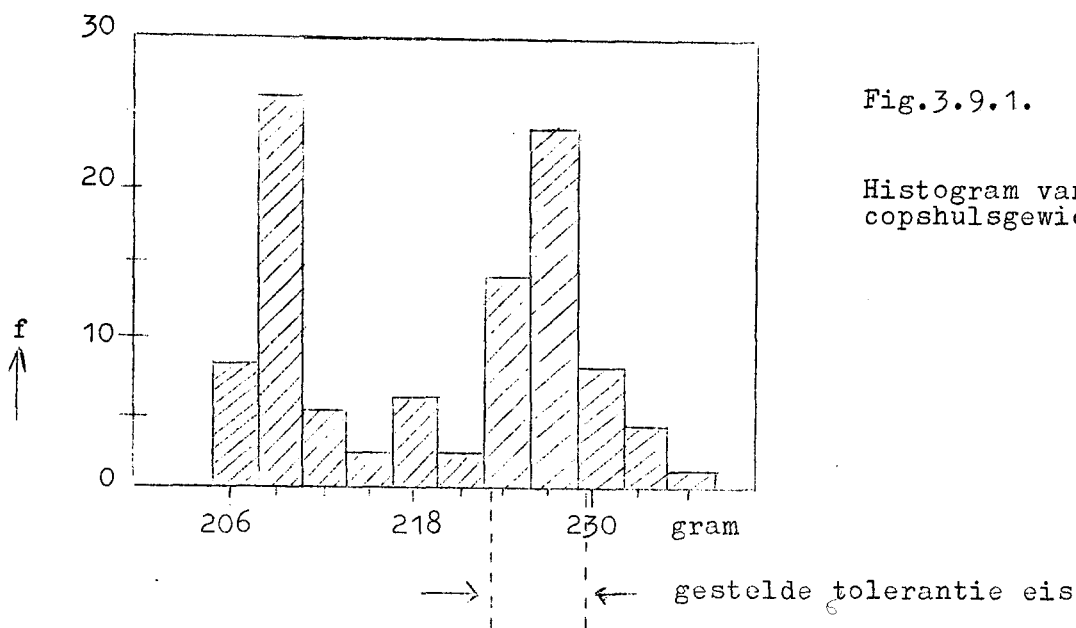


Fig.3.9.1.

Histogram van 100 copshulsgewichten

weer te geven. In figuur 3.9.1. is tevens de tolerantieëis aangegeven die leverancier en afnemer na onderling overleg waren overeengekomen.

Conclusies : Het histogram is tweetoppig ; vermoedelijk zijn de producten van twee machines, die niet gelijk zijn afgesteld, gemengd. Door betere controle op de machine afstelling kan men een homogener product maken.

Ook dan echter is de spreiding nog te groot om aan de gestelde tolerantie eis

$$223 \leq \text{gewicht} \leq 229 \text{ gram}$$

te voldoen. Men moet of het productieproces verbeteren of de tolerantie verwijderen. Dit laatste bleek mogelijk.

Voorbeeld 3.9.2. (Philips)

Een staalstaaf van 6 mm diameter geleverd door een vaste leverancier voldeed niet aan de verlangde toleranties, hetgeen een nabewerking noodzakelijk maakte. Een onderzoek in de fabriek werd ingesteld ; daarbij werd de diameter gemeten van 100 staven uit een voor verzending gereed liggende partij, en van 100 staven die direct van de trekbank werden genomen. Fig.3.9.2. geeft de beide histogrammen.

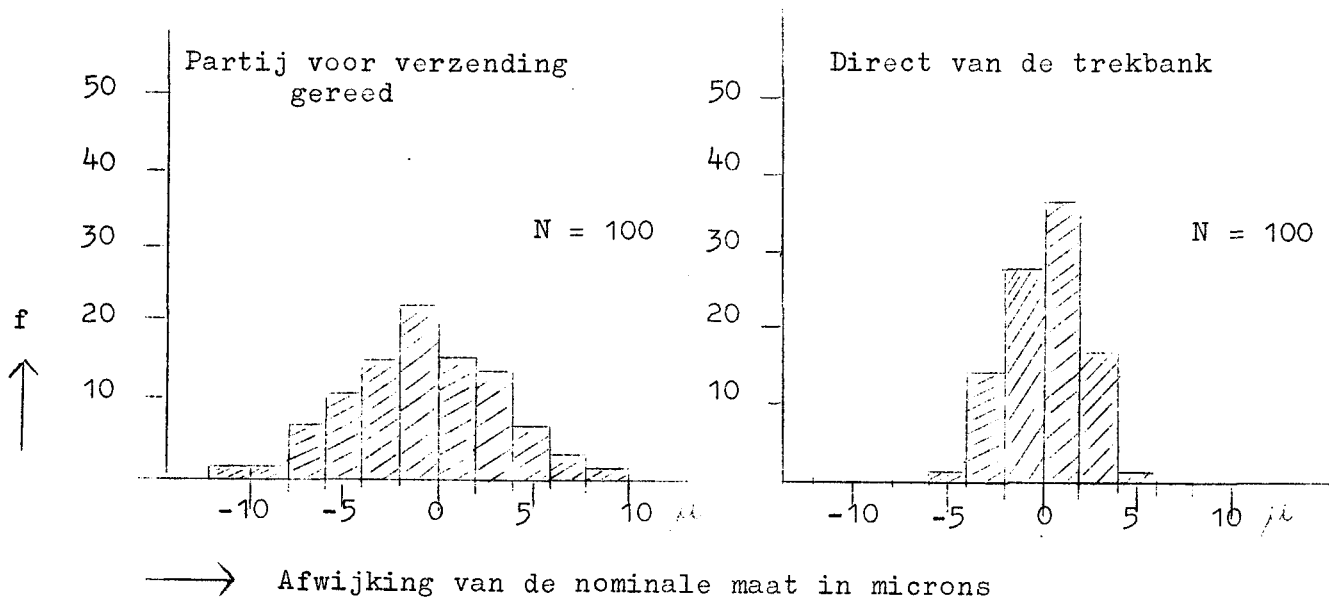


Fig.3.9.2. Histogrammen van 100 diameter bepalingen aan 6 mm staalstaaf uit een voor verzending gereed liggende partij en aan staven direct van één trekbank genomen.

Conclusie : De spreiding tussen direct na elkaar op één machine geproduceerde staven is aanzienlijk kleiner dan de spreiding in de verzendpartij. Vermoedelijk zijn in deze laatste producten van verschillende machines, of op verschillende dagen geproduceerd, doorengemengd. Met weinig extra zorg kan een belangrijke verbetering worden bereikt. De verschillen tussen machines, of tussen dagen, zijn hier niet zó groot, dat een uitgesproken tweetoppigheid optreedt als in Fig.1.7.1. Zij manifesteren zich door een verbreding van de verdeling.

Voorbeeld III. (Philips).

De buigvastheid werd gemeten aan monsters genomen van 232 klossen fosforbrons van 0.8 mm dikte. Het materiaal wordt om een scherpe rechtehoekige rand gebogen en daarbij wordt waargenomen bij welke buighoek barsten optreden. Fig.3.9.3. geeft het histogram.

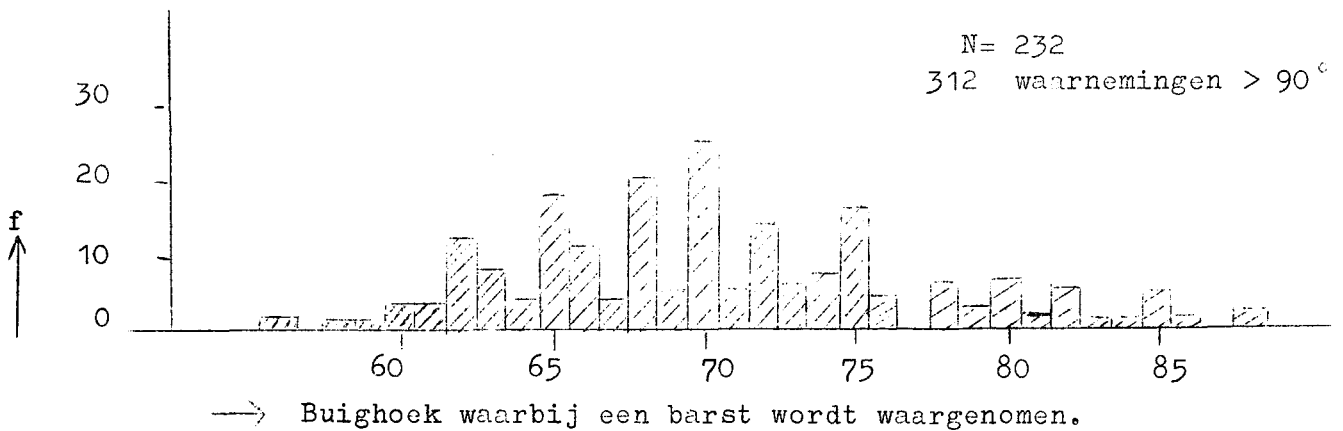


Fig.3.9.3. Histogram van buigvastheidsbepalingen aan een partij fosforbronsband van 0.8 mm dikte.

Conclusie : De waarnemer heeft een uitgesproken voorkeur voor even uitkomsten en/of uitkomsten die op een 5 eindigen. De waarneming is onnauwkeurig. Het verdient aanbeveling de waarnemingen op 5° af te ronden en de meetopstelling hiervoor in te richten. Dit vereenvoudigt het waarnemingsmateriaal.

Bij de constructie van frequentietabellen en histogrammen dient men op de volgende punten te letten.

- 1) Kies de interval- of klasse-indeling zó, dat iedere waarneming ondubbelzinnig in één bepaalde interval thuis hoort, en niet op de grens tussen twee klassen kan vallen.
- 2) Kies de klassebreedte zo dat het gehele gebied van x-waarden in 10 tot 15 klassen wordt onderverdeeld.
- 3) Kies bij voorkeur klassen van gelijke breedte. De hoogte van een kolom in een histogram als in fig.3.9.1., 3.9.2., en 3.9.3. kan dan eenvoudig gelijk aan de frequentie worden genomen. Soms is men echter genoodzaakt klassen van ongelijke breedte te gebruiken ; dan dient men erop te letten dat het oppervlak aan de kolom evenredig met de frequentie moet zijn.
- 4) Moet men histogrammen vergelijken die gebaseerd zijn op ongelijke aantallen waarnemingen dan verdient het de voorkeur in de histogrammen de relatieve frequenties, d.w.z. de frequenties, $\frac{f}{N}$ in % uitgedrukt weer te geven. Op het histogram dient men dan ook steeds het totaal aantal waarnemingen N te vermelden.

3.10. Vraagstukken.

3.10.1. B

Diameters van 60 glazen knikkers in mm.

12.05	12.29	12.53	12.57	12.72	12.06
12.71	12.04	12.50	12.20	12.26	12.16
12.25	12.30	12.20	11.77	12.07	12.18
12.40	12.29	12.65	12.35	12.41	11.94
13.15	12.18	12.53	12.54	11.62	12.43
12.94	12.49	12.00	12.36	12.41	12.30
12.00	12.55	12.20	12.69	12.71	12.39
12.40	12.07	12.35	12.40	12.61	12.56
12.49	12.31	12.32	12.52	12.10	12.67
12.33	11.61	12.89	12.38	12.42	12.28

- a) Geef bovenstaande metingen weer in een frequentietabel.
- b) Teken een net histogram dat als deel van een rapport kan worden gebruikt.
- c) Zet de gegevens uit op waarschijnlijkheidspapier en bepaal grafisch het gemiddelde en de standaardafwijking.

HOOFDSTUK IV. TWEE- EN MEERDIMENSIONALE VERDELINGEN

4.1. Twee-dimensionale verdelingen

De voorgaande beschouwingen kunnen we naar twee en meer dimensies uitbreiden. We beginnen weer met een praktisch voorbeeld ; aan 60 hardpapieren schijfjes zijn dikte en gewicht gemeten, met het in Tabel 4.1.1 vermelde resultaat.

TABEL 4.1.1

Dikte en gewicht gemeten aan 60 hardpapieren schijfjes
 x = dikte in mm. y = gewicht in mgr.

x	y	x	y	x	y	x	y	x	y	x	y
2.12	902	2.09	777	2.04	755	2.19	787	2.18	783	2.18	811
2.16	796	2.05	774	2.03	770	2.10	787	2.03	772	2.02	767
2.20	797	2.13	806	2.05	773	2.05	763	2.21	812	2.16	802
2.14	796	2.04	755	2.13	802	2.04	782	2.14	776	2.14	813
2.06	768	2.17	796	2.20	795	2.17	802	2.09	782	2.08	782
2.10	799	2.07	788	2.15	803	2.08	781	2.14	803	2.20	800
2.10	773	2.20	805	2.17	775	2.07	780	2.04	771	2.16	804
2.17	791	2.09	768	2.21	806	2.10	765	2.13	799	2.15	799
2.17	809	2.11	784	2.04	764	2.21	801	2.19	807	2.04	766
2.24	796	2.04	784	2.07	785	2.06	780	2.20	789	2.04	772

Een eerste methode om enig inzicht te krijgen in de aard van dergelijk materiaal is door dikte en gewicht tegen elkaar uit te zetten. Dit is in figuur 4.1.1. geschied voor de eerste paar waarnemingen

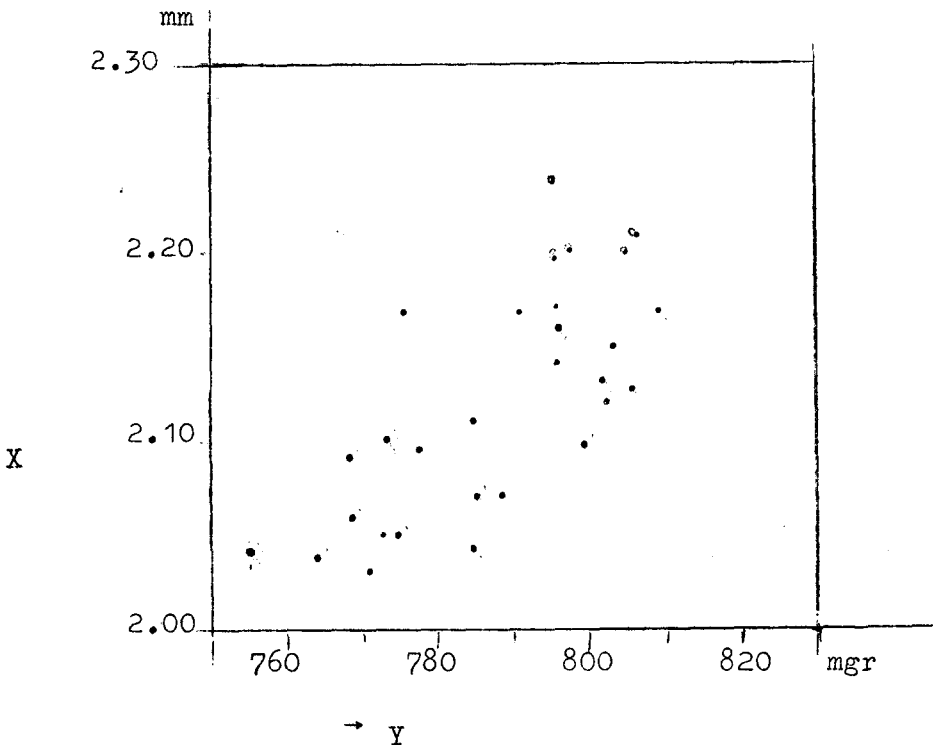


Fig.4.1.1. Dikte en gewicht tegen elkaar uitgezet voor de eerste 30 paar waarnemingen uit Tabel 4.1.1 ; X en Y zijn afhankelijk

Dikte en gewicht zijn beide toevalsvariabelen. Het gewicht wordt niet volledig door de dikte bepaald, maar we zien wel dat het gewicht in doorsnee toeneemt met de dikte. De verdeling van het gewicht hangt af van de dikte, of de verdeling van de dikte hangt af van het gewicht. We hebben hier een voorbeeld van zgn. stochastische afhankelijkheid. Een voorbeeld van onafhankelijkheid levert Fig. 4.1.2. Hier zijn de 60 gewichtsmetingen uit Tabel 4.1. in 30 paren gesplitst en is de tweede waarneming van ieder paar uitgezet tegen de eerste. Er is geen samenhang te verwachten tussen gewichten van verschillende willekeurig bijeengenomen schijfjes en de gelijkmatige spreiding van de punten over de grafiek is hiermee in overeenstemming

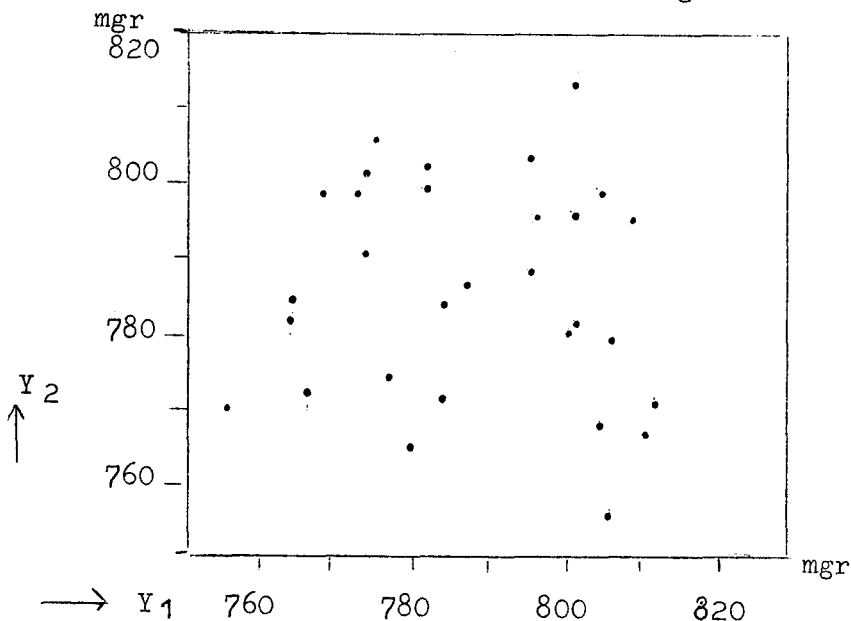


Fig.4.1.2. Twee opeenvolgende gewichtswaarnemingen uit Tabel 4.4.1 tegen elkaar uitgezet ; Y_1 en Y_2 zijn onafhankelijk.

Een andere manier om dergelijk waarnemingsmateriaal weer te geven, vooral wanneer het om grote aantallen waarnemingen gaat is in een twee-dimensionale frequentie tabel als tabel 4.1.2. (zie volgende bladzijde)

Zowel dikte als gewicht zijn in een aantal klassen verdeeld en in ieder vakje van de tabel is de frequentie n_{ij} vermeld van schijfjes die in de bijbehorende dikte- en gewichtsklasse vallen.

Sommeren we in horizontale richting dan vinden we de z.g. marginale frequentie $n_{i.}$ voor de dikten, d.w.z. de frequenties in de dikteklassen ongeacht het gewicht. Sommeren in verticale richting geeft evenzo de marginale frequenties $n_{.j}$ voor de gewichten. De totale frequentie is $n_{..} = n = 200$.

TABEL 4.1.2.

Een twee-dimensionale frequentietabel voor dikte en gewicht van 200 hardpapieren schijfjes.

i ↓	Gewichtsklassen in mgr.							
	750- 759	760- 769	770- 779	780- 789	790- 799	800- 809	810 819	
Dikte klassen in mm	frequenties n_{ij}							$n_{i.} =$
2.01 - 2.03		2	4					6
2.04 - 2.06	4	13	13	8				39
2.07 - 2.09	1	2	5	14	4			26
2.10 - 2.12	1	4	4	6	7	3		25
2.13 - 2.15		1	3	3	15	16	3	41
2.16 - 2.18			2	3	11	23	7	46
2.19 - 2.21				2	2	10	2	17
2.22 - 2.24					2			2
$n_{.j} =$	7	22	31	36	41	52	13	200= $n_{..}$ = n

De klasseindeling in Tabel 4.1.2. is natuurlijk kunstmatig ; in wezen zijn dikte en gewicht continue variabelen en men neemt daarom aan dat het proces dat de schijfjes heeft voortgebracht in principe kan worden beschreven door een twee-dimensionale kansdichtheid of verdelingsfunctie.

$$\Pr(x < \underline{x} < x + dx, y < \underline{y} < y + dy) = f(x,y) dx dy. \quad (4.1.1.)$$

De somregel leidt dan tot

$$\Pr(x_1 < \underline{x} < x, y_1 < \underline{y} < y) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(\xi, \eta) d\xi d\eta, \quad (4.1.2.)$$

terwijl natuurlijk moet gelden

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) d\xi d\eta = 1. \quad (4.1.3)$$

De marginale verdeling van \underline{x} , d.w.z. de verdeling van \underline{x} , d.w.z. de verdeling van \underline{x} ongeacht de waarde van y wordt

$$f_{\underline{x}}(x) = \int_{-\infty}^{+\infty} f(x, \eta) d\eta \quad (4.1.4a)$$

en evenzo de marginale verdeling van \underline{y}

$$f_{\underline{y}}(y) = \int_{-\infty}^{+\infty} f(\xi, y) d\xi. \quad (4.1.4b)$$

De indices x en y aan deze functie geven aan dat $f_{\underline{x}}(x)$ en $f_{\underline{y}}(y)$ geenszins dezelfde functies behoeven te zijn.

Verder onderscheiden we de voorwaardelijke verdeling van \underline{x} als y gegeven is :

$$f(x|y) = f(x, y) / \int_{-\infty}^{+\infty} f(\xi, y) d\xi \quad (4.1.5a)$$

en de voorwaardelijke verdeling van \underline{y}

$$f(y|x) = f(x, y) / \int_{-\infty}^{+\infty} f(x, \eta) d\eta. \quad (4.1.5b)$$

Dit zijn als het ware kansdichtheden die corresponderen met één bepaalde rij of één bepaalde kolom in Tabel 4.1.2.

Combinatie van (4.1.4) en (4.1.5) geeft verder

$$f(x, y) = f_{\underline{x}}(x) \times f(y|x) = f_{\underline{y}}(y) \times f(x|y). \quad (4.1.6)$$

De variabelen \underline{x} en \underline{y} heten onderling onafhankelijk indien geldt

$$\boxed{f(x, y) = f_{\underline{x}}(x) \times f_{\underline{y}}(y)} \quad (4.1.7)$$

dus indien de kansdichtheid van x en y gelijk is aan het product van de beide marginale kansdichtheden.

Uit (4.1.7) volgt dat bij onafhankelijkheid

$$f(x|y) = f_{\underline{x}}(x) \quad \text{en} \quad f(y|x) = f_{\underline{y}}(y). \quad (4.1.8)$$

In woorden : de voorwaardelijke verdeling van \underline{x} is onafhankelijk van y , en die van \underline{y} onafhankelijk van x . De verdeling in Tabel 4.1.2. is dan in iedere kolom hetzelfde. Het E-symbool bij een twee-dimensionale verdeling betekent de operator

$$E\{g(\underline{x}, \underline{y})\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(\xi, \eta) f(\xi, \eta) d\xi d\eta. \quad (4.1.9)$$

We noemen $E\{g(\underline{x}, \underline{y})\}$ het populatiegemiddelde van de toevalsvariabele $g(\underline{x}, \underline{y})$.

We hebben bv.

$$\underline{x} = E(\underline{x}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \xi f(\xi, \eta) d\xi d\eta = \int_{-\infty}^{+\infty} \xi f_{\underline{x}}(\xi) d\xi. \quad (4.1.10)$$

voor het populatiegemiddelde van \underline{x} , en evenzo voor \underline{y} .

Verder

$$\sigma^2(\underline{x}) = \text{var}(\underline{x}) = E\{\underline{x} - E(\underline{x})\}^2 = E(\underline{x}^2) - \{E(\underline{x})\}^2. \quad (4.1.11)$$

Formule blijft dus ook bij een twee-dimensionale populatie geldig.

Naast de variantie, een maat voor de spreiding van \underline{x} (of van \underline{y}) onderscheiden we thans ook de covariantie gedefinieerd door

$$\text{cov}(\underline{x}, \underline{y}) = E\{[\underline{x} - E(\underline{x})] [\underline{y} - E(\underline{y})]\} = E(\underline{x} \underline{y}) - E(\underline{x}) E(\underline{y}) \quad (4.1.12)$$

De covariantie is een maat voor de graad waarin \underline{x} en \underline{y} samengaan. Neemt in het algemeen \underline{y} toe met \underline{x} als in Fig. 4.1.1. dan is $\text{cov}(\underline{x}, \underline{y})$ positief; neemt \underline{y} af met toenemende \underline{x} dan is $\text{cov}(\underline{x}, \underline{y})$ negatief. Zijn \underline{x} en \underline{y} onderling onafhankelijk dan geldt

$$E(\underline{x} \underline{y}) = E(\underline{x}) E(\underline{y}) \quad (4.1.13)$$

en is

$$\text{cov}(\underline{x}, \underline{y}) = 0. \quad (4.1.14)$$

Het omgekeerde is niet waar; wanneer aan (4.1.14) is voldaan zijn \underline{x} en \underline{y} niet noodzakelijk onafhankelijk. Is niet aan (4.1.14) voldaan, dan zijn \underline{x} en \underline{y} in ieder geval afhankelijk.

Algemeen zal de covariantie toenemen wanneer de spreiding in \underline{x} en \underline{y} toenemen.

Stellen we bv.

$$\underline{x}_1 = a\underline{x} \quad \text{en} \quad \underline{y}_1 = b\underline{y}$$

dan zal gelden

$$\text{var}(\underline{x}_1) = a^2 \text{var}(\underline{x}), \quad \text{var}(\underline{y}_1) = b^2 \text{var}(\underline{y})$$

en
$$\text{cov}(\underline{x}_1, \underline{y}_1) = ab \text{cov}(\underline{x}, \underline{y}).$$

In plaats van de covariantie beschouwt men daarom vaak de zg. correlatie coëfficiënt ρ gedefinieerd door

$$\rho = \frac{\text{cov}(\underline{x}, \underline{y})}{\sqrt{\text{var}(\underline{x}) \text{var}(\underline{y})}} \quad (4.1.15)$$

Dit is een dimensieloze grootte die een maat geeft voor de samenhang tussen \underline{x} en \underline{y} onafhankelijk van de schaal waarin \underline{x} en \underline{y} worden uitgedrukt. Naast gemiddelden en varianties is de correlatie coëfficiënt een belangrijke parameter van een twee-dimensionale kansdichtheidsverdeling.

Stel

$$\underline{x}' = \underline{x} - E(\underline{x}), \quad \underline{y}' = \underline{y} - E(\underline{y}).$$

We beschouwen nu de kwadratische vergelijking in t .

$$E \{t\underline{x}' + \underline{y}'\}^2 = t^2 \text{var}(\underline{x}) + 2t \text{cov}(\underline{x}, \underline{y}) + \text{var}(\underline{y}) = 0. \quad (4.1.15)$$

De linkerzijde is kennelijk een positief definitie kwadratische functie van t , zij kan nooit negatief worden en de vergelijking heeft dus of geen of twee samenvallende reële wortels. Dan moet de discriminant kleiner of hoogstens gelijk aan nul zijn ; d.w.z.

$$\{\text{cov}(\underline{x}, \underline{y})\}^2 - \text{var}(\underline{x}) \text{var}(\underline{y}) \leq 0, \quad (4.1.17)$$

hetgeen equivalent is aan

$$\rho^2 \leq 1, \text{ of } -1 \leq \rho \leq +1. \quad (4.1.18)$$

De correlatie coëfficiënt is dus in absolute waarde hoogstens gelijk aan 1.

De waarden $\rho = +1$ of $\rho = -1$ worden alleen bereikt wanneer de discriminant van (4.1.15) nul is. Dan heeft deze vergelijking twee reële samenvallende wortels. Dit kan alleen indien

$$\frac{\underline{y}'}{\underline{x}'} = a = \text{constant}$$

of als

$$\underline{y} - E(\underline{y}) = a\{\underline{x} - E(\underline{x})\}. \quad (4.1.19)$$

Wanneer $\rho = +1$ of $\rho = -1$ is \underline{y} een lineaire functie van \underline{x} ; \underline{y} wordt dan door \underline{x} volledig bepaald.

4.2. Meer-dimensionale verdelingen

We kunnen aan hard papieren schijfjes naast dikte en gewicht nog meer grootheden meten, bv. volume, vochtgehalte, chemische componenten etc. Zij zullen allen fluctuaties vertonen en we kunnen ze als toevalsvariabelen opvatten. We hebben dan bv. k variabelen.

$$\underline{x}_1, \underline{x}_2, \dots, \underline{x}_k$$

met een gezamenlijke kansdichtheidsfunctie

$$\Pr(x_i < \underline{x}_i < x_i + dx_i, i = 1 \dots k) = f(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k. \quad (4.2.1)$$

Voor ieder dezer variabele kunnen we weer een marginale verdeling definiëren, bv.

$$f_{\underline{x}_1}(\underline{x}_1) = \int_{-\infty}^{+\infty} d\xi_2 \int_{-\infty}^{+\infty} d\xi_3 \dots \int_{-\infty}^{+\infty} d\xi_k f(\xi_1, \xi_2, \dots, \xi_k) \quad (4.2.2)$$

Het E-symbool op populatiegemiddelde is nu

$$E\{g(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_k)\} = \int_{-\infty}^{+\infty} d\xi_1 \int_{-\infty}^{+\infty} d\xi_2 \dots \int_{-\infty}^{+\infty} d\xi_k g(\xi_1, \xi_2, \dots, \xi_k) f(\xi_1, \xi_2, \dots, \xi_k) \quad (4.2.3)$$

Voor iedere variabele hebben we een gemiddelde $E(\underline{x}_i)$ en een variantie $\text{var}(\underline{x}_i)$, en voor ieder paar een covariantie $\text{cov}(\underline{x}_i, \underline{x}_j)$.

De variabelen $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_k$ heten nu weer onderling onafhankelijk, indien

$$f(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_k) = f_{\underline{x}_1}(\underline{x}_1) f_{\underline{x}_2}(\underline{x}_2) \times \dots \times f_{\underline{x}_k}(\underline{x}_k). \quad (4.2.4)$$

Zijn $\underline{x}_1, \dots, \underline{x}_k$ onderling onafhankelijk dan is

$$\text{cov}(\underline{x}_i, \underline{x}_j) = 0, \quad i \neq j. \quad (4.2.5)$$

Van bijzonder praktisch belang zijn de volgende algemene betrekkingen.

$$\text{Is } \underline{y} = \sum_i a_i \underline{x}_i, \quad i = 1, \dots, k \quad (4.2.6)$$

dan is

$$E(\underline{y}) = \sum_i a_i E(\underline{x}_i), \quad (4.2.7)$$

en

$$\text{var}(\underline{y}) = \sum_i a_i^2 \text{var}(\underline{x}_i) + 2 \sum_{i < j} a_i a_j \text{cov}(\underline{x}_i, \underline{x}_j), \quad (4.2.8)$$

of, wanneer de \underline{x}_i 's allen onderling onafhankelijk zijn

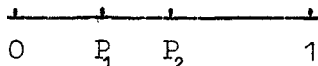
$$\boxed{\text{var}(\underline{y}) = \sum_i a_i^2 \text{var}(\underline{x}_i).} \quad (4.2.9)$$

Wegens deze fundamentele formule, die bij tal van problemen toepassing vindt, speelt in de statistische theorie de variantie zo'n belangrijke rol. Bij praktische toepassingen heeft de variantie echter het nadeel dat haar dimensie het kwadraat is van de dimensie van de uitgevoerde metingen ; bij wegingen in mgr wordt de variantie uitgedrukt in mgr ; bij de experimentator spreekt dat slecht aan. Bij praktische toepassingen verdienen daarom standaardafwijkingen de voorkeur. Voor deze geldt

$$\sigma(\underline{y}) = \sqrt{\sum_i a_i^2 \sigma^2(\underline{x}_i)}. \quad (4.2.10)$$

VRAAGSTUKKEN

4.3.1. Op een lijnstuk met de lengte 1 worden volkomen aselekt 2 punten, P_1 en P_2 , aangewezen (zie fig.)



Stel $OP_1 = u$, $P_1P_2 = v$.

De gezamenlijke verdeling van \underline{u} en \underline{v} is

$$f(u,v) = 2 \quad \text{voor } 0 \leq u \leq 1, \quad v_0 \leq v \leq 1 \\ \text{en } 0 \leq u+v \leq 1.$$

Vragen :

- a) Wat is de marginale verdeling van u ?
- b) Hoe groot is $E(\underline{u})$, $E(\underline{u}^2)$ en $\text{var}(\underline{u})$?
- c) Wat is de voorwaardelijke verdeling van u als v een gegeven waarde bezit ?
- d) Hoe groot zijn $E(\underline{u}, \underline{v})$, $\text{cov}(\underline{u}, \underline{v})$ en $\rho(\underline{u}, \underline{v})$?

4.3.2. \underline{x}_1 en \underline{x}_2 zijn twee onderling onafhankelijke variabelen met gemiddelden μ_1, μ_2 en varianties σ_1^2 en σ_2^2 .

Wat is de waarde van

- a) $E\{(\underline{x}_1 - \mu_1)^2(\underline{x}_2 - \mu_2)\}$,
- b) $E\{(\underline{x}_1 - \mu_1)^2(\underline{x}_2 - \mu_2)^2\}$,
- c) $E(\underline{x}_1^2 \underline{x}_2)$,
- d) $E(\underline{x}_1^2 \underline{x}_2^2)$.

4.3.3. \underline{x}_1 en \underline{x}_2 zijn twee onderling onafhankelijke variabelen. Bewijs dat

$$\text{var}(\underline{x}_1 \underline{x}_2) = \mu_1^2 \text{var} \underline{x}_2 + \mu_2^2 \text{var} \underline{x}_1 + \text{var} \underline{x}_1 \text{var} \underline{x}_2$$

Gegeven $\underline{x}_i \quad i=1, \dots, n$ zijn n toevalsvariabelen met eenzelfde standaardafwijking σ .

Gevraagd de covariantie en de correlatiecoëfficiënt van \underline{y} en \underline{z} wanneer.

- a) $\underline{y} = \underline{x}_1 + \underline{x}_2$, $\underline{z} = \underline{x}_3 + \underline{x}_4$,
- b) $\underline{y} = \underline{x}_1 + \underline{x}_2$, $\underline{z} = \underline{x}_2 + \underline{x}_3$,
- c) $\underline{y} = \underline{x}_1 + \underline{x}_2 + \underline{x}_3$, $\underline{z} = \underline{x}_2 + \underline{x}_3 + \underline{x}_4$,
- d) $\underline{y} = \underline{x}_1$; $\underline{z} = \underline{x} = \frac{\sum \underline{x}_i}{n}$.

HOOFDSTUK V. WAARNEMINGSREEKSEN

5.1. Enkelvoudige waarnemingsreeksen

5.1.1. Onder een enkelvoudige waarnemingsreeks zullen we verstaan een reeks waarnemingsuitkomsten verkregen door herhaling van een waarneming. Bij de statistische interpretatie vatten we een enkelvoudige waarnemingsreeks steeds op als een serie onderling onafhankelijke trekkingen, of als aselechte steekproef, uit eenzelfde populatie. Zouden we de frequentie verdeling van die populatie kennen dan zou daarmee het experiment, statistisch gezien, volledig zijn beschreven. Doch doorgaans is deze verdeling niet bekend. Het doel van een reeks waarnemingen is dan de kenmerken van deze onbekende verdeling zo goed mogelijk te bepalen.

Doorgaans kunnen wel op grond van onze algemene ervaringen redelijke onderstellingen worden gemaakt aangaande de algemene vorm van de verdeling. In de meeste gevallen zal men redelijkerwijs mogen onderstellen dat dit een normale verdeling is; in andere gevallen zal men een exponentiële of een log normale verdeling aannemen. Deze verdelingen zijn door een of twee parameters volledig bepaald. De statistiek ziet daarom als zijn doel uit een waarnemingsreeks (of aselechte steekproef) de onbekende populatie parameters zo goed mogelijk te schatten en de nauwkeurigheid van de gevonden schattingen aan te geven.

Voorlopig zullen we ons beperken tot het meest voorkomende geval, waarbij men onderstelt dat de waarnemingen door een normale verdeling voldoende kunnen worden beschreven. We hebben dan met twee parameters te maken

$$\begin{aligned} \text{het populatie gemiddelde} &= \mu, \\ \text{de populatie standaardafwijking} &= \sigma. \end{aligned}$$

5.2. Numerieke bewerking van een waarnemingsreeks

Hiermee corresponderen

$$\begin{aligned} \text{een gemiddelde} &= \bar{x}, \\ \text{en een standaardafwijking} &= s, \end{aligned}$$

uit een steekproef bepaald. Daarnaast zijn echter nog een aantal andere grootheden van belang die in Tabel 5.2.1. zijn samengevat.

Tabel 5.2.1.

Kenmerkende grootheden voor een enkelvoudige waarnemingsreeks.

Benaming	Symbool en definitie	Toelaatbaar afrondingsinterval
Aantal waarnemingen	n	
Waarnemingen	$X_i, i = 1, \dots, n$	$\frac{R}{2\sqrt{n}}$ of $\frac{s}{2}$
Som of totaal	$T = \sum X_i$	
Gemiddelde	$\bar{X} = \sum X_i / n$	$\frac{R}{2n}$ of $\frac{s}{2\sqrt{n}}$
Kwadratensom	$KS = \sum (X_i - \bar{X})^2$	
Aantal vrijheidsgraden	$v = (n-1)$	
Variantie	$s^2 = \frac{KS}{v} = \frac{\sum (X_i - \bar{X})^2}{n-1}$	$\frac{s^2}{\sqrt{2v}}$
Standaardafwijking	$s = \sqrt{s^2}$	$\frac{s}{\sqrt{8v}}$
Standaardafwijking van het gemiddelde	$s(\bar{x}) = \frac{s}{\sqrt{n}}$	$\frac{s(\bar{x})}{\sqrt{8v}}$
Spreadingsbreedte	$R = X_{\max} - X_{\min}$	

Toelichtingen.

Gegeven een serie van 8 diktemetingen aan hardpapieren schijfjes:

$X_i = 2.08 \quad 2.14 \quad 2.13 \quad 2.04 \quad 2.17 \quad 2.08 \quad 2.06 \quad 2.08 \text{ mm.} \quad (5.2.1.)$

Afronding.

De spreiding in de dikte is vrij groot. De spreidingsbreedte bedraagt

$$R = 2.17 - 2.04 = 0.13 \text{ mm.}$$

Een directe praktische vraag is of men waarnemingen mag vereenvoudigen door ze af te ronden en hoever men hiermee mag gaan. Het is bijvoorbeeld duidelijk, dat gezien de grote spreiding meting van de dikte tot op microns weinig zin zou hebben.

Het toelaatbare afrondingsinterval bedraagt

$$\frac{R}{2\sqrt{n}} = \frac{0.13}{2\sqrt{8}} = 0.023 \text{ mm.}$$

De waarnemingen zijn op 0.01 mm. juist afgerond. Ze mogen desnoods op 0.02 mm. worden afgerond doch dit heeft weinig zin.

Een juiste afronding is belangrijk omdat men daardoor de numerieke berekeningen zeer vereenvoudigt.

Is de standaardafwijking s reeds bekend, dan kan men deze in plaats van R gebruiken om het toelaatbare afrondingsinterval te bepalen.

Berekening van de kwadratensom.

We hebben twee formules.

$$KS = \sum (X_i - \bar{X})^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n} \quad (5.2.2.)$$

$$KS = \sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2 \quad (5.2.3.)$$

(5.2.3.) is van belang bij theoretische beschouwingen; doch is ongeschikt voor numerieke berekeningen omdat kleine afrondingsfouten in \bar{X} soms relatief grote fouten in de KS kunnen geven. Voor de numerieke berekening passen we daarom bij voorkeur (5.2.2.) toe.

Met een rekenmachine kunnen we de sommen $\sum X_i$ en $\sum X_i^2$ in één bewerking worden berekend.

De kwadratensom verandert niet wanneer we van alle waarnemingen een constant bedrag aftrekken; dit vereenvoudigt de berekening echter aanmerkelijk zodat zij uit het hoofd kan worden uitgevoerd. Het verminderen met een constant bedrag noemt men het coderen van de waarnemingen; bij statistisch werk wordt dit veel toegepast ten einde het waarnemingsmateriaal zo veel mogelijk te vereenvoudigen.

Voorbeeld.

Voor de waarnemingsreeks (5.2.1.) vinden we met een rekenmachine

$$\begin{aligned}\Sigma X_i &= 16.78 \quad \text{mm} \\ \Sigma X_i^2 &= 35.2098 \quad \text{mm}^2 \\ (\Sigma X_i)^2 / 8 &= 35.1960 \quad \text{mm}^2 \\ \text{KS} &= 0.0138 \quad \text{mm}^2\end{aligned}$$

en hieruit

$$\begin{aligned}\bar{X} &= \frac{16.78}{8} = 2.0975 \quad \text{mm}, \\ s^2 &= \frac{0.0138}{7} = 0.0020 \quad \text{mm}^2, \\ s &= 0.0447 \quad \text{mm}, \\ n &= 8, \quad \nu = 7.\end{aligned}$$

Trekken we van de waarnemingen 2.10 mm af dan vinden we

$$Y_i = X_i - 2.10 = -0.02 \quad 0.04 \quad 0.03 \quad -0.06 \quad 0.07 \quad -0.02 \quad -0.04 \quad -0.02 \quad \text{mm}$$

en hieruit uit het hoofd

$$\begin{aligned}\Sigma Y_i &= -0.02 \quad \text{mm} \\ \Sigma Y_i^2 &= 0.0138 \\ (\Sigma Y_i)^2 / 8 &= 0.00005 \\ \text{KS} &= 0.0138 \quad \text{mm}^2\end{aligned}$$

en hieruit

$$\bar{X} = 2.10 + \bar{Y} = 2.10 - \frac{0.02}{8} = 2.0975 \quad \text{mm}$$

en de andere waarden als boven.

Ook bij gebruik van een rekenmachine wordt de berekening vaak door aftrekken van een constant bedrag vereenvoudigd. Dit bedrag moet dan liefst zo gekozen worden dat alle verschillen $X_i - c$ positief zijn. In het beschouwde geval zal men bijvoorbeeld 2.00 mm aftrekken.

Afronding. Blijkens de laatste kolom van tabel mag het gemiddelde op

$$\frac{R}{2n} = \frac{0.13}{16} = 0.008 \text{ mm worden afgerond. Afronding op } 0.01 \text{ mm is dus}$$

iets te grof; men kan het gemiddelde als

$$\bar{X} = 2.098 \text{ mm}$$

weergeven.

Wanneer, als hier, het af te ronden getal op een 5 eindigt, rondt men af naar een even eindcijfer, dus 2.0975 wordt 2.098.

De standaardafwijking s mag worden afgerond op

$$\frac{s}{\sqrt{8v}} = \frac{0.045}{\sqrt{56}} = 0.006 \text{ mm.}$$

Men kan dus met

$$s = 0.045 \text{ mm volstaan.}$$

De waarnemingsreeks kan dus in zijn geheel met

$$n = 8, \quad \bar{X} = 2.098 \text{ mm,} \quad s = 0.045 \text{ mm}$$

worden samengevat. Zo nodig kan men ook $s(\bar{X})$ vermelden.

De originele waarnemingen kunnen terstond worden afgerond. Het verdient aanbeveling bij de verdere berekeningen een of twee decimalen extra mee te nemen en pas tot afronding van \bar{X} , s , en $s(\bar{X})$ over te gaan wanneer de berekeningen geheel voltooid zijn.

Het aantal vrijheidsgraden.

Bij een enkelvoudige waarnemingsreeks is het aantal vrijheidsgraden v steeds $=(n-1)$ en behoeft niet afzonderlijk vermeld. Doch het aantal vrijheidsgraden speelt wel een aparte rol; het is maatgevend voor de nauwkeurigheid van de varianties of de standaardafwijking s .

Vaak beschikken we over verscheidene waarnemingsreeksen waarvan mag worden ondersteld dat zij tot populatie met een zelfde standaardafwijking σ behoren. Een voorbeeld geeft tabel 5.2. De wegingen werden alle met dezelfde balans

Tabel 5.2.

Bepaling van de weegnauwkeurigheid uit 4 series van 3 wegingen aan 4 verschillende hardpapier schijfjes.

Schijfje	Wegingen in mgr.	KS_i mgr. ²	v_i	s_i^2 mgr. ²
1	807 807 802	16.7	2	8.35
2	699 696 698	4.7	2	2.35
3	793 795 792	4.7	2	2.35
4	805 800 804	14.0	2	7.00
		$\Sigma KS_i = 40.1$	$\Sigma v_i = 8$	
Eindresultaat:				
$s^2 = \frac{40.1}{8} = 5.01, \quad s = 2.2 \text{ mgr.}, \quad v = 8$				

uitgevoerd en we mogen aannemen dat de populatiespreiding in de 4 reeksen steeds dezelfde is geweest. In plaats van voor iedere serie apart een variantie of standaardafwijking te berekenen kan men de resultaten samenvoegen. De algemene formule luidt

$$s^2 = \frac{\Sigma KS_i}{\Sigma v_i} = \frac{\Sigma v_i s_i^2}{\Sigma v_i}, \quad v = \Sigma v_i = N-k \quad (5.2.4.)$$

wanneer N het totaal aantal waarnemingen en k het aantal samengevoegde series weergeven. De aldus door samenvoeging verkregen variantie of standaardafwijking is nauwkeuriger dan de waarden uit de afzonderlijke series afgeleid. De aparte betekenis van het aantal vrijheidsgraden komt bij dit samenvoegen duidelijk aan het licht.

In de natuurkunde noemt men het aantal vrijheidsgraden het aantal parameters dat gegeven moet zijn om een bepaalde geometrische configuratie volledig te bepalen. Hierbij aansluitend moet aan een serie van n waarnemingen n vrijheidsgraden worden toegekend, want de serie is slechts dan volledig bepaald als alle n waarnemingen bekend zijn.

Doch we kunnen ons de serie ook gegeven denken door het gemiddelde \bar{X} en de n afwijkingen $X_i - \bar{X}$. Aan het gemiddelde moet dan 1 vrijheidsgraad worden toegekend en aan de afwijkingen n-1, want hun som moet steeds nul zijn. Het zijn alleen de afwijkingen die de waarde van s^2 of s bepalen, het gemiddelde \bar{X} heeft hierop geen invloed.

Bewerking van een omvangrijke waarnemingsreeks.

Deze geschiedt het best met behulp van een frequentie tabel; we nemen tabel 3.1.2. (pag.24) als voorbeeld. De berekening van \bar{X} en s is uitgevoerd in tabel 5.3.

We nemen daarbij aan dat de waarnemingen die in één klasse vallen allen samenvallen met het klasse midden. Dit komt in feite neer op een afronding en wel bij tabel 5.3 of een afronding met een afrondingsinterval van 7 mgr. (= de klassebreedte). Waarnemingen die in een frequentietabel worden weergegeven behoeven daarom niet vooraf reeds te worden afgerond.

Vervolgens nemen we het midden van een van de middelste klassen als 0-punt, en nummeren de klassen van hieruit als aangegeven in kolom 4. Dit komt neer op een codering waarbij we de waarnemingen X vervangen door

Tabel 5.3

Berekening van \bar{X} , s^2 en s door middel van een frequentietabel. Gegevens van tabel 3.1.2: Gewichten van hardpapier schijfjes.

Gewichtsklasse mgr.	Middenklasse mgr.	Frequentie f	Code Y	fY	fY ²
750 - 756	753	1	-6	-6	36
757 - 763	760	4	-5	-20	100
764 - 770	767	0	-4	0	0
771 - 777	774	4	-3	-12	36
778 - 784	781	7	-2	-14	28
785 - 791	788	15	-1	-15	15
792 - 798	795	14	0	0	0
799 - 805	802	26	1	26	25
806 - 812	809	16	2	32	64
813 - 819	816	2	3	6	18
820 - 826	823	1	4	4	16
		90 = N = Σf		+ 1 = ΣfY	338 = ΣfY^2

Uit bovenstaande gegevens volgt:

$$\bar{Y} = \frac{+1}{90} = 0.011, \quad \bar{X} = 795 + 7 \bar{Y} = 795.1 \text{ mgr.}$$

$$KS(Y) = 338 - \frac{(+1)^2}{90} = 338, \quad KS(X) = 7^2 \times 338 = 16562,$$

$$s^2(X) = \frac{16562}{89} = 186 \text{ mgr}^2, \quad s = 13.6 \text{ mgr.}$$

$$Y = \frac{X-795}{7}$$

We berekenen nu voor Y het gemiddelde en de kwadratensom en vinden hieruit het gemiddelde en de kwadratensom voor X: volgens

$$\bar{X} = 795 + 7 \bar{Y}, \quad KS(X) = 7^2 KS(Y).$$

5.3. Schatting van de standaardafwijking uit de spreidingsbreedte

Een eenvoudiger methode om de standaardafwijking te schatten is door de spreidingsbreedte R te vermenigvuldigen met een factor A(n) die alleen afhangt van het aantal waarnemingen in de reeks, n. Waarden van deze factor zijn gegeven in Tabel 19, kolom 2. Voor de waarnemingsreeks (5.21) is $R = 2.17 - 2.04 = 0.13$ mm, $n = 8$ en wordt

$$s = A(n) R = 0.351 \times 0.13 = 0.046 \text{ mm} \tag{5.31}$$

terwijl uit de kwadratensom $s = 0.044$ mm werd gevonden. Beschikken we over verscheidene waarnemingsreeksen alle met eenzelfde aantal waarnemingen dan kan men de gemiddelde spreidingsbreedte gebruiken. De 4 reeksen van 3 waarnemingen in Tabel 5.2. geven

$$R = 5, 3, 3, 5 \text{ mgr. ; } \bar{R} = 4.0 \text{ mgr.}$$

en

$$s = 0.59 \times 4.0 = 2.36 \text{ mgr.}$$

tegen $s = 2.24$ mgr. door samenvoegen van de kwadratensommen gevonden. Het samenvoegen van een aantal waarnemingsreeksen gaat bij gebruik van de spreidingsbreedte alleen goed indien alle waarnemingsreeksen uit hetzelfde aantal waarnemingen bestaan (als in Tabel 5.2.).

De waarden van de factor A(n) gelden voor een normale verdeling en de berekening van een schatting s uit R of \bar{R} is daarom alleen toepasbaar bij verdelingen die redelijk door een normale verdeling kunnen worden weergegeven. De schatting van s uit de kwadratensom is bij iedere verdeling toepasbaar.

De schatting van s uit R of \bar{R} is iets minder nauwkeurig dan die uit de kwadratensom. Voegen we de kwadratensommen van k reeksen met ieder n waarnemingen samen, dan is het totaal aantal vrijheidsgraden

$$v = k(n-1) ;$$

in Tabel 5.2 $v = 4(3-1) = 8$. Berekenen we s uit \bar{R} dan kunnen we aan deze waarde van s ongeveer

$$v = 0.9 k(n-1) \tag{5.3.2}$$

vrijheidsgraden toekennen, dit getal naar beneden tot een geheel getal afgerond. In het geval van Tabel 5.2 dus

$$v = 0.9 \times 8 = 7.2 = 7.$$

Wat dit praktisch precies voor gevolgen heeft wordt later uiteengezet. De factor A(n) verschilt weinig van $1/\sqrt{n}$ zoals het volgende tabelletje laat zien

	A(n)	$\frac{1}{\sqrt{n}}$
n = 2	0.886	0.707
4	0.486	0.500
6	0.395	0.408
8	0.351	0.354
10	0.325	0.316

Heeft men geen tabel bij de hand dan kan $s = R/\sqrt{n}$ in een snelle ruwe schatting dienst doen.

5.4. Rechtvaardiging van de afrondingsregels

Ronden we de waarnemingen af met een afrondingsinterval a dan worden daarbij afrondingsfouten \underline{x} toegevoegd met een rechthoekige frequentieverdeling

$$f(x) = \frac{1}{a}, \quad -\frac{1}{2}a \leq x \leq +\frac{1}{2}a.$$

De variantie van deze afrondingsfouten bedraagt

$$\sigma^2(x) = \frac{a^2}{12}.$$

Is de variantie van de niet afgeronde waarnemingen σ_0^2 en van de afgeronde waarnemingen σ_1^2 dan zal gelden

$$\sigma_1^2 = \sigma_0^2 + \frac{a^2}{12} \quad (5.4.1)$$

want we mogen aannemen dat de oorspronkelijke waarnemingen en de afrondingsfouten onafhankelijk zijn.

Schrijven we nu voor

$$a \leq \frac{\sigma_0}{2}$$

dan is

$$\sigma_1^2 \leq \sigma_0^2 \left(1 + \frac{1}{48}\right).$$

Door het afronden neemt de variantie dus hoogstens met 2% toe, de standaardafwijking hoogstens met 1%. Dit is een zeer geringe toename die gezien de numerieke vereenvoudiging die erdoor wordt bereikt alleszins gerechtvaardigd is.

Waar doorgaans σ onbekend is ligt het voor de hand

$$a \leq \frac{s}{2} \quad \text{of} \quad a \leq \frac{R}{2\sqrt{n}}$$

voor te schrijven zoals in Tabel 5.2.1. is gedaan, daar $s = R/\sqrt{n}$ een snelle ruwe schatting van s is.

Verder is $\sigma(\bar{x}) = \sigma(x)/\sqrt{n}$ en het toelaatbare afrondingsinterval voor het gemiddelde \bar{x} is daarom $1/\sqrt{n}$ maal dat voor de individuele waarnemingen.

De afrondingsintervallen voor s^2 en s in Tabel 5.2.1. zijn gebaseerd op soortgelijke overwegingen, uitgaande van theoretisch afgeleide waarden voor de standaardafwijkingen van deze grootheden.

5.3. VRAAGSTUKKEN

5.3.1. Rond de volgende waarnemingsreeksen op de juiste wijze af en bereken uit het hoofd gemiddelde, variantie, en standaardafwijking.

a)	12	14	17	15	15	18	12	13
b)	153	175	180	162	155	160		
c)	762	885	700	926	910	775	815	821

5.3.2. Voer de berekeningen van vraagstuk 1 uit met behulp van een rekenmachine.

5.3.3. Aan 4 stalen kogels werd in een aantal verschillende richtingen de diameter gemeten met de navolgende uitkomsten.

Kogel No	Meting in microns						
1	5520	5529	5530	5527			
2	5528	5528	5528	5525	5528	5526	
3	5522	5522	5521	5521	5520		
4	5521	5523	5520				

Bepaal door samenvoegen van de variantie de gezamenlijke variantie en standaardafwijking.

Welke spreidingsbronnen zijn in deze standaardafwijking opgenomen ?

Is het samenvoegen van de varianties in dit geval volkomen juist ?

5.3.4. Past men de formule voor het samenvoegen van varianties toe op n paren waarnemingen (zg. duplo bepalingen) dan vindt men

$$s^2 = \frac{\sum d_i^2}{2n} \quad \text{met} \quad v = n,$$

waarin de d_i 's de verschillen tussen de duplo's zijn. Bewijs dat.

5.3.5. Bereken gemiddelde en standaardafwijking voor de gegevens van vraagstuk 3.10.1., en vergelijk de uitkomst met de waarden die met behulp van de grafische methode werden gevonden.

5.3.6. Bereken de gemiddelde spreidingsbreedte voor de 9 series van 10 waarnemingen waarin Tabel 3.1.1 (pag.24) is verdeeld. Bereken hieruit s en vergelijk de gevonden waarde met die berekend in Tabel 5.3.(pag 63).

HOOFDSTUK VI. STEEKPROEFVERDELINGEN EN HUN TOEPASSINGEN

6.1. Steekproefverdelingen

\bar{x} en s worden in de statistiek opgevat als schattingen van de parameters μ en σ . De beoordeling van de waarde van deze schattingen en de conclusie die men erop mag baseren geschiedt op grond van hun zg. steekproefverdelingen.

We nemen daarbij in eerste instantie aan, dat de populatie bekend is, en we denken ons dat we daaruit niet één maar bij herhaling een steekproef van n elementen trekken en daaruit \bar{x} en s berekenen. Dit is een procedure die we onbepaald kunnen herhalen. De steekproef in zijn geheel kunnen we opvatten als een element uit een oneindige populatie van steekproeven van n stuks. Daarmee correspondeert dan ook een oneindige populatie van gemiddelden \bar{x} en standaardafwijkingen s ; de verdelingen op deze populatie noemen we de steekproefverdeling van \bar{x} en s . Het onderstrepen van de symbolen geeft daarbij afdoende aan dat we ze nu als elementen van een populatie beschouwen.

Is de oorspronkelijke populatie bekend dan kan men daaruit de steekproefverdeling van \bar{x} , s of andere steekproefkenmerken in principe bepalen.

Verder betekent

$$\underline{x}_i = \mu + \underline{u}_i \sigma \quad , \quad i=1, \dots, n, \quad (6.2.3)$$

dat \underline{x}_i , $i=1, \dots, n$ een steekproef van n onderling onafhankelijke waarnemingen uit een normale verdeling is met gemiddelde μ en standaardafwijking σ . Evenzo geeft

$$\underline{x}_i = \mu + \underline{w}_i \sigma \quad , \quad i=1, \dots, n \quad (6.2.4)$$

hetzelfde voor een willekeurige verdeling.

Tenzij nadrukkelijk anders vermeld onderstellen de notaties (6.2.3) en (6.2.4) steeds onderlinge onafhankelijkheid.

6.3. De χ^2 -verdeling

Is

$$\underline{x}_i = \underline{u}_i, \quad i=1, \dots, v. \quad (6.3.1)$$

Dan is

$$\underline{\chi}_v^2 = \sum_{i=1}^v \underline{x}_i^2 = \sum_{i=1}^v \underline{u}_i^2, \quad \text{met } v \text{ vrijheidsgraden.} \quad (6.3.2)$$

In woorden: $\underline{\chi}_v^2$ is de som van de kwadraten van v onderling onafhankelijke standaard-normale toevalsvariabelen.

Daar

$$E(\underline{u}^2) = 1$$

is

$$E(\underline{\chi}_v^2) = v \quad (6.3.3)$$

Verder vindt men door partiële integratie gemakkelijk dat

$$E(\underline{u}^4) = 3, \quad (6.3.4)$$

en dus

$$\text{var}(\underline{u}^2) = 3 - 1^2 = 2.$$

HOOFDSTUK VI. STEEKPROEFVERDELINGEN EN HUN TOEPASSINGEN

6.1. Steekproefverdelingen

\bar{x} en s worden in de statistiek opgevat als schattingen van de parameters μ en σ . De beoordeling van de waarde van deze schattingen en de conclusie die men erop mag baseren geschiedt op grond van hun zg. steekproefverdelingen.

We nemen daarbij in eerste instantie aan, dat de populatie bekend is, en we denken ons dat we daaruit niet één maar bij herhaling een steekproef van n elementen trekken en daaruit \bar{x} en s berekenen. Dit is een procedure die we onbeperkt kunnen herhalen. De steekproef in zijn geheel kunnen we opvatten als een element uit een oneindige populatie van steekproeven van n stuks. Daarmee correspondeert dan ook een oneindige populatie van gemiddelden \bar{x} en standaardafwijkingen s ; de verdelingen op deze populatie noemen we de steekproefverdeling van \bar{x} en s . Het onderstrepen van de symbolen geeft daarnij afdoende aan dat we ze nu als elementen van een populatie beschouwen.

Is de oorspronkelijke populatie bekend dan kan men daaruit de steekproefverdeling van \bar{x} , s of andere steekproefkenmerken in principe bepalen.

Algemeen geschiedt dit steeds volgens eenzelfde methode. De kans op een aselechte steekproef waarvan \underline{x}_i valt in het interval x_i, dx_i is

$$\Pr(\underline{x}_i \leq x_i \leq x_i + dx_i ; i=1, \dots, n) = f(x_1) f(x_2) \dots f(x_n) dx_1 dx_2 \dots dx_n. \quad (6.1.1)$$

De kans

$$\Pr(\bar{x} < \underline{\bar{x}} < \bar{x} + d\bar{x}) \quad (6.1.2)$$

dat $\underline{\bar{x}}$ in het interval $\bar{x}, d\bar{x}$ valt vinden we dan door (6.1.1) te integreren over het gebied waarvoor

$$\bar{x} < \frac{\sum x_i}{n} < \bar{x} + d\bar{x}. \quad (6.1.3)$$

Evenzo vinden we de verdeling van s door (6.1.1) te integreren over het gebied waarvoor

$$s < \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} < s + ds. \quad (6.1.4)$$

In principe is deze methode steeds toepasbaar al leidt zij soms tot integralen die niet gemakkelijk oplosbaar zijn.

In het navolgende zullen we echter niet op de wiskundige complicaties ingaan, doch alleen de resultaten weergeven.

6.2. Een nuttige notatie

Voortaan betekent de schrijfwijze

$$\underline{x} = \mu + \underline{u}\sigma \quad (6.2.1)$$

\underline{x} heeft een normale verdeling met gemiddelde μ en standaardafwijking σ . Evenzo betekent

$$\underline{x} = \mu + \underline{w} \sigma \quad (6.2.2)$$

\underline{x} heeft een willekeurige verdeling met gemiddelde μ en standaardafwijking σ .

Verder betekent

$$\underline{x}_i = \mu + \underline{u}_i \sigma, \quad i=1, \dots, n, \quad (6.2.3)$$

dat \underline{x}_i , $i=1, \dots, n$ een steekproef van n onderling onafhankelijke waarnemingen uit een normale verdeling is met gemiddelde μ en standaardafwijking σ . Evenzo geeft

$$\underline{x}_i = \mu + \underline{w}_i \sigma, \quad i=1, \dots, n \quad (6.2.4)$$

hetzelfde voor een willekeurige verdeling.

Tenzij nadrukkelijk anders vermeld onderstellen de notaties (6.2.3) en (6.2.4) steeds onderlinge onafhankelijkheid.

6.3. De χ^2 -verdeling

Is

$$\underline{x}_i = \underline{u}_i, \quad i=1, \dots, v. \quad (6.3.1)$$

Dan is

$$\chi_v^2 = \sum_{i=1}^v \underline{x}_i^2 = \sum_{i=1}^v \underline{u}_i^2, \quad \text{met } v \text{ vrijheidsgraden.} \quad (6.3.2)$$

In woorden : χ_v^2 is de som van de kwadraten van v onderling onafhankelijke standaard-normale toevalsvariabelen.

Daar

$$\begin{aligned} \text{is} \quad E(\underline{u}^2) &= 1 \\ E(\chi_v^2) &= v \end{aligned} \quad (6.3.3)$$

Verder vindt men door partiële integratie gemakkelijk dat

$$E(\underline{u}^4) = 3, \quad (6.3.4)$$

en dus

$$\text{var}(\underline{u}^2) = 3 - 1^2 = 2.$$

Dan is

$$\text{var}(\chi^2_v) = 2v. \quad (6.3.5)$$

Fig.6.3.1 brengt enige χ^2 -verdelingen in beeld. Voor een gering aantal vrijheidsgraden is dit een sterk asymmetrische verdeling doch naarmate v toeneemt convergeert zij, in overeenstemming met de centrale limietstelling naar een normale verdeling. Voor $v > 30$ kan men de χ^2 -verdeling dan ook zonder bezwaar door een normale verdeling met gemiddelde v en standaardafwijking $\sqrt{2v}$ vervangen.

De exacte formule voor de χ^2 -verdeling luidt

$$f(\chi^2_v) d\chi^2_v = A (\chi^2_v)^{\frac{1}{2}v-1} e^{-\chi^2_v/2} d\chi^2_v, \quad (6.3.6)$$

waarin

$$A = \left\{ 2^{v/2} \Gamma\left(\frac{v}{2}\right) \right\}^{-1}, \quad (6.3.7)$$

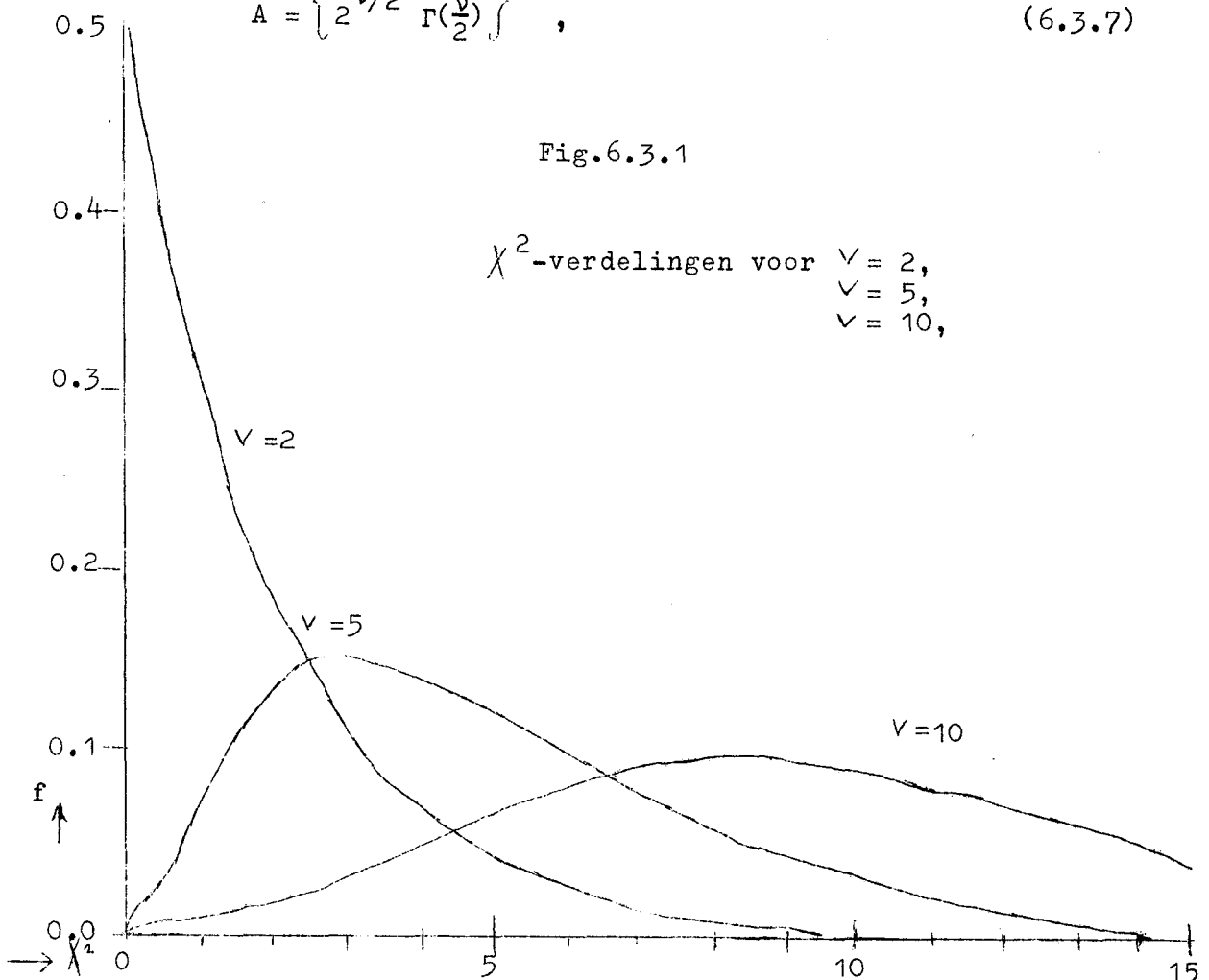


Fig.6.3.1

χ^2 -verdelingen voor $v = 2,$
 $v = 5,$
 $v = 10,$

waarin Γ de bekende gamma-functie voorstelt. De uitdrukking (6.3.6) wordt verkregen door

$$\left(\frac{1}{\sqrt{2\pi}} \right)^v \exp \left\{ - \frac{\sum u_i^2}{2} \right\} du_1 du_2 \dots du_v$$

te integreren over het gebied waarvoor

$$\chi_v^2 < \sum u_i^2 < \chi_v^2 + d\chi_v^2 .$$

De cumulatieve χ^2 -verdeling is getabelleerd in Tabel 3. Boven langs de tabel staan de cumulatieve kansen

$$\Pr(\underline{\chi}_v^2 < \chi_v^2),$$

rechts het aantal vrijheidsgraden en in de tabel de bijbehorende waarde van χ^2 . We lezen uit deze tabel bv.

$$\Pr(\underline{\chi}_4^2 < 9.49) = 0.95, \quad (6.3.8)$$

$$\Pr(\underline{\chi}_4^2 < 0.711) = 0.05, \quad (6.3.9)$$

en dus

$$\Pr(0.711 < \underline{\chi}_4^2 < 9.49) = 0.90. \quad (6.3.10)$$

Uit de definitie 6.3.2. van χ^2 volgt onmiddellijk

$$\underline{\chi}_{v_1}^2 + \underline{\chi}_{v_2}^2 + \dots + \underline{\chi}_{v_k}^2 = \chi_v^2 \quad \text{met } v = v_1 + v_2 + \dots + v_k. \quad (6.3.11)$$

In woorden : de som van een aantal onderling onafhankelijke χ^2 -variabelen is een χ^2 -variabele met een aantal vrijheidsgraden gelijk aan de som van de aantallen vrijheidsgraden der componenten.

6.4. Een toepassing van de χ^2 -verdeling.

$$\text{Is } \underline{x}_i = \mu + \underline{u}_i \sigma, \quad i=1, \dots, n \quad (6.4.1)$$

dan is

$$\frac{\underline{x}_i - \mu}{\sigma} = \underline{u}_i,$$

en dus

$$\frac{\sum (\underline{x}_i - \mu)^2}{\sigma^2} = \chi_n^2 \quad (6.4.2)$$

een χ^2 -variabele met n vrijheidsgraden.

Meestal kennen we het ware populatiegemiddelde μ niet en we moeten daarom in (6.4.2) μ vervangen door het steekproefgemiddelde \bar{x} . Het kan nu worden bewezen dat

$$\frac{\sum (\underline{x}_i - \bar{x})^2}{\sigma^2} = \chi_v^2 = \chi_{n-1}^2,$$

en dus

$$\frac{v \sigma^2}{\sigma^2} = \chi_v^2 \quad \text{met } v = n-1,$$

of
$$\frac{\underline{s}^2}{\sigma^2} = \frac{\chi_v^2}{v}, \quad (6.4.3)$$

of
$$\frac{\underline{s}}{\sigma} = \frac{\chi_v}{\sqrt{v}}. \quad (6.4.4)$$

Met andere woorden de verhoudingen \underline{s}^2/σ^2 of \underline{s}/σ hebben steekproefverdelingen die we direct uit de χ^2 -verdeling kunnen afleiden. Voor een \underline{s} met 61 vrijheidsgraden kunnen we (6.3.8) nu herschrijven als

of
$$\Pr\left(\frac{\underline{s}^2}{\sigma^2} < \frac{9.49}{4} = 2.37\right) = 0.95 \quad (6.4.5)$$

of
$$\Pr\left(\frac{\underline{s}}{\sigma} < \sqrt{2.37} = 1.54\right) = 0.95 \quad (6.4.6)$$

of
$$\Pr\left(\frac{\underline{s}}{1.54} = 0.65 \underline{s} < \sigma\right) = 0.95. \quad (6.4.7)$$

Deze laatste vergelijking geeft een zg. linker betrouwbaarheidsgrens voor de onbekende σ . In plaats van (6.4.7) schrijft men ook

$$0.65 \underline{s} < \sigma \text{ met } 95\% \text{ betrouwbaarheid}$$

of
$$\text{of met onbetrouwbaarheid } \alpha = 5\% \quad (6.4.8)$$

Wanneer we uit normale verdelingen steekproeven van 5 stuks nemen en voor iedere steekproef de grenswaarde $0.65 \underline{s}$ berekenen, dan zal deze grens in 95% van de gevallen kleiner zijn dan σ .

Men kan op precies dezelfde wijze uit (6.3.9) een rechter betrouwbaarheidsgrens voor σ geven. Deze luidt voor $v = 4$

$$\sigma < 2.37 \underline{s} \text{ met } 95\% \text{ betrouwbaarheid} \quad (6.4.9)$$

en door combinatie van (6.4.8) en (6.4.9) krijgen we een tweezijdige betrouwbaarheidsinterval :

$$0.65 \underline{s} < \sigma < 2.37 \underline{s} \text{ met } 90\% \text{ betrouwbaarheid} \\ \text{of met een onbetrouwbaarheid } \alpha = 10\% \quad (6.4.10)$$

In dit laatste geval geeft 10% de totale onbetrouwbaarheid, d.w.z. de kans dat aan één van beide ongelijkheden in (6.4.10) niet zal zijn voldaan. De kans voor iedere ongelijkheid afzonderlijk is 5%, daarom noemen we (6.4.10) een symmetrisch betrouwbaarheidsinterval. Men kan ook asymmetrische intervallen construeren door voor de linker- en rechtergrens verschillende onbetrouwbaarheden te kiezen, doch dit is niet gebruikelijk en we gaan hier niet verder op in.

De onbetrouwbaarheid α kan willekeurig gekozen worden; zeer gebruikelijke waarden zijn $\alpha = 0.01 = 1\%$ en $\alpha = 0.05 = 5\%$. Hoe kleiner α des te breder wordt het interval.

Algemeen kan men de betrouwbaarheidsintervallen voor σ weergeven met

$$a_1 \underline{s} < \sigma \quad \text{met éézijdige onbetrouwbaarheid } \alpha, \quad (6.4.11)$$

$$a_2 \underline{s} > \sigma \quad \text{met éézijdige onbetrouwbaarheid } \alpha, \quad (6.4.12)$$

$$a_1 \underline{s} < \sigma < a_2 \underline{s} \quad \text{met tweezijdige onbetrouwbaarheid } \alpha, \quad (6.4.13)$$

De waarden van de factoren a_1 en a_2 hangen af van ν en van de verlangde betrouwbaarheid α . Gangbare waarden zijn gegeven in Tabel 5.

De samenhang met de Tabel 3 voor de χ^2 -verdeling is gegeven door

$$a_1 = \sqrt{\frac{\nu}{\chi_{\nu}^2}} \quad \text{voor } P = (1 - \alpha), \quad (6.4.14)$$

$$a_2 = \sqrt{\frac{\nu}{\chi_{\nu}^2}} \quad \text{voor } P = \alpha .$$

6.4. Aanvulling

Uit (6.3.11) volgt dat deze betrouwbaarheidsintervallen ook van toepassing zijn op een variantie s^2 die door samenvoegen van een aantal waarnemingsreeksen volgens de formule

$$s^2 = \frac{\sum v_i s_i^2}{\sum v_i}, \quad v = \sum v_i$$

zijn verkregen, mits deze reeksen aselechte steekproeven vormen uit populaties met eenzelfde standaardafwijking σ .

6.5. Diverse eigenschappen van de verdeling van \underline{s}^2 en van \underline{s}

Daar $E(\chi_v^2) = v$ volgt uit (6.4.3)

$$E(\underline{s}^2) = \sigma^2. \quad (6.5.1)$$

In woorden : Het populatiegemiddelde van \underline{s}^2 is gelijk aan σ^2 . Deze betrekking geldt niet alleen voor de normale verdeling, doch geldt algemeen voor iedere verdeling. Dit volgt uit

$$\begin{aligned} E\{(n-1) \underline{s}^2\} &= E\{\sum (\underline{x}_i - \bar{\underline{x}})^2\} = E\{\sum \underline{x}_i^2 - n\bar{\underline{x}}^2\} = \\ &= E\{\sum (\underline{x}_i - \mu)^2 - n(\bar{\underline{x}} - \mu)^2\} = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2. \end{aligned} \quad (6.5.2)$$

Op grond van (6.5.1) noemt men \underline{s}^2 een zuivere schatting van σ^2 . Deze benaming is ook in andere gevallen van toepassing. Berekent men uit een steekproef een schatting \underline{k} van een populatieparameter κ dan heet \underline{k} een zuivere schatting van κ indien

$$E(\underline{k}) = \kappa. \quad (6.5.3)$$

Uit (6.5.1) volgt dat \underline{s} geen zuivere schatting is van σ . Immers we hebben

$$\text{var } \underline{s} = E(\underline{s}^2) - \{E(\underline{s})\}^2 = \sigma^2 - \{E(\underline{s})\}^2$$

$$\text{of} \quad \{E(\underline{s})\}^2 = \sigma^2 - \text{var } \underline{s}, \quad (6.5.4)$$

en daar $\text{var } \underline{s} > 0$ moet steeds gelden

$$E(\underline{s}) < \sigma. \quad (6.5.5)$$

Het verschil is echter niet groot en neemt met toenemend aantal vrijheidsgraden snel af, zoals Tabel 6.5.1 laat zien.

TABEL 6.5.1.

Verhouding $E(\underline{s})/\sigma$ als functie van het aantal vrijheidsgraden bij een normale populatie.

$\nu =$	$E(\underline{s})/\sigma$	ν	$E(\underline{s})/\sigma$
1	0.798	10	0.975
2	0.886	20	0.988
3	0.921	30	0.992
4	0.940	40	0.994
5	0.952	50	0.995

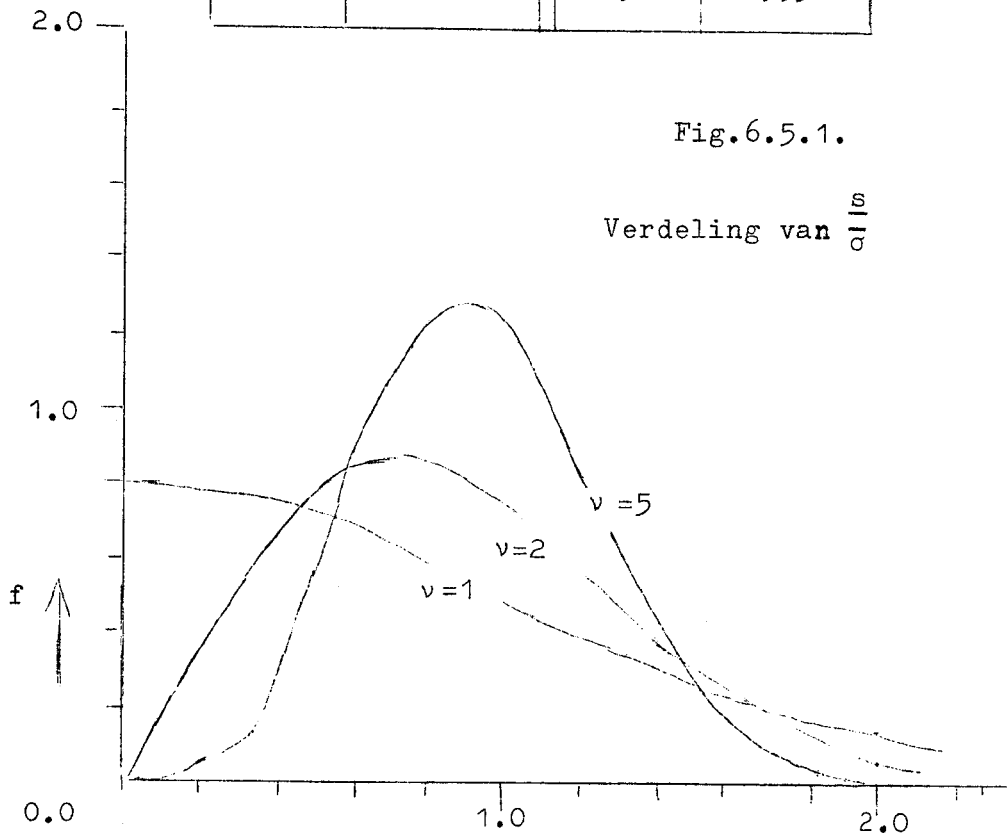


Fig.6.5.1.

Verdeling van $\frac{s}{\sigma}$

In fig.6.5.1. zijn enige verdelingen van \underline{y}/σ in beeld gebracht. Zoals uit deze figuur te zien is convergeert de verdeling van \underline{s}/σ zeer snel naar een normale verdeling met toenemende ν , veel sneller dan de χ^2 -verdeling. Reeds bij $\nu = 5$ wordt de normale vorm vrij nauwkeurig benaderd.

Is het aantal vrijheidsgraden voldoende groot dan kan men zonder bezwaar aannemen dat \underline{s} een normale verdeling bezit met een gemiddelde σ . De variantie van \underline{s} vinden we op de volgende wijze. We gaan uit van

$$\underline{s}^2 = \sigma^2 + 2\sigma(\underline{s} - \sigma) + (\underline{s} - \sigma)^2. \quad (6.5.6)$$

Is het aantal vrijheidsgraden niet te klein dan ligt \underline{s} in de buurt van σ en is de term $(\underline{s} - \sigma)^2$ te verwaarlozen t.o.v. de term $2\sigma(\underline{s} - \sigma)$. Dan geldt daarom in eerste benadering:

$$\text{var } \underline{s}^2 = 4\sigma^2 \text{ var } \underline{s}. \quad (6.5.7)$$

Uit $\underline{s}^2 = \sigma^2 \chi_v^2 / v$ en $\text{var } \chi_v^2 = 2v$ volgt verder

$$\text{var } \underline{s}^2 = \frac{2\sigma^4}{v} \quad (6.5.8)$$

en met (6.5.7) dus

$$\text{var } \underline{s} = \frac{\sigma^2}{2v}, \quad (6.5.9)$$

en

$$\sigma(\underline{s}) = \frac{\sigma}{\sqrt{2v}}. \quad (6.5.10)$$

Bij het vaststellen van de in Tabel 5.2.1. voor s gegeven afrondingsintervallen is van deze functie gebruik gemaakt. Het afrondingsinterval mag hoogstens $\frac{1}{2} \sigma(\underline{s}) = \sigma / \sqrt{8v}$ bedragen, daar we σ niet kennen vervangen we σ door de schatting s , om het toelaatbare afrondingsinterval te bepalen.

Samenvattend kunnen we voor grote v bij benadering aannemen dat

$$\underline{s} = \sigma + \underline{u}\sigma / \sqrt{2v} = \sigma(1 + \underline{u} / \sqrt{2v}). \quad (6.5.11)$$

Maken we hiervan gebruik dan vinden we voor de factoren a_1 en a_2 in de betrouwbaarheidsgrenzen (6.4.11), (6.4.13).

$$a_1 = \frac{\sqrt{2v}}{\sqrt{2v} + u_F}, \quad a_2 = \frac{\sqrt{2v}}{\sqrt{2v} - u_F} \quad (6.5.12)$$

waarbij u_F uit Tabel 1 kan worden opgezocht.

Hierbij geldt

$$F = 1 - \alpha \quad \text{voor ééNZijddige grenzen}$$

$$\text{en } F = 1 - \frac{1}{2}\alpha \quad \text{voor een betrouwbaarheidsinterval.}$$

De formules (6.5.12) kunnen ook dienst doen om het aantal vrijheidsgraden te bepalen vereist om σ met een vooraf voorgeschreven nauwkeurigheid te bepalen. Stel, we verlangen σ met 95% betrouwbaarheid op $\pm 10\%$ te kennen; d.w.z. we verlangen ongeveer $a_1 = 0.90$ en $a_2 = 1.10$ of

$$\frac{a_2}{a_1} = \frac{1.10}{0.90} = 1.22.$$

Dan moet

$$\frac{\sqrt{2v} + u_F}{\sqrt{2v} - u_F} = 1.22$$

of

$$2.22 u_F = 0.22 \sqrt{2v}.$$

Voor een tweezijdige betrouwbaarheid van 95% geldt $u_F = 1.96$ en wordt dus vereist

$$v = \frac{1}{2} \left(\frac{2.22 \times 1.96}{0.22} \right)^2 = 196. \quad (6.5.13)$$

De eis σ op $\pm 10\%$ nauwkeurig te bepalen is op zichzelf niet voldoende. Statistisch dient men er tevens bij te vermelden met welke betrouwbaarheid aan dit verlangen moet worden voldaan.

Verder vestigen we de aandacht op het zeer grote aantal vrijheidsgraden, en dus ook waarnemingen, dat vereist wordt om $\pm 10\%$ met 95% betrouwbaarheid te realiseren. Zonder toepassing van statistische methoden heeft men doorgaans de neiging de waarde van verrichte waarnemingen te overschatten.

Tenslotte zij er op gewezen dat we in het voorafgaande het begrip "aantal vrijheidsgraden" in tweeërlei zin gebruikt hebben.

- 1) Als het aantal vrijheidsgraden van een reeks of reeksen waarnemingen dat wordt gebruikt bij de bepaling van een s^2 of s .
- 2) Als een parameter waardoor de χ^2 -verdeling wordt bepaald.

Wanneer we de χ^2 -verdeling toepassen voor de interpretatie van een s^2 komen beide betekenissen overeen; het aantal vrijheidsgraden van s^2 is dan tevens de parameter van de ermee corresponderende χ^2 -verdeling. Doch de χ^2 -verdeling wordt in de statistiek ook toegepast als benadering van de verdeling van tal van andere statistische grootheden en in dat geval speelt het aantal vrijheidsgraden als parameter van de χ^2 -verdeling een geheel eigen rol. Men moet de beide betekenissen daarom goed uiteenhouden.

Bij de bepaling van een standaardafwijking s uit een gemiddelde spreidingsbreedte \bar{R} volgens (5.3.1.) kan volgens (5.3.2.) een aantal vrijheidsgraden

$$v = 0.9 k(n-1) \text{ naar beneden afgerond} \quad (6.5.14)$$

worden toegeschreven. Dat betekent dat, indien \underline{s} uit \bar{R} wordt bepaald, in goede benadering geldt

$$\frac{\underline{s}}{\sigma} = \frac{\underline{\chi}_v}{\sqrt{v}} \quad (6.5.15)$$

mits we voor v het door (6.5.14) voorgeschreven aantal vrijheidsgraden gebruiken; v functioneert hier alleen nog als de parameter van de χ^2 -verdeling.

6.6. Een betrouwbaarheidsinterval voor een gemiddelde .

Is

$$\underline{x}_i = \mu + \underline{w}_i \sigma \quad i = 1, \dots, n \quad (6.6.1)$$

dan geldt algemeen

$$E(\bar{x}) = \mu \quad , \quad (6.6.2)$$

$$\text{var}(\bar{x}) = \frac{\sigma^2}{n} \quad , \quad (6.6.3)$$

Volgens de centrale limietstelling geldt dan in goede benadering voor niet te kleine n

$$\bar{x} = \mu + \underline{u} \frac{\sigma}{\sqrt{n}}$$

of

$$\frac{(\bar{x} - \mu) \sqrt{n}}{\sigma} = \underline{u} \quad (6.6.4)$$

Hieruit volgt dan bijv.

$$\Pr(-u < \frac{(\bar{x} - \mu) \sqrt{n}}{\sigma} < +u) = (1 - \alpha) \quad (6.6.5)$$

wanneer u die waarde uit Tabel 1 is waarvoor

$$2 \{1 - F(u)\} = \alpha \quad (6.6.6)$$

We kunnen (6.6.5) herschrijven als

$$\Pr(\bar{x} - u \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + u \frac{\sigma}{\sqrt{n}}) = (1 - \alpha) \quad (6.6.7)$$

Is σ bekend dan geeft deze vergelijking een betrouwbaarheidsinterval voor het gemiddelde μ . Evenzo kunnen we éézijdige betrouwbaarheidsintervallen construeren.

Een moeilijkheid bij de toepassing van (6.6.7) is echter dat we σ doorgaans niet kennen doch alleen een schatting \underline{s} , die op zichzelf aan statistische fluctuaties onderhevig is.

Het interval (6.6.7) is gebaseerd op de relatie (6.6.4). Het ligt nu voor de hand om, voor het geval σ onbekend is te vragen naar de verdeling van de grootte (6.6.4) wanneer we daarin σ door \underline{s} vervangen, dus van

$$\underline{t} = \frac{(\bar{x} - \mu) \sqrt{n}}{\underline{s}} \quad (6.6.8)$$

Deze zgn. t -verdeling kan wederom volgens het algemene principe van § 1 worden gevonden wanneer de verdeling van de waarnemingen bekend is.

Hebben we met normaal verdeelde waarnemingen te maken, dus geldt

$$\underline{x}_i = \mu + \underline{u}_i \sigma \quad (6.6.9)$$

dan vindt men

$$f(t) dt = \frac{A dt}{(1 + t^2/v)^{\frac{1}{2}(v+1)}} \quad (6.6.10)$$

waarbij de constante A zo moet worden bepaald dat

$$\int_{-\infty}^{+\infty} f(\tau) d\tau = 1.$$

De exacte wiskundige uitdrukking voor A laten we hier achterwege. We zien uit (6.6.10) dat de verdeling van t symmetrisch is t.o.v. het punt t = 0, en verder uitsluitend afhangt van het aantal vrijheidsgraden v bij de schatting \underline{s} behorend.

Op grond van de t-verdeling geldt nu bijv.

$$\Pr(-t < \frac{(\bar{x} - \mu) \sqrt{n}}{\underline{s}} < +t) = (1 - \alpha),$$

$$\text{of} \quad \Pr(\bar{x} - t \frac{\underline{s}}{\sqrt{n}} < \mu < \bar{x} + t \frac{\underline{s}}{\sqrt{n}}) = (1 - \alpha) \quad (6.6.11)$$

wanneer t die waarde is waarvoor

$$\Pr(\underline{t} > t) = \frac{1}{2}\alpha. \quad (6.6.12)$$

Daarmee is een tweezijdig betrouwbaarheidsinterval voor μ gevonden. Eenzijdige betrouwbaarheidsintervallen vinden we op geheel analoge wijze.

De waarden van t behoevend voor het construeren van deze betrouwbaarheidsintervallen zijn getabelleerd in Tabel 2. In § 5.2 vonden we uit een serie diktemetingen aan 8 hardpapier schijfjes

$$\begin{aligned} \bar{x} &= 2.098 \text{ mm} ; n = 8 \\ s &= 0.045 \text{ mm} ; v = 7. \end{aligned}$$

Met behulp van Tabel 2 vinden we dan bijv.

$$2.098 - 2.36 \frac{0.045}{\sqrt{8}} < \mu < 2.098 + 2.36 \frac{0.045}{\sqrt{8}}$$

$$\text{of} \quad 2.060 < \mu < 2.136 \text{ mm met } 95\% \text{ betrouwbaarheid,} \quad (6.6.13)$$

En evenzo

$$2.098 - 3.00 \frac{0.045}{\sqrt{8}} < \mu$$

$$2.050 \text{ mm} < \mu \text{ met } 99\% \text{ betrouwbaarheid} \quad (6.6.14)$$

Enige verdere eigenschappen van de t-verdeling en haar toepassingen.

We kunnen \underline{t} als volgt schrijven

$$\underline{t} = \frac{(\bar{x} - \mu) \sqrt{n}}{\underline{s}} = \frac{(\bar{x} - \mu) \sqrt{n}}{\sigma} : \frac{\underline{s}}{\sigma} \quad (6.6.15)$$

Nu is $\frac{(\bar{x} - \mu) \sqrt{n}}{\sigma} = \underline{u}$ en $\frac{\underline{s}}{\sigma} = \sqrt{\frac{\chi_v^2}{v}} = \frac{\chi_v}{\sqrt{v}}$.

Bovendien zijn voor steekproeven uit een normale verdeling \bar{x} en \underline{s} onderling onafhankelijk, hetgeen we hier echter niet bewijzen kunnen. Tezamen geeft dit

$$\underline{t} = \underline{u} : \frac{\chi_v}{\sqrt{v}}, \quad (6.6.16)$$

of in woorden : de t-verdeling is de verdeling van het quotient van twee onafhankelijke toevalsvariabelen met respectievelijk een standaard normale en een χ_v/\sqrt{v} verdeling.

De onafhankelijkheid tussen \bar{x} en \underline{s} is belangrijk omdat daaruit volgt dat het bij gebruik van de t-verdeling onverschillig is of men een schatting \underline{s} gebruikt die uit dezelfde waarnemingen is bepaald als \bar{x} ; men kan evengoed een schatting \underline{s} gebruiken ten dele of geheel uit andere series waarnemingen gevonden.

In Tabel 5.2. (pag.60) geeft bijv. 4 series van 3 wegingen aan 4 hard papieren schijfjes waaruit we vinden

$$\begin{aligned} \bar{x}_1 &= 805.3 \text{ mgr, } s_1 = 2.9 \text{ mgr, } n=3, v=2 ; \\ \bar{x}_2 &= 697.7 \text{ " , } s_2 = 1.5 \text{ " , " , " ; } \\ \bar{x}_3 &= 793.3 \text{ " , } s_3 = 1.5 \text{ " , " , " ; } \\ \bar{x}_4 &= 803.0 \text{ " , } s_4 = 2.6 \text{ " , " , " . } \end{aligned}$$

We kunnen nu bijv. op grond van \bar{x}_1 en s_1 vaststellen dat

$$798.1 < \mu_1 < 812.5 \text{ mgr. met 95\% betrouwbaarheid} \quad (6.6.17)$$

Daar echter $s_1, s_2, s_3,$ en s_4 allen schattingen zijn van eenzelfde σ kunnen zij worden samengevoegd tot

$$s = 2.2 \text{ mgr, } v = 8$$

en hiermee vinden we

$$802.4 < \mu_1 < 808.1 \text{ mgr met 95\% betrouwbaarheid} \quad (6.6.18)$$

een veel nauwer interval dan (6.6.17). Ten dele is de winst aan nauwkeurigheid te wijten aan een lagere waarde van s , doch een belangrijker deel is toe te schrijven aan de lagere waarde van t die behoort bij het groter aantal vrijheidsgraden ($t = 4.30$ voor $v = 2$ tegen $t = 2.31$ voor $v = 8$). Het beste in een dergelijk geval is de individuele waarden s_1, \dots, s_4 te vergeten en alleen de door samenvoeging verkregen waarde s te gebruiken.

Alleen wanneer het aantal vrijheidsgraden gering is heeft het een sterke invloed op de waarde van t . Bij hogere waarden van ν wordt de invloed betrekkelijk gering en komt het aantal vrijheidsgraden er niet meer zo op aan. In de limiet $\nu \rightarrow \infty$ wordt $\underline{s} \equiv \sigma$ en dan valt de t -verdeling met de u -verdeling samen.

In de praktijk komt het op de exacte waarde van een betrouwbaarheidsinterval niet zo zeer aan; het gaat meer om de orde van grootte. Algemeen rekent men dan ook dat men het verschil tussen t en u zonder bezwaar kan verwaarlozen wanneer $\nu > 30$. Men kan dan (6.6.7) gebruiken i.p.v. (6.6.11). Een eenvoudige vuistregel is in dat geval dat het 95% betrouwbaarheidsinterval voor μ wordt gegeven door $\bar{x} \pm 2s/\sqrt{n}$. Ten slotte zij nog opgemerkt dat de onafhankelijkheid van \bar{x} en \underline{s} uit een zelfde waarnemingsreeks alleen geldt voor normaal verdeelde waarnemingen. Bij niet normale verdelingen treedt wel degelijk een zekere afhankelijkheid op, doch deze is gering wanneer de verdeling ten naaste bij normaal is.

Uitgebreide theoretische onderzoeken hebben overigens aangetoond, dat de t -verdeling robust is, d.w.z. weinig beïnvloed wordt door de vorm van de verdeling van de oorspronkelijke waarnemingen. De betrouwbaarheidsintervallen op de t -verdeling gebaseerd geven dan ook nog een redelijk resultaat bij verdelingen die vrij ver van de normale afwijken. Voor grote waarden van ν is gemakkelijk in te zien dat dit zo moet zijn. De steekproeffluctuaties van \underline{s}/σ spelen dan een relatief geringe rol en de teller van (6.6.15) heeft een u -verdeling op grond van de centrale limietstelling.

Het betrouwbaarheidsinterval kan dienst doen bij de kritische beoordeling van de waarnemingsresultaten. Wanneer men een $\frac{1}{2}$ pond boter koopt zal de klant verlangen dat de netto inhoud gemiddeld ten minste 250 gram bedraagt. Weegt men nu bijv. 10 pakjes na dan kan men op grond van deze metingen een betrouwbaarheidsinterval voor het gemiddelde gewicht per pakje aangeven. Ligt dit interval in zijn geheel boven de 250 gram dan is daarmee, statistisch beschouwd, bewezen dat μ groter dan 250 gram is. Ligt het interval in zijn geheel beneden de 250 gram dan bewijst dit dat $\mu < 250$ gram. Sluit het interval de waarde 250 gram in dan is het niet mogelijk een duidelijke uitspraak te doen. Zo nodig zal men meer waarnemingen moeten verrichten teneinde μ binnen nauwer perken vast te leggen. Daarbij kan men in eerste benadering aannemen dat de breedte van een betrouwbaarheidsinterval afneemt omgekeerd evenredig met de wortel uit het aantal waarnemingen.

Welke betrouwbaarheid men verlangt en welke conclusies op grond van een betrouwbaarheidsinterval men trekt hangt sterk van de omstandigheden af. Bij wetenschappelijk onderzoek zal men vaak een betrouwbaarheidsinterval weergeven als maat voor de bereikte nauwkeurigheid, echter zonder hieraan uitgesproken conclusies te verbinden. Een betrouwbaarheid van 95% is dan algemeen gebruikelijk.

In het geval van het botergewicht zou men daarentegen op grond van een betrouwbaarheidsinterval de conclusie kunnen trekken dat men niet het volle half pond ontvangt en dan tot maatregelen overgaan; bijv. in het uiterste geval tot een gerechtelijke vervolging overgaan. In zulke gevallen zal men natuurlijk voorzichtiger zijn en een hogere betrouwbaarheid eisen.

Vaste regels laten zich hier niet geven.

6.7. Een betrouwbaarheidsinterval voor het verschil van twee gemiddelden

Vaak beschikken we bij experimenteel werk over twee series waarnemingen onder enigzins verschillende omstandigheden genomen, zodat het redelijk is te onderstellen dat de standaardafwijking σ voor beide series hetzelfde is geweest maar dat mogelijk een verschil tussen de populatie-gemiddelden μ_1 en μ_2 kan zijn opgetreden. Men meet bijv. een aantal producten afkomstig van twee verschillende machines van hetzelfde type teneinde na te gaan in hoeverre deze gelijk zijn afgesteld. Of men laat twee analisten eenzelfde chemische analyse enige malen herhalen teneinde te onderzoeken of door mogelijke systematische verschillen in de uitvoering van de analyse ook systematische verschillen in de uitkomsten optreden.

Op grond van de t-verdeling kan ook in dit geval een betrouwbaarheidsinterval voor het verschil $\mu_1 - \mu_2$ worden aangegeven. Daartoe maken we gebruik van de volgende hulpstelling.

Hulpstelling : Zij

$$\underline{x}_1 = \mu_1 + \underline{u}_1 \sigma_1, \quad \text{en} \quad \underline{x}_2 = \mu_2 + \underline{u}_2 \sigma_2 \quad (6.7.1)$$

en \underline{x}_1 en \underline{x}_2 onderling onafhankelijk dan is

$$(\underline{x}_1 + \underline{x}_2) = \mu_1 + \mu_2 + \underline{u} \sqrt{\sigma_1^2 + \sigma_2^2}. \quad (6.7.2)$$

We zullen dit niet uitvoerig bewijzen.

Algemeener geldt : Is aan (6.7.1) voldaan en

$$\underline{y} = a_1 \underline{x}_1 + a_2 \underline{x}_2 \quad (6.7.3)$$

dan is

$$\underline{y} = (a_1 \mu_1 + a_2 \mu_2) + \underline{u} \sqrt{a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2}. \quad (6.7.4)$$

Deze formule laat zich gemakkelijk generalizeren tot de volgende.

Gegeven n onafhankelijke normaal verdeelde toevalsvariabelen

$$\left. \begin{array}{l} \underline{x}_i = \mu_i + \underline{u}_i \sigma_i, \quad i = 1, \dots, n \\ \text{en} \quad \underline{y} = \sum a_i \underline{x}_i, \\ \text{dan is} \quad \underline{y} = \mu + \underline{u} \sigma, \\ \text{met} \quad \mu = \sum a_i \mu_i \quad \sigma = \sqrt{\sum a_i^2 \sigma_i^2} \end{array} \right\} \quad (6.7.5)$$

Een lineaire functie van onderling onafhankelijke normaal verdeelde toevalsvariabelen is weer een normaal verdeelde variabele, waarvoor gemiddelde en standaardafwijking op eenvoudige wijze uit die der componenten kunnen worden berekend.

Keren we nu terug tot ons eigenlijke probleem.
Zijn

$$\begin{aligned} \underline{x}_{1i} &= \mu_1 + \underline{u}_i \sigma \quad , \quad i = 1, \dots, n_1 \\ \text{en} \quad \underline{x}_{2j} &= \mu_2 + \underline{u}_j \sigma \quad , \quad j = 1, \dots, n_2 \end{aligned} \tag{6.7.6}$$

de beide waarnemingsreeksen met resp. n_1 en n_2 waarnemingen, dan is op grond van de hulpstelling :

$$\begin{aligned} \bar{\underline{x}}_1 &= \mu_1 + \underline{u}_1 \sigma \sqrt{1/n_1} \\ \bar{\underline{x}}_2 &= \mu_2 + \underline{u}_2 \sigma \sqrt{1/n_2} \\ \text{en} \quad \bar{\underline{x}}_1 - \bar{\underline{x}}_2 &= (\mu_1 - \mu_2) + \underline{u} \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ \text{of} \quad \underline{u} &= \frac{\{(\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - (\mu_1 - \mu_2)\}}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \end{aligned} \tag{6.7.7}$$

Is σ bekend dan kunnen we met behulp van (6.7.7) aan de hand van de Tabel van de cumulatieve u-verdeling een betrouwbaarheidsinterval voor $\mu_1 - \mu_2$ aangeven. Is σ onbekend en alleen een schatting s van σ bekend dan ligt het voor de hand i.p.v. (6.7.7) te beschouwen de variabele

$$\begin{aligned} \underline{t} &= \frac{(\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - (\mu_1 - \mu_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \\ &= \frac{(\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} : \frac{s}{\sigma} \end{aligned} \tag{6.7.8}$$

Dit is weer het quotient van twee onafhankelijke variabelen, een teller met een \underline{u} -verdeling en een noemer met een χ_v / \sqrt{v} verdeling. De \underline{t} volgens (6.7.8) heeft dus dezelfde verdeling als de \underline{t} die we in de voorgaande paragraaf beschouwden. Dit leidt tot een betrouwbaarheidsinterval.

$$(\bar{x}_1 - \bar{x}_2) - t_s \sqrt{\frac{n_1 + n_2}{n_1 n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_s \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad (6.7.9)$$

met betrouwbaarheid $(1 - \alpha)$,
waarbij t in Tabel 2 wordt opgezocht.
Al naar behoefte kan men zich ook tot een éézijdig interval beperken.

Voorbeeld

Men heeft producten gemeten afkomstig van twee machines met de volgende uitkomsten.

Machine 1 :	5.31	5.34	5.30	5.31	mm			
2 :	5.33	5.34	5.37	5.40	5.34	5.32	5.35	mm

Uit deze gegevens vinden we

$$\left. \begin{array}{l} \bar{x}_1 = 5.315 \text{ mm} \quad n_1 = 4 \\ \bar{x}_2 = 5.350 \text{ mm} \quad n_2 = 7 \end{array} \right\} s = 0.024 \text{ mm} \quad ; \quad v = 9$$

en hiermee met Tabel 2

$$\bar{x}_1 - \bar{x}_2 = -0.035 \text{ mm},$$

$$-0.035 - 2.26 \times 0.024 \sqrt{\frac{11}{28}} < \mu_1 - \mu_2 < -0.035 + 2.26 \times 0.024 \sqrt{\frac{11}{28}},$$

of $-0.069 < \mu_1 - \mu_2 < -0.001 \text{ mm}$ met $\alpha = 5\%$. (6.7.10)

Men mag dus met vrij grote zekerheid besluiten dat de machines gemiddeld enigszins verschillen.
De opmerking aan het eind van § 6.6 gemaakt over het gebruik van betrouwbaarheidsintervallen gelden ook hier.

6.8. Het toetsen van hypothesen

Een andere techniek, nauw verwand aan de betrouwbaarheids intervallen, staat bekend als het toetsen van een hypothese. Men gaat hierbij uit van bepaalde onderstellingen en stelt een hypothese. Een voorbeeld kan dit het beste toelichten.

In het geval van par. 6.7 onderstelt men dat beide series waarnemingen aselechte steekproeven zijn uit normale populaties met eenzelfde standaardafwijking σ . De zg. nulhypothese die men wenst te toetsen is dan

$$H_0 : \quad = \mu_1 = \mu_2 \quad (6.8.1)$$

D.w.z. men wenst te toetsen of de beide gemiddelden gelijk zijn. Dit geschiedt op grond van de volgende overwegingen.

Is aan de onderstellingen en de nulhypothese voldaan dan zal de grootheid

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\underline{s}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (6.8.2)$$

een t-verdeling met ν vrijheidsgraden bezitten.

Is de numerieke waarde van t , die we uit de waarnemingen becijferen zó groot, dat een dergelijke waarde door de zuiver toevallige fluctuaties slechts zelden zou optreden, dan acht men de waarnemingen strijdig met de gestelde hypothese, en zal men deze hypothese verwerpen.

De beoordeling geschiedt op grond van de bij de gevonden waarde van t behorende overschrijdingskans. Is deze overschrijdingskans kleiner dan een vooraf vastgestelde kritieke drempelwaarde, de onbetrouwbaarheidsdrempel α , dan acht men waarnemingen en hypothese strijdig.

Bijv. voor het voorbeeld (6.7.10) vinden we

$$t = \frac{5.315 - 5.350}{0.024} \sqrt{\frac{28}{11}} = -2.325 ; \quad \nu = 9. \quad (6.8.3)$$

We zien nu uit Tabel 2 dat bij $\nu = 9$ de waarde $t = 2.26$ een tweezijdige overschrijdingskans van 5% bezit. De overschrijdingskans van de gevonden waarde $t = -2.325$ is dus kleiner dan 5% en men zal daarom de hypothese $\mu_1 = \mu_2$ bij een onbetrouwbaarheidsdrempel van 5% verwerpen.

Men onderscheidt één- en tweezijdige toetsen. Luidt de nul-hypothese als (6.8.1) dan zal men de hypothese strijdig met de waarnemingen verklaren zowel bij een grote positieve als bij een grote negatieve waarde van t . Soms echter stelt men als hypothese

$$H_0 : \quad \mu_1 \leq \mu_2 \quad (6.8.4)$$

Zo kan bijv. μ_2 betrekking hebben op een nieuw procédé en men is geïnteresseerd in de vraag of dit een verbetering geeft. Is het nieuwe procédé slechter dan het bestaande dan hebben we er geen belang bij. Stel nu dat $\mu_2 < \mu_1$ een verbetering betekent, dan stelt men juist de tegengestelde hypothese. Zijn de waarnemingen strijdig met deze hypothese dan acht men $\mu_2 < \mu_1$ bewezen. Bij de éénzijdige hypothese (6.8.4)

hoort dan een eenzijdige toets ; de waarnemingen zijn alleen strijdig met de hypothese wanneer t een grote positieve waarde bezit, niet bij een negatieve waarde.

Bij tweezijdig toetsen moet men uit Tabel 2 die waarden van t gebruiken die behoren bij een tweezijdige betrouwbaarheidsinterval ; bij eenzijdig toetsen de waarde die behoort bij een eenzijdig betrouwbaarheidsinterval. De onbetrouwbaarheidsdrempel is gelijk aan de bijbehorende onbetrouwbaarheid α .

In plaats van (6.8.1) kan men evengoed de hypothese

$$H_0 : \mu_1 - \mu_2 = b \quad (6.8.5)$$

toetsen waarbij b een te voren gegeven constante is. Heeft men een enkele reeks waarnemingen dan kan men evenzo de hypothesen

$$H_0 : \mu = b, \quad (6.8.6)$$

of
$$H_0 : \mu > b, \quad (6.8.7)$$

toetsen. Het principe is daarbij steeds hetzelfde. De hypothese (6.8.1) is de meest gangbare en deze is daarom als voorbeeld genomen.

Het zal uit het voorafgaande duidelijk zijn, dat het vaststellen van een betrouwbaarheidsinterval en het toetsen van hypothesen nauw verwante methoden zijn. Inderdaad omvat een betrouwbaarheidsinterval met onbetrouwbaarheid α al die waarden van b waarvoor de hypothesen (6.8.5) of (6.8.6) niet strijdig zijn met de waarnemingen, bij toetsen met een onbetrouwbaarheidsdrempel α , terwijl voor alle waarden van b buiten het interval strijdigheid zal worden geconstateerd. Het toetsen van een hypothese komt dus eenvoudig neer op het vaststellen of een gegeven waarde voor een gemiddelde μ , of voor het verschil $\mu_1 - \mu_2$ van twee gemiddelden in of buiten een bepaald betrouwbaarheidsinterval ligt.

Daar waar methoden voor het construeren van een betrouwbaarheidsinterval bestaan kan men daarom hiermee volstaan. Er zijn echter ook gevallen waar het toetsen van een hypothese wel mogelijk is, maar het aangeven van een betrouwbaarheidsinterval niet, of althans moeilijk uitvoerbaar is. Daarom speelt het toetsen van hypothesen in de toegepaste statistiek een rol.

Zijn de waarnemingen strijdig met de gestelde hypothesen, dan drukt men dit vaak zo uit, dat men de hypothese verwerpt. Dit is echter een weinig zeggende uitdrukking die men m.i. beter kan vermijden. Strijdigheid ontstaat niet altijd noodzakelijkerwijs omdat aan de hypothese niet is voldaan. Zij kan ook daaruit voortspruiten dat aan de onderstellingen die aan een toets ten grondslag liggen niet is voldaan ; dat bijv. de waarnemingen geen aselechte steekproef vormen, of de populatie geen normale verdeling bezit. Meestal past men de toetsen echter toe onder omstandigheden waarbij men op grond van de experimentele omstandigheden in de eerste plaats kan verwachten, dat eventueel niet aan de nulhypothese zal zijn voldaan, en waar men afwijkingen van de onderstellingen minder behoeft te vrezen. Ook zijn goede toetsen meestal veel minder gevoelig voor afwijkingen in de onderstellingen, dan voor een afwijking van de nul-hypothese. De t -toets bijv. is zg. robust, d.w.z. weinig gevoelig voor afwijkingen van de normale verdeling. Zij kan zonder bezwaar worden toegepast ook al wijken de verdelingen vrij sterk van de normale af.

6.9. De F-toets en de F-verdeling

De F-toets dient voor het vergelijken van twee varianties. Hier is het toetsen de meest gebruikelijke toepassing en daarmee zullen we dus beginnen.

Stel we ontvangen een partij kunstmest in balen en we willen hiervan het gehalte aan ammoniumsulfaat bepalen. In het algemeen zal dit gehalte niet volmaakt constant zijn en het is dus gewenst meerdere monsters te onderzoeken op verschillende plaatsen uit de partij genomen.

We verrichten nu twee series metingen :

a) Een serie van 10 monsters uit 10 verschillende a-select aangewezen balen met als resultaat

$$\bar{x}_1 = 7.32 \% , \quad s_1^2 = 41.8(\%)^2 , \quad v_1 = 9 .$$

b) Een serie analyses aan 7 monsters op verschillende plaatsen uit eenzelfde baal ; waaruit gevonden wordt

$$\bar{x}_2 = 7.27 \% , \quad s_2^2 = 24.1(\%)^2 , \quad v_2 = 6 .$$

Nu is het denkbaar dat tussen balen systematische verschillen optreden. Is dat het geval dan moeten we verwachten dat de variantie σ_1^2 groter zal zijn dan σ_2^2 . Het feit dat s_1^2 groter is dan s_2^2 wijst reeds in deze richting, maar het verschil kan ook door zuiver toevallige fluctuaties zijn veroorzaakt. We willen daarom toetsen de nulhypothese

$$H_0 \quad \sigma_1^2 = \sigma_2^2 = \sigma^2 . \quad (6.9.1)$$

Is deze hypothese juist dan geldt, wanneer we beide waarnemingsreeksen als aselechte steekproeven uit normale populaties opvatten

$$\frac{s_1^2}{\sigma^2} = \frac{\chi^2_{v_1}}{v_1} , \quad \frac{s_2^2}{\sigma^2} = \frac{\chi^2_{v_2}}{v_2} \quad (6.9.2)$$

en dus

$$\frac{s_1^2}{s_2^2} = \frac{v_2 \chi^2_{v_1}}{v_1 \chi^2_{v_2}} = \underline{F} \frac{v_1}{v_2} \quad (6.9.3)$$

waarbij $\chi^2_{v_1}$ en $\chi^2_{v_2}$ kennelijk onafhankelijk van elkaar zijn. De grootheid

\underline{F} is dus een toevalsvariabele met een verdeling die volledig door de aantallen vrijheidsgraden v_1 en v_2 is bepaald. De exacte formule voor de \underline{F} -verdeling laten we hier achterwege.

Waarden voor \underline{F} met een overschrijdingskans van 5% en van 1% zijn getabelleerd in Tabel 4 en het is deze Tabel die we bij de praktische toepassing steeds gebruiken.

In het beschouwde numerieke voorbeeld vinden we

$$\frac{s_1^2}{s_2^2} = \frac{41.8}{24.1} = 1.73 = F_6^9 \quad (6.9.4)$$

Uit Tabel 4 vinden we nu dat bij $v_1 = 9$, $v_2 = 6$ de waarde

$$F_6^9 = 4.10 \quad \text{een } 5\%$$

en $F_6^9 = 7.98 \quad \text{een } 1\% \text{ overschrijdingskans bezit.}$

De gevonden waarde $F_6^9 = 1.72$ is aanzienlijk geringer en kan dus gemakkelijk door zuiver toeval tot stand komen ook wanneer de hypothese (6.9.1) juist is. Er is dus geen enkel afdoend bewijs dat tussen balen werkelijk grotere verschillen optreden dan binnen balen. Er zijn bij toepassing van de F -toets een aantal punten waarop men goed dient acht te geven. In de notatie

$$F_{v_2}^{v_1}$$

betekent de bovenste waarde v_1 steeds het aantal vrijheidsgraden van de teller in (6.9.4) en v_2 steeds dat van de noemer. Voor $\alpha = 5\%$ is dus bijv.

$$F_6^9 = 4.10 \quad \text{en} \quad F_6^9 = 3.37.$$

De F -toets is oorspronkelijk ontworpen als een éézijdige toets. In het behandelde geval is nl. $\sigma_1^2 < \sigma_2^2$ uitgesloten; de fluctuaties tussen monsters uit verschillende balen kan nooit kleiner zijn dan tussen monsters uit één baal. Zouden we dus $s_1^2 < s_2^2$ vinden dan zou geen toets beschreven behoeven te worden uitgevoerd.

Dit is niet steeds het geval. Het had ook kunnen zijn dat de varianties s_1^2 en s_2^2 waren verkregen uit monsters genomen uit verschillende balen maar uit twee verschillende partijen kunstmest. Doel van het onderzoek is dan na te gaan of de inhomogeniteit van deze partijen duidelijk verschilt of niet.

Er is dan geen reden om $\sigma_1^2 < \sigma_2^2$ a priori uit te sluiten. In dit geval voert men de F -toets uit door steeds de grootste variantie in de teller te zetten. Is dus $s_1^2 > s_2^2$ dan vergelijkt men

$$\frac{s_1^2}{s_2^2} \quad \text{met} \quad F_{v_2}^{v_1} = F_6^9.$$

Is echter $s_1^2 < s_2^2$ dan vergelijkt men

$$\frac{s_2^2}{s_1^2} \quad \text{met} \quad F_{v_1}^{v_2} = F_9^6.$$

Dit is een tweezijdige toets en de overschrijdingskans of onbetrouwbaarheidsdrempel is nu twee maal zo groot als bij Tabel 4 is aangegeven. De F -tabel kan ook dienen om een betrouwbaarheidsinterval te geven voor de verhouding van twee onbekende varianten.

Stel $\sigma_1^2 \neq \sigma_2^2$ dan geldt toch

$$\frac{s_1^2}{\sigma_1^2} = \frac{\chi^2_{v_1}}{v_1}, \quad \frac{s_2^2}{\sigma_2^2} = \frac{\chi^2_{v_2}}{v_2}$$

en dus

$$\frac{s_1^2}{s_2^2} \frac{\sigma_2^2}{\sigma_1^2} = \frac{F_{v_1}}{F_{v_2}} \quad (6.9.5)$$

en

$$\frac{s_2^2}{s_1^2} \frac{\sigma_1^2}{\sigma_2^2} = \frac{F_{v_2}}{F_{v_1}} \quad (6.9.6)$$

Op grond van deze betrekkingen geldt

$$\frac{1}{F_{v_2}} \frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < F_{v_1} \frac{s_1^2}{s_2^2} \quad (6.9.7)$$

met 90% betrouwbaarheid wanneer we de F-waarden in Tabel 4 (pag.10) opzoeken en met 98% betrouwbaarheid wanneer we pag.11 gebruiken.

In het behandelde voorbeeld

$$\frac{1}{4.10} \frac{41.8}{24.1} < \frac{\sigma_1^2}{\sigma_2^2} < 3.37 \frac{41.8}{24.1}$$

of $0.42 < \frac{\sigma_1^2}{\sigma_2^2} < 5.83, \quad \text{met } \alpha = 10\%,$

of $0.65 < \frac{\sigma_1}{\sigma_2} < 2.42, \quad \text{met } \alpha = 10\%.$

Al bewijzen de waarnemingen niet afdoende dat $\sigma_1^2 > \sigma_2^2$ toch is het zeer wel mogelijk dat dit wel waar is en alleen niet sprekend tot uiting komt omdat het aantal waarnemingen te gering is.

HOOFDSTUK VII. VARIANTIE ANALYSE

7.1. Een eenvoudige situatie

Wanneer we als in par.6.9 de inhomogeniteit van een partij kunstmest zouden willen onderzoeken door de spreiding tussen monsters uit één baal genomen te vergelijken met de spreiding tussen monsters uit verschillende balen, zouden we de proef doorgaans niet zó uitvoeren als in par. 6.9 beschreven. Een veel betere proefopzet is als weergegeven in Tabel 7.1.

TABEL 7.1

Ammoniumsulfaatgehalten bepaald aan telkens 3 monsters uit 8 verschillende balen.

Codering : $X = (\text{Analyse in } \% - 7.00) \times 100.$

Monster :	j = 1	2	3	som	gemiddelde
Baal : i =	$x_{ij} =$			$x_{i\cdot}$	$\bar{x}_{i\cdot}$
1	30	41	37	108	36
2	25	30	29	84	28
3	26	37	24	87	29
4	40	43	31	114	38
5	30	35	37	102	34
6	25	24	32	81	27
7	38	48	40	126	42
8	26	30	34	90	30
				$x_{\cdot\cdot} = 792$	$\bar{x}_{\cdot\cdot} = 33$

Variantie analyse

$$1) A = \sum_i \sum_j x_{ij}^2 = 30^2 + 41^2 + \dots + 30^2 + 34^2 = 27146$$

$$2) B = \frac{\sum_i x_{i\cdot}^2}{n} = \frac{108^2 + 84^2 + \dots + 90^2}{3} = 26742$$

$$3) C = \frac{x_{\cdot\cdot}^2}{mn} = \frac{792^2}{24} = 26136$$

Resultaat

Variatiebron	Kwadr. Som	Vrijh. Gr.	Gem. Kw $10^{-4} (\%)^2$	Schatting van
Totaal	$27146 - 26136 = 1010$	23	86.6	$\sigma_0^2 + 3\sigma_1^2$
Tussenbalen	$26742 - 26136 = 606$	7		
Binnenbalen	$27146 - 26742 = 404$	16	25.2	σ_0^2

$$\sigma_0 = \sqrt{86.6 \times 10^{-4}} = 0.093\% \quad , \quad \hat{\sigma}_1 = \sqrt{\frac{86.6 - 25.2}{3}} \times 10^{-4} = 0.078\%$$

Tot nu toe hebben we een waarnemingsreeks steeds weergegeven door een formule van de vorm

$$\underline{x}_i = \mu + \underline{w}_i \sigma, \quad i = 1, \dots, n, \quad (7.1.1)$$

of

$$\underline{x}_i = \mu + \underline{u}_i \sigma, \quad i = 1, \dots, n.$$

Een dergelijke formule heet het Model. Het model van een waarnemingsreeks beschrijft het mechanisme waardoor we ons denken dat de waarnemingsreeks tot stand is gekomen. En het doel van een statistische analyse is de model parameters μ en σ , te schatten. De in (7.1.1) gegeven modellen zijn daarbij de eenvoudigst denkbare modellen die van toepassing zijn op enkelvoudige waarnemingsreeksen.

In Tabel 7.1 hebben we nu een serie waarnemingen van een gecompliceerder structuur die door de in (7.1.1) gegeven modellen niet bevredigend kan worden weergegeven. We moeten nog een derde element in het model opnemen en schrijven

$$\underline{x}_{ij} = \mu + \underline{w}_i \sigma_1 + \underline{w}_{ij} \sigma_0, \quad i = 1, \dots, m \quad (7.1.2)$$

of

$$\underline{x}_{ij} = \mu + \underline{u}_i \sigma_1 + \underline{u}_{ij} \sigma_0, \quad j = 1, \dots, n \quad (7.1.3)$$

Hierin is μ het gemiddeld gehalte voor de gehele partij. Daarnaast kunnen we een gemiddeld gehalte per baal onderscheiden en deze baalgemiddelden kunnen door toevallige fluctuaties van het algemeen gemiddelde μ afwijken. Deze fluctuaties worden weergegeven door de term $\underline{w}_i \sigma_1$ (of $\underline{u}_i \sigma_1$) waarbij $i = 1, \dots, m$ de m balen aangeven waaruit monsters zijn genomen. Wanneer we deze fluctuaties tussen balen als toevallige fluctuaties opvatten houdt dat in, dat de balen, waaruit monsters worden genomen aselekt worden aangewezen. Vatten we de verdeling tussen balen als een normale verdeling op ($\underline{u}_i \sigma_1$) dan houdt dit in dat we in principe een oneindige populatie van balen beschreven. In de praktijk wil dit zeggen dat het aantal balen dat we in het onderzoek betrekken slecht een klein gedeelte is van de gehele partij.

Tenslotte denken we ons ook nog enige inhomogeniteit binnen balen en deze worden in het model weergegeven door de term $\underline{w}_{ij} \sigma_0$ (of $\underline{u}_{ij} \sigma_0$)

met $j = 1, \dots, n$; n is daarbij het aantal monsters dat uit een baal genomen wordt, In Tabel 7.1 is $m = 8$ en $n = 3$.

De modellen (7.1.2) en (7.1.3) bevatten nu 3 parameters μ , σ_1 en σ_0 , en ons volgende probleem is uit de waarnemingen in Tabel 7.1 die drie parameters te schatten.

Wat σ_0 betreft is dit niet moeilijk. We beschikken in Tabel 7.1 over 8 series van 3 waarnemingen. Uit ieder van deze kan een schatting s^2 van σ_0^2 met twee vrijheidsgraden worden berekend en door deze samen te voegen krijgen we een eindwaarde met 16 vrijheidsgraden. Algemeen vinden we uit één serie de kwadratensom.

$$K_{S_i} = \sum_j x_{ij}^2 - \frac{x_{i.}^2}{n} \quad (7.1.4)$$

wanneer we met $x_{i.}$ de som

$$x_{i.} = \sum_j x_{ij} \quad (7.1.5)$$

weergegeven. Door samenvoegen krijgen we de totale kwadratensom

$$K_{S_0} = \sum_i \sum_j x_{ij}^2 - \frac{\sum_i x_{i.}^2}{n}; \quad v = m(n-1). \quad (7.1.6)$$

Een schatting van de variantie σ_1^2 tussen balen verkrijgen we uit de variantie tussen balen verkrijgen we uit de variantie tussen de sommen $x_{i.}$ of de gemiddelden $\bar{x}_{i.} = x_{i.}/n$. Voor de sommen geldt op grond van (7.1.2)

$$\underline{x}_i = n\mu + n\underline{w}_i\sigma_1 + \sigma_0 \sum_j \underline{w}_{ij} \quad (7.1.7)$$

en hieruit volgt bij toepassing van (4.2.9 ; pag.50)

$$\text{var}(\underline{x}_i) = n^2\sigma_1^2 + n\sigma_0^2 \quad (7.1.8)$$

en dan voor het gemiddelde

$$\text{var}(\bar{x}_i) = \sigma_1^2 + \frac{\sigma_0^2}{n}. \quad (7.1.9)$$

De kwadratensom corresponderend met $\text{var}(\underline{x}_i)$ is

$$KS = \sum_i x_{i.}^2 - \frac{x_{..}^2}{m}, \quad v = m-1, \quad (7.1.10)$$

waarin

$$x_{..} = \sum_i x_{i.} = \sum_i \sum_j x_{ij} \quad (7.1.11)$$

de som van alle waarnemingen weergeeft.

In de variantie analyse berekenen we echter niet een schatting van $\text{var}(\underline{x}_i)$ of $\text{var}(\bar{x}_i)$ doch van

$$n \text{var}(\bar{x}_i) = \frac{\text{var } \underline{x}_i}{n} = n\sigma_1^2 + \sigma_0^2. \quad (7.1.12)$$

De hiermee corresponderende kwadratensom is

$$K_{S_1} = \frac{\sum_i x_{i.}^2}{n} - \frac{x_{..}^2}{mn}, \quad v = m-1. \quad (7.1.13)$$

De variantie analyse van Tabel 7.1 verloopt nu volgens het volgende recept.

Berekening van de variantie analyse van een Tabel als 7.1 met m rijen van n waarnemingen

1. Bereken de sommen $x_{i.}$ en hieruit $x_{..}$.
2. Bereken achtereenvolgens

$$A = \sum_i \sum_j x_{ij}^2,$$

$$B = \frac{\sum_i x_{i.}^2}{n}$$

$$C = \frac{x_{..}^2}{mn}$$

3. Dan luidt de variantie analyse

Tabel 7.2.

Variantie analyse van m series van waarnemingen.

Variatiebron	Kwadratensom	Vrijheidsgraden	Gemiddeld kwadraat	Schatting van
Totaal	KS = A - C	$v = mn - 1$		$n\sigma_1^2 + \sigma_0^2$
Tussenrijen	KS ₁ = B - C	$v_1 = m - 1$	KS ₁ /v ₁	
Binnenrijen	KS ₀ = A - B	$v_0 = m(n - 1)$	KS ₀ /v ₀	σ_0^2

Numeriek is deze bewerking in Tabel 7.1 reeds uitgevoerd. Hierboven hebben we de variatiebronnen neutraal met de benamingen "tussen rijen" en "binnen rijen" weergegeven. In een daadwerkelijk uitgevoerde analyse is het beter te verwijzen naar de effecten die we bestuderen, zoals in Tabel 7.1 waar we een effect "tussen balen" en "binnen balen" onderscheiden.

Uit de gemiddelde kwadraten vinden we dan de schattingen $\hat{\sigma}_0$ en $\hat{\sigma}_1$ van σ_0 en σ_1 als in Tabel 7.1 berekend.

7.2. Variantie-analyse (vervolg)

We kunnen Tabel 7.1 en de bijbehorende analyse ook nog op een andere manier beschouwen. De analyse berust op een splitsing van de waarnemingen in een som van 3 componenten met een bijpassende splitsing van het aantal vrijheidsgraden :

$$x_{ij} = \bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - x_{i.}), \quad (7.2.1)$$

$$v = mn = 1 + (m-1) + m(n-1) \quad (7.2.2)$$

De waarnemingen uit Tabel 7.1 zijn hieronder in Tabel 7.2 nog eens weergegeven met de erbijbehorende splitsing in componenten.

TABEL 7.2.

Splitsing in componenten van de waarnemingen van Tabel 7.1.

		Originele waarn			Splitsing in 3 componenten					
j =		1	2	3						
i =		x_{ij}			$\bar{x}_{i.}$	$\bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.})$				
1		30	41	37	36		+3	-6	+5	+1
2		25	30	29	28		-5	-3	+2	+1
3		26	37	24	29	33	-4	-3	+8	-5
4		40	43	31	38		+5	+2	+5	-7
5		30	35	37	34		+1	-4	+1	+3
6		25	24	32	27		-6	-2	-3	+5
7		38	48	40	42		+9	-4	+6	-2
8		26	30	34	30		-3	-4	0	+4

De component $\bar{x}_{..}$ bezit 1, de component $(\bar{x}_{i.} - \bar{x}_{..})$ (m-1) en de component $(x_{ij} - \bar{x}_{i.})$, $m(n-1)$ graden van vrijheid. Deze splitsing komt overeen met de structuur van het model.

De 3 componenten zijn verder orthogonaal, d.w.z. de som van hun onderlinge producten is nul, indien over de gehele tabel wordt gesommeerd : dus

$$\sum_i \sum_j \bar{x}_{..} (\bar{x}_{i.} - \bar{x}_{..}) = 0$$

$$\sum_i \sum_j \bar{x}_{i.} (\bar{x}_{ij} - \bar{x}_{i.}) = 0 \tag{7.2.3}$$

en

$$\sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..}) (x_{ij} - \bar{x}_{i.}) = 0.$$

Ten gevolge van deze orthogonaliteit geldt bijv.

$$\sum_i \sum_j x_{ij}^2 = \sum_i \sum_j \bar{x}_{..}^2 + \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2, \tag{7.2.4}$$

en evenzo

$$\sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 \quad (7.2.5)$$

De variantie analyse in Tabel 7.1 of 7.2 is nu in wezen gebaseerd op deze laatste vergelijking. De totale kwadraten som

$$\sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = 1010 \quad \text{met} \quad v = 23$$

is nl. gesplitst in de kwadratensommen voor twee orthogonale componenten:

$$\text{tussen balen} = KS_1 = \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})^2 = 606, \quad \text{met} \quad v = 7,$$

$$\text{en binnen balen} = KS_0 = \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 = 404, \quad \text{met} \quad v = 16.$$

De kwadratensom KS_0 wordt uitsluitend bepaald door de verschillen tussen monsters binnen balen, en verschillen tussen de balen onderling hebben alleen invloed op de kwadratensom KS_1 . Deze laatste bevat ook een bijdrage van de verschillen tussen monsters binnen balen.

7.3. Toepassing van de F-toets

Een bij de variantie veel gebruikte toets is de F.toets. Haar toepassing berust op de volgende algemene stelling, die we hier niet zullen bewijzen :

Stelling ; Gegeven :

$$\underline{x}_i = \mu + \underline{u}_i \sigma, \quad i = 1, \dots, N.$$

Indien we deze waarnemingen in een aantal, k , orthogonale componenten splitsen met kwadratensommen KS_1, \dots, KS_k en bijbehorende aantallen vrijheidsgraden v_1, \dots, v_k , zodat dus geldt

$$KS = \sum_i (x_i - \bar{x})^2 = KS_1 + KS_2 + \dots + KS_k,$$

$$v = N-1 = v_1 + v_2 + \dots + v_k,$$

dan zijn de k gemiddelde kwadraten

$$GK_j = s_j^2 = \frac{KS_j}{v_j}, \quad j = 1, \dots, k$$

onderling onafhankelijke schattingen van de variantie σ^2 .

Op verhoudingen tussen deze gemiddelde kwadraten is dan de F-verdeling van toepassing.

De variantie analyse voor tabel 7.1 luidde

Variatiebron	KS	v	GK	Schatting van
Tussenbalen	606	7	86.6	$\sigma_0^2 + 3\sigma_1^2$
binnenbalen	404	16	25.2	σ_0^2

Deze analyse is van toepassing op het algemene model

$$\underline{x}_{ij} = \mu + \underline{w}_i \sigma_1 + \underline{w}_{ij} \sigma_0 \quad (7.2.1)$$

Specialiseren we het model tot

$$\underline{x}_{ij} = \mu + \underline{w}_i \sigma_1 + \underline{u}_{ij} \sigma_0 \quad (7.2.2)$$

d.w.z. onderstellen we dat de verschillen tussen monsters binnen balen een normale verdeling bezitten en stellen we als nul-hypothese

$$H_0 : \sigma_1 = 0$$

dan geldt onder deze nul-hypothese het model

$$\underline{x}_{ij} = \mu + \underline{u}_{ij} \sigma_0,$$

en zijn volgens genoemde stelling de kwadraten sommen 86,6 en 25,2 onafhankelijke schattingen van σ_0^2 . Passen we nu de F-toets toe dan is

$$F_{46}^7 = \frac{86.6}{25.2} = 3.44 ;$$

terwijl voor

$$\alpha = 0.05 : F_{16}^7 = 2.66 \quad \text{en voor } \alpha = 0.01 : F_{16}^7 = 4.20.$$

De overschrijdingskans van de waargenomen F-waarde ligt dys tussen de 5% en 1% in en er is een sterke aanwijzing dat de waarnemingen strijdig zijn met de hypothese ; dus dat $\sigma_1^2 > 0$ en dat tussen balen inderdaad systematische verschillen bestaan. Voor dit soort toepassingen werd de F-toets aanvankelijk ontwikkeld ; vandaar dat de F-toets als een éézijdige toets wordt Al te veel betekenis moet men overigens aan de F-toets niet toekennen. Wanneer men bv. uit vroegere ervaring reeds weet, dat tussen balen verschillen optreden dan zal men dat ook aannemen, al is de F-waarde die een serie waarnemingen oplevert niet significant. In andere gevallen kan men een significante F-waarde wel eens negeren, indien het effect dat zij zou bewijzen, uit hoofde van onze bestaande kennis in hoge mate onaannemelijk is. Hoe we de F-toets hanteren en welke conclusies we uit de variantie analyse trekken hangt in de practijk mede, en in sterke mate, af van de kennis van het probleem waarmee we te maken hebben.

Is $\sigma_1 > 0$ en met het model

$$\underline{x}_{ij} = \mu + \underline{u}_i \sigma_1 + \underline{u}_{ij} \sigma_0$$

kunnen we voor de verhouding σ_1 / σ_0 een betrouwbaarheidsinterval aangeven. De gemiddelde kwadraten zijn nu onafhankelijke schattingen van σ_0^2 en $\sigma_0^2 + 3\sigma_1^2$ en met de methode van par. 6.9. vinden we

$$\frac{1}{2.66} \frac{86.6}{25.2} < \frac{\sigma_0^2 + 3\sigma_1^2}{\sigma_0^2} < 3.50 \frac{86.6}{25.2}, \quad \alpha = 10\%$$

of

$$1.29 < \frac{\sigma_0^2 + 3\sigma_1^2}{\sigma_0^2} < 12.04, \quad \alpha = 10\%$$

of

$$0.31 < \frac{\sigma_0}{\sigma_0} < 1.92, \quad \alpha = 10\%$$

Vraagstukken.

7.3.1. Militairen behalen bij schietoefeningen een maand na elkaar de volgende scores:

1^e proef: 50 41 57 55 33 63 54 56

2^e proef: 53 55 58 58 42 56 67 68

De scores staan in dezelfde volgorde.

Vragen:

- a) Construeer een 95% betrouwbaarheidsinterval voor het verschil in gemiddelde score tussen de 1e en 2e oefening, aannemend dat beide series onafhankelijk zijn.
- b) Construeer een 95% betrouwbaarheidsinterval voor het gemiddeld verschil tussen de 2 oefeningen, door eerst voor ieder militair het verschil te berekenen en deze verschillen te behandelen als één enkelvoudige serie waarnemingen.
- c) Het tweede interval is kleiner dan het eerste. Waarom? Welke van de twee is het juiste?
- d) Toets de hypothese dat er geen vooruitgang in gemiddelde score is door middel van een variantie analyse.
- e) Is er een verband tussen de F-toets in d) en een van de betrouwbaarheidsintervallen uit a) en b)?
- f) We kunnen onderscheiden tussen een toevallige fout in de scores (σ_0^2) en toevallige fluctuaties tussen personen (σ_1^2). Geef een schatting van deze variantie?
- g) Construeer een betrouwbaarheidsinterval voor σ_0^2 en voor de verhouding σ_1^2 / σ_0^2 .

7.3.2. Bij de bepaling van de p_H van grondsoorten weegt men 10 gram van het bodemmateriaal af, schudt dit enige tijd met een zekere hoeveelheid gedestilleerd water, en meet dan de p_H . Ten einde na te gaan of de hoeveelheid water invloed heeft is de volgende proef genomen:
Men heeft zes monsters van 10 gram afgewogen (M_1, \dots, M_6) en hiervan 3 bewerkt met 20 ml en 3 met 30 ml water. De splitsing in twee groepen van 3 geschiedde aselekt. Iedere p_H meting werd 1 x herhaald. De resultaten waren

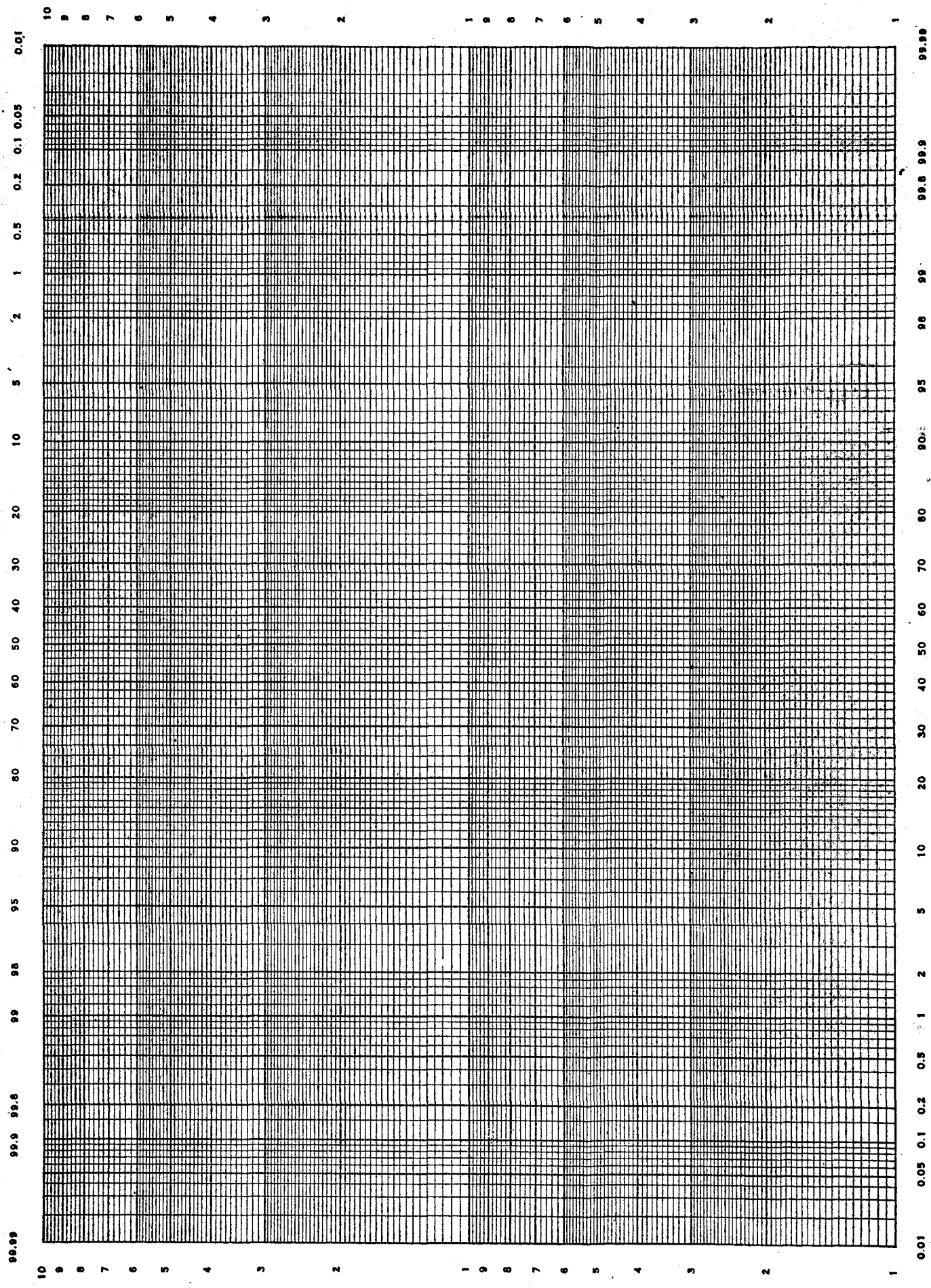
20 ml water			30 ml water		
Monster	1 ^e meting	2 ^e meting	Monster	1 ^e meting	2 ^e meting
2	5,45	5,42	4	5,54	5,54
5	5,32	5,39	6	5,47	5,48
1	5,37	5,42	3	5,41	5,45

De variantie-analyse voor deze gegevens luidt:

Variatiebron	$\nu =$
Tussen 20 en 30 ml	1
Tussen monster bij eenzelfde hoeveelheid water	4
Binnenmonsters	6

- Voer de variantie-analyse volledig uit.
- Welk model is hier van toepassing?
- Geef een schatting van de parameters die in het model voorkomen.
- Toets de hypothese dat de hoeveelheid water geen invloed heeft.
- Geef een betrouwbaarheidsinterval voor het gemiddelde verschil in p_H bij gebruik van 20 en 30 ml water.

Fig.3.7.2. Een voorbeeld van logaritmisch waarschiijnlijkheidspapier



ERRATA

Pag.88 : De onderste regel van Tabel 7.1 moet luiden

$$\hat{\sigma}_0 = \sqrt{25.2 \times 10^{-4}} = 0.050 \%, \quad \hat{\sigma}_1 = \sqrt{\frac{86.6 - 25.2}{3} \times 10^{-4}} = 0.045 \%$$

Pag.94 : De beide modellen op deze pagina moeten worden genummerd (7.3.1) en (7.3.2) i.p.v. (7.2.1) en (7.2.2)

Pag.95 : Laatste regel $\frac{\sigma_0}{\sigma_0}$ moet zijn $\frac{\sigma_1}{\sigma_0}$.

7.4. Een betrouwbaarheidsinterval voor het gemiddelde

Een ander probleem is het vaststellen van een betrouwbaarheidsinterval voor het gemiddelde ammoniumsulfaat gehalte voor de gehele partij. Dit kan belangrijk zijn in verband met het vaststellen of het product aan wettelijk gestelde eisen voldoet. In andere gevallen kan de prijs van de partij hierdoor worden bepaald.

We moeten in dit kader nog een verdere uitbreiding geven aan de modellen die we kunnen toepassen. In Tabel 7.1. onderzochten we monsters genomen uit 8 balen en pasten het algemene model

$$\underline{x}_{ij} = \mu + \underline{w}_i \sigma_1 + \underline{w}_{ij} \sigma_0 \quad (7.4.1)$$

hierop toe. Dit onderstelt dat we de 8 balen opvatten als een aselechte steekproef uit een grotere partij. Het is echter ook denkbaar, dat de 8 balen samen de gehele partij vormen. In dat geval zijn de variaties tussen balen geen toevalsvariabelen en schrijven we het model.

$$\underline{x}_{ij} = \mu + a_i + \underline{w}_{ij} \sigma_0, \quad (7.4.2)$$

met de nevenvoorwaarde

$$\sum a_i = 0. \quad (7.4.3)$$

Statistische uitspraken zijn waarschijnlijkheidsuitspraken ; zij hebben dus betrekking op experimenten die in principe onbeperkt kunnen worden herhaald. De vorm waarin we het model schrijven hangt nauw samen met wat we als herhaling van een experiment zien.

De onderstreepte symbolen geven die componenten die bij herhaling van het experiment aan toevallige fluctuaties onderhevig zijn ; de niet onderstreepte symbolen geven constanten die bij herhaling van het experiment niet aan variaties onderhevig zijn.

Onder model (7.4.2) treden alleen de fluctuaties tussen monsters binnen balen als toevallige fluctuaties op. Het gemiddelde $\bar{x}_{..}$ omvat 24 waarnemingen en de variantie van het gemiddelde is dus

$$\sigma^2(\bar{x}_{..}) = \frac{\sigma_0^2}{24} .$$

Een schatting van deze variantie vinden we uit het G.K. binnen balen, nl.

$$s^2(\bar{x}_{..}) = \frac{25.2}{24} \times 10^{-4} = 1.05 \times 10^{-4} (\%)^2 ,$$

waaruit $s(\bar{x}_{..}) = 0.0103 \%$ met $v = 16$.
Hieruit volgt

of $7.33 - 2.12 \times 0.0103 < \mu < 7.33 + 2.12 \times 0.0103$,

$7.306 < \mu < 7.354 \%$; $\alpha = 5 \%$.	(7.4.4)
--	---------

Onder model (7.4.1) is $\bar{x}_{..}$ onderhevig aan fluctuaties binnen balen en tussen balen. We hebben nu

$$\sigma^2(\bar{x}_{..}) = \frac{\sigma_0^2}{24} + \frac{\sigma_1^2}{8} ,$$

Een schatting hiervan vinden we uit het G.K. tussen balen

$$s^2(\bar{x}_{..}) = \frac{86.6}{24} \times 10^{-4} = 3.61 \times 10^{-4} (\%)^2$$

waaruit $s(\bar{x}_{..}) = 0.019 \%$, met $v = 7$.
Hieruit volgt

of $7.33 - 2.36 \times 0.019 < \mu < 7.33 + 2.36 \times 0.019$

$7.285 < \mu < 7.375 \%$; $\alpha = 5 \%$	(7.4.5)
--	---------

Het interval is nu breder wegens grotere toevallige fluctuaties en minder vrijheidsgraden

Feitelijk onderstellen de betrouwbaarheidsintervallen (7.4.4) en (7.4.5) normale verdelingen, doch ook hier is de centrale limietstelling van toepassing. Bij een enigszins omvangrijk waarnemingsmateriaal komt de normaliteit er niet zo erg op aan.

In principe onderstelt het interval (7.4.5) verder dat de partij die we bemonsteren uit een oneindig aantal balen bestaat. Is dit niet het geval dan moeten we onze theorie enigszins herzien.

Stel we hebben een populatie bestaande uit een eindig aantal N elementen waaraan we een eigenschap x_i , $i = 1, \dots, N$ kunnen meten. Dan definiëren we gemiddelde en variantie van deze populatie

$$\mu = \frac{\sum_{i=1}^N x_i}{N}, \quad (7.4.6)$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N - 1}. \quad (7.4.7)$$

Nemen we hieruit een aselechte steekproef zonder teruglegging van n stuks dan geldt voor het gemiddelde \bar{x}

$$\sigma^2(\bar{x}) = \sigma^2 \left(\frac{1}{n} - \frac{1}{N} \right) = \frac{\sigma^2}{n} \left(1 - \frac{n}{N} \right). \quad (7.4.8)$$

Voor $N \rightarrow \infty$ gaat dit over in $\sigma^2(\bar{x}) = \sigma^2/n$ een formule die reeds eerder werd gegeven. Voor $n = N$ wordt $\sigma^2(\bar{x}) = 0$; dit is begrijpelijk want dan nemen we de gehele populatie waar. De algemene formule (7.4.8) kunnen we hier niet bewijzen, maar wel toepassen.

Hebben we met een partij van N balen kunstmest te maken en is σ_1^2 de variantie tussen balen, dan zal wanneer we uit m balen n monsters nemen de variantie van het gemiddelde van de mn waarnemingen luiden

$$\sigma^2(\bar{x}_{..}) = \frac{\sigma_1^2}{m} \left(1 - \frac{m}{N} \right) + \frac{\sigma_0^2}{mn}. \quad (7.4.9)$$

We kunnen in deze formule de schattingen $\hat{\sigma}_1$ en $\hat{\sigma}_0$ invoeren, zoals die uit de variantie-analyse in Tabel 7.1 werden berekend, en krijgen dan een schatting van de variantie van het gemiddelde. Een moeilijkheid is echter dat de schatting $\hat{\sigma}_1^2$ wordt verkregen uit het verschil van twee G.K. en dat het dientengevolge niet mogelijk is aan $\hat{\sigma}_1^2$ en aan de schatting volgens (7.4.9) een bepaald aantal vrijheidsgraden toe te kennen

Voor $m = N$ valt σ_1^2 in (7.4.9) weg en kunnen we de schatting van $GK_0 = \hat{\sigma}_0^2$ met $m(n-1)$ vrijheidsgraden gebruiken. Voor $N = \infty$ kunnen we daarentegen de schatting van $GK_1 = \sigma_0^2 + m\sigma_1^2$ met $(m-1)$ vrijheidsgraden gebruiken, doch voor tussenliggende gevallen bestaat geen exacte theorie. Wanneer we een betrouwbaarheidsinterval baseren op $v = (m-1)$ zijn we vermoedelijk aan de veilige kant. Passen we formule (7.4.9) toe in het geval van de balen kunstmest (Tabel 7.1) voor een partijgrootte $N = 20$ dan is

$$\hat{\sigma}^2(\bar{x}_{..}) = \frac{(0.045)^2}{8} \left(1 - \frac{8}{20}\right) + \frac{0.050^2}{24} = 0.000258,$$

of $\hat{\sigma}(\bar{x}_{..}) = 0.016 \%$

7.5. Een variantie-analyse in twee richtingen

We beschouwen nog eenmaal hetzelfde waarnemingsmateriaal als in Tabel 7.1. Uit 8 balen kunstmest zijn telkens 3 monsters onderzocht. We denken ons nu echter de proef zó gewijzigd dat de monsters 1, 2, en 3 zijn onderzocht door 3 verschillende analisten (of in 3 verschillende laboratoria). Het is dan denkbaar dat niet alleen tussen de rijen of balen onderlinge verschillen optreden, maar evenzeer tussen de kolommen of analisten bv. door een verschil in tempo van werken, door verschil in de gebruikte apparatuur; en dergelijke.

We moeten dan nog een verdere term in ons model opnemen en schrijven

$$\underline{x}_{ij} = \mu + \underline{w}_i \sigma_1 + \underline{w}_j \sigma_2 + \underline{w}_{ij} \sigma_0 \quad (7.5.1)$$

indien we zowel balen als analisten opvatten als aselechte steekproeven uit een populatie van balen en van analisten respectievelijk. Alternatieve modellen zijn

$$\underline{x}_{ij} = \mu + a_i + \underline{w}_j \sigma_2 + \underline{w}_{ij} \sigma_0, \quad \Sigma a_i = 0 \quad (7.5.2)$$

$$\underline{x}_{ij} = \mu + \underline{w}_i \sigma_1 + b_j + \underline{w}_{ij} \sigma_0, \quad \Sigma b_j = 0, \quad (7.5.3)$$

en

$$\underline{x}_{ij} = \mu + a_i + b_j + \underline{w}_{ij} \sigma_0, \quad \Sigma a_i = 0, \quad \Sigma b_j = 0 \quad (7.5.4)$$

steeds met $i = 1, \dots, m, \quad j = 1, \dots, n.$

Onderstellen we normale verdelingen dan moeten we in deze modellen de w 's vervangen door u 's. In de modellen (7.5.2), (7.5.3) en (7.5.4) vatten we de 8 balen, de 3 analisten, of beide balen en analisten op als de volledige populatie(s) waarvoor we ons interesseren. In aansluiting op de modellen (7.5.1)-(7.5.4) splitsen we de waarnemingen nu in 4 orthogonale componenten:

$$\underline{x}_{ij} = \bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + (\bar{x}_{.j} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}), \quad (7.5.5)$$

$$v = mn = 1 + (m - 1) + (n - 1) + (m - 1)(n - 1),$$

De variantie analyse is gebaseerd op splitsing van de kwadratensom :

$$\begin{aligned} \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 &= \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_i \sum_j (\bar{x}_{.j} - \bar{x}_{..})^2 + \\ &\sum_i \sum_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2. \end{aligned} \quad (7.5.6)$$

De uitvoering van de analyse is geïllustreerd in Tabel (7.5.1). Het recept luidt nu:

Bereken :

$$A = \sum_i \sum_j x_{ij}^2,$$

$$B = \sum_i x_{i.}^2 / n,$$

$$C = \sum_j x_{.j}^2 / m,$$

$$D = x_{..}^2 / mn.$$

Dan luidt de variantie analyse

Variatiebron	K.S.	$\nu =$	GK	Schatting van
Totaal	$A - D = KS_t$	$(mn-1)$		
Tussen rijen(balen)	$B - D = KS_1$	$(m - 1)$	GK_1	$\sigma_0^2 + n\sigma_1^2$
Tussen kolommen(analysten)	$C - D = KS_2$	$(n - 1)$	GK_2	$\sigma_0^2 + m\sigma_2^2$
Residu	$(A-B-C+D) = KS_0$	$(m-1)(n-1)$	GK_0	σ_0^2

Numeriek is dit weergegeven in Tabel 7.5.1 en de overeenkomstige splitsing in orthogonale componenten in Tabel 7.5.2.

TABEL 7.5.1

Variantie analyse van Tabel 7.1. in twee richtingen, naar balen en analysten

j =	1	2	3		
	$x_{ij} =$			$x_{i.}$	$\bar{x}_{i.}$
i=1	30	41	37	106	36
2	25	30	29	84	28
3	26	37	24	87	29
4	40	43	31	114	38
5	30	35	37	102	34
6	25	24	32	81	27
7	38	48	40	126	42
8	26	30	34	90	30
$x_{.j} =$	240	288	264	792 = $x_{..}$	
$\bar{x}_{.j}$	30	36	33		33 = $\bar{x}_{..}$
Berekening : $A = \sum_i \sum_j x_{ij}^2 = 27146$ $B = \sum_i x_{i.}^2 / 3 = 26742$ $C = \sum_j x_{.j}^2 / 8 = 26280$ $D = x_{..}^2 / 24 = 26136$					

Variantie Analyse

Variatiebron	KS	v	GK	Schatting van
Totaal	A - D = 1010	23		
Tussen balen	B - D = 606	7	86.6	$\sigma_0^2 + 3\sigma_1^2$
Tussen analysten	C - D = 144	2	72.0	$\sigma_0^2 + 8\sigma_2^2$
Residu	A B C D = 260	14	18.6	σ_0^2

TABEL 7.5.2

Splitsing van de gegevens van Tabel 7.1 in orthogonale componenten overeenkomende met de variantie analyse van Tabel 7.5.1.

$\bar{x}_{..}$	$\bar{x}_{i.} - \bar{x}_{..}$	$\bar{x}_{.j} - \bar{x}_{..}$	$x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}$		
33	- 3		-3	+2	+1
	- 5		0	-1	+1
	- 4		0	+5	-5
	+ 5	-3 +3 0	+5	+2	-7
	+ 1		-1	-2	+3
	- 6		+1	-6	+5
	+ 9		-1	+3	-2
	- 3		-1	-3	+4

Nemen we nu het model (7.5.1) dan geeft σ_1^2 de variantie tussen de balen onderling, en σ_2^2 de variantie tussen analysten. Daarenboven treden in alle waarnemingen toevallige fluctuaties op met variantie σ_0^2 . We hebben bij de variantie analyse de waarnemingen nu gesplitst in 3 orthogonale componenten zodanig dat de verschillen tussen balen, of tussen analysten, ieder slechts in één van de componenten tot uiting komt. Deze beide componenten bevatten ook een gedeelte van de toevallige fluctuaties. De derde component bevat uitsluitend de toevallige fluctuaties.

Onderstellen we dat de toevallige fluctuaties normaal verdeeld zijn dan kunnen we nu de hypothesen dat tussen balen of tussen analysten geen verschillen bestaan, dus de hypothesen $H_0 : \sigma_1 = 0$ of $H_0 : \sigma_2 = 0$, separaat toetsen. We hebben voor

$$H_0 : \sigma_1 = 0 \quad F_{14}^7 = \frac{86.6}{18.6} = 4.55,$$

en voor

$$H_0 : \sigma_2 = 0 \quad F_{14}^2 = \frac{72.0}{18.6} = 3.87.$$

Voor $\alpha = 0.05$ is $F_{14}^7 = 2,76$ en $F_{14}^2 = 3.74$.

De eerste hypothese is dus vrij duidelijk strijdig met de waarnemingen; er bestaan systematische verschillen tussen balen. De tweede hypothese is ook strijdig maar niet zo sterk. Er zijn dus aanwijzingen dat de analysten systematisch in hun uitkomsten met elkaar verschillen. Zoals Tabel 7.5.1 laat zien ligt de analyst 1 gemiddeld iets te laag, analyst 2 iets te hoog en analyst 3 juist goed. Op complicaties in de interpretatie van Tabel 7.5.1 wanneer we een van de modellen (7.5.2), (7.5.3) of (7.5.4) gebruiken, of wanneer we voor het totaal gemiddelde een betrouwbaarheidsinterval willen aangeven, gaan we verder niet in.

7.6. Variantie analyse

We zullen de aspecten van de variantie analyse nog eenmaal met een voorbeeld recapitulieren. Stel een groep van 8 militairen behalen bij twee schietoefeningen de volgende scores :

50 41 57 55 33 63 54 56
53 55 58 58 42 54 67 68

In principe kunnen we op deze gegevens op 3 verschillende wijzen de variantie analyse toepassen.

Analyse A

	KS	v	GK	schatting van
Totaal	1264	15		
Tussen Oefeningen	132	1	132	$\sigma_o^2 + 8\sigma_a^2$
Binnen Oefeningen	1132	14	80.8	σ_o^2

Analyse B

	KS	v	GK	schatting van
Totaal	1264	15		
Tussen Militairen	919	7	131	$\sigma_o^2 + 2\sigma_b^2$
Binnen Militairen	345	8	43	σ_o^2

Analyse C

	KS	v	GK	Schatting van
Totaal	1264			
Tussen Oefeningen	132	1	131	$\sigma_o^2 + 8\sigma_a^2$
Tussen Militairen	919	7	132	$\sigma_o^2 + 2\sigma_b^2$
Residu	213	7	30.5	σ_o^2

Geven we met de index $i = 1, 2$ de beide oefeningen weer en met $j = 1, \dots, 8$ de militairen, dan komen deze 3 analyses overeen met de volgende splitsingen in orthogonale componenten :

$$A : x_{ij} - \bar{x}_{..} = (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.}),$$

$$B : x_{ij} - \bar{x}_{..} = (\bar{x}_{.j} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{.j}),$$

$$C : x_{ij} - \bar{x}_{..} = (\bar{x}_{i.} - \bar{x}_{..}) + (\bar{x}_{.j} - \bar{x}_{..}) + (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})$$

Bij A zijn de 2e en 3e component uit het rechterlid van C tot één component samengevoegd, en bij B de 1e en 3e component. Daar dit orthogonale componenten zijn kunnen dan ook de kwadratensommen en bijbehorende aantallen vrijheidsgraden bij elkaar worden opgeteld.

Aan de hand van deze splitsing in componenten is het niet moeilijk de bijbehorende modellen op te schrijven :

$$A : \underline{x}_{ij} = \mu + a_i + u_{ij}\sigma \quad , \quad \sum_i a_i = 0$$

$$B : \underline{x}_{ij} = \mu + b_j + u_{ij}\sigma \quad , \quad \sum_j b_j = 0,$$

$$C : \underline{x}_{ij} = \mu + a_i + b_j + u_{ij}\sigma, \quad \sum_i a_i = 0, \quad \sum_j b_j = 0.$$

Hierin stellen a_i ($i = 1,2$) de verschillen tussen de rijen voor en b_j ($j = 1, \dots, 8$) de verschillen tussen de kolommen.

Model A en analyse A is van toepassing wanneer we beide rijen opvatten als twee onderling onafhankelijke waarnemingsreeksen. Dus bijv. indien de 8 militairen die aan de beide schietoefeningen deelnemen beide keren aselekt uit een grotere compagnie werden aangewezen ; het zijn dan twee verschillende steekproeven uit deze populatie ; of ook indien de 2 groepen van 8 militairen uit twee verschillende compagnieën afkomstig zijn en men de schietwaardigheid van deze compagnieën onderling wil vergelijken.

Past men de F-toets toe met

$$F'_{14} = \frac{132}{80.8} = 1.63$$

dan toetst men de hypothesen

$$H_0 : a_i = 0, \quad (i = 1,2)$$

of wel de hypothese dat er in werkelijkheid geen verschillen zijn tussen de eerste en tweede rij uitkomsten.

Dit kan men niet alleen met de F-toets toetsen maar ook door de t-toets. Het blijkt dan dat voor een tweezijdige t-toets geldt

$$F = t^2$$

zodat deze toetsen in feite identiek zijn.

De lezer kan dit zelf verifiëren.

Model en analyse B vatten de waarnemingen op als 8 paren, waarbij de verschillen binnen de paren van zuiver toevallige aard zijn maar tussen paren systematisch verschillen kunnen voorkomen. Deze analyse is bijv. van toepassing wanneer men de kwaliteit van 8 militairen wil vergelijken. Een enkele serie resultaten heeft dan geen waarde. Uit de bovenste rij getallen mag men bijv. niet besluiten dat no.5 die een score 33 haalt over het algemeen slechter schiet dan no.6 met een score 63 ; want men weet uit één stel metingen niet, hoe groot de toevallige fluctuaties zijn. Deze kan men pas vaststellen wanneer men door waarnemingen te herhalen en bij beschouwde analyse onderstelt dat de tweede waarneming een herhaling is van de eerste.

Met de F-toets

$$F_8^7 = \frac{131}{43} = 3.05$$

toetst men de hypothese

$$H_0 : b_j = 0, \quad (j = 1, \dots, 8)$$

dat er tussen de militairen geen verschillen bestaan. Hier toetst men alle b_j 's gelijktijdig en deze toets kan niet door een t-toets worden

vervangen. Want met een t-toets toetst men slechts 2 series tegen elkaar ; de F-toets toetst hier 8 series van waarnemingen tezamen.

Model en analyse C tenslotte gaan er van uit dat er zowel systematische verschillen tussen rijen of oefeningen als tussen kolommen of militairen kunnen optreden. Deze analyse is van toepassing indien dezelfde 8 militairen hebben deelgenomen aan twee schietoefeningen enige tijd na elkaar gehouden. We splitsen nu twee orthogonale componenten af, waarvan de ene de verschillen tussen oefeningen en de andere de verschillen tussen oefeningen en de andere de verschillen tussen militairen bevat. Of dergelijke verschillen aanwezig zijn kan nu worden getoetst met twee afzonderlijke F-toetsen, nl.

$$H_0 : a_i = 0 \quad \text{met} \quad F_7^i = \frac{131}{30.5} = 4.29 \quad \text{en}$$

$$H_0 : b_j = 0 \quad \text{met} \quad F_7^j = \frac{132}{30.5} = 4.33.$$

De eerste van deze F-toetsen kan ook met behulp van een t-toets worden uitgevoerd. Men berekent daartoe eerst de verschillen tussen de scores bij de eerste en de tweede oefening gehaald, te weten

3 14 1 3 9 -9 13 12

Door deze verschillen te nemen elimineren we de verschillen tussen militairen onderling. Men kan nu met een t-toets toetsen of het gemiddelde van deze 8 verschillen al dan niet van nul afwijkt. De tweezijdige t-toets geeft dan weer hetzelfde resultaat als de toets F_7^* hierboven vermeld. De lezer ga dit zelf na.

7.7. Enkele opmerkingen over de variantie analyse

De variantie-analyse is in de experimentele wetenschap een machtig hulpmiddel, omdat in de meeste technische metingen meerdere bronnen van fluctuaties tegelijk optreden en men deze door middel van de variantie-analyse kan scheiden en hun belangrijkheid bepalen. Enkele voorbeelden mogen dit toelichten.

Metten we een bepaalde volume vloeistof af met behulp van een pipet, dan kunnen we bv. de volgende bronnen van fluctuaties onderscheiden

- a) de fout van de pipetering,
- b) verschillen tussen pipetten onderling,
- c) verschillen tussen personen die de pipet hanteren.

Door nu met ieder van een aantal pipetten van eenzelfde type een paar maal een volume vloeistof af te meten en de afgemeten hoeveelheid door weging nauwkeurig te bepalen, en op deze gegevens variantie analyse toe te passen kunnen we de pipeteerfouten, of de variatie binnen pipetten scheiden van de systematische verschillen tussen pipetten onderling. Of door met éénzelfde pipet een aantal verschillende personen enige malen achtereenvolgend een volume vloeistof te laten afmeten kunnen we de invloed van de persoon die de pipet hanteert onderzoeken. Zo nodig kan men in een grotere proef een aantal pipetten, en een aantal herhalingen tezamen uitvoeren en op deze hieruit tegelijk de grootte orde van verschillen tussen pipetten en tussen personen ineens bepalen.

Een ander voorbeeld leveren diametermetingen aan een partij stalen kogels. Dat men aan iedere kogel slechts één meting dan kan men onmogelijk uitmaken of verschillen tussen kogels onderling te wijten zijn aan

- a) meetfouten
- b) onrondheid van de kogels
- c) verschillen in de gemiddelde diameters van de kogels.

Is a) het belangrijkste dan moet men eventueel de meetmethode verbeteren, is b) de hoofdoorzaak dan zal men de machine die de kogels hun ronde vorm geeft moeten verbeteren, en is c) de belangrijkste, dan moet men de hoeveelheid materiaal beter doseren. Het heeft daarom alle zin de verschillende mogelijke oorzaken van fluctuaties door doelmatig opgezette metingen te scheiden, want de technische maatregelen die men eventueel dient te nemen om het product te verbeteren hangen er sterk van af, welke bron van fluctuaties de belangrijkste is. Maar al te vaak wordt tegen deze leefregel gezondigd. Men meet een serie producten en schrijft de fluctuaties in metingen zonder meer aan het productieproces toe; dan volgen, soms kostbare, maatregelen om het proces te verbeteren, terwijl eerst later bij een doelmatiger onderzoek blijkt, dat niet het productieproces de fluctuaties veroorzaakt, maar dat deze aan onvolgende nauwkeurigheid van de meetmethode moeten worden toegeschreven. Men zij dus gewaarschuwd en op zijn hoede.

Aan de toepassing van de F-toets moet geen te grote waarde worden toegekend. Hij geeft een nuttige indicatie of bepaalde affecten al dan niet aanwezig zijn ; maar niet meer dan een indicatie en nooit een absoluut bewijs. Bij de beoordeling van de resultaten van een experiment is de F-toets niet alleen bepalend, doch zal men ook gebruik maken van zijn kennis betreffende het probleem dat moet worden onderzocht. Dit geldt trouwens algemeen bij de toepassing van statistische analyse methoden. Zij vormen een nuttig hulpmiddel bij de interpretatie van waarnemingsuitkomsten.

I N H O U D

		<u>BLZ.</u>
	INLEIDING	3 t/m 5
<u>TABEL 1</u>	NORMALE VERDELING	6 - 7
<u>TABEL 2</u>	t - VERDELING	8
<u>TABEL 3</u>	χ^2 - VERDELING	9
<u>TABEL 4</u>	F - VERDELING	10 - 11
<u>TABEL 5</u>	BETROUWBAARHEIDSINTERVALLEN VOOR DE STANDAARDAFWIJKING, σ , BIJ NORMALE POPULATIES	12
<u>TABEL 6</u>	HET BEPALEN VAN EEN ZUIVERE SCHATTING s^2 VAN DE VARIANTIE, σ^2 , UIT EEN GEMIDDELDE SPREIDINGSBREEDTE, \bar{R} , EN HET BENADEREN VAN DE VERDELING VAN \bar{R}^2 DOOR DE χ^2 - VERDELING	13
<u>TABEL 7</u>	VERDELING VAN DE STEEKPROEFCORRELATIECOEFFICIENT INDIEN $\rho = 0$	14
<u>TABEL 8</u>	BETROUWBAARHEIDSINTERVALLEN VOOR HET UNIVERSUMGEMIDDELDE, μ , VAN EEN EXPONENTIELE VERDELING	15
<u>TABEL 9</u>	BINOMIALE VERDELING	16 - 17
<u>TABEL 10</u>	BETROUWBAARHEIDSINTERVALLEN VOOR DE PARAMETER p VAN DE BINO- MIALE VERDELING MET GEMIDDELDE BETROUWBAARHEID	18 t/m 21
<u>TABEL 11</u>	POISSON - VERDELING	22
<u>TABEL 12</u>	BETROUWBAARHEIDSINTERVALLEN VOOR DE PARAMETER μ VAN DE POISSON- VERDELING MET GEMIDDELDE BETROUWBAARHEID	23
<u>TABEL 13</u>	GEBIEDEN WAARIN DE BINOMIALE VERDELING BENADERD KAN WORDEN DOOR DE NORMALE VERDELING OF DOOR DE POISSON - VERDELING BIJ GEGEVEN BENADERINGSNAUWKEURIGHEID	24
<u>TABEL 14</u>	VERDELINGSVRIJ BETROUWBAARHEIDSINTERVAL VOOR DE MEDIAAN; TEKENTOETS	25
<u>TABEL 15</u>	WILCOXON' S VERDELINGSVRIJE TOETS VOOR TWEE STEEKPROEVEN	26
<u>TABEL 16</u>	KENDALL' S VERDELINGSVRIJE CORRELATIE TOETS	27
<u>TABEL 17</u>	STEEKPROEFVOORSCHRIFTEN VOOR PARTIJKEURING	28
<u>TABEL 18</u>	HET BEREKENEN VAN KEURINGSKARAKTERISTIEKEN	29
<u>TABEL 19</u>	HET BEREKENEN VAN REGELGRENZEN VOOR EEN BEHEERST FABRIKAGEPROCES	30
<u>TABEL 20</u>	ASELECTE TREKKINGEN UIT EEN NORMALE (0, 1) VERDELING	31
<u>TABEL 21</u>	TOEVALSCIJFERS	32
<u>TABEL 22</u>	VERLOTINGSTABELLEN	33
<u>TABEL 23</u>	KWADRATEN, WORTELS, FACULTEITEN EN ANDERE FUNCTIES VAN DE GETALLEN 1 t/m 100	34
	VERANTWOORDING EN UITGEBREIDER TABELLEN	35 - 36
	10 BLADEN NORMAAL WAARSCHIJNLIJKHEIDSPAPIER	
	5 BLADEN LOGARITHMISCH PAPIER	

<p>ENGELSE VERTALING VAN ENKELE VEEL GEBRUIKTE WOORDEN EN BEGRIPPEN.</p>	<p>ENGLISH EQUIVALENTS OF SOME TERMS AND CONCEPTS FREQUENTLY USED.</p>
<p>A: AFHANKELIJK AFLEESVOORBEELD ASELECT</p>	<p>DEPENDENT EXAMPLE OF READING-OFF THE TABLES AT RANDOM</p>
<p>B: BENADERING BEREKENEN BETROUWBAARHEIDS NIVEAU BETROUWBAARHEIDSINTERVAL</p>	<p>APPROXIMATION TO CALCULATE ; TO COMPUTE CONFIDENCE LEVEL CONFIDENCE INTERVAL</p>
<p>C: CORRELATIE-COEFFICIENT</p>	<p>COEFFICIENT OF CORRELATION</p>
<p>E: EENZIJDIG</p>	<p>ONE-SIDED</p>
<p>F: FACTOR FRACTIEL</p>	<p>FACTOR FRACTILE</p>
<p>G: GEBRUIK GEMIDDELDE</p>	<p>USE MEAN, MEAN VALUE</p>
<p>I: INTERVAL MET 'GEMIDDELDE BETROUWBAARHEID'</p>	<p>'AVERAGE CONFIDENCE' INTERVAL</p>
<p>K: KEURINGSKARAKTERISTIEK KRITIEKE WAARDE</p>	<p>OPERATING CHARACTERISTIC CRITICAL VALUE</p>
<p>M: MEDIAAN MODUS</p>	<p>MEDIAN MODE</p>
<p>N: NOEMER</p>	<p>DENOMINATOR</p>
<p>O: ONAFHANKELIJK ONBETROUWBAARHEIDS NIVEAU OPPERVLAK OVERSCHRIJDINGSKANS</p>	<p>INDEPENDENT SIGNIFICANCE LEVEL AREA EXCEEDANCE PROBABILITY</p>
<p>P: PARAMETER POPULATIE</p>	<p>PARAMETER POPULATION; UNIVERSE</p>
<p>R: REGELGRENS</p>	<p>CONTROL LIMIT</p>
<p>S: SCHATTING SPREIDINGSBREEDTE STANDAARDAFWIJKING STEEKPROEF</p>	<p>ESTIMATE RANGE STANDARD DEVIATION SAMPLE</p>
<p>T: TELLER TOEPASSING TOETS TOETSEN TOEVALSCIJFERS TREKKING TWEEZIJDIG</p>	<p>NUMERATOR APPLICATION TEST TO TEST RANDOM NUMBERS RANDOM DRAWING TWO-SIDED</p>
<p>V: VARIANTIE VERDELING(NORMALE) VERDELINGSVRIJE TOETS VERKLARING VAN SYMBOLEN VERLOTINGSSERIES VRIJHEIDSGRADEN</p>	<p>VARIANCE DISTRIBUTION(NORMAL) DISTRIBUTION-FREE TEST EXPLANATION OF SYMBOLS RANDOM SERIES DEGREES OF FREEDOM</p>
<p>W: WAARDE WAARNEMING</p>	<p>VALUE OBSERVATION</p>

I N L E I D I N G

1. I N D E L I N G D E R T A B E L L E N

DEZE TABELLEN ZIJN VERZAMELD, BEWERKT EN IN VERSCHILLENDE GEVALLEN BEREKEND DOOR DE STATISTISCHE GROEP VAN HET NATUURKUNDIG LABORATORIUM.

DE VOLGORDE DER TABELLEN IS ENIGSZINS ARBITRAIR: MET NORMALE VERDELINGEN VERWANTE EN CONTINUE VERDELINGEN WORDEN GEVOLGD DOOR DISCRETE VERDELINGEN, TABELLEN VOOR PARTIJKERING EN REGELGRENZEN, EN TABELLEN MET ASELECTE TREKKINGEN EN VOLGORDEN. AAN HET EINDE DER VERZAMELING VOLGT EEN LITERatuurVERWIJZING NAAR GEBRUIKTE EN UITGEBREIDERE TABELLENVERZAMELINGEN. IN PARAGRAAF 3 VAN DEZE INLEIDING WORDEN TOEPASSINGEN VAN VERSCHILLENDE TABELLEN VERMELD, BENEVENS ENKELE BELANGRIJKE EIGENSCHAPPEN, W.O. RELATIES MET ANDERE VERDELINGEN. OP DE OMSLAG WORDEN VERTALINGEN GEGEVEN VAN ENKELE DER GEBRUIKTE TERMEN.

2. N O T A T I E

IN DE TABELLEN WORDT MET P STEEDS DE KANS AANGEDUID, DAT DE BESCHOUWDE VARIABLE KLEINER IS DAN OF GELIJK IS AAN DE OPGEGEVEN WAARDE.

DOOR ONDERSTREPING VAN EEN SYMBOOL WORDT AANGEGEVEN, DAT EEN GROOTHEID EEN WAARSCHIJNLIJKHEIDSVERDELING VOLGT |

DE KANS DAT \underline{x} KLEINER IS DAN x , GEGEVEN EEN AANTAL PARAMETERWAARDEN $\theta_1, \theta_2, \dots$ WORDT SYMBOLISCH AANGEGEVEN MET

$$P(\underline{x} < x \mid \theta_1, \theta_2, \dots)$$

HET TEKEN \cong (SPREEK UIT: 'IS ISOMOOR MET') WORDT GEBRUIKT OM AAN TE GEVEN, DAT DE UITDRUKKINGEN IN LINKER- EN RECHTERLID DEZELFDE VERDELING VOLGEN.

EEN GROOTHEID DIE EEN NORMALE VERDELING BEZIT MET $\mu = 0$ EN $\sigma = 1$ WORDT AANGEDUID MET \underline{u} .

VOOR DE χ^2 - EN t -VERDELING WORDT HET AANTAL VRIJHEIDSGRADEN ALS ONDERINDEX AAN HET SYMBOOL TOEGEVOEGD; VOOR DE F -VERDELING WORDT HET AANTAL VRIJHEIDSGRADEN VAN TELLER EN NOEMER ALS BOVEN- RESP. ONDERINDEX AAN HET SYMBOOL TOEGEVOEGD, B.V. $F_{v_2}^{v_1}$ IS EEN VARIABLE DIE EEN F -VERDELING VOLGT MET v_1 EN v_2 VRIJHEIDSGRADEN. EEN FRACTIEL WORDT AANGEGEVEN DOOR TUSSEN HAAKJES ACHTER HET SYMBOOL DE BETROUWBAARHEID AAN TE GEGEVEN. B.V. $F_{v_2}^{v_1}(P)$.

3. T O E L I C H T I N G B I J D E T A B E L L E N

TABEL 1. EIGENSCHAP: $\underline{u}^2 \cong \chi_{-1}^2$.

TABEL 2. DEFINITIE: $\underline{t}_v \cong \underline{u} / \sqrt{\chi_v^2 / v}$, ALS \underline{u} EN χ_v^2 ONDERLING ONAFHANKELIJK ZIJN. HIERUIT VOLGT $\underline{t}_v^2 \cong F_v^1$.

TOEPASSINGEN: DEZE TABEL KAN O.A. GEBRUIKT WORDEN OM BIJ ONBEKENDE σ 's TWEESZIJDIGE BETROUWBAARHEIDSINTERVALLEN MET ONBETROUWBAARHEID P TE VORMEN VOOR:

- a. EEN POPULATIE-GEMIDDELDE, μ , VAN EEN NORMALE VERDELING OP GROND VAN EEN STEEKPROEF VAN n ONAFHANKELIJKE WAARNEMINGEN:

$$\bar{x} - t_{n-1} \left\{ (1+P) / 2 \right\} \cdot s / \sqrt{n} < \mu < \bar{x} + t_{n-1} \left\{ (1+P) / 2 \right\} \cdot s / \sqrt{n}$$

- b. HET VERSCHIL $\delta = \mu_1 - \mu_2$ VAN TWEE POPULATIEGEMIDDELDEN VAN NORMALE VERDELINGEN OP GROND VAN TWEE STEEKPROEVEN VAN n_1 RESP. n_2 ONAFHANKELIJKE WAARNEMINGEN

$$\bar{x}_1 - \bar{x}_2 - t_{n-1} \left\{ (1+P) / 2 \right\} \cdot s \sqrt{(n_1+n_2) / n_1 n_2} < \delta < \bar{x}_1 - \bar{x}_2 + t_{n-1} \left\{ (1+P) / 2 \right\} \cdot s \sqrt{(n_1+n_2) / n_1 n_2}$$

$$\text{WAARIN } s^2 = \left\{ (n_1-1) \cdot s_1^2 + (n_2-1) \cdot s_2^2 \right\} / (n_1+n_2-2) .$$

- c. DE POPULATIEWAARDE, β , VAN DE HELLING VAN EEN LINEAIRE REGRESSIELIJK OF EEN LINEAIR REGRESSIEOPPERVLAK OP GROND VAN EEN AANTAL WAARNEMINGEN (y_1, \dots, y_n) BEHOURENDE BIJ DE WAARDEN VAN DE ONAFHANKELIJK VARIABELE (x_1, \dots, x_n) RESPECTIEVELIJK:

$$b - t_{n-2} \left\{ (1+P) / 2 \right\} \cdot s_b < \beta < b + t_{n-2} \left\{ (1+P) / 2 \right\} \cdot s_b ,$$

$$\text{WAARIN } b = s_{xy} / s_x^2 \text{ EN } s_b = s_y^2 \cdot (1-r^2) / s_x^2 \cdot (n-2); \quad r = s_{xy} / s_x \cdot s_y .$$

TABEL 3. DEFINITIE EN EIGENSCHAPPEN:

ALS u_1, \dots, u_n ONDERLING ONAFHANKELIJK ZIJN, DAN IS $\chi_n^2 \cong u_1^2 + \dots + u_n^2$.

ALS $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)$ EN x_1, \dots, x_n ZIJN ONAFHANKELIJKE WAARNEMINGEN UIT EEN NORMALE VERDELING, DAN IS $\chi_{n-1}^2 \cong s^2 \cdot (n-1) / \sigma^2$.

ALS $\chi_{v_1}^2$ EN $\chi_{v_2}^2$ ONDERLING ONAFHANKELIJK ZIJN, DAN IS $\chi_{v_1+v_2}^2 \cong \chi_{v_1}^2 + \chi_{v_2}^2$.

BOVENDIEN GELDT: $\chi_{v_1}^2 / v_1 \cong F_{\infty}^{v_1}$.

ALS Γ_n HET SYMBOOL IS VOOR DE SOM VAN n EXPONENTIEEL VERDEELDE GROOTHEDEN, DAN

$$\text{IS } \Gamma_n \cong \frac{1}{2} \chi_{2n}^2 .$$

ALS x EEN POISSONVERDELING VOLGT MET GEMIDDELDE μ , DAN IS

$$P(x \geq x | \mu) = P(\chi_{2x}^2 < 2\mu) .$$

TOEPASSINGEN: DE χ^2 -VERDELING WORDT GEBRUIKT OM WAARDEN VAN σ^2 VAN EEN NORMALE VERDELING TE TOETSEN OF HIERVOOR BETROUWBAARHEIDSINTERVALLEN TE VORMEN (ZIE TABEL 5), ZOWEL OP GROND VAN REGRESSIE- EN VARIANTIEANALYSEGEGEVENS ALS OP GROND VAN EEN OF TWEE STEEKPROEVEN. VOOR DE RESTVARIANTIESCHATTING s_r^2 VAN EEN LINEAIRE REGRESSIE MET ÉÉN ONAFHANKELIJK VARIABELE EN IN TOTAAL n WAARNEMINGEN GELDT B.V.:

$$\chi_{n-2}^2 \cong s_r^2 \cdot (n-2) / \sigma^2 .$$

DE χ^2 -VERDELING WORDT OOK TOEGEPAST ALS BENADERING VOOR EEN ANDERE VERDELING. DE BELANGRIJKSTE TOETSEN ZIJN DE ZGN. AANPASSINGSTOETSEN EN DE χ^2 -TOETS VOOR DE $m \times n$ TABEL. VOOR EEN BESCHRIJVING VAN DEZE TOETSEN ZIJ VERWEZEN NAAR DE HANDBOEKEN.

TABEL 4. DEFINITIE EN EIGENSCHAPPEN:

ALS $\chi_{v_1}^2$ EN $\chi_{v_2}^2$ ONDERLING ONAFHANKELIJK ZIJN, DAN IS $F_{v_2}^{v_1} \cong (\chi_{v_1}^2 / v_1) / (\chi_{v_2}^2 / v_2)$.

HIERUIT VOLGT ALS s_1^2 EN s_2^2 ONAFHANKELIJKE SCHATTINGEN ZIJN VAN DEZELFDE σ^2 , MET v_1 EN v_2 VRIJHEIDSGRADEN RESP. DAN IS $F_{v_2}^{v_1} \cong s_1^2 / s_2^2$.

BIJZONDERE GEVALLEN VAN DE F-VERDELING WORDEN o.a. GEGEVEN DOOR

$$F_{\infty}^{v_1} \cong \chi_{v_1}^2 \quad F_{v_1}^1 \cong \chi_{v_1}^2 \quad F_{\infty}^1 \cong u^2 .$$

ALS Γ_n / n EN Γ_m / m TWEE ONDERLING ONAFHANKELIJKE GEMIDDELDEN ZIJN VAN WAARNEMINGEN VAN TWEE STEEKPROEVEN UIT DEZELFDE EXPONENTIELE VERDELING, DAN IS

$$F_{-2m}^{2n} = (\Gamma_n / n) / (\Gamma_m / m)$$

ALS x BINOMIAAL VERDEELD IS, DAN IS $P(x \geq x | n, p) = P\left\{F_{2(n-x+1)}^{2x} < (n-x+1)p / xq\right\}$
 DE NIET GETABELLEERDE LINKERSTAARTKANSEN VAN DE F-VERDELING KAN MEN VINDEN MET BE-
 HULP VAN DE SYMMETRIEBETREKKING: $F_{v_2}^{v_1}(p) = 1 / F_{v_1}^{v_2}(1-p)$.

TOEPASSINGEN: MET BEHULP VAN DEZE TABEL KAN O.A. OP GROND VAN TWEE ONAF-
 HANKELIJKE VARIANTIESCHATTINGEN s_1^2 EN s_2^2 EEN BETROUWBAARHEIDSINTERVAL GEVORMD WORDEN
 VOOR DE VERHOUDING, $c = \sigma_1^2 / \sigma_2^2$, VAN DE BIJBEHORENDE POPULATIEWAARDEN:

$$s_1^2 / s_2^2 F_{v_2}^{v_1} \left\{ (1+p) / 2 \right\} < c < s_1^2 F_{v_1}^{v_2} \left\{ (1+p) / 2 \right\} / s_2^2$$

TABEL 5. TOEPASSING: DEZE TABEL KAN MEN ZOWEL GEBRUIKEN VOOR HET BEREKENEN VAN BE-
 TROUWBAARHEIDSINTERVALLEN VAN σ OP GROND VAN EEN STEEKPROEF UIT EEN NORMALE VERDELING,
 ALS VOOR INTERVALLEN OP GROND VAN k STEEKPROEVEN UIT k VERSCHILLENDE NORMALE VER-
 DELINGEN MET $\sigma_1^2 = \dots = \sigma_k^2 = \sigma^2$, HET AANTAL VRIJHEIDSGRADEN IS IN DIT GEVAL
 $(\sum n_i - k)$.

OOK BIJ LINEAIRE REGRESSIE EN BIJ VARIANTIEANALYSE KAN MEN UIT DE RESTVA-
 RIANTIESCHATTING EEN BETROUWBAARHEIDSINTERVAL BEREKENEN. VOOR HET AANTAL VRIJHEIDSGRADEN
 IN DEZE GEVALLEN ZIJ VERWEZEN NAAR DE HANDBOEKEN.

TABEL 7. EIGENSCHAP: VOOR HET BEREKENEN VAN IN DE TABEL ONTBREKENDE WAARDEN KAN MEN
 GEBRUIK MAKEN VAN DE RELATIE: $t_{n-2} \approx \frac{r}{n} \sqrt{n-2} / \sqrt{1 - r^2}$.

TABEL 8. EIGENSCHAP: ALS x EXPONENTIEEL VERDEELD IS MET GEMIDDELDE μ , DAN IS
 $x / \mu \approx \frac{1}{2} \chi_{2}^2$ EN DUS IS $m \cdot \bar{x} / \mu \approx \frac{1}{2} \chi_{2m}^2$.

TOEPASSING: ALS \bar{x} BEREKEND IS UIT EEN ONVOLLEDIG WAARGENOMEN STEEKPROEF VAN
 ONAFHANKELIJKE WAARNEMINGEN UIT EENZELFDE EXPONENTIELE VERDELING, D.W.Z. ALS DE WAAR-
 DEN VAN DE KLEINSTE k WAARNEMINGEN VAN DE m ZIJN WAARGENOMEN, DAN KAN MEN DE TA-
 BEL OOK GEBRUIKEN ALS MEN $m = k$ STELT EN $\bar{x} = (x_1 + x_2 + \dots + x_{k-1} + (n-k)x_k) / k$.

OOK IN ANDERE DAN HET BOVENGENOEMDE MEEST VOORKOMENDE GEVAL IS DE TABEL TE
 GEBRUIKEN.

TABEL 9. EIGENSCHAPPEN: ALS x EN y ONDERLING ONAFHANKELIJKE TREKKINGEN ZIJN UIT BINO-
 MIALE VERDELINGEN MET PARAMETERS n_1, p EN n_2, p RESPECTIEVELIJK, DAN HEEFT $x + y$ EEN
 BINOMIALE VERDELING MET PARAMETERS $(n_1 + n_2), p$.
 BOVENDIEN GELDT

$$P\left\{F_{v_2}^{v_1} < F\right\} = P\left\{x \geq v_1/2 \mid n = (v_1 + v_2 - 2) / 2; p = v_1 F / (v_2 + v_1 F)\right\}$$

DEZELFDE RELATIE UITGEDRUKT IN DE PARAMETERS VAN DE BINOMIALE VERDELING IS VERMELD
 ONDER TABEL 4.

DE BINOMIALE VERDELING IS HIER GETABELLEERD VOOR $p < \frac{1}{2}$. OM KANSEN TE VIN-
 DEN VOOR $p > \frac{1}{2}$, KAN MEN GEBRUIK MAKEN VAN DE RELATIE:

$$P(x \geq x | n, p) = 1 - P(x \geq n - x | n, q), \text{ met } q = 1 - p.$$

ALS r, p PARAMETERS ZIJN VAN EEN NEGATIEF BINOMIAAL VERDEELDE GROOTHEID y ,
 DAN GELDT: $P(y \geq y | r, p) = P(x < r | n = r + y - 1, p)$.

BENADERINGEN: VOOR n GROOT EN p NIET AL TE DICHT BIJ 0 OF 1 KAN DE
 BINOMIALE VERDELING BENADERD WORDEN DOOR DE NORMALE VERDELING:

a. ZONDER CONTINUITEITSCORRECTIE: $u \approx (x - np) / \sqrt{npq}$

b. MET CONTINUITEITSCORRECTIE: $P(x \leq x | n, p) = P(u < (x - np + \frac{1}{2}) / \sqrt{npq})$

VOOR KLEINE p EN ENIGSZINS GROTE n KAN DE BINOMIALE VERDELING BENADERD
 WORDEN DOOR DE POISSON-VERDELING.

VOOR DE NAUWKEURIGHEID VAN BEIDE BENADERINGEN ZIJ VERWEZEN NAAR TABEL 13.

TABEL 10. DEFINITIE: DE DEFINITIE VAN EEN BETROUWBAARHEIDSINTERVAL, ALS DE VERZAMELING VAN ALLE PARAMETERWAARDEN, DIE OP GROND VAN EEN TOETS MET EEN BETROUWBAARHEID P NIET VERWORPEN KUNNEN WORDEN IS VOOR DISCRETE VERDELINGEN NIET EQUIVALENT MET DE OMSCHRIJVING, DAT EEN BETROUWBAARHEIDSINTERVAL MET EEN KANS $P' = P$ DE WARE PARAMETERWAARDE BEVAT. DEZE LAATSTE KANS IS GROTER DAN OF GELIJK AAN P , AFHANKELIJK VAN DE WERKELIJKE WAARDE VAN DE PARAMETER. HET IS NU MOGELIJK VOOR HET CONSTRUEREN VAN EEN BETROUWBAARHEIDSINTERVAL DE TWEEDE OMSCHRIJVING TE GEBRUIKEN MET VOOR P' EEN GEMIDDELDE WAARDE, N.L. GEMIDDELD OVER SOMMIGE MOGELIJKE PARAMETERWAARDEN. DE WAARDE VAN P' WORDT DE GEMIDDELDE BETROUWBAARHEID GENOEMD.

TOEPASSING: DE VERSCHILLEN TUSSEN DE BETROUWBAARHEIDSINTERVALLEN VOLGENS DE EERSTE OMSCHRIJVING EN DE IN DE TABEL VERMELDE INTERVALLEN ZIJN ALLEEN GROOT VOOR KLEINE n . IN HET ALGEMEEN IS HET HIER GEGEVEN BETROUWBAARHEIDSINTERVAL IETS SMALLER. IN DIE GEVALLEN, WAARIN MEN DE TWEEDE OMSCHRIJVING VAN EEN BETROUWBAARHEIDSINTERVAL, WELKE IN DE PRAKTIJK MEER AANSPREEKT, DE VOORKEUR GEEFT, DIENT MEN BOVENSTAANDE TABEL TE GEBRUIKEN, WAARBIJ EEN EXACTE BETROUWBAARHEID P' VERVANGEN MOET WORDEN DOOR EEN GEMIDDELDE BETROUWBAARHEID.

TABEL 11. EIGENSCHAPPEN: VOOR RELATIE MET DE χ^2 -VERDELING EN Γ -VERDELING ZIE ONDER TABEL 3. VOOR RELATIE MET BINOMIALE VERDELING ZIE ONDER TABEL 9.

BENADERING: DE POISSON-VERDELING KAN VOOR GROTE μ BENADERD WORDEN DOOR DE NORMALE VERDELING:

- a. ZONDER CONTINUITEITSCORRECTIE: $u \cong (x - \mu) / \sqrt{\mu}$
- b. MET CONTINUITEITSCORRECTIE: $P(x \leq x | \mu) = P\left\{u < (x - \mu + \frac{1}{2}) / \sqrt{\mu}\right\}$.

TABEL 12.

ZIE ONDER TABEL 10.

TABEL 13.

IN DE FIGUUR IS AANGEGEVEN BIJ WELKE WAARDEN VAN n EN p DE NORMALE BENADERING (MET CONTINUITEITSCORRECTIE) EN DE POISSON-BENADERING TENMINSTE EEN VOORGESCHREVEN NAUWKEURIGHEID HEBBEN IN DE BUURT VAN HET 0,025 PUNT IN DE LINKER- EN IN DE RECHTERSTAART VAN DE VERDELING.

HIER IS ALLEEN DE NAUWKEURIGHEID VAN DE BENADERING IN DE BUURT VAN DEZE FRACTIELEN ONDERZOEKT, OMDAT DEZE FRACTIELEN BIJ TOETSING VEEL GEBRUIKT WORDEN.

VOOR IEDERE BENADERING ZIJN TWEE GRENSLIJNEN GETROKKEN, EEN WAARVOOR DE AFWIJKING TUSSEN DE EXACTE STAARTKANS EN DE BENADERDE 0,0025 IS, DIT IS EEN VRIJ GOEDE BENADERING, EN EEN WAARVOOR DE AFWIJKING 0,0125 IS, DIT IS EEN VRIJ SLECHTE BENADERING.

UIT DE GRAFIEK BLIJKT, DAT DE BEKENDE VUISTREGELS GEEN ERG GOEDE BENADERING OPLEVEREN. HET LIGT IN DE BEDOELING IN EEN PUBLIKATIE DEZE GRAFIEK UITVOERIG TOE TE LICHTEN.

TABEL 14. TOEPASSING: DEZE TABEL IS OOK TE GEBRUIKEN VOOR DE TEKENTOETS. DE EXACTE STAARTKANSEN VOOR DE BINOMIALE VERDELING MET $p = \frac{1}{2}$ WORDEN GEGEVEN; DE GETABELLEERDE KANS IS GELIJK AAN: $P(\underline{m} \geq m | n, p = \frac{1}{2})$.

TABEL 15. BENADERING: DE GROOTHEID \underline{w} KAN BENADERD WORDEN DOOR EEN NORMALE VERDELING ALS n EN m BEIDE B.V. GROTER ZIJN DAN 7:

$$P(\underline{w} \leq \underline{w} \mid n, m) = P\left\{ \underline{u} < (w - m \cdot n + 1) \sqrt{m \cdot n (n + m + 1) / 3} \right\}.$$

TABEL 17.

HET PHILIPS STANDAARD STEEKPROEF SYSTEEM ONDERSCHIEDT ZICH VAN ANDERE STEEKPROEFSYSTEMEN, ZOALS B.V. BESCHREVEN IN MIL.STD.105 A, DOORDAT DE KEURINGSKARAKTERISTIEK GEKARAKTERISEERD WORDT DOOR HET CONTROLEPUNT, WAARBIJ 50% VAN DE PARTIJEN GOEDGEKEURD EN 50% AFGEKEURD WORDT, EN DOOR DE STEILHEID IN HET CONTROLEPUNT. IN DE TABELLEN WORDT DE STEILHEID BEPAALD DOOR DE PARTIJGROOTTE, NAARMATE DE PARTIJ GROTER IS, DIENT DE KEURINGSKARAKTERISTIEK STEILER TE ZIJN. HET VERBAND TUSSEN PARTIJGROOTTE EN STEILHEID IN DEZE TABEL IS ONGEVEER GELIJK AAN DIT VERBAND VOOR ANDERE STEEKPROEFSYSTEMEN.

TABEL 18. TOEPASSING: MEN KAN UIT DEZE TABEL VOOR IEDERE WAARDE VAN m HET GOEDKEURCRITERIUM (BIJ m OF MINDER DEFECTIEVEN IN DE STEEKPROEF WORDT DE PARTIJ GOEDGEKEURD), EN VOOR IEDERE GOEDKEURKANS P , DE BIJBEHORENDE FRACTIE DEFECTIEVEN IN DE PARTIJ BEREKENEN, ALS AANGEGEVEN ONDERAAN DE TABEL.

DE TABEL IS EEN TABEL VAN DE POISSON-VERDELING, WAARVAN DIE μ GETABELLEERD IS, WAARVOOR GELDT: $P(\underline{m} \leq m \mid \mu) = P_A$.

TABEL 22. TOEPASSING: ALS MEN 10 EXEMPLAREN DUSDANIG WIL VERDELLEN OVER 3 BEHANDELINGSWIJZEN, A, B EN C, DAT 3 EXEMPLAREN BEHANDELINGSWIJZE A ONDERGAAN, DRIE B EN VIER C, DAN KAN MEN DIT DOEN MET BEHULP VAN TOEVALSCIJFERS. VERLOTINGSSERIES ZIJN HIERVOOR ECHTER MAKKELIJKER, OMDAT EEN VERLOTINGSSERIE VAN n GETALLEN IEDER VAN DIE GETALLEN JUIST EEN KEER BEVAT. MEN NUMMERT DE EXEMPLAREN VAN 0 TOT 9, MEN KIEST EEN VERLOTINGSSERIE VAN 0-9 EN GEEFT DE EXEMPLAREN MET DE EERSTE DRIE NUMMERS VAN DE VERLOTINGSSERIE BEHANDELING A, ENZ.

4. HET AFRONDEN VAN WAARNEMINGEN

BIJ HET UITVOEREN VAN STATISTISCHE BEREKENINGEN OP GROND VAN EEN SERIE WAARNEMINGEN, LEIDT AFRONDEN VAN WAARNEMINGEN TOT VEREENVOUDIGING VAN HET REKENWERK, TE GROVE AFRONDING KAN ECHTER LEIDEN TOT ONJUISTE UITKOMSTEN. IN HET ALGEMEEN WORDT DE VARIANTIE VAN DE WAARNEMINGEN HIERDOOR VERGROOT.

VOOR WAARNEMINGEN UIT EEN NORMALE OF BIJNA - NORMALE POPULATIE MET BEKENDE STANDAARDAFWIJKING σ IS EEN MAXIMALE AFRONDING OP EENHEDEN VAN $0,6 \sigma$ NOG TOELAATBAAR.

IS σ NIET BEKEND, MAAR MOET MEN DEZE SCHATTEN, DAN IS DEZELFDE REGEL TE GEBRUIKEN, ALS DE STEEKPROEFVARIANTIE TENMINSTE 4 VRIJHEIDSGRADEN HEEFT. SCHAT MEN σ MET BEHULP VAN SPREIDINGSBREEDTEN DAN DIENT MEN NOG IETS VOORZICHTIGER TE ZIJN.

5. VERANTWOORDING

OP BLZ. 35 WORDT NAAR DE HERKOMST VAN DE GEPUBLICEERDE TABELLEN VERWEZEN EN NAAR UITGEBREIDERE TABELLEN.

DE UITGEVERS VAN DE TABELLENVERZAMELINGEN, WAARAAN TABELLEN ONTLEEND ZIJN, ZIJN ZO VRIENDELIJK GEWEEST HUN TOESTEMMING VOOR HET OPNEMEN TE VERLENEN.

DE TABELLEN 2 EN 7 ZIJN ONTLEEND AAN: R.A.FISHER AND F.YATES, 'STATISTICAL TABLES FOR BIOLOGICAL, AGRICULTURAL AND MEDICAL RESEARCH', OLIVER AND BOYD, Ltd., EDINBURGH, 1957.

TABEL 1

NORMALE VERDELING

$F(u)$, HET OPPERVLAK ONDER DE NORMALE CURVE VAN $-\infty$ TOT u ; EN DE ORDINAAT $f(u)$ BEHORENDE BIJ u .

u	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	$f(u)$
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359	.3989
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753	.3970
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141	.3910
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517	.3814
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879	.3683
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224	.3521
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549	.3332
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852	.3123
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133	.2897
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389	.2661
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621	.2420
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830	.2179
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015	.1942
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177	.1714
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319	.1497
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	.1295
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545	.1109
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633	.0940
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706	.0790
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767	.0656
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	.0540
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	.0440
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	.0355
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	.0283
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	.0224
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952	.0175
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964	.0136
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974	.0104
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981	.0079
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986	.0060
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990	.0044
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993	.0033
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995	.0024
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997	.0017
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998	.0012
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.0009
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.0006

AFLEESVOORBEELD, VERKLARING VAN SYMBOLEN

ALS BIJVOORBEELD $u = 2.02$, DAN NEEMT MEN DE RIJ AANGEDUID MET 2.0 EN DE KOLOM MET .02. HET GEVONDEN GETAL GEEFT DAN AAN: $F(u)$ D.I. HET OPPERVLAK ONDER DE NORMALE CURVE VAN $-\infty$ TOT u . IN FORMULE:

$$F(u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} \exp \frac{-u^2}{2} du, \text{ WAARIN } u = \frac{x-\mu}{\sigma};$$




μ IS HET GEMIDDELDE, σ IS DE STANDAARDAFWIJKING.

VOOR NEGATIEVE WAARDEN VAN u GELDT: $F(-u) = 1 - F(u)$.


$f(u)$ IS DE ORDINAAT BEHORENDE BIJ DE WAARDEN VAN u IN DE EERSTE KOLOM.

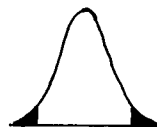
NORMALE VERDELING (VERVOLG)

HET OPPERVLAK $F(u)$ EN OVERSCHRIJDINGSKANSEN VOOR ENIGE WAARDEN VAN u

u	 $F(u)$	 EENZIJDIGE OVERSCHRIJDINGSKANS $1 - F(u)$	 TWEEZIJDIGE OVERSCHRIJDINGSKANS $2 [1 - F(u)]$
1.00	0.8413	0.1587	0.3174
1.20	0.8849	0.1151	0.2302
1.40	0.9192	0.0808	0.1616
1.60	0.9452	0.0548	0.1096
1.80	0.9641	0.0359	0.0718
2.00	0.9772	0.0228	0.0456
2.20	0.9861	0.0139	0.0278
2.40	0.9918	0.0072	0.0144
2.60	0.9953	0.0047	0.0094
2.80	0.9974	0.0026	0.0052
3.00	0.9987	0.0013	0.0026
3.20	0.9993	0.0007	0.0014

ENKELE FRACTIELEN; WAARDEN VAN u VOOR GEBUIKELIJKE WAARDEN VAN DE OVERSCHRIJDINGSKANSEN

 $1 - F(u)$	u
0.20	0.842
0.10	1.282
0.05	1.645
0.02	2.054
0.01	2.326
0.005	2.576
0.002	2.878
0.001	3.090
0.0001	3.219
0.00001	4.265

 $2 [1 - F(u)]$	u
0.20	1.282
0.10	1.645
0.05	1.960
0.02	2.326
0.01	2.576
0.005	2.807
0.002	3.090
0.001	3.290
0.0001	3.897
0.00001	4.417

VERKLARING VAN SYMBOLEN

ZIE BLADZIJDE 6 HIERNAAST

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE

TABEL 2

t - VERDELING

FRACTIELEN

EENZIJDIG	BETROUWBAARHEID				
	0.90	0.95	0.975	0.99	0.995
TWEEZIJDIG	0.80	0.90	0.95	0.98	0.99
v					
1	3.08	6.31	12.7	31.8	63.7
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.01	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.41	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.06
13	1.35	1.77	2.16	2.65	3.01
14	1.34	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
20	1.32	1.72	2.09	2.53	2.85
25	1.32	1.71	2.06	2.48	2.79
30	1.31	1.70	2.04	2.46	2.75
40	1.30	1.68	2.02	2.42	2.70
50	1.30	1.68	2.01	2.40	2.68
100	1.29	1.66	1.98	2.36	2.63
200	1.29	1.65	1.97	2.34	2.60
∞	1.28	1.64	1.96	2.33	2.58

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN ZOEKT IN DE TABEL HET GETAL VOORKOMENDE IN DE RIJ MET HET GEGEVEN AANTAL VRIJHEIDSGRADEN (v) EN IN DE KOLOM BEHOURENDE BIJ DE GEWENSTE BETROUWBAARHEID (P). DIT GETAL IS DE GEWENSTE WAARDE VAN t_v .

$$\text{VOORBEELD: } P(t_{5} < +2.01) = 0.95$$

$$P(-2.57 < t_{5} < +2.57) = 0.95$$

TOEPASSING

ZIE INLEIDING VOOR HET VORMEN VAN BETROUWBAARHEIDSGRENZEN MET BEHULP VAN DE t-VERDELING VOOR EEN GEMIDDELDE, VOOR HET VERSCHIL VAN TWEE GEMIDDELDEN, VOOR HELLINGEN VAN REGRESSIELIJNEN EN ANDERE TOEPASSINGEN.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

χ^2 - VERDELING

FRACTIELEN

v	EENZIJDIGE BETROUWBAARHEID														v
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995	.999	
1	-	-	.001	.004	.016	.102	.455	1.32	2.71	3.84	5.02	6.63	7.88	10.8	1
2	.010	.020	.051	.103	.211	.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6	13.8	2
3	.072	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8	16.3	3
4	.207	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9	18.5	4
5	.412	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7	20.5	5
6	.676	.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5	22.5	6
7	.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3	24.3	7
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0	26.1	8
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6	27.9	9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2	29.6	10
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8	31.3	11
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3	32.9	12
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8	34.5	13
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3	36.1	14
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8	37.7	15
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3	39.3	16
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7	40.8	17
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2	42.3	18
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6	43.8	19
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0	45.3	20
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4	46.8	21
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8	48.3	22
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2	49.7	23
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6	51.2	24
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9	52.6	25
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	54.1	26
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	55.5	27
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	56.9	28
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	58.3	29
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	59.7	30
u	-2.58	-2.33	-1.96	-1.64	-1.28	-0.67	0.00	+0.67	+1.28	+1.64	+1.96	+2.33	+2.58	+3.09	u

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN ZOEKT IN DE TABEL HET GETAL VOORKOMENDE IN DE RIJ MET HET GEGEVEN AANTAL VRIJHEIDSGRADEN (v) EN IN DE KOLOM BEHORENDE BIJ DE GEWENSTE BETROUWBAARHEID (P). DIT GETAL IS DE GEWENSTE WAARDE VAN χ^2_v .

$$\text{VOORBEELD: } P(\chi^2_{14} < 23.7) = .95.$$

TOEPASSING

ZIE INLEIDING VOOR HET TOETSEN EN HET VORMEN VAN BETROUWBAARHEIDSINTERVALLEN MET BEHULP VAN DE χ^2 -VERDELING. ZIE OOK TABEL 5.

BENADERING

VOOR WAARDEN VAN v > 30 KAN χ^2 VIA DE NORMALE VERDELING WORDEN BEREKEND UIT:

$$\chi^2_v(P) = \frac{(\sqrt{2v} + u)^2}{2}.$$

DE WAARDEN VAN u, BEHORENDE BIJ DE BENODIGDE BETROUWBAARHEID ZIJN ONDERAAN DE TABEL GEGEVEN.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

F - VERDELING

FRACTIËLEN; EENZIJDIGE BETROUWBAARHEID .95

		VRIJHEIDSGRADEN VAN DE TELLER																			
		1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60	120	∞	
VRIJHEIDSGRADEN VAN DE NOEMER	1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254	
	2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
	3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	8.53
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	5.63
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.01	2.97	2.93
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.62	2.58	2.54
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.49	2.45	2.40
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.69	2.62	2.54	2.50	2.47	2.43	2.38	2.34	2.30
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.53	2.46	2.39	2.34	2.31	2.27	2.22	2.18	2.13
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.48	2.40	2.33	2.28	2.25	2.20	2.16	2.11	2.07
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.42	2.35	2.28	2.23	2.19	2.15	2.11	2.06	2.01
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.38	2.31	2.25	2.18	2.15	2.10	2.06	2.01	1.96
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.34	2.27	2.19	2.14	2.11	2.06	2.02	1.97	1.92
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.23	2.15	2.07	2.02	1.98	1.94	1.89	1.84	1.78	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.20	2.13	2.05	2.00	1.96	1.91	1.86	1.81	1.76	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.09	2.01	1.93	1.88	1.84	1.79	1.74	1.68	1.62	
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	2.00	1.92	1.84	1.78	1.74	1.69	1.64	1.58	1.51	
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.92	1.84	1.75	1.69	1.65	1.59	1.53	1.47	1.39	
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.83	1.75	1.66	1.60	1.55	1.50	1.43	1.35	1.25	
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.75	1.67	1.57	1.51	1.46	1.39	1.32	1.22	1.00	

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN ZOEKT IN DE TABEL HET GETAL VOORKOMENDE IN DE RIJ MET HET GEGEVEN AANTAL VRIJHEIDSGRADEN VAN DE NOEMER (v_2) EN IN DE KOLOM BEHORENDE BIJ HET GEGEVEN AANTAL VRIJHEIDSGRADEN VAN DE TELLER (v_1). DIT GETAL IS DE GEWENSTE WAARDE VAN F_{v_1, v_2}

VOORBEELD: $P(F_{10, 7} < 3.14) = .95$

TOEPASSING

ZIE INLEIDING VOOR HET VORMEN VAN BETROUWBAARHEIDSGRADEINTERVALLEN MET BEHULP VAN DE F-VERDELING.

ZIE VERDER PAG. 11, HIERNAAST.

F - VERDELING (VERVOLG)

FRACTIELEN; EENZIJDIGE BETROUWBAARHEID .99

		VRIJHEIDSGRADEN VAN DE TELLER																			
		1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60	120	∞	
1	4052	5000	5403	5625	5764	5859	5928	5982	6023	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366		
2	98.5	99.0	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5		
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1		
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.6	13.5	13.5		
5	16.3	13.5	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.45	9.38	9.29	9.20	9.11	9.02		
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.06	6.97	6.88		
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.82	5.74	5.65		
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.27	5.20	5.12	5.03	4.95	4.86		
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.48	4.40	4.31		
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.32	4.25	4.17	4.08	4.00	3.91		
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.78	3.69	3.60		
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.77	3.70	3.62	3.54	3.45	3.36		
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.58	3.51	3.43	3.34	3.25	3.17		
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.42	3.35	3.27	3.18	3.09	3.00		
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.05	2.96	2.87		
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.17	3.10	3.02	2.93	2.84	2.75		
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.83	2.75	2.65		
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	2.99	2.92	2.84	2.75	2.66	2.57		
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.91	2.84	2.76	2.67	2.58	2.49		
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.85	2.78	2.69	2.61	2.52	2.42		
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.79	2.72	2.64	2.55	2.46	2.36		
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.74	2.67	2.58	2.50	2.40	2.31		
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.69	2.62	2.54	2.45	2.35	2.26		
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.65	2.58	2.49	2.40	2.31	2.21		
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.61	2.53	2.45	2.36	2.27	2.17		
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.46	2.39	2.30	2.21	2.11	2.01		
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.28	2.20	2.11	2.02	1.92	1.80		
60	7.06	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.10	2.03	1.94	1.84	1.73	1.60		
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.94	1.86	1.76	1.66	1.53	1.38		
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.78	1.70	1.59	1.47	1.32	1.00		

VRIJHEIDSGRADEN VAN DE NOEMER

BENADERING

VOOR v_1 EN v_2 GROTER DAN 30 KUNNEN DE WAARDEN VAN F (.95) ALS VOLGT BEREKEND WORDEN

$$10 \log F_{v_2}^{v_1} (.95) \approx \frac{1.4287}{\sqrt{h-0.95}} - 0.681 \left(\frac{1}{v_1} - \frac{1}{v_2} \right), \text{ WAARIN } \frac{1}{h} = \frac{1}{2} \left(\frac{1}{v_1} + \frac{1}{v_2} \right).$$

EN F (.99) UIT

$$10 \log F_{v_2}^{v_1} (.99) \approx \frac{2.0206}{\sqrt{h-1.40}} - 1.073 \left(\frac{1}{v_1} - \frac{1}{v_2} \right), \text{ WAARIN } \frac{1}{h} = \frac{1}{2} \left(\frac{1}{v_1} + \frac{1}{v_2} \right).$$

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

TABEL 5

BETROUWBAARHEIDSINTERVALLEN VOOR EEN
STANDAARDAFWIJKING σ , BIJ NORMALE POPULATIES

FAKTOREN a_1 EN a_2 VOOR HET BEREKENEN VAN DE BETROUWBAARHEIDSGRENZEN

	BETROUWBAARHEID							
	.95		.975		.99		.995	
EENZIJDIG								
TWEEZIJDIG	.90		.95		.98		.99	
v	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
1	0.51	15.9	0.45	31.9	0.39	79.8	0.36	160.
2	0.58	4.41	0.52	6.28	0.47	9.98	0.43	14.1
3	0.62	2.92	0.57	3.73	0.51	5.10	0.48	6.47
4	0.65	2.37	0.60	2.87	0.55	3.62	0.52	4.40
5	0.67	2.09	0.62	2.45	0.58	3.00	0.55	3.48
6	0.69	1.92	0.64	2.20	0.60	2.62	0.57	2.98
7	0.71	1.80	0.66	2.04	0.62	2.38	0.59	2.66
8	0.72	1.71	0.68	1.92	0.63	2.20	0.60	2.44
9	0.73	1.65	0.69	1.83	0.64	2.08	0.62	2.28
10	0.74	1.59	0.70	1.75	0.66	1.98	0.63	2.15
15	0.77	1.44	0.74	1.55	0.70	1.69	0.68	1.81
20	0.80	1.36	0.77	1.44	0.73	1.56	0.71	1.64
30	0.83	1.27	0.80	1.34	0.77	1.42	0.75	1.48
40	0.85	1.23	0.82	1.28	0.79	1.34	0.77	1.39
50	0.86	1.20	0.84	1.24	0.81	1.30	0.79	1.34
60	0.86	1.18	0.85	1.22	0.82	1.27	0.81	1.30
70	0.88	1.16	0.86	1.20	0.83	1.24	0.82	1.27
80	0.89	1.15	0.87	1.18	0.84	1.22	0.83	1.25
90	0.89	1.14	0.87	1.17	0.85	1.21	0.84	1.23
100	0.90	1.13	0.88	1.16	0.86	1.19	0.84	1.22
u	1.64		1.96		2.33		2.58	

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

BIJ EEN TWEEZIJDIGE BETROUWBAARHEID VAN $P = .95$ EN EEN AANTAL VRIJHEIDSGRADEN, $v = 9$, VINDT MEN IN DE TABEL DE BIJBEHORENDE WAARDEN VAN $a_1 = 0.69$ EN $a_2 = 1.83$. HET BETROUWBAARHEIDSINTERVAL VOOR σ WORDT ALS VOLGT GEVORMD:

EENZIJDIG: $a_1 s < \sigma$ OF $a_2 s > \sigma$

TWEEZIJDIG: $a_1 s < \sigma < a_2 s$.

v IS HET AANTAL VRIJHEIDSGRADEN VAN DE VARIANTIESCHATTING (ZIE INLEIDING).

TOEPASSING

ZIE INLEIDING VOOR HET VORMEN VAN BETROUWBAARHEIDSINTERVALLEN MET BEHULP VAN DE GEGEVEN WAARDEN VAN a_1 EN a_2 .

BENADERING

VOOR $v > 30$ KUNNEN DE GRENZEN BEREKEND WORDEN UIT :

$$a_1 = \frac{\sqrt{2v}}{\sqrt{2v} + u} \quad \text{EN} \quad a_2 = \frac{\sqrt{2v}}{\sqrt{2v} - u}$$

DE WAARDEN VAN u , BEHORENDE BIJ DE GEWENSTE BETROUWBAARHEID ZIJN ONDERAAN DE TABEL GEGEVEN.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE

TABEL 6

HET BEPALEN VAN EEN ZUIVERE SCHATTING, s^2 , VAN DE VARIANTIE σ^2 UIT EEN GEMIDDELDE SPREIDINGSBREEDTE \bar{R} EN HET BENADEREN VAN DE VERDELING VAN \bar{R}^2 DOOR DE χ^2 -VERDELING.

TABEL A: DE FACTOR B^2

k \ n	2	3	4	5	6	7	8	9	10
1	.503	.272	.199	.163	.140	.125	.114	.105	.099
2	.610	.307	.216	.174	.148	.130	.118	.110	.102
3	.661	.319	.222	.177	.150	.132	.120	.110	.103
4	.683	.327	.225	.178	.151	.133	.121	.111	.104
5	.706	.330	.227	.180	.153	.134	.121	.112	.104
10	.743	.338	.231	.183	.154	.135	.122	.113	.105
∞	.783	.350	.236	.184	.156	.137	.123	.113	.105

TABEL B: HET AANTAL VRIJHEIDSGRADEN, v

k \ n	2	3	4	5	6	7	8	9	10
1	1.0	2.0	2.9	3.8	4.7	5.5	6.3	7.0	7.7
2	1.9	3.8	5.7	7.5	9.2	10.8	12.3	13.8	15.1
3	2.8	5.7	8.4	11.1	13.6	16.0	18.3	20.5	22.6
4	3.7	7.5	11.2	14.7	18.1	21.3	24.4	27.3	30.1
5	4.6	9.3	13.9	18.4	22.6	26.6	30.4	34.0	37.5
10	9.0	18.4	27.6	36.5	44.9	52.9	60.6	67.8	74.8
$\frac{\Delta v}{\Delta k}$	0.88	1.82	2.74	3.62	4.47	5.27	6.03	6.76	7.45

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

IS \bar{R} DE GEMIDDELDE SPREIDINGSBREEDTE GEVONDEN UIT k SERIES IEDER VAN n WAARNE-
MINGEN, DAN IS $s^2 = \bar{R}^2 B^2$ EEN ZUIVERE SCHATTING VAN σ^2 . DE VERDELING VAN
 $\frac{\sqrt{R^2 B^2}}{\sigma^2}$ KAN MEN BENADEREN DOOR EEN χ^2 -VERDELING MET v VRIJHEIDSGRADEN. $\frac{\Delta v}{\Delta k}$ IS DE
TOENAME VAN v GEDEELD DOOR DE TOENAME VAN k , WAARMEE MEN HET AANTAL VRIJHEIDS-
GRADEN VOOR $k > 5$ KAN BEREKENEN.

TOEPASSING

HET AANTAL VRIJHEIDSGRADEN UIT TABEL B, DIE ALHOEWEL GEBROKEN, KUNNEN WORDEN
AFGEROND, EN DE VARIANTIESCHATTING VERKREGEN MET BEHULP VAN TABEL A KUNNEN
GEBRUIKT WORDEN VOOR EEN, ZIJ HET BENADERDE TOETSING MET BEHULP VAN DE F -
VERDELING OF OM MET BEHULP VAN TABEL 5 EEN BETROUWBAARHEIDSINTERVAL VOOR σ
FAST TE STELLEN.

VOOR EEN ZUIVERE SCHATTING VAN DE STANDAARDAFWIJKING UIT \bar{R} ZIE TABEL 19, BLZ. 30

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

TABEL 7

VERDELING VAN DE STEEKPROEF-CORRELATIE-
COEFFICIENT, r , INDIEN $\rho = 0$.

FRACTIELEN

EENZIJDIG TWEEZIJDIG	BETROUWBAARHEID				EENZIJDIG TWEEZIJDIG	BETROUWBAARHEID			
	.975	.99	.995	.9995		.975	.99	.995	.9995
n					n				
3	0.997	1.000	1.000	1.000	18	0.468	0.543	0.590	0.708
4	0.950	0.980	0.990	0.999	19	0.456	0.529	0.575	0.693
5	0.878	0.934	0.959	0.991	20	0.444	0.516	0.561	0.679
6	0.811	0.882	0.917	0.974	25	0.396	0.462	0.505	0.618
7	0.754	0.833	0.875	0.951	30	0.361	0.423	0.463	0.568
8	0.707	0.789	0.834	0.925	35	0.334	0.392	0.430	0.533
9	0.666	0.750	0.798	0.898	40	0.312	0.367	0.403	0.501
10	0.632	0.715	0.765	0.872	45	0.294	0.346	0.380	0.475
11	0.602	0.685	0.735	0.847	50	0.279	0.329	0.361	0.452
12	0.576	0.658	0.708	0.823	55	0.266	0.313	0.345	0.432
13	0.553	0.634	0.684	0.801	60	0.254	0.300	0.330	0.415
14	0.532	0.612	0.661	0.780	70	0.235	0.278	0.306	0.386
15	0.514	0.592	0.641	0.760	80	0.220	0.260	0.287	0.362
16	0.497	0.574	0.623	0.742	90	0.208	0.245	0.271	0.342
17	0.482	0.558	0.606	0.725	100	0.197	0.233	0.257	0.325
u	1.96	2.33	2.58	3.30	u	1.96	2.33	2.58	3.30
u ²	3.84	5.43	6.63	10.89	u ²	3.84	5.43	6.63	10.89

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN ZOEKT IN DE TABEL HET GETAL VOORKOMENDE IN DE RIJ MET HET GEGEVEN AANTAL PAREN WAARNEMINGEN (n) EN IN DE KOLOM MET DE GEWENSTE BETROUWBAARHEID (P). DIT GETAL IS DE GEWENSTE WAARDE VAN r_n .

VOORBEELD: $P(r_8 < 0.789) = .99$.

TOEPASSING

DEZE TABEL KAN GEBRUIKT WORDEN OM TE TOETSEN OF DE POPULATIE-CORRELATIE-COEFFICIENT ρ GELIJK AAN NUL IS.

BENADERING

VOOR $n > 20$ IS r TE BENADEREN MET:

$$r = u / \sqrt{n - 2 + u^2}$$

DE WAARDEN VAN u EN u^2 (BEHORENDE BIJ DE GEWENSTE BETROUWBAARHEID) ZIJN ONDERAAN DE TABEL GEGEVEN.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

TABEL 8

BETROUWBAARHEIDSINTERVALLEN VOOR HET POPULATIE-
GEMIDDELTE μ VAN EEN EXPONENTIELE VERDELINGFAKTOREN a_1 EN a_2 VOOR HET BEREKENEN VAN HET BETROUWBAARHEIDSINTERVAL

EENZIJDIG TWEEZIJDIG	BETROUWBAARHEID							
	.95		.975		.99		.995	
	.90		.95		.98		.99	
m	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
1	0.33	19.42	0.27	39.53	0.22	99.50	0.19	200.00
2	0.42	5.63	0.36	8.26	0.30	13.47	0.27	19.32
3	0.48	3.66	0.42	4.84	0.36	6.88	0.32	8.88
4	0.52	2.93	0.46	3.70	0.40	4.85	0.36	5.97
5	0.55	2.54	0.49	3.08	0.43	3.91	0.40	4.63
6	0.57	2.29	0.52	2.73	0.46	3.36	0.42	3.91
7	0.59	2.13	0.54	2.49	0.48	3.00	0.45	3.44
8	0.61	2.01	0.56	2.32	0.50	2.75	0.47	3.11
9	0.62	1.92	0.57	2.19	0.52	2.57	0.48	2.88
10	0.64	1.84	0.58	2.09	0.53	2.42	0.50	2.69
11	0.65	1.79	0.60	2.00	0.55	2.31	0.51	2.55
12	0.66	1.74	0.61	1.94	0.56	2.20	0.53	2.43
13	0.67	1.69	0.62	1.88	0.57	2.13	0.54	2.32
14	0.68	1.66	0.63	1.83	0.58	2.06	0.55	2.24
15	0.68	1.62	0.64	1.79	0.59	2.00	0.56	2.17
16	0.69	1.59	0.65	1.75	0.60	1.95	0.57	2.12
17	0.70	1.57	0.65	1.72	0.61	1.91	0.58	2.06
18	0.71	1.55	0.66	1.69	0.61	1.88	0.58	2.01
19	0.71	1.53	0.67	1.66	0.62	1.84	0.59	1.97
20	0.72	1.51	0.67	1.64	0.63	1.80	0.60	1.93
22	0.73	1.48	0.69	1.59	0.64	1.75	0.61	1.86
24	0.74	1.45	0.70	1.56	0.65	1.70	0.62	1.81
26	0.74	1.43	0.70	1.53	0.66	1.67	0.63	1.76
28	0.75	1.41	0.71	1.51	0.67	1.63	0.64	1.72
30	0.76	1.39	0.72	1.48	0.68	1.60	0.65	1.69
u	1.64		1.96		2.33		2.58	

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

BIJ EEN TWEEZIJDIGE BETROUWBAARHEID VAN $P = .95$ EN EEN AANTAL WAARNEMINGEN, m , WAAROP HET STEEKPROEFGEMIDDELTE, \bar{x} , GEBASEERD IS, VINDT MEN ALS $m = 7$ GEKOZEN WORDT, IN DE TABEL DE BIJBEHORENDE WAARDEN VAN

$$a_1 = 0.54 \text{ EN } a_2 = 2.49.$$

HET BETROUWBAARHEIDSINTERVAL VOOR μ WORDT ALS VOLGT GEVORMD:

$$\text{EENZIJDIG: } a_1 \bar{x} < \mu \text{ OF } a_2 \bar{x} > \mu$$

$$\text{TWEEZIJDIG: } a_1 \bar{x} < \mu < a_2 \bar{x}$$

TOEPASSING

ZIE INLEIDING VOOR DE TOEPASSINGEN.

BENADERING

VOOR $m > 30$ KUNNEN DE GRENZEN BEREKEND WORDEN UIT: $a_{1,2} = \frac{4m}{(\sqrt{4m} \pm u)^2}$

DE WAARDEN VAN u , BEHORENDE BIJ DE GEWENSTE BETROUWBAARHEID ZIJN ONDERAAN DE TABEL GEGEVEN.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

BINOMIALE VERDELING

DE KANS DAT EEN TREKKING, m , UIT EEN BINOMIALE VERDELING $\leq c$ ALS FUNCTIE VAN n EN p .

n	c	p										$\frac{1}{6}$	$\frac{1}{3}$	
		.01	.05	.10	.15	.20	.25	.30	.35	.40	.45			.50
2	0	.9801	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500	.6944	.4444
	1	.9999	.9975	.9900	.9775	.9600	.9375	.9100	.8775	.8400	.7975	.7500	.9722	.8889
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	.9703	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250	.5787	.2963
	1	.9997	.9928	.9720	.9392	.8960	.8438	.7840	.7182	.6480	.5748	.5000	.9259	.7407
	2	1.0000	.9999	.9990	.9966	.9920	.9844	.9730	.9571	.9360	.9089	.8750	.9954	.9630
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	.9606	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625	.4822	.1975
	1	.9994	.9860	.9477	.8905	.8192	.7383	.6517	.5630	.4752	.3910	.3125	.8681	.5926
	2	1.0000	.9995	.9963	.9880	.9728	.9492	.9163	.8735	.8208	.7585	.6875	.9838	.8889
	3	1.0000	1.0000	.9999	.9995	.9984	.9961	.9919	.9850	.9744	.9590	.9375	.9992	.9877
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	.9510	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0312	.4019	.1317
	1	.9990	.9774	.9185	.8352	.7373	.6328	.5282	.4284	.3370	.2562	.1875	.8038	.4609
	2	1.0000	.9988	.9914	.9734	.9421	.8965	.8369	.7648	.6826	.5931	.5000	.9645	.7901
	3	1.0000	1.0000	.9995	.9978	.9933	.9844	.9692	.9460	.9130	.8688	.8125	.9967	.9547
	4	1.0000	1.0000	.9999	.9999	.9997	.9990	.9976	.9948	.9898	.9816	.9688	.9999	.9959
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	.9415	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156	.3349	.0878
	1	.9985	.9672	.8857	.7765	.6554	.5339	.4202	.3191	.2333	.1636	.1094	.7368	.3512
	2	1.0000	.9978	.9842	.9527	.9011	.8306	.7443	.6471	.5443	.4415	.3438	.9377	.6804
	3	1.0000	.9999	.9987	.9941	.9830	.9624	.9295	.8726	.8208	.7447	.6562	.9913	.8999
	4	1.0000	1.0000	.9999	.9996	.9984	.9954	.9891	.9777	.9590	.9308	.8906	.9993	.9822
	5	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.9993	.9982	.9959	.9917	.9844	1.0000	.9986
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	.9321	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078	.2791	.0585
	1	.9980	.9556	.8503	.7166	.5767	.4450	.3294	.2338	.1586	.1024	.0625	.6698	.2634
	2	1.0000	.9962	.9743	.9262	.8520	.7564	.6471	.5323	.4199	.3164	.2266	.9042	.5706
	3	1.0000	.9998	.9973	.9879	.9667	.9294	.8740	.8002	.7102	.6083	.5000	.9824	.8267
	4	1.0000	1.0000	.9998	.9988	.9953	.9871	.9712	.9444	.9037	.8471	.7734	.9980	.9547
5	5	1.0000	1.0000	1.0000	.9999	.9996	.9987	.9962	.9910	.9812	.9643	.9375	.9999	.9931
	6	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.9994	.9984	.9963	.9922	1.0000	.9995
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

BINOMIALE VERDELING (VERVOLG)

n	c	P										$\frac{1}{6}$	$\frac{1}{3}$	
		.01	.05	.10	.15	.20	.25	.30	.35	.40	.45			.50
8	0	.9227	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039	.2326	.0391
	1	.9973	.9428	.8131	.6572	.5033	.3671	.2553	.1691	.1064	.0632	.0352	.6047	.1951
	2	1.000	.9942	.9619	.8948	.7969	.6786	.5518	.4278	.3154	.2201	.1445	.8652	.4682
	3	1.000	.9996	.9950	.9786	.9437	.8862	.8059	.7064	.5941	.4770	.3633	.9693	.7413
	4	1.000	1.000	.9996	.9972	.9896	.9727	.9420	.8939	.8263	.7396	.6367	.9954	.9121
9	0	1.000	1.000	1.000	.9998	.9988	.9958	.9887	.9747	.9502	.9115	.8555	.9996	.9803
	1	1.000	1.000	1.000	1.000	.9999	.9996	.9987	.9964	.9915	.9819	.9648	1.000	.9974
	2	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9998	.9993	.9983	.9961	1.000	.9998
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	1.000	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020	.2038	.0260
10	0	.9135	.9288	.7748	.5995	.4362	.3004	.1960	.1211	.0705	.0385	.0195	.5427	.1431
	1	.9966	.9288	.7748	.5995	.4362	.3004	.1960	.1211	.0705	.0385	.0195	.5427	.1431
	2	.9999	.9916	.9470	.8592	.7382	.6007	.4628	.3373	.2318	.1495	.0898	.8217	.3772
	3	1.000	.9994	.9917	.9661	.9144	.8343	.7297	.6089	.4826	.3614	.2539	.9520	.6503
	4	1.000	1.000	.9991	.9944	.9804	.9511	.9012	.8283	.7334	.6214	.5000	.9910	.8552
10	5	1.000	1.000	.9999	.9994	.9969	.9900	.9747	.9464	.9007	.8342	.7461	.9989	.9576
	6	1.000	1.000	1.000	1.000	.9997	.9987	.9957	.9888	.9750	.9502	.9102	.9999	.9917
	7	1.000	1.000	1.000	1.000	1.000	.9999	.9996	.9986	.9962	.9909	.9805	1.000	.9990
	8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9997	.9992	.9980	1.000	1.000
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0	.9044	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010	.1615	.0173
	1	.9957	.9139	.7361	.5443	.3758	.2440	.1493	.0860	.0464	.0233	.0107	.4845	.1040
	2	.9999	.9885	.9298	.8202	.6778	.5256	.3828	.2616	.1673	.0996	.0547	.7752	.2991
	3	1.000	.9990	.9872	.9500	.8791	.7759	.6496	.5138	.3823	.2670	.1719	.9303	.5593
	4	1.000	.9999	.9984	.9901	.9672	.9219	.8497	.7515	.6331	.5044	.3770	.9845	.7969
10	5	1.000	1.000	.9998	.9986	.9936	.9803	.9526	.9051	.8338	.7384	.6230	.9976	.9234
	6	1.000	1.000	1.000	.9999	.9991	.9965	.9894	.9740	.9452	.8980	.8281	.9997	.9803
	7	1.000	1.000	1.000	1.000	.9999	.9996	.9984	.9952	.9877	.9726	.9453	1.000	.9966
	8	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9995	.9983	.9955	.9893	1.000	.9996
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9997	.9990	1.000	1.000
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

$P(m < 2 | n = 6, p = .10) = .9842$ BETEKENT, DAT BIJ EEN STEEKPROEFGROOTTE VAN 6 STUKS UIT EEN BINOMIAAL VERDEELDE POPULATIE DE KANS OP 2 OF MINDER DEFECTIEVEN BIJ EEN WAARDE VAN $p = 0.10$ GELIJK IS AAN 0.9842. MEN VINDT DEZE WAARDE DOOR HET VAKJE BEHORENDE BIJ $n = 6$, $c = 2$ EN $p = 0.10$ OP TE ZOEKEN.

BENADERING

ZIE VOOR BENADERING MET DE NORMALE VERDELING OF DE POISSON-VERDELING TABEL 13, BLZ.24.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

TABEL 10

BETROUWBAARHEIDSGRENZEN VOOR DE PARAMETER p
VAN DE BINOMIALE VERDELING MET GEMIDDELDE
BETROUWBAARHEID.

BETROUWBAARHEIDSGRENZEN a_1 EN a_2 IN PROCENTEN

EENZIJDIG TWEEZIJDIG		GEMIDDELDE BETROUWBAARHEID							
		.95		.975		.99		.995	
		.90		.95		.98		.99	
n	m	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
5	0	0.0	28	0.0	31	0.0	34	0.0	37
	1	4.3	55	2.9	62	1.8	69	1.2	73
	2	13	73	10	78	7.0	84	5.3	87
	3	27	87	22	90	16	93	13	95
	4	45	96	38	97	31	98	27	99
	5	72	100	69	100	66	100	63	100
10	0	0.0	16.	0.0	19	0.0	22	0.0	24
	1	2.1	33	1.4	38	0.8	44	0.6	48
	2	6.2	45	4.6	50	3.1	56	2.4	60
	3	12	56	9.4	60	7.0	66	5.7	69
	4	19	65	15	70	12.	74	10	77
	5	26	74	22	78	18	82	16	84
	6	35	81	30	85	26	88	23	90
	7	44	88	40	91	34	93	31	94
	8	55	94	50	95	44	97	40	98
	9	67	98	62	99	56	99	52	99
10	84	100	81	100	78	100	76	100	
15	0	0.0	12	0.0	13	0.0	16	0.0	18
	1	1.4	23	0.9	27	0.6	32	0.4	35
	2	4.0	32	3.0	36	2.0	41	1.6	44
	3	7.6	40	6.1	44	4.5	49	3.6	52
	4	12	48	9.8	52	7.7	56	6.4	59
	5	17	54	14	58	12	63	9.9	66
	6	22	61	19	65	16	69	14	72
	7	27	67	24	71	20	74	18	77
	8	33	73	29	76	26	80	23	82
	9	39	78	35	81	31	84	28	86
	10	46	83	42	86	37	88	34	90
	11	52	88	48	90	44	92	41	94
	12	60	92	56	94	51	96	48	96
	13	68	96	64	97	59	98	56	98
	14	77	99	73	99	68	99	65	100
15	88	100	87	100	84	100	82	100	
20	0	0.0	8.9	0.0	10	0.0	12	0.0	13
	1	1.0	18	0.7	21	0.4	25	0.3	28
	2	3.0	25	2.2	28	1.5	32	1.1	35
	3	5.6	31	4.5	35	3.3	39	2.7	42
	4	8.8	37	7.2	41	5.6	45	4.7	48
	5	12	43	10	46	8.3	50	7.1	53
	6	16	48	14	52	11	56	9.9	58
	7	20	53	17	57	15	61	13	63
	8	24	58	21	62	18	65	16	68
	9	28	63	25	66	22	70	20	72
	10	32	68	29	71	26	74	24	76
	11	37	72	34	75	30	78	28	80
	12	42	76	38	79	35	82	32	84
	13	47	80	43	83	39	85	37	87
	14	52	84	48	86	44	89	42	90
	15	57	88	54	90	50	92	47	93
	16	63	91	59	93	55	94	52	95
	17	69	94	65	96	61	97	58	97
	18	75	97	72	98	68	98	65	99
	19	82	99	79	99	75	100	72	100
20	91	100	90	100	88	100	87	100	

BETROUWBAARHEIDSGRENZEN a_1 EN a_2 IN PROCENTEN

EENZIJDIG TWEEZIJDIG		GEMIDDELDE BETROUWBAARHEID							
		.95		.975		.99		.995	
		.90		.95		.98		.99	
n	m	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
25	0	0.0	7.2	0.0	8.5	0.0	10	0.0	11
	1	0.8	14	0.5	17	0.3	20	0.2	23
	2	2.4	20	1.8	23	1.2	27	0.9	29
	3	4.6	26	3.5	29	2.6	32	2.1	35
	4	7.0	31	5.7	34	4.4	37	3.7	40
	5	9.6	35	8.1	38	6.5	42	5.6	45
	6	12	40	11	43	8.9	47	7.8	49
	7	16	44	14	47	12	51	10	54
	8	19	48	16	51	14	55	13	58
	9	22	52	20	56	17	59	15	62
	10	25	56	23	59	20	63	18	65
	11	29	60	26	63	23	67	21	69
	12	32	64	30	67	26	70	24	72
	13	36	68	33	70	30	74	28	76
	14	40	71	37	74	33	77	31	79
	15	44	75	41	77	37	80	35	82
	16	48	78	44	80	41	83	38	85
	17	52	81	49	84	45	86	42	87
	18	56	85	53	86	49	89	46	90
	19	60	88	57	89	53	91	51	92
20	65	90	62	92	58	93	55	94	
30	0	0.0	6.1	0.0	7.6	0.0	8.6	0.0	9.6
	1	0.7	12	0.4	15	0.2	17	0.2	19
	2	2.0	17	1.5	20	1.0	23	0.8	25
	3	3.8	22	2.6	24	2.2	28	1.7	30
	4	5.8	26	4.4	29	3.7	32	3.1	34
	5	8.0	30	6.5	33	5.4	36	4.6	38
	6	10	34	8.7	37	7.3	40	6.4	42
	7	13	38	11	40	9.4	44	8.3	46
	8	15	41	14	44	12	48	10	50
	9	18	45	16	48	14	51	12	53
	10	21	48	19	51	16	54	15	57
	11	24	52	21	54	19	58	17	60
	12	26	55	24	58	21	61	20	63
	13	29	58	27	61	24	64	22	66
	14	32	61	30	64	27	67	25	69
	15	35	65	33	67	30	70	28	72
	16	39	68	36	70	33	73	31	75
	17	42	71	39	73	36	76	34	78
	18	45	74	42	76	39	78	37	80
	19	48	76	46	79	42	81	40	83
20	52	79	49	81	46	84	43	85	

ZIE VERDER BLADZIJDE 20.

TABEL 10 (VERVOLG) BETROUWBAARHEIDSGRENZEN a_1 EN a_2 IN PROCENTEN

EENZIJDIG TWEEZIJDIG		GEMIDDELDE BETROUWBAARHEID							
		.95		.975		.99		.995	
		.90		.95		.98		.99	
n	m	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
40	0	0.0	4.6	0.0	5.8	0.0	6.6	0.0	7.4
	1	0.5	9.2	0.3	12	0.2	13	0.1	15
	2	1.5	13	1.1	15	0.7	17	0.6	19
	3	2.8	17	2.2	19	1.6	21	1.3	23
	4	4.2	20	3.5	22	2.7	25	2.3	27
	5	5.8	23	4.9	25	4.0	28	3.4	30
	6	7.6	26	6.5	28	5.4	31	4.7	33
	7	9.4	29	8.2	31	6.9	34	6.1	36
	8	11	32	10	34	8.5	37	7.6	39
	9	13	35	12	37	10	40	9.1	42
	10	15	37	14	40	12	43	11	45
	11	17	40	16	42	14	45	13	48
	12	19	43	18	45	16	48	14	50
	13	22	45	20	48	18	51	16	53
	14	24	48	22	50	20	53	18	55
	15	26	50	24	53	22	56	20	58
	16	28	53	26	55	24	58	22	60
	17	30	55	28	58	26	61	25	63
	18	33	58	30	60	28	63	27	65
	19	35	60	33	63	30	65	28	67
	20	37	63	35	65	32	68	31	69
50	0	0.0	3.7	0.0	4.5	0.0	5.3	0.0	6.0
	1	0.4	7.4	0.3	9.1	0.2	11	0.1	12
	2	1.2	11	0.9	12	0.6	14	0.4	16
	3	2.2	13	1.7	15	1.3	17	1.0	19
	4	3.4	16	2.8	18	2.2	20	1.8	22
	5	4.6	19	3.9	20	3.2	23	2.7	25
	6	6.0	21	5.2	23	4.3	25	3.7	27
	7	7.4	24	6.5	26	5.5	28	4.8	30
	8	9.0	26	7.9	28	6.7	30	6.0	32
	9	10	28	9.3	30	8.0	33	7.2	35
	10	12	30	11	33	9.4	35	8.5	37
	11	14	33	12	35	11	37	9.9	39
	12	15	35	14	37	12	40	11	42
	13	17	37	15	39	14	42	13	44
	14	19	39	17	42	15	44	14	46
	15	20	41	19	44	17	46	16	48
	16	22	43	20	46	18	48	17	50
	17	24	46	22	48	20	50	19	52
	18	26	48	24	50	22	52	20	54
	19	27	50	26	52	23	54	22	56
	20	29	52	27	53	25	56	24	58

ZIE VERDER BLADZIJDE 21.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

BETROUWBAARHEIDSGRENZEN a_1 EN a_2 IN PROCENTEN

EENZIJDIG TWEEZIJDIG		GEMIDDELDE BETROUWBAARHEID							
		.95		.975		.99		.995	
		.90		.95		.98		.99	
n	m	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
75	0	0.0	2.5	0.0	3.0	0.0	3.6	0.0	4.1
	1	0.3	5.0	0.2	6.0	0.1	7.2	0.1	8.2
	2	0.8	7.2	0.6	8.2	0.4	9.6	0.3	11
	3	1.4	9.1	1.2	10	0.8	12	0.7	13
	4	2.2	11	1.8	12	1.4	14	1.2	15
	5	3.0	13	2.6	14	2.1	16	1.8	17
	6	4.0	14	3.4	15	2.8	17	2.4	19
	7	4.9	16	4.3	17	3.6	19	3.2	20
	8	5.9	18	5.2	19	4.4	21	3.9	22
	9	6.9	19	6.1	21	5.2	23	4.7	24
	10	7.9	21	7.0	22	6.1	24	5.5	26
	11	9.0	22	8.0	24	7.0	26	6.4	27
	12	10	24	9.0	26	8.0	28	7.3	29
	13	11	25	10	27	9.0	29	8.2	30
	14	12	27	11	28	9.9	31	9.2	32
	15	13	28	12	30	11	32	10	34
	16	14	30	13	31	12	34	11	35
	17	16	31	14	33	13	35	12	37
	18	17	33	15	34	14	37	13	38
	19	18	34	16	36	15	38	14	40
20	19	36	18	38	16	40	15	41	
100	0	0.0	1.9	0.0	2.6	0.0	2.7	0.0	3.1
	1	0.2	3.8	0.1	5.1	0.1	5.4	0.1	6.2
	2	0.6	5.4	0.4	6.2	0.3	7.3	0.2	8.1
	3	1.1	6.8	0.9	7.8	0.6	8.9	0.5	9.7
	4	1.7	8.2	1.4	9.2	1.0	10	0.9	11
	5	2.3	9.6	1.9	11	1.5	12	1.3	13
	6	3.0	11	2.5	12	2.1	13	1.8	14
	7	3.6	12	3.2	13	2.7	15	2.4	16
	8	4.4	13	3.8	14	3.3	16	2.9	17
	9	5.2	15	4.5	15	3.9	17	3.5	18
	10	5.9	16	5.2	17	4.6	19	4.1	20
	11	6.6	17	6.0	18	5.3	20	4.8	21
	12	7.4	18	6.7	19	6.0	21	5.4	22
	13	8.2	19	7.5	21	6.7	22	6.1	23
	14	9.0	20	8.2	22	7.4	23	6.7	25
	15	9.8	22	9.0	23	8.1	25	7.5	26
	16	11	23	9.8	24	8.8	26	8.2	27
	17	12	24	11	25	9.6	27	8.8	28
	18	12	25	11	26	10	28	9.6	29
	19	13	26	12	28	11	29	10	30
20	14	27	13	29	12	30	12	32	

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN VINDT IN DE TABEL BIJ EEN GEGEVEN STEEKPROEFGROOTTE n , BIJ EEN GEGEVEN AANTAL DEFECTIEVEN, m , EN BIJ EEN GEWENSTE GEMIDDELDE BETROUWBAARHEID, p , HET BETROUWBAARHEIDSINTERVAL VOOR DE PARAMETER p . HET BETROUWBAARHEIDSINTERVAL VOOR p WORDT ALS VOLGT GEVORMD:

$$\text{EENZIJDIG: } a_1 < p \quad \text{OF} \quad a_2 > p$$

$$\text{TWEEZIJDIG: } a_1 < p < a_2$$

TOEPASSING

ZIE INLEIDING VOOR DE TOEPASSING VAN DEZE TABEL.

BENADERING

VOOR NIET GETABELLEERDE WAARDEN VAN n KAN HET INTERVAL BEREKEND WORDEN UIT

$$na_{1,2} = m + u^2 \left(\frac{1}{2} - \frac{m}{n} \right) \pm u \sqrt{m \left(1 - \frac{m}{n} \right)}$$

TABEL 11

POISSON-VERDELING

DE KANS, DAT EEN TREKKING, m , UIT EEN POISSONVERDELING $\leq c$ ALS FUNCTIE VAN μ

$\mu \backslash c$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
.05	.9512	.9988	1.000												
.10	.9048	.9953	.9998	1.000											
.15	.8607	.9898	.9995	1.000											
.20	.8187	.9825	.9988	.9999	1.000										
.25	.7788	.9735	.9978	.9999	1.000										
.30	.7408	.9631	.9964	.9997	1.000										
.35	.7047	.9513	.9945	.9995	1.000										
.40	.6703	.9384	.9921	.9992	.9999	1.000									
.45	.6376	.9246	.9891	.9988	.9999	1.000									
.50	.6065	.9098	.9856	.9982	.9998	1.000									
.55	.5770	.8943	.9815	.9975	.9997	1.000									
.60	.5488	.8781	.9769	.9966	.9996	1.000									
.65	.5220	.8614	.9717	.9956	.9994	.9999	1.000								
.70	.4966	.8442	.9659	.9942	.9992	.9999	1.000								
.75	.4724	.8266	.9595	.9927	.9989	.9999	1.000								
.80	.4493	.8088	.9526	.9909	.9986	.9998	1.000								
.85	.4274	.7907	.9451	.9889	.9982	.9998	1.000								
.90	.4066	.7725	.9371	.9865	.9977	.9997	1.000								
.95	.3867	.7541	.9287	.9839	.9970	.9995	.9999	1.000							
1.00	.3679	.7358	.9197	.9810	.9963	.9994	.9999	1.000							
1.10	.3329	.6990	.9004	.9743	.9946	.9990	.9999	1.000							
1.20	.3012	.6626	.8795	.9662	.9922	.9985	.9998	1.000							
1.30	.2725	.6268	.8571	.9569	.9893	.9978	.9996	.9999	1.000						
1.40	.2466	.5918	.8335	.9463	.9858	.9968	.9994	.9999	1.000						
1.50	.2231	.5578	.8088	.9344	.9814	.9955	.9991	.9998	1.000						
1.60	.2019	.5249	.7834	.9212	.9763	.9940	.9987	.9997	1.000						
1.70	.1827	.4932	.7572	.9068	.9704	.9920	.9981	.9996	.9999	1.000					
1.80	.1653	.4628	.7306	.8913	.9636	.9896	.9974	.9994	.9999	1.000					
1.90	.1496	.4338	.7037	.8747	.9559	.9868	.9966	.9992	.9998	1.000					
2.00	.1353	.4060	.6767	.8571	.9474	.9834	.9955	.9989	.9998	1.000					
2.20	.1108	.3546	.6227	.8194	.9275	.9751	.9925	.9980	.9995	.9999	1.000				
2.40	.0907	.3084	.5697	.7787	.9041	.9643	.9884	.9967	.9991	.9998	1.000				
2.60	.0743	.2674	.5184	.7360	.8774	.9510	.9828	.9947	.9985	.9996	.9999	1.000			
2.80	.0608	.2311	.4694	.6919	.8477	.9349	.9756	.9919	.9976	.9993	.9998	1.000			
3.00	.0498	.1992	.4232	.6472	.8153	.9161	.9665	.9881	.9962	.9989	.9997	.9999	1.000		
3.20	.0408	.1712	.3799	.6025	.7806	.8946	.9554	.9832	.9943	.9982	.9995	.9999	1.000		
3.40	.0334	.1468	.3397	.5584	.7442	.8705	.9422	.9769	.9917	.9973	.9992	.9998	.9999	1.000	
3.60	.0273	.1257	.3028	.5152	.7064	.8441	.9267	.9692	.9883	.9960	.9987	.9996	.9999	1.000	
3.80	.0224	.1074	.2689	.4735	.6678	.8156	.9091	.9599	.9840	.9942	.9981	.9994	.9998	1.000	
4.00	.0183	.0916	.2381	.4335	.6288	.7851	.8893	.9489	.9786	.9919	.9972	.9991	.9997	.9999	1.000
4.20	.0150	.0780	.2102	.3954	.5898	.7531	.8675	.9361	.9721	.9889	.9959	.9986	.9996	.9999	1.000
4.40	.0123	.0663	.1851	.3594	.5512	.7199	.8436	.9214	.9642	.9851	.9943	.9980	.9993	.9998	.9999
4.60	.0100	.0563	.1626	.3257	.5132	.6858	.8180	.9050	.9549	.9805	.9922	.9971	.9990	.9997	.9999
4.80	.0082	.0477	.1425	.2942	.4763	.6510	.7908	.8867	.9442	.9749	.9896	.9960	.9986	.9995	.9999
5.00	.0067	.0404	.1246	.2650	.4405	.6160	.7622	.8666	.9319	.9682	.9863	.9946	.9980	.9993	.9998

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

$P(m \leq 3, \mu = 1.3) = .9569$ BETEKENT, DAT BIJ EEN STEEKPROEF UIT EEN POISSON-VERDEELDE POPULATIE MET GEMIDDELDE, $\mu = 1.3$ DE KANS OP 3 OF MINDER DEFECTIEVEN GELIJK IS AAN 0.9569.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

BETROUWBAARHEIDSINTERVALLEN VOOR DE
PARAMETER μ VAN DE POISSONVERDELING
MET GEMIDDELTE BETROUWBAARHEID.
BETROUWBAARHEIDSGRENZEN a_1 EN a_2

	GEMIDDELTE BETROUWBAARHEID							
	.95		.975		.99		.995	
EENZIJDIG								
TWEEZIJDIG	.90		.95		.98		.99	
m	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
0	0.00	1.92	0.00	2.51	0.00	3.32	0.00	3.94
1	0.18	3.91	0.11	4.67	0.06	5.6	0.04	6.4
2	0.58	5.50	0.42	6.42	0.28	7.6	0.21	8.4
3	1.09	7.03	0.85	8.00	0.62	9.2	0.49	10.2
4	1.66	8.46	1.35	9.51	1.04	10.9	0.87	11.8
5	2.29	9.84	1.91	11.0	1.52	12.4	1.30	13.4
6	2.94	11.2	2.50	12.4	2.06	13.9	1.79	14.9
7	3.63	12.5	3.13	13.3	2.62	15.3	2.30	16.4
8	4.34	13.8	3.78	15.1	3.20	16.7	2.85	17.9
9	5.10	15.1	4.46	16.4	3.82	18.1	3.42	19.3
10	5.80	16.3	5.15	17.8	4.45	19.4	4.02	20.7
11	6.55	17.6	5.85	19.0	5.10	20.8	4.63	22.1
12	7.30	18.8	6.55	20.3	5.75	22.2	5.25	23.4
13	8.10	20.1	7.30	21.6	6.45	23.5	5.90	24.8
14	8.85	21.3	8.00	22.8	7.15	24.8	6.55	26.2
15	9.65	22.5	8.75	24.1	7.85	26.1	7.25	27.5
16	10.4	23.7	9.50	25.4	8.55	27.4	7.90	28.8
17	11.3	24.9	10.3	26.6	9.25	28.7	8.60	30.2
18	12.0	26.1	11.1	27.9	10.0	30.0	9.30	31.4
19	12.8	27.3	11.9	29.0	10.7	31.2	10.0	32.7
20	13.6	28.5	12.6	30.3	11.4	32.5	10.7	34.0
u	1.64		1.96		2.33		2.58	
u^2	2.70		3.84		5.43		6.63	

AFLIEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN VINDT IN DE TABEL BIJ EEN GEGEVEN AANTAL DEFECTIEVEN, m , EN BIJ EEN GEWENSTE GEMIDDELTE BETROUWBAARHEID, p , HET BETROUWBAARHEIDSINTERVAL VOOR DE PARAMETER μ . HET BETROUWBAARHEIDSINTERVAL VOOR μ WORDT ALS VOLGT GEVORMD:

$$\text{EENZIJDIG: } a_1 < \mu \text{ OF } a_2 > \mu$$

$$\text{TWEEZIJDIG: } a_1 < \mu < a_2.$$

TOEPASSING

ZIE INLEIDING VOOR DE TOEPASSING VAN DEZE TABEL.

BENADERING

VOOR WAARDEN VAN m GROTER DAN 20 KUNNEN DE GRENZEN WORDEN BEREKEND UIT:

$$a_{1,2} = m + \frac{u^2}{2} \pm u \sqrt{m + \frac{u^2}{4}} = \frac{(u \pm \sqrt{4m + u^2})^2}{4}$$

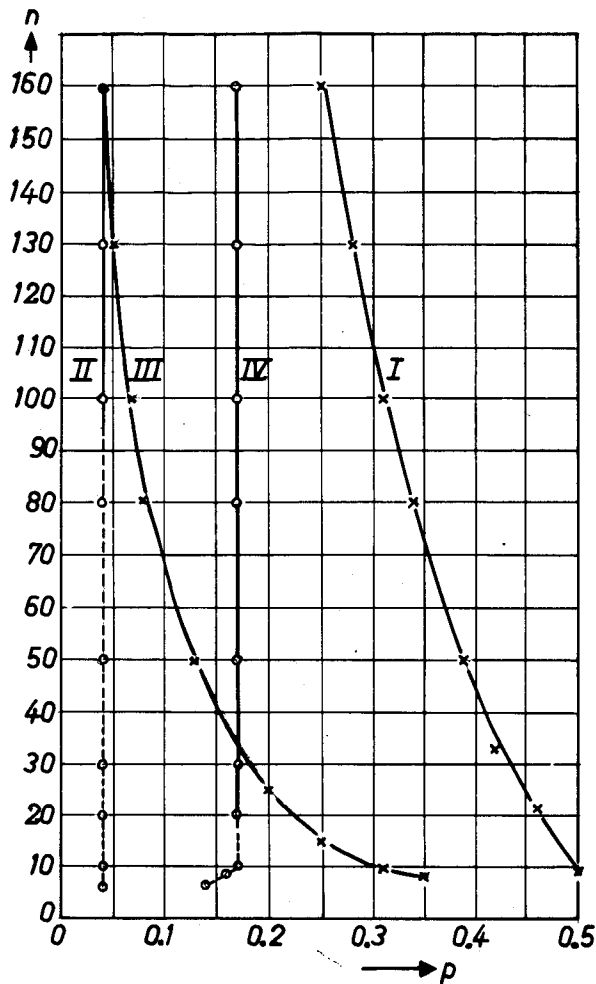
DE WAARDEN VAN u EN u^2 BEHORENDE BIJ DE GEWENSTE BETROUWBAARHEID ZIJN ONDERAAN DE TABEL GEGEVEN.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

TABEL 13 (GRAFIEK)

GEBIEDEN WAARIN DE BINOMIALE VERDELING BENADERD KAN WORDEN DOOR DE NORMALE VERDELING OF DOOR DE POISSON-VERDELING BIJ GEGEVEN MAXIMALE BENADERINGSNAUWKEURIGHEID.



AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

n = STEEKPROEFGROOTTE; p = FRACTIE IN DE BINOMIALE POPULATIE.

- I : GRENS VOOR GEBIED VAN NORMALE BENADERING; LINKS VAN DE LIJN IS DE AFWIJKING OP DE 0.025 EN 0.975 NIVEAU'S GROTER DAN 0.0025, RECHTS IS DE AFWIJKING KLEINER.
- II : GRENS VOOR GEBIED VAN POISSON-BENADERING; RECHTS VAN DE LIJN IS DE AFWIJKING OP DE 0.025 EN 0.975 NIVEAU'S GROTER DAN 0.0025, LINKS IS DE AFWIJKING KLEINER.
- III: GRENS VOOR GEBIED VAN NORMALE BENADERING; LINKS VAN DE LIJN IS DE AFWIJKING OP DE 0.025 EN 0.975 NIVEAU'S GROTER DAN 0.0125, RECHTS IS DE AFWIJKING KLEINER.
- IV : GRENS VOOR GEBIED VAN POISSON-BENADERING; RECHTS VAN DE LIJN IS DE AFWIJKING OP DE 0.025 EN 0.975 NIVEAU'S GROTER DAN 0.0125, LINKS IS DE AFWIJKING KLEINER.

N.B. EEN GESTIPPelde LIJN GEEFT AAN DAT HET 0.025 NIVEAU ONTBREEKT BIJ DE BINOMIALE VERDELING. DE LIJNEN ZIJN CONTINU GETROKKEN; IN FEITE ZIJN ZIJ DISCONTINU, OMDAT EEN PUNT VAN DE BINOMIALE VERDELING WORDT GENOMEN TUSSEN 1% EN 5% EN TUSSEN 95% EN 99% ZO DICHT MOGELIJK BIJ DE TE BESCHOUWEN CUMULATIEVE NIVEAU'S.

VERDELINGSVRIJ BETROUWBAARHEIDSINTERVAL VOOR
DE MEDIAAN $\pi_{.50}$; TEKENTOETS.

BETROUWBAARHEDEN

n \ m	1	2	3	4	5	6	7	8	9	10
4	.9375	.6875	.3125	.0625						
5	.9688	.8125	.5000	.1875	.0312					
6	.9844	.8906	.6562	.3438	.1094	.0156				
7	.9922	.9375	.7734	.5000	.2266	.0625	.0078			
8	.9961	.9648	.8555	.6367	.3633	.1445	.0352	.0039		
9	.9980	.9805	.9102	.7461	.5000	.2539	.0898	.0195	.0020	
10	.9990	.9893	.9453	.8281	.6230	.3770	.1719	.0547	.0107	.0010
11	.9995	.9941	.9673	.8867	.7256	.5000	.2744	.1133	.0327	.0059
12	.9998	.9968	.9807	.9270	.8062	.6128	.3872	.1938	.0730	.0193
13	.9999	.9983	.9888	.9539	.8666	.7095	.5000	.2905	.1334	.0461
14	.9999	.9991	.9935	.9713	.9102	.7880	.6047	.3953	.2120	.0898
15		.9995	.9963	.9824	.9408	.8491	.6964	.5000	.3036	.1509
16		.9997	.9979	.9894	.9616	.8949	.7728	.5982	.4018	.2272
17		.9999	.9988	.9936	.9755	.9283	.8338	.6855	.5000	.3145
18		.9999	.9993	.9962	.9846	.9519	.8811	.7597	.5927	.4073
19			.9996	.9978	.9904	.9682	.9165	.8204	.6762	.5000
20			.9998	.9987	.9941	.9793	.9423	.8684	.7483	.5881
21			.9999	.9993	.9964	.9867	.9608	.9054	.8083	.6682
22			.9999	.9996	.9978	.9916	.9738	.9331	.8569	.7383
23				.9998	.9987	.9947	.9827	.9534	.8950	.7976
24				.9999	.9992	.9967	.9887	.9680	.9242	.8463
25				.9999	.9995	.9980	.9927	.9784	.9461	.8852

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN VINDT IN DE TABEL IN DE RIJ BEHORENDE BIJ DE GEGEVEN STEEKPROEFGROOTTE (n) EN IN DE KOLOM BEHORENDE BIJ HET GEGEVEN AANTAL PLUS- OF MINTEKENS (m) DE KANS OP HET AANTREFFEN VAN m OF MEER PLUSTEKENS (RESP. MINTEKENS) RECHTS (RESP. LINKS) VAN DE VERONDERSTELDE MEDIAANWAARDE. HIERBIJ WORDT AANGENOMEN DAT EEN WAARNEMING EEN GELIJKE KANS HEEFT OM LINKS OF RECHTS VAN DE MEDIAAN TE LIGGEN. DE WAARNEMINGEN DIENEN EERST NAAR GROOTTE GERANGSCHIKT TE WORDEN ZODAT $x_1 \leq x_2 \leq \dots \leq x_n$. HET BETROUWBAARHEIDSINTERVAL VOOR DE MEDIAAN, $\pi_{.50}$ WORDT ALS VOLGT GEVORMD:

EENZIJDIG: $\pi_{.50} \geq x_m$ OF $\pi_{.50} \leq x_{n-m+1}$ MET BETROUWBAARHEID P

TWEEZIJDIG: $x_m \leq \pi_{.50} \leq x_{n-m+1}$ MET BETROUWBAARHEID $(2P - 1)$

N.B. $P(m \leq c) = 1 - P(m+1 \geq c)$.

TOEPASSING

ZIE INLEIDING VOOR HET TOEPASSEN VAN DEZE TABEL.

BENADERING

VOOR WAARDEN VAN $n > 25$ KAN P GEVONDEN WORDEN UIT TABEL 1 (NORMALE VERDELING)

MET:

$$u = \frac{n - 2m + 1}{\sqrt{n}}$$

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

TABEL 15

WILCOXON'S VERDELINGSVRIJE TOETS VOOR TWEE STEEKPROEVEN

LINKER KRITIEKE WAARDEN VAN W BIJ TWEEZIJDIGE BETROUWBAARHEID .90

n \ m	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	-	0																	
4	-	0	2																
5	0	2	4	8															
6	0	4	6	10	14														
7	0	4	8	12	16	22													
8	2	6	10	16	20	26	30												
9	2	8	12	18	24	30	36	42											
10	2	8	14	22	28	34	40	48	54										
11	2	10	16	24	32	38	46	54	62	68									
12	4	10	18	26	34	42	52	60	68	76	84								
13	4	12	20	30	38	48	56	66	74	84	94	102							
14	6	14	22	32	42	52	62	72	82	92	102	112	122						
15	6	14	24	36	46	56	66	78	88	100	110	122	132	144					
16	6	16	28	38	50	60	72	84	96	108	120	130	142	154	166				
17	6	18	30	40	52	66	78	90	102	114	128	140	152	166	178	192			
18	8	18	32	44	56	70	82	96	110	122	136	150	164	176	190	204	218		
19	8	20	34	46	60	74	88	102	116	130	144	160	174	188	202	216	232	246	
20	8	22	36	50	64	78	94	108	124	138	154	168	184	200	214	230	246	260	276

LINKER KRITIEKE WAARDEN VAN W BIJ TWEEZIJDIGE BETROUWBAARHEID .95

n \ m	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	-	-																	
4	-	-	0																
5	-	0	2	4															
6	-	2	4	6	10														
7	-	2	6	10	12	16													
8	0	4	8	12	16	20	26												
9	0	4	8	14	20	24	30	34											
10	0	6	10	16	22	28	34	40	46										
11	0	6	12	18	26	32	38	46	52	60									
12	2	8	14	22	28	36	44	52	58	66	74								
13	2	8	16	24	32	40	48	56	66	74	82	90							
14	2	10	18	26	34	44	52	62	72	80	90	100	110						
15	2	10	20	28	38	48	58	68	78	88	98	108	118	128					
16	2	12	22	30	42	52	62	74	84	94	106	118	128	138	150				
17	4	12	22	34	44	56	68	78	90	102	114	126	138	150	162	174			
18	4	14	24	36	48	60	72	84	96	110	122	134	146	160	172	186	198		
19	4	14	26	38	50	64	76	90	104	116	130	142	156	170	184	198	210	224	
20	4	16	28	40	54	68	82	96	110	124	138	152	166	180	194	210	224	238	254

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN ZOEKT IN DE TABEL HET GETAL VOORKOMENDE IN DE RIJ MET HET GEGEVEN AANTAL, n , ONAFHANKELIJKE WAARNEMINGEN VAN DE VARIABLE x EN IN DE KOLOM MET HET GEGEVEN AANTAL, m , ONAFHANKELIJKE WAARNEMINGEN VAN DE VARIABLE y . DIT GETAL IS, K , DE KRITIEKE WAARDE VAN W , WAAR BENEDEN DE KANS DAT x EN y UIT EENZELFDE VERDELING KOMEN KLEINER IS DAN DE AANGEGEVEN TWEEZIJDIGE BETROUWBAARHEID. W WORDT ALS VOLGT BEREKEND: MEN VERGELIJKT IEDERE x MET IEDERE y EN KENT DAARBIJ ALS VOLGT PUNTEN TOE:

WANNEER $x > y$ 2 PUNTEN; WANNEER $x = y$ 1 PUNT; WANNEER $x < y$ 0 PUNTEN; DE SOM VAN ALLE PUNTEN GEEFT DE WAARDE VAN w . DE NULHYPOTHESE WORDT VERWORPEN INDIEN $w < K$ OF $w > 2m \cdot n - K$, BIJ TWEEZIJDIG TOETSEN. BIJ EENZIJDIG TOETSEN IS DE BETROUWBAARHEID GELIJK AAN $1/2 + 1/2 P$. DE RECHTS EENZIJDIGE KRITIEKE WAARDE VINDT MEN DOOR K VAN $2m \cdot n$ AF TE TREKKEN.

DE STEEKPROEFGROOTTE VAN m EN n IS ONDERLING VERWISSELBAAR. ZO WORDEN DE KRITIEKE WAARDEN VOOR $m = 13$ EN $n = 7$ GEVONDEN BIJ $m = 7$ EN $n = 13$.

TOEPASSING

DEZE TOETS WORDT GEBRUIKT OM NA TE GAAN OF TWEE STEEKPROEVEN UIT EENZELFDE POPULATIE ZIJN GETROKKEN. DEZE TOETS HANGT NIET AF VAN DE VORM VAN DE VERDELING.

BENADERING

ZIE INLEIDING.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

KENDALL'S VERDELINGSVRIJE CORRELATIE TOETS

KRITIEKE WAARDEN VAN S BIJ ENKELE BETROUWBAARHEDEN

	BETROUWBAARHEID					BETROUWBAARHEID			
	.95	.975	.99	.995		.95	.975	.99	.995
EENZIJDIG	.95	.975	.99	.995	EENZIJDIG	.95	.975	.99	.995
TWEEZIJDIG	.90	.95	.98	.99	TWEEZIJDIG	.90	.95	.98	.99
n					n				
4	6	-	-	-	21	56	66	78	86
5	8	10	10	-	22	61	71	83	91
6	11	13	13	15	23	65	75	89	99
7	13	15	17	19	24	68	80	94	104
8	16	18	20	22	25	72	86	100	110
9	18	20	24	26	26	77	91	107	117
10	21	23	27	29	27	81	95	113	125
11	23	27	31	33	28	86	100	118	130
12	26	30	36	38	29	90	106	126	138
13	28	34	40	44	30	95	111	131	145
14	33	37	43	47	31	99	117	137	151
15	35	41	49	53	32	104	122	144	160
16	38	46	52	58	33	108	128	152	166
17	42	50	58	64	34	113	133	157	175
18	45	53	63	69	35	117	139	165	181
19	49	57	67	75	36	122	146	172	190
20	52	62	72	80	37	128	152	178	198
u	1.64	1.96	2,33	2.58	38	133	157	185	205
					39	139	163	193	213
					40	144	170	200	222
					u	1.64	1.96	2.33	2.58

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

BIJ 20 PAREN WAARNEMINGEN ($n = 20$) WORDT DE HYPOTHESE VAN ONAFHANKELIJKHEID VERWORPEN INDIEN $S > 80$ OF $S < -80$ MET TWEEZIJDIGE BETROUWBAARHEID, $P = .99$.

TOEPASSING

OP GROND VAN WAARNEMINGSMATERIAAL, n ONDERLING ONAFHANKELIJKE WAARNEMINGSPAREN (x_i, y_i) ($i = 1, 2, \dots, n$) VAN x EN y , WELKE GROOTHEDEN EEN SIMULTANE VERDELING BEZITTEN, WORDT DE HYPOTHESE H_0 GETOETST DAT x EN y ONDERLING ONAFHANKELIJK VERDEELD ZIJN. DEZE TOETS HANGT NIET AF VAN DE VORM VAN DE VERDELING.

BENADERING

WANNEER $n > 40$ KAN S BEREKEND WORDEN VIA EEN NORMALE BENADERING:
 $|S| = u \sqrt{n(n-1)(2n+5) / 18} + 1$. DE WAARDEN VAN u , BEHORENDE BIJ DE GEWENSTE BETROUWBAARHEID ZIJN ONDERAAN DE TABEL GEGEVEN.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

TABEL 17

STEEKPROEFVOORSCHRIFTEN VOOR PARTIJKEURING

PHILIPS STANDAARD STEEKPROEF SYSTEEM (S.S.S.)
 KEUZE VAN DE STEEKPROEFGROOTTE n EN HET AFKEURCRITERIUM c

PARTIJ- GROOTTE	CONTROLEPUNT $P_{.50}$																							
	$1/4$ %		$1/2$ %		1%		2%		3%		5%		7%		10%									
ENKELVOUDIGE STEEKPROEF	n	c	n	c	n	c	n	c	n	c	n	c	n	c	n	c								
≤ 100	A	-	A	-	45	0	30	0	19	0	12	0	9	0	6	0								
101-200	A	-	90	0	55	0	30	0	50	1	30	1	23	1	16	1								
201-500	195	0	110	0	150	1	80	1	55	1	35	1	35	2	25	2								
501-1000	225	0	305	1	160	1	80	1	90	2	55	2	50	3	35	3								
DUBBELE STEEKPROEF	n_1	c_1	c_2	n_1	c_1	c_2	n_1	c_1	c_2	n_1	c_1	c_2	n_1	c_1	c_2	n_1	c_1	c_2						
1001- 2000	330	0	1	150	0	1	110	0	2	55	0	2	45	0	3	25	0	3	30	1	5	22	1	5
2001- 5000	425	0	2	200	0	2	135	0	3	70	0	3	70	1	5	45	1	5	55	2	10	40	2	10
5001-10000	525	0	3	260	0	3	220	1	5	110	1	5	125	2	10	75	2	10	75	3	15	55	3	15
10001-20000	875	1	5	440	1	5	380	2	10	190	2	10	180	3	15	110	3	15	100	4	20	75	4	20
20001-50000	1500	2	10	750	2	10	540	3	15	270	3	15	240	4	20	140	4	20	120	5	25	85	5	25
50001 EN MEER	2200	3	15	1100	3	15	700	4	20	350	4	20	290	5	25	175	5	25	145	6	30	150	6	30

FRACTIES DEFECTIEVEN $P_{.95}$ EN $P_{.10}$ VOOR HET P.S.S.S.

PARTIJ- GROOTTE	CONTROLEPUNT															
	$1/4$ %		$1/2$ %		1%		2%		3%		5%		7%		10%	
ENKELVOUDIG	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$
≤ 100					0.25	2.1	0.50	5.4	0.50	8.3	0.65	14.0	0.85	19.6	1.1	28.0
101-200			0.10	1.2	0.13	2.7	0.20	5.6	0.75	6.3	1.2	10.5	1.7	14.7	2.4	21.0
201-500	0.04	0.65	0.06	1.3	0.28	2.1	0.50	4.2	0.70	6.6	1.2	11.0	2.3	13.6	3.3	19.5
501-1000	0.03	0.79	0.15	1.0	0.25	2.1	0.50	4.4	1.0	5.9	1.7	9.8	2.8	12.6	4.0	18.0
DUBBEL	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$	$P_{.95}$	$P_{.10}$
1001- 2000	0.04	0.61	0.09	1.2	0.25	2.1	0.55	4.2	1.0	5.8	1.7	9.7	2.9	12.6	4.1	18.0
2001- 5000	0.07	0.55	0.14	1.1	0.35	2.0	0.70	3.9	1.2	5.4	2.1	9.0	3.9	10.2	5.5	14.5
5001-10000	0.09	0.50	0.17	1.0	0.40	1.8	0.80	3.6	1.7	4.4	2.8	7.3	4.4	9.4	6.3	13.5
10001-20000	0.10	0.45	0.20	0.90	0.55	1.5	1.1	2.9	1.9	4.1	3.1	6.8	4.8	9.1	6.8	13.0
20001-50000	0.14	0.35	0.30	0.75	0.65	1.4	1.3	2.7	2.1	3.9	3.4	6.5	5.0	8.8	7.1	12.5
50001 EN MEER	0.16	0.35	0.30	0.70	0.70	1.3	1.4	2.6	2.1	3.8	3.6	6.5	5.1	8.7	7.3	12.4

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

BIJ DUBBELE STEEKPROEFSHEMA'S NEEMT MEN EERST EEN STEEKPROEF GROOT n_1 STUKS EN WORDT DE PARTIJ GOEDGEKEURD WANNEER $m_1 < c_1$ EN AFGEKEURD WANNEER $m_1 > c_2$. IS $c_1 < m_1 < c_2$ DAN WORDT EEN TWEEDE STEEKPROEF GENOMEN VAN $n_2 = 2n_1$ STUKS EN DE PARTIJ UITEINDELIJK GOEDGEKEURD WANNEER $m < c_2$ EN AFGEKEURD WANNEER $m > c_2$. HIERBIJ IS m_1 HET AANTAL DEFECTIEVEN IN DE EERSTE STEEKPROEF EN m HET AANTAL IN BEIDE STEEKPROEVEN TEZAMEN.

$P_{.50}$ IS HET FOUTENPERCENTAGE WAARBIJ DE PARTIJ 50% KANS HEEFT TE WORDEN GOEDGEKEURD.

$P_{.95}$ IS HET FOUTENPERCENTAGE WAARBIJ DE PARTIJ 95% KANS HEEFT TE WORDEN GOEDGEKEURD (AQL).

$P_{.10}$ IS HET FOUTENPERCENTAGE WAARBIJ DE PARTIJ 10% KANS HEEFT TE WORDEN GOEDGEKEURD (LTPD).

A BETEKENT, DAT GEEN STEEKPROEF MAG WORDEN GENOMEN MAAR DAT DE GEHELE PARTIJ MOET WORDEN GEINSPECTEERD.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

HET BEREKENEN VAN KEURINGSKARAKTERISTIEKEN

WAARDEN VAN u VAN DE POISSONVERDELING WAARVOOR DE KANS OP EEN WAARGENOMEN WAARDE VAN m OF MINDER DEFECTIEVEN GELIJK IS AAN EEN GEGEVEN WAARDE VAN p (GOEDKEUR-

KANS)

m	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
0	.00501	.0101	.0253	.0513	.105	.288	.693	1.386	2.303	2.996	3.689	4.605	5.298
1	.103	.149	.242	.355	.532	.961	1.678	2.693	3.890	4.744	5.572	6.638	7.430
2	.338	.436	.619	.818	1.102	1.727	2.674	3.920	5.322	6.296	7.224	8.406	9.274
3	.672	.823	1.090	1.366	1.745	2.535	3.672	5.109	6.681	7.754	8.768	10.045	10.978
4	1.078	1.279	1.623	1.970	2.433	3.369	4.671	6.274	7.994	9.154	10.242	11.605	12.594
5	1.537	1.785	2.202	2.613	3.152	4.219	5.670	7.423	9.275	10.513	11.668	13.108	14.150
6	2.037	2.330	2.814	3.286	3.895	5.083	6.670	8.558	10.532	11.842	13.060	14.571	15.660
7	2.571	2.906	3.454	3.981	4.656	5.956	7.669	9.684	11.771	13.148	14.422	16.000	17.134
8	3.132	3.507	4.115	4.695	5.432	6.838	8.669	10.802	12.995	14.434	15.763	17.403	18.578
9	3.717	4.130	4.795	5.426	6.221	7.726	9.669	11.914	14.206	15.705	17.085	18.783	19.998
10	4.321	4.771	5.491	6.169	7.021	8.620	10.668	13.020	15.407	16.962	18.390	20.145	21.398
11	4.943	5.428	6.201	6.924	7.829	9.519	11.668	14.121	16.598	18.208	19.682	21.490	22.779
12	5.580	6.099	6.922	7.690	8.646	10.422	12.668	15.217	17.782	19.442	20.962	22.821	24.145
13	6.231	6.782	7.654	8.464	9.470	11.329	13.668	16.310	18.958	20.668	22.230	24.139	25.496
14	6.893	7.477	8.396	9.246	10.300	12.239	14.668	17.400	20.128	21.886	23.490	25.446	26.836
15	7.566	8.181	9.144	10.035	11.135	13.152	15.668	18.486	21.292	23.098	24.741	26.743	28.166
16	8.249	8.895	9.902	10.831	11.976	14.068	16.668	19.570	22.452	24.302	25.984	28.031	29.484
17	8.942	9.616	10.666	11.633	12.822	14.986	17.668	20.652	23.606	25.500	27.220	29.310	30.792
18	9.644	10.346	11.438	12.442	13.672	15.907	18.668	21.731	24.756	26.692	28.448	30.581	32.092
19	10.353	11.082	12.216	13.254	14.525	16.830	19.668	22.808	25.902	27.879	29.671	31.845	33.383
20	11.069	11.825	12.999	14.072	15.383	17.755	20.668	23.883	27.045	29.062	30.888	33.103	34.668
21	11.791	12.574	13.787	14.894	16.244	18.682	21.668	24.956	28.184	30.241	32.102	34.355	35.947
22	12.520	13.329	14.580	15.719	17.108	19.610	22.668	26.028	29.320	31.416	33.309	35.601	37.219
23	13.255	14.088	15.377	16.548	17.975	20.540	23.668	27.098	30.453	32.586	34.512	36.841	38.485
24	13.995	14.853	16.178	17.382	18.844	21.471	24.668	28.167	31.584	33.752	35.710	38.077	39.745
25	14.740	15.623	16.984	18.218	19.717	22.404	25.667	29.234	32.711	34.916	36.905	39.308	41.000
26	15.490	16.397	17.793	19.058	20.592	23.338	26.667	30.300	33.836	36.077	38.096	40.535	42.252
27	16.245	17.175	18.606	19.900	21.469	24.273	27.667	31.365	34.959	37.234	39.284	41.757	43.497
28	17.004	17.957	19.422	20.746	22.348	25.209	28.667	32.428	36.080	38.389	40.468	42.975	44.738
29	17.767	18.742	20.241	21.594	23.229	26.147	29.667	33.491	37.198	39.541	41.649	44.190	45.976
30	18.534	19.532	21.063	22.444	24.113	27.086	30.667	34.552	38.315	40.690	42.827	45.401	47.210

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN ZOEKT IN DE TABEL HET GETAL VOORKOMENDE IN DE RIJ MET HET GEVONDEN AANTAL DEFECTIEVEN, m , EN IN DE KOLOM BEHORENDE BIJ DE GEGEVEN GOEDKEURKANS (p). DIT GETAL IS DE WAARDE VAN $u = np$. IS $P(k < 8) = 0.975$ DAN IS $u = 4.115 = np$. ALS $n = 100$ IS $p = .04115$. $n =$ STEEKPROEFGROOTTE; $p =$ FRACTIE DEFECTIEVEN.

TOEPASSING

DE TABEL KAN WORDEN GEBRUIKT VOOR HET BEREKENEN VAN DE KEURINGSKARAKTERISTIEKEN ALS $p < 0.1$.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

TABEL 19

HET BEREKENEN VAN REGELGRENZEN VOOR EEN BEHEERST
FABRIKAGEPROCES.

FAKTOREN VOOR HET BEREKENEN VAN REGELGRENZEN (ENGLSE METHODE).

STEEKPROEF- GROOTTE n	$\frac{\sigma}{E(\bar{R})}$	BENEDENGRENS VOOR R		BOVENGRENS VOOR R		GRENZEN VOOR \bar{x}		GRENZEN VOOR DE MEDIAAN		GRENZEN VOOR x	
	A	D _{0.001}	D _{0.025}	D _{0.975}	D _{0.999}	A _{0.001}	A _{0.025}	M _{0.001}	M _{0.025}	E _{0.001}	E _{0.025}
2	0.886	0.00	0.04	2.81	4.12	1.94	1.23	1.94	1.23	2.74	1.74
3	0.591	0.04	0.18	2.17	2.98	1.05	0.67	1.22	0.78	1.83	1.16
4	0.486	0.10	0.29	1.93	2.57	0.75	0.48	0.82	0.52	1.50	0.95
5	0.430	0.16	0.37	1.81	2.34	0.59	0.38	0.71	0.45	1.33	0.84
6	0.395	0.21	0.42	1.72	2.21	0.50	0.32	0.57	0.36	1.22	0.77
7	0.370	0.26	0.46	1.66	2.11	0.43	0.27	0.52	0.33	1.14	0.73
8	0.351	0.29	0.50	1.62	2.04	0.38	0.24	0.44	0.28	1.08	0.69
9	0.337	0.32	0.52	1.58	1.99	0.35	0.22	0.42	0.27	1.04	0.66
10	0.325	0.35	0.54	1.56	1.93	0.32	0.20	0.37	0.24	1.00	0.64
11	0.315	0.38	0.56	1.53	1.91	0.29	0.19	0.36	0.23	0.97	0.62
12	0.307	0.40	0.58	1.51	1.87	0.27	0.17	0.33	0.21	0.95	0.60

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

MEN VINDT IN DE TABEL BIJ EEN GEGEVEN STEEKPROEFGROOTTE, n, DE FAKTOREN VOOR HET BEREKENEN VAN REGELGRENZEN VOLGENS DE ENGLSE METHODE. UIT EEN VRIJ GROOT AANTAL STEEKPROEFJES VAN n WAARNEMINGEN IEDER WORDT \bar{R} BEREKEND. DE GRENZEN WORDEN ALS VOLGT GEVORMD:

GRENZEN VOOR R:

$$\begin{aligned} \text{TWEEZIJDIG, } P = .98 : \quad & \bar{R} \cdot D_{0.001} < R < \bar{R} \cdot D_{0.999} \\ \text{TWEEZIJDIG, } P = .95 : \quad & \bar{R} \cdot D_{0.025} < R < \bar{R} \cdot D_{0.975} \end{aligned}$$

GRENZEN VOOR \bar{x} :

$$\begin{aligned} \text{TWEEZIJDIG, } P = .98 : \quad & \bar{x} - \bar{R} \cdot A_{0.001} < \bar{x} < \bar{x} + \bar{R} \cdot A_{0.001} \\ \text{TWEEZIJDIG, } P = .95 : \quad & \bar{x} - \bar{R} \cdot A_{0.025} < \bar{x} < \bar{x} + \bar{R} \cdot A_{0.025} \end{aligned}$$

GRENZEN VOOR M:

$$\begin{aligned} \text{TWEEZIJDIG, } P = .98 : \quad & \bar{x} - \bar{R} \cdot M_{0.001} < M < \bar{x} + \bar{R} \cdot M_{0.001} \\ \text{TWEEZIJDIG, } P = .95 : \quad & \bar{x} - \bar{R} \cdot M_{0.025} < M < \bar{x} + \bar{R} \cdot M_{0.025} \end{aligned}$$

GRENZEN VOOR x:

$$\begin{aligned} \text{TWEEZIJDIG, } P = .98 : \quad & \bar{x} - \bar{R} \cdot E_{0.001} < x < \bar{x} + \bar{R} \cdot E_{0.001} \\ \text{TWEEZIJDIG, } P = .95 : \quad & \bar{x} - \bar{R} \cdot E_{0.025} < x < \bar{x} + \bar{R} \cdot E_{0.025} \end{aligned}$$

SCHATTING VAN σ :

$$s = \bar{A} \bar{R}.$$

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

ASELECTE TREKKINGEN UIT EEN NORMALE VERDELING

MET $\mu = 0$, $\sigma = 1$.

-0.84	0.68	0.11	1.05	0.09	0.83	1.35	1.01	-0.54	0.28	0.88	-1.01	0.30
1.37	-0.77	-0.89	0.61	-2.54	1.55	-0.92	-0.20	1.11	-1.52	-0.62	-0.43	0.95
-0.18	-1.39	0.99	-0.01	-0.22	0.36	1.27	0.45	0.59	1.52	-0.36	-0.76	-1.00
0.35	-1.91	0.50	-0.92	-0.30	0.36	1.28	-0.69	-1.64	0.42	0.86	0.13	1.43
2.82	0.05	-0.95	1.17	1.22	0.14	-2.13	-0.76	0.24	-1.07	0.98	-0.87	1.06
2.12	1.32	0.81	-1.91	-1.23	-0.17	-0.38	-0.03	-1.65	-0.35	-0.73	1.69	-0.66
-0.99	-0.67	-0.28	-0.48	-0.12	-2.21	-0.10	0.40	-0.83	-0.37	-0.52	0.02	0.60
0.34	-0.81	-1.64	-0.21	-0.19	0.66	0.12	0.39	1.06	-0.38	-0.72	0.33	1.11
0.64	0.96	-1.44	-2.47	-1.69	-0.45	2.31	1.39	-0.17	0.09	-0.97	-0.79	-0.08
0.11	-0.24	0.01	-0.28	0.04	1.28	0.47	0.77	-1.31	-0.38	1.28	1.41	1.41
-0.02	-0.66	0.24	-0.43	1.09	-0.41	0.42	-0.37	-0.56	0.38	-0.18	1.79	0.78
-0.86	1.08	0.46	0.08	1.03	-1.73	-1.07	0.35	0.66	0.88	-0.27	1.38	-2.01
0.01	0.27	0.02	-0.60	-0.61	-1.01	0.22	-1.25	1.43	-1.92	1.17	0.71	-1.53
-0.03	0.46	-2.09	-1.60	-0.56	-1.36	-1.33	-0.70	-0.18	1.11	0.91	-0.28	0.30
-0.41	-0.77	2.08	-0.98	0.36	-0.36	-2.24	1.41	-0.07	-1.59	0.15	-0.96	0.87
0.62	0.00	-0.80	-0.62	0.35	0.31	-0.20	-0.04	-0.41	-0.33	0.72	-0.09	-0.77
0.83	1.01	0.27	1.47	0.29	-2.79	0.20	0.42	0.51	-0.63	-1.25	-0.60	2.32
-0.78	-0.92	0.21	-0.60	-0.59	0.14	0.09	-0.09	0.97	0.96	-1.08	-0.09	0.62
0.03	0.98	-1.33	0.66	-0.57	-0.68	0.02	-0.59	-1.20	1.05	-1.18	-0.51	-0.93
-2.30	0.24	1.70	-0.52	-0.15	1.13	-0.23	0.77	-0.10	-0.48	-0.89	-0.29	1.65
0.78	1.01	-0.92	1.82	-0.67	-1.30	-1.07	-0.69	1.53	-0.60	0.30	1.50	0.13
0.66	1.49	1.73	-0.27	-0.87	-0.64	1.43	0.14	-0.68	0.69	-0.65	-0.47	1.35
-0.34	1.16	0.66	0.59	0.35	-1.63	0.22	-0.34	1.70	-0.46	0.88	-2.34	-1.14
1.16	0.48	-1.69	0.07	0.70	-0.49	-1.24	-0.51	1.40	-0.87	0.98	-1.09	0.35
0.76	0.40	2.13	-0.25	-0.15	-0.25	0.91	0.54	0.17	-0.45	1.97	1.18	-2.15
-1.00	-0.12	-2.20	0.75	1.30	-0.64	-1.88	1.35	-1.98	-0.14	-0.82	-0.53	0.43
-0.32	-0.85	1.57	0.25	0.75	-0.61	0.08	0.06	-0.61	-1.02	0.60	-0.73	-1.41
0.18	-0.46	-1.85	2.24	1.08	0.24	2.19	0.66	1.03	-0.58	-0.37	1.47	-0.14
0.08	0.35	-2.24	-0.27	0.67	-1.01	0.17	-0.35	0.66	1.38	0.68	-0.78	1.69
1.98	0.23	-0.15	-0.22	-0.56	-0.62	-0.61	-0.53	-0.62	1.11	-0.88	-0.30	0.84
0.91	0.20	-0.45	0.06	-1.10	-0.23	-0.12	-0.56	-0.03	-1.49	1.73	1.73	-2.14
-0.32	-0.17	-0.03	-0.75	0.93	0.42	0.70	-0.02	0.15	-0.14	0.38	1.02	-0.75
-1.77	0.12	1.01	-2.27	0.55	0.32	-1.72	0.66	-1.86	0.50	-0.65	-0.31	-0.58
0.16	-1.48	1.19	-0.23	0.59	0.41	-1.46	0.10	-1.55	-0.35	1.07	1.28	-0.25
0.79	0.52	-1.03	-1.18	-0.85	1.67	-0.36	0.52	0.34	0.94	-0.05	-1.14	0.29
0.20	1.56	0.35	-0.62	-0.15	-0.19	-1.39	-0.46	-0.63	0.50	-0.54	-0.27	-1.00
1.31	1.04	-0.82	0.47	0.81	0.26	0.01	0.47	0.30	0.83	-0.28	-1.14	-1.69
0.35	-0.18	1.76	0.92	-1.21	-0.48	0.52	0.40	0.31	0.04	-0.61	1.68	0.63
0.74	-0.26	-0.68	-0.32	-0.44	0.01	-0.47	-1.15	1.13	-0.12	-0.22	-1.70	-1.34
0.27	1.00	-0.88	-0.23	-0.47	-1.61	-0.01	-0.29	0.68	0.30	0.15	0.27	-1.76
1.39	-0.51	-0.14	-0.89	0.76	0.14	-0.49	0.82	-0.01	-1.45	-0.86	0.69	0.73
-0.14	0.05	0.77	1.33	-0.58	-0.76	0.19	0.23	-0.45	-0.23	0.75	-0.85	-0.10
-0.76	-0.69	-1.61	0.33	-0.13	0.30	-0.05	-0.87	0.03	0.77	0.82	0.50	0.69
0.37	-0.58	-0.67	0.68	0.17	-0.54	1.07	-0.12	0.03	-0.37	-0.73	-0.05	0.63
-2.28	-0.69	0.58	-0.38	-1.65	-0.86	-1.39	0.64	-0.64	1.26	0.40	-0.15	-1.33
0.36	-1.41	-1.06	-0.89	-0.28	-0.75	0.96	-0.03	0.34	-0.62	0.17	1.66	-0.63
1.37	0.24	0.51	-0.28	0.32	0.32	1.32	0.17	0.01	-0.00	1.36	0.76	0.11
1.04	0.41	-0.24	0.19	-2.10	-0.69	-0.98	1.49	0.49	0.83	-0.98	-1.38	-1.08
-0.67	0.15	-1.60	0.84	-0.07	-0.06	-0.25	-0.51	-0.63	0.88	-1.30	0.53	1.05
0.39	-0.24	0.29	-0.03	0.83	0.67	-1.20	0.62	-2.07	1.21	0.06	-0.87	0.48
-0.80	0.10	0.59	-0.89	-0.59	-0.94	0.88	0.53	-0.02	-0.90	1.09	-0.56	0.33
-0.45	0.24	-0.99	-0.95	-0.00	0.41	1.40	-0.85	0.58	-1.06	0.98	1.38	-2.54
-0.40	-0.00	-0.03	-1.07	0.64	1.55	0.58	0.15	-1.37	-1.68	-0.70	-1.03	0.45
-1.19	0.97	1.28	0.52	0.10	-0.10	-0.00	1.10	-1.19	-0.89	-0.12	-1.26	0.16
0.44	0.19	-0.85	-0.60	-0.42	0.71	-1.15	2.59	2.09	1.31	-2.00	0.28	0.93
0.00	-0.84	-0.32	0.45	-1.14	1.59	1.55	1.10	0.38	0.46	0.96	-1.81	-0.55
0.31	0.29	0.38	0.16	0.79	0.98	-0.50	0.48	1.03	0.68	0.97	-0.40	0.38
0.86	0.39	-1.08	-0.48	2.57	-0.51	0.79	-0.93	1.62	-1.21	1.11	0.52	-0.05
1.23	-2.64	-2.03	-0.09	1.77	0.40	-0.49	-0.60	-0.06	1.06	-2.19	0.14	-0.06
-0.05	0.53	-0.19	-1.75	-0.80	0.13	0.48	-0.77	-0.65	1.42	0.02	1.66	-2.01

TABEL 21

T O E V A L S C I J F E R S

36 06 46 13 25	09 48 47 41 91	67 84 19 55 21	50 91 77 92 73	88 72 39 04 12
07 25 44 57 45	93 84 57 29 55	15 26 69 92 45	58 35 75 23 89	75 12 47 61 54
02 33 89 14 71	52 02 72 25 35	37 87 44 05 16	12 44 49 48 66	11 92 85 53 37
51 87 11 59 95	20 22 19 05 86	83 24 04 61 26	98 22 47 90 97	87 69 50 48 09
89 72 70 32 02	04 75 48 97 75	40 01 83 62 31	07 88 73 17 20	96 70 57 98 53
94 67 32 91 07	71 49 59 59 91	47 87 85 95 53	46 07 76 82 59	36 54 27 42 77
40 20 54 27 91	89 00 67 90 84	38 00 98 13 70	19 92 81 24 17	71 34 00 88 53
94 52 73 33 96	41 96 69 66 75	56 39 87 31 16	97 48 97 77 46	20 72 59 95 56
86 07 79 95 74	82 02 57 50 23	19 74 72 98 77	17 76 89 99 03	22 32 05 76 83
22 17 48 61 18	01 25 82 30 42	76 06 24 63 95	90 52 87 51 78	00 39 80 65 49
23 17 23 20 75	87 18 04 28 99	21 80 69 36 04	34 70 20 77 76	18 30 97 48 62
24 27 10 67 76	39 87 93 02 76	56 57 64 66 20	15 71 48 22 11	25 06 98 68 69
35 38 80 01 79	44 29 85 66 28	77 39 38 88 10	27 29 86 32 44	71 65 98 92 32
22 63 70 64 07	21 26 88 40 28	98 63 35 73 42	60 49 50 18 91	38 36 27 02 93
04 43 05 32 11	94 39 12 37 38	87 21 17 35 09	98 60 40 65 18	12 50 25 99 16
06 63 30 04 99	83 20 18 44 79	66 41 94 44 93	48 29 35 28 82	49 94 22 01 50
28 28 40 79 08	21 86 42 98 35	05 94 38 66 41	98 37 04 87 99	42 22 62 29 47
30 05 78 69 83	43 68 37 15 73	36 57 33 23 96	43 21 68 17 76	81 18 85 25 94
54 83 43 03 57	78 78 17 89 41	06 99 65 47 75	63 92 29 26 09	11 85 81 53 65
41 61 55 84 94	75 19 63 23 60	06 44 29 77 02	79 41 69 93 61	96 53 45 98 39
06 01 22 15 95	22 23 83 26 29	39 95 50 23 53	87 00 55 83 49	24 76 90 24 80
20 56 94 22 81	07 86 11 61 30	81 70 61 89 74	83 56 28 71 58	81 18 45 40 94
79 51 06 83 63	01 03 56 59 04	26 05 83 06 01	16 51 72 44 99	98 41 73 86 43
69 58 97 33 79	22 16 00 65 91	12 92 55 89 73	19 07 06 41 38	34 73 12 43 45
04 50 94 95 99	48 14 54 12 97	49 86 70 98 56	06 93 58 74 94	55 92 16 91 54
39 93 94 40 33	81 09 23 42 98	56 50 79 19 25	23 07 84 81 05	56 68 68 57 69
17 79 56 31 98	27 97 82 33 62	61 52 59 10 26	70 98 60 39 42	90 75 46 94 86
80 54 04 48 41	05 79 16 40 07	17 26 94 82 80	68 08 09 64 53	37 58 99 36 10
79 86 00 59 48	35 04 48 39 46	04 71 43 88 01	10 41 56 45 66	43 00 83 69 67
14 72 30 79 13	95 96 59 31 70	82 43 92 35 11	16 85 17 53 73	54 82 35 65 82
28 45 78 47 60	52 78 55 17 11	83 93 55 20 47	28 22 38 32 06	44 39 10 13 49
42 70 25 26 46	83 22 00 23 87	97 32 26 12 70	55 15 62 88 31	89 23 84 59 38
11 84 85 18 88	73 20 07 72 47	66 44 92 10 27	29 17 13 12 33	27 85 59 31 76
26 59 28 19 74	29 90 19 71 19	68 14 07 13 36	01 31 68 09 28	98 60 59 30 31
42 33 03 34 94	42 79 42 13 28	31 77 52 02 05	94 78 45 93 07	65 56 49 47 64
03 83 10 84 20	23 08 20 55 02	87 00 35 28 30	35 16 81 40 12	95 83 80 52 81
02 73 79 02 38	74 75 56 37 47	89 79 16 81 66	75 83 90 35 61	18 45 61 80 48
63 90 04 21 64	59 23 19 69 88	46 31 26 75 55	87 41 93 10 64	90 06 27 26 38
19 74 52 13 57	00 54 60 02 12	63 02 32 30 27	72 98 93 55 13	62 98 93 54 79
15 18 14 82 28	21 77 74 95 64	63 45 16 53 70	68 77 68 19 13	96 91 42 32 97
03 66 02 96 69	92 73 90 78 64	29 95 05 30 38	10 78 48 69 44	01 74 96 98 45
34 18 99 00 10	38 21 44 18 01	88 22 54 59 36	12 99 85 93 51	73 06 46 92 32
88 58 63 52 43	58 32 01 29 25	74 39 96 20 36	64 75 40 85 26	24 04 67 34 33
56 58 69 07 85	85 06 75 95 54	87 11 73 21 18	58 04 97 21 86	75 29 21 16 06
64 81 69 43 66	28 72 33 17 85	44 99 88 90 86	87 48 59 72 99	80 83 81 31 54
90 38 91 52 74	41 90 46 38 82	05 56 03 19 28	84 81 18 28 73	94 77 76 21 89
28 49 79 39 95	30 43 23 12 16	55 99 69 63 48	40 02 15 15 24	84 49 29 40 39
50 87 17 56 96	34 07 37 63 91	41 65 91 70 82	78 29 40 71 59	47 97 64 69 58
35 65 04 30 42	82 42 37 71 93	01 43 95 08 01	48 00 55 88 43	47 12 01 57 23
27 01 05 53 39	60 93 79 14 62	84 06 26 57 43	76 12 15 08 53	67 00 95 81 33

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

VERLOTINGSSERIES

SERIE- BEREIK	
0-9	2 5 9 1 7 8 0 3 4 6 7 9 2 3 6 5 4 0 8 1 2 0 4 6 3 7 9 8 1 5 8 2 3 9 5 1 6 0 4 7 8 9 7 0 2 1 6 4 5 3 3 0 9 8 7 4 5 1 6 2 1 9 8 7 4 3 0 2 6 5 1 7 4 3 8 2 9 0 5 6 8 4 1 9 7 0 3 5 2 6 8 6 2 4 9 5 3 7 1 0 2 1 0 8 3 9 5 4 6 7 9 4 1 7 8 2 5 3 0 6 7 0 9 1 6 5 3 4 8 2 4 1 8 9 2 6 5 0 3 7 3 5 6 7 2 0 4 8 9 1 0 1 4 8 9 7 5 2 6 3 4 2 0 7 5 8 1 6 3 9 7 9 5 1 3 6 0 4 2 8 7 6 1 4 2 8 9 3 0 5 2 6 1 9 0 7 8 3 4 5
0-24	22 06 04 21 14 12 08 10 02 09 05 03 07 23 24 00 17 11 20 01 15 19 18 16 13 17 08 13 00 02 06 09 20 01 03 23 19 10 04 16 18 07 22 24 14 21 15 05 12 11 02 20 09 18 10 22 23 05 04 12 21 16 15 08 13 24 03 01 17 06 07 00 11 14 19 24 02 08 20 01 03 15 19 10 07 18 17 22 12 21 04 00 23 14 16 05 09 06 13 11 14 06 12 21 17 04 10 07 24 08 22 20 01 15 19 09 16 05 00 03 23 02 18 13 11 21 06 14 05 13 02 01 15 16 04 07 08 12 23 00 17 03 11 24 09 19 18 20 22 10 06 22 19 07 02 13 00 20 15 21 05 10 11 24 03 18 12 08 09 23 16 17 14 04 01 06 01 10 24 22 11 14 16 00 09 08 13 12 05 07 19 15 23 20 18 04 02 03 17 21 13 20 02 06 22 12 24 15 01 16 05 09 11 07 18 08 19 03 14 21 23 00 04 17 10 16 05 13 14 02 21 08 00 23 04 18 09 10 19 03 24 20 15 01 11 07 12 17 22 06
0-49	42 16 32 40 17 36 03 23 41 10 34 25 07 47 44 24 29 30 04 22 13 26 11 37 00 31 02 33 06 12 45 05 14 27 21 15 01 20 19 38 35 43 39 40 28 46 18 48 08 09 37 21 38 30 19 02 43 01 04 20 24 45 00 11 27 47 32 16 36 25 07 17 41 44 23 31 34 12 42 29 18 08 49 03 40 33 15 39 28 05 35 14 09 26 13 48 46 22 06 10 22 38 27 02 35 20 37 30 26 45 43 47 12 17 36 44 15 25 09 31 41 06 40 49 10 39 08 19 28 05 01 18 21 34 24 23 29 11 42 07 00 48 32 16 04 14 03 33 46 13 00 04 33 35 22 25 07 28 21 14 32 34 45 44 42 12 48 43 23 46 16 49 47 36 15 13 40 31 19 10 30 27 11 41 38 37 09 08 20 06 05 17 24 29 26 02 39 18 03 01 24 46 41 21 49 08 12 39 48 29 27 25 11 16 37 18 01 30 02 19 28 33 09 04 00 38 10 42 23 34 45 26 47 44 05 36 07 17 03 31 06 22 13 43 15 35 14 20 40 32 35 00 39 07 48 14 02 19 21 11 29 49 47 04 09 34 30 22 33 27 01 40 06 03 37 26 05 31 41 23 25 10 15 46 44 24 20 12 17 38 42 36 43 18 13 08 16 28 45 32 46 07 32 12 08 15 00 44 22 03 02 13 25 14 23 29 47 31 30 28 11 41 48 04 49 19 18 21 37 16 24 27 26 34 38 17 45 40 35 01 06 05 39 09 42 43 10 33 20 36 06 00 10 13 04 08 24 27 48 28 44 19 32 25 05 46 09 07 47 17 39 49 38 35 21 14 31 16 34 26 42 37 33 12 18 03 43 20 41 29 22 30 15 23 02 11 40 36 01 45
0-99	81 74 67 95 70 61 37 42 62 93 52 07 16 29 39 43 08 77 25 72 56 65 32 27 40 63 51 54 50 53 96 34 18 88 13 73 98 20 05 68 33 15 12 41 17 47 48 60 28 02 22 89 04 71 14 76 78 49 86 03 83 45 57 97 44 84 82 87 21 58 11 23 80 31 90 55 35 26 79 75 66 99 09 06 10 30 01 00 36 24 91 19 64 46 69 85 94 59 92 38 03 16 15 40 71 77 07 86 46 61 92 67 78 72 66 43 75 91 02 04 69 95 29 53 39 25 48 74 35 19 62 09 34 44 55 20 97 21 22 58 89 80 36 47 06 82 00 49 56 50 38 87 12 14 17 65 76 93 30 85 59 96 88 08 64 81 28 83 13 99 94 52 18 37 42 79 10 41 26 60 45 01 11 84 27 73 05 32 23 54 51 98 63 57 68 24 90 33 31 70 84 05 56 07 91 26 98 74 12 45 80 32 73 54 69 79 86 46 57 99 09 00 48 36 06 14 55 16 41 71 77 75 35 94 28 21 51 64 93 18 20 25 04 76 59 44 89 90 96 68 95 63 01 42 30 19 23 82 65 67 24 52 29 70 11 27 50 81 15 43 92 39 02 83 38 60 97 37 66 40 49 85 33 13 58 72 08 47 61 88 62 22 31 34 53 10 03 17 87 78 97 06 67 70 33 03 51 43 96 02 63 52 94 10 92 21 34 48 39 13 41 28 75 19 64 15 23 77 59 93 62 69 98 29 14 74 53 57 11 08 65 37 31 88 18 87 45 60 49 32 35 99 58 86 42 50 05 07 47 27 76 09 79 56 30 78 95 38 24 80 89 54 71 81 85 36 91 46 17 00 90 82 01 12 55 40 04 61 68 26 72 84 16 25 83 44 73 22 20 66

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

EEN VERLOTINGSSERIE IS EEN ASELECTE PERMUTATIE VAN DE GETALLEN 0 TOT EN MET n.
 OM EEN VOLLEDIGE SERIE TE KRIJGEN DIENT MEN HORIZONTAAL TE LEZEN.
 IN HET GEDEELTE VAN 0-9 EN VAN 0-24 DIENT MEN TUSSEN IEDERE RIJ EEN SPATIE TE
 LEZEN.

TOEPASSING

ZIE INLEIDING VOOR TOEPASSINGEN

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

TABEL 23

KWADRATEN, WORTELS, FACULTEITEN EN ANDERE FUNCTIES
VAN DE GETALLEN 1 TOT EN MET 100.

n	n^2	\sqrt{n}	$\sqrt{10n}$	n!	log n!	n	n^2	\sqrt{n}	$\sqrt{10n}$	n!	log n!
1	1	1.000	3.162	1.00	0.0000	51	2 601	7.141	22.583	1.55	66.1906
2	4	1.414	4.472	2.00	0.3010	52	2 704	7.211	22.804	8.07	67.9066
3	9	1.732	5.477	6.00	0.7782	53	2 809	7.280	23.022	4.27	69.6309
4	16	2.000	6.325	2.40	1.3802	54	2 916	7.348	23.238	2.31	71.3633
5	25	2.236	7.071	1.20	2.0792	55	3 025	7.416	23.452	1.27	73.1037
6	36	2.449	7.746	7.20	2.8573	56	3 136	7.483	23.664	7.11	74.8519
7	49	2.646	8.367	5.04	3.7024	57	3 249	7.550	23.875	4.05	76.6077
8	64	2.828	8.944	4.03	4.6055	58	3 364	7.616	24.083	2.35	78.3712
9	81	3.000	9.487	3.63	5.5598	59	3 481	7.681	24.290	1.39	80.1420
10	100	3.162	10.000	3.63	6.5598	60	3 600	7.746	24.495	8.32	81.9202
11	121	3.317	10.488	3.99	7.6012	61	3 721	7.810	24.698	5.08	83.7055
12	144	3.464	10.954	4.79	8.6803	62	3 844	7.874	24.900	3.15	85.4979
13	169	3.606	11.402	6.23	9.7943	63	3 969	7.937	25.100	1.98	87.2972
14	196	3.742	11.832	8.72	10.9404	64	4 096	8.000	25.298	1.27	89.1034
15	225	3.873	12.247	1.31	12.1165	65	4 225	8.062	25.495	8.25	90.9163
16	256	4.000	12.649	2.09	13.3206	66	4 356	8.124	25.690	5.44	92.7359
17	289	4.123	13.038	3.56	14.5511	67	4 489	8.185	25.884	3.63	94.5619
18	324	4.243	13.416	6.40	15.8063	68	4 624	8.246	26.077	2.48	96.3945
19	361	4.359	13.784	1.22	17.0851	69	4 761	8.307	26.268	1.71	98.2333
20	400	4.472	14.142	2.43	18.3861	70	4 900	8.367	26.458	1.20	100.0784
21	441	4.583	14.491	5.11	19.7083	71	5 041	8.426	26.646	8.50	101.9297
22	484	4.690	14.832	1.12	21.0508	72	5 184	8.485	26.833	6.12	103.7870
23	529	4.796	15.166	2.58	22.4125	73	5 329	8.544	27.019	4.47	105.6503
24	576	4.899	15.492	6.20	23.7927	74	5 476	8.602	27.203	3.31	107.5196
25	625	5.000	15.811	1.55	25.1906	75	5 625	8.660	27.386	2.48	109.3946
26	676	5.099	16.125	4.03	26.6056	76	5 776	8.718	27.568	1.88	111.2754
27	729	5.196	16.432	1.09	28.0370	77	5 929	8.775	27.749	1.45	113.1619
28	784	5.292	16.733	3.05	29.4841	78	6 084	8.832	27.928	1.13	115.0540
29	841	5.385	17.029	8.84	30.9465	79	6 241	8.888	28.107	8.95	116.9516
30	900	5.477	17.321	2.65	32.4237	80	6 400	8.944	28.284	7.16	118.8547
31	961	5.568	17.607	8.22	33.9150	81	6 561	9.000	28.460	5.80	120.7632
32	1 024	5.657	17.889	2.63	35.4202	82	6 724	9.055	28.636	4.75	122.6770
33	1 089	5.745	18.166	8.68	36.9387	83	6 889	9.110	28.810	3.94	124.5961
34	1 156	5.831	18.439	2.95	38.4702	84	7 056	9.165	28.983	3.31	126.5204
35	1 225	5.916	18.708	1.03	40.0142	85	7 225	9.220	29.155	2.82	128.4498
36	1 296	6.000	18.974	3.72	41.5705	86	7 396	9.274	29.326	2.42	130.3843
37	1 369	6.083	19.235	1.38	43.1387	87	7 569	9.327	29.496	2.11	132.3238
38	1 444	6.164	19.494	5.23	44.7185	88	7 744	9.381	29.665	1.85	134.2683
39	1 521	6.245	19.748	2.04	46.3096	89	7 921	9.434	29.833	1.65	136.2177
40	1 600	6.325	20.000	8.16	47.9116	90	8 100	9.487	30.000	1.49	138.1719
41	1 681	6.403	20.248	3.34	49.5244	91	8 281	9.539	30.166	1.35	140.1310
42	1 764	6.481	20.494	1.40	51.1477	92	8 464	9.592	30.332	1.24	142.0948
43	1 849	6.557	20.736	6.04	52.7811	93	8 649	9.644	30.496	1.16	144.0632
44	1 936	6.633	20.976	2.66	54.4246	94	8 836	9.695	30.659	1.09	146.0364
45	2 025	6.708	21.213	1.20	56.0778	95	9 025	9.747	30.822	1.03	148.0141
46	2 116	6.782	21.448	5.50	57.7406	96	9 216	9.798	30.984	9.92	149.9964
47	2 209	6.856	21.679	2.59	59.4127	97	9 409	9.849	31.145	9.62	151.9831
48	2 304	6.928	21.909	1.24	61.0939	98	9 604	9.899	31.305	9.43	153.9744
49	2 401	7.000	22.136	6.08	62.7841	99	9 801	9.950	31.464	9.33	155.9700
50	2 500	7.071	22.361	3.04	64.4831	100	10.000	10.000	31.623	9.33	157.9700

AFLEESVOORBEELD; VERKLARING VAN SYMBOLEN

$6! = 7.20 \times 10^2$, WAARBIJ DE EXPONENT VAN DE TIENMACHT GEGEVEN IS IN DE KOLOM log n!
EN WEL VÓÓR DE PUNT.

VERANTWOORDING EN UITGEBREIDER TABELLEN

ZIE BLADZIJDE 35 EN VOLGENDE.

VERANTWOORDING VAN BIJ DE TABELLEN GEBRUIK-
TE LITERATUUR EN VERWIJZING NAAR UITGEBREIDER
TABELLEN

TABEL No.	WERD SAMENGESTELD MET BEHULP VAN No.	WERD BEREKEND MET BEHULP VAN No.	UITGEBREIDER TABELLEN IN
1			1 EN 2
2	2 , TABEL IV EN 3 , TABEL III		2 EN 3
3	1 , TABEL 8		1 , 2 EN 3
4	1 , TABEL 18		1 , 2 EN 3
5		1 , TABEL 8	
6	1 , TABEL 30	1 , TABEL 30	1
7	3 , TABEL VII	3 , TABEL III	
8		1 , TABEL 18	
9		4	4
10	5	1 , TABEL 18	5
11	1 , TABEL 7		1
12		1 , TABEL 18	
13		1 , TABEL 8, EN 4	
14	4		4
15	6		6
16	7		
17	8		8
18	9		9
19	1 , TABEL 22	1 , TABEL 22	
20	10		10 EN 11
21			11 EN 12
22			13
23			14

|1| E.S.PEARSON AND H.O.HARTLEY:

BIOMETRIKA TABLES FOR STATISTICIANS, VOL.1 ;
CAMBRIDGE UNIVERSITY PRESS, 1956.

|2| A.HALD:

STATISTICAL TABLES AND FORMULAS ;
JOHN WILEY AND SONS, INC., NEW YORK, 1952.

|3| R.A.FISHER AND F.YATES:

STATISTICAL TABLES FOR BIOLOGICAL, AGRICULTURAL AND MEDICAL RESEARCH;
OLIVER AND BOYD, EDINBURGH, 1956.

|4| THE STAFF OF THE COMPUTATION LABORATORY:

TABLES OF THE CUMULATIVE BINOMIAL PROBABILITY DISTRIBUTION;
CAMBRIDGE UNIVERSITY PRESS, 1955.

- [5] H.C.HAMAKER:
"AVERAGE CONFIDENCE" LIMITS FOR BINOMIAL PROBABILITIES;
REVIEW OF THE INTERNATIONAL STATISTICAL INSTITUTE 21, 1953.
- [6] DORALINE WABEKE EN CONSTANCE VAN EEDEN:
HANDLEIDING VOOR DE TOETS VAN WILCOXON;
REPORT S 176 (M 65) OF THE STATISTICAL DEPARTMENT OF THE
MATHEMATICAL CENTRE, AMSTERDAM, 1955.
- [7] MEJ.L.KAARSEMAKER EN A.VAN WIJNGAARDEN:
TABLES FOR USE IN RANK CORRELATION;
REPORT R 73 OF THE STATISTICAL DEPARTMENT OF THE MATHEMATICAL
CENTRE, AMSTERDAM, 1952.
- [8] A.H.SCHAAFSMA EN F.G.WILLEMZE:
MODERN KWALITEITSBELEID;
PHILIPS TECHNISCHE BIBLIOTHEEK, 1954.
- [9] J.M.CAMERON:
TABLES FOR CONSTRUCTING AND FOR COMPUTING THE OPERATING
CHARACTERISTICS OF SINGLE-SAMPLING PLANS;
INDUSTRIAL QUALITY CONTROL VOL.IX, No.1, PART 1, 1952.
- [10] H.WOLD:
TRACTS FOR COMPUTERS, No.XXV - RANDOM NORMAL DEVIATES;
CAMBRIDGE UNIVERSITY PRESS, 1954.
- [11] ANON.
A MILLION RANDOM DIGITS WITH 100,000 NORMAL DEVIATES;
THE FREE PRESS, GLENCOE, ILL., U.S.A., 1955.
- [12] M.G.KENDAL EN B.BABINGTON-SMITH
TRACTS FOR COMPUTERS, No.XXIV - TABLES OF RANDOM SAMPLING NUMBERS;
CAMBRIDGE UNIVERSITY PRESS, 1954.
- [13] H.C.HAMAKER:
VERLOTINGSSERIES;
AAN TE VRAGEN BIJ: GROEP STATISTIEK, NATUURKUNDIG LABORATORIUM DER
N.V.PHILIPS' GLOEILAMPENFABRIEKEN - EINDHOVEN.
- [14] C.D.HODGMAN:
STANDARD MATHEMATICAL TABLES, 11th EDITION;
CHEMICAL RUBBER PUBLISHING CO., CLEVELAND, OHIO, U.S.A., 1957.