

Statistische Afdeling
van het
Mathematisch Centrum,
Amsterdam

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

Rapport S 83
door
T.J.Terpstra en
Dr J.Hemelrijk Jr

Een onderzoek naar correlaties tussen opeenvolgende generaties
van drosophila.

Maart 1952.

1. Inleiding.

Het verstrekte materiaal (zie bijlage 1) is verkregen op een wijze, die hieronder, voor zoverre het de genen-structuur betreft, in zeer vereenvoudigde vorm beschreven wordt. Een stam drosophila met een abnormaal abdomen is gekruist met een stam drosophila met een normaal abdomen. Indien de genen G bepalend zijn voor een abnormaal abdomen en de gegen g voor een normaal abdomen, wordt deze kruising voorgesteld door

$$\begin{array}{ccc} G G & & g g \\ (\text{abn.abd.}) & & (\text{norm.abd.}) \end{array} \quad (1)$$

(eerste generatie:
F₁)

In de groep F₁ (zie bijlage) zijn uit de met een pijl aangegeven cultures 1, 3, 5, 15 en 17 dieren gekozen, die na kruising de generatie F₂ geven. Hiervoor geldt het schema

$$G g \quad G g \quad (\text{eerste gen.: F}_1)$$

$$\begin{array}{ccc} G G & G g & g g \text{ (tweede gen.: F}_2) \\ (\text{abn.abd.}) & & (\text{norm.abd.}) \end{array} \quad (2)$$

In de groep F₂ zijn vervolgens uit de met pijlen aangegeven cultures 1^b, 1^e, 3^d, 5^e, 5^d, 15^d, en 15^e weer dieren gekozen en gekruist.

De generatie F₃, die dan ontstaat, bezit de genencombinaties

$$\begin{array}{ccc} G G & G g & g g \text{ (derde gen.: F}_3) \\ (\text{abn.abd.}) & & (\text{norm.abd.}) \end{array} \quad (3)$$

De groep G G in F₃ ontstaat door kruising van 2 dieren uit F₂ met genencombinaties G G, de groep g g in F₃ uit 2 dieren in F₂ met genencombinaties g g, en de groep G g in F₃ kan ontstaan zijn uit 2 dieren in F₂ met resp. g g en G G, g g en G g, G G en G g, en G g met G g. Wanneer de genen G en g inderdaad bepalend zijn voor de kenmerken "abnormaal abdomen" en "normaal abdomen", dan zullen de drosophilae in de cultures van generatie F₂ met kleine percentages A.A. (we schrijven in het vervolg kortheidshalve A.A. voor "abnormaal abdomen" en N.A. voor "normaal abdomen") in het algemeen weinig genen G bezitten, en derhalve corresponderen met culturen van F₃, die eveneens kleine percentages A.A. bevatten.

Hetzelfde geldt voor de culturen met grote percentages A.A. Er zal dan dus een positieve correlatie bestaan tussen de percentages A.A. in de cultures van de generaties F₂ en F₃. Deze positieve correlatie is echter niet te verwachten

tussen de percentages A.A. in de cultures van F_1 en F_2 , daar de genen-samenstelling voor alle dieren in F_1 dezelfde is (zie schema's (1) en (2)).

We mogen bovendien nog verwachten, dat de aantallen exemplaren met A.A., voorkomende in ieder der cultures van F_1 (zie bijlage 1), een binomiale verdeling bezitten, waarbij bovendien in iedere culture van F_1 de waarschijnlijkheid van het optreden van een exemplaar met het kenmerk A.A. dezelfde is. Immers de genen-structuur is bij de onderparen, waaruit deze cultures zijn ontstaan, steeds dezelfde. Hetzelfde geldt t.a.v. de aantallen exemplaren met A.A. in de cultures F_2 , daar ieder onderpaar van een culture F_2 dezelfde genenstructuur bezit.

Dit geldt echter niet t.a.v. de aantallen exemplaren met A.A. in de cultures F_3 .

Zoals boven reeds opgemerkt is, is het hier beschreven erfelijkheidsschema een sterk vereenvoudigd model. Er is geen sprake van, dat er slechts één paar genen is, dat bepalend is voor het al of niet abnormaal zijn voor het abdomen. Het is daarom zeer wel mogelijk, dat de hierboven vermelde verschijnselen niet te constateren zijn.

Het doel van het onderzoek is, aan de hand van het verstrekte waarnemingsmateriaal na te gaan of bovenstaande verschijnselen inderdaad toch nog geconstateerd kunnen worden.

2. Het toetsen van de correlaties.

Beschouwen we allereerst de percentages A.A. in de cultures van de generaties F_2 en F_3 .

Indien een positieve correlatie bestaat tussen de percentages A.A. in de twee generaties, zullen, indien de cultures van F_2 naar opklimmende percentages A.A. gerangschikt worden, de percentages A.A. in de cultures van F_3 een stijgend verloop vertonen.

Het al dan niet bestaan van een positieve correlatie kunnen we dus toetsen door een toets tegen verloop (Engels:trend) toe te passen op de gevonden percentages A.A. voor de cultures van F_3 , nadat deze zodanig gerangschikt zijn, dat ze corresponderen met een rangschikking naar opklimmende grootte van de percentages A.A. in die cultures van F_2 , waaruit zij ontstaan zijn. (Zie voor een beschrijving van deze toets de aan het einde van dit rapport toegevoegde

Memoranda S 47 (M 6) en B 73 (M 28)). Deze toets geeft een eenzijdige overschrijdingskans $k=0,08$, welke vrij klein is, maar toch nog te groot om te kunnen concluderen, dat het

f. betekenis

optredende stijgende verloop (d.w.z. optredende positieve correlatie) niet aan het toeval kan worden geweten. (Vgl. Memorandum S 47 (M 6)). In grafiek 1 zijn de verschillende percentages A.A. uitgezet. Ook deze grafiek geeft niet de indruk, dat er van een correlatie van enige sprake zou zijn, hoewel men er een zwakke positieve correlatie uit zou kunnen vermoeden. Indien we, ter controle, de toets tegen verloop toepassen op de percentages A.A. van de cultures van F_2 , waarbij deze weer zodanig gerangschikt zijn, dat ze corresponderen met een rangschikking naar opklimmende grootte van de percentages A.A. in de cultures van F_2 , waaruit zij zijn voortgekomen, dan vinden wij een eenzijdige overschrijdingskans $k=0,20$ (vgl. grafiek 2).

Conclusie: Deze toetsingen geven geen aanleiding, om aan te nemen, dat de correlaties tussen F_1 en F_2 en tussen F_2 en F_3 één van beide van 0 zouden verschillen. Wel is er bij F_2 , F_3 een zwakke suggestie in die richting te bespeuren.

3. De waarschijnlijkheidsverdelingen van de aantallen exemplaren met het kenmerk "Abnormaal Abdomen" in de cultures F_1 , F_2 en F_3

Zoals in §1 reeds is uiteengezet, kunnen we op grond van de geneschema's (1), (2) en (3) verwachten, dat het optreden van A.A. binnen ieder der cultures van generatie F_1 dezelfde is en dat dit eveneens geldt t.a.v. de cultures van F_2 maar niet t.a.v. die van F_3 .

Wij kunnen de gelijkheid van deze waarschijnlijkheden voor ieder der generaties apart toetsen met behulp van de in het toegevoegde Memorandum S 43 (M 31) beschreven χ^2 -toets. Noodzakelijk voor het toepassen van deze χ^2 -toets is echter, dat de verwachtingen van de aantallen exemplaren met het kenmerk A.A. in ieder der cultures (in het Memorandum "steekproeven" genoemd) tenminste vijf is, wat hier niet het geval is. Echter kunnen we voor de generaties F_2 en F_3 op een bij de opzet van het experiment passende wijze, nieuwe steekproeven vormen, door alle cultures, die uit één culture in de vorige generatie voortgekomen zijn, bij elkaar te voegen. Op deze wijze vinden we voor F_2 en F_3 de hieronder getabelleerde steekproeven en aantallen exemplaren met het kenmerk "Abnormale Abdomen".

Generatie F₂

nummer van de steekproef 1)	totaal aantal exemplaren	aantallen met "Abn. Abd."
1	543	14
3	522	24
5	700	30
15	639	24
17	547	13

Generatie F₃

nummer van de steekproef 1)	totaal aantal exemplaren	aantallen met "Abn. Abd."
1 ^b	306	17
1 ^e	213	4
3 ^d	238	7
5 ^a	198	5
5 ^d	272	3
15 ^d	237	6
15 ^e	232	13

Voor de generatie F₁ beschikken we helaas niet over een dergelijke "natuurlijke" groepering, zodat de toets op deze generatie niet kan worden toegepast.

De χ^2 -toets, toegepast op de bovenstaande getabelleerde steekproeven voor F₂ en F₃ geeft resp. de overschrijdingskansen $k=0,18$ en $k=0,11$.

De onderstelling, dat het optreden van het kenmerk "Abnormaal Abdomen" in de verschillende cultures van F₂ even groot is, kan dus op grond van de gedane waarnemingen niet verworpen worden, evenmin als voor de cultures van Generatie F₃.

4. Conclusies en aanwijzingen voor een eventuele nieuwe proefopzet.

Aan de hand van bovenstaande resultaten komen we tot de volgende conclusies:

- 1) Deze nummering komt overeen met de nummering van de cultures in de vorige generatie (zie bijlage 1), waaruit de samengevoegde steekproeven zijn voortgekomen.

1. Er is geen reden, om aan te nemen, dat de kans op het optreden van "Abnormaal Abdomen" in de verschillende cultures van generatie F_2 verschillend is. Hetzelfde geldt voor F_3 .
2. In overeenstemming hiermee is, dat geen positieve correlatie te constateren is tussen de fracties exemplaren met het kenmerk "Abnormaal Abdomen" in de cultures van F_1 en van F_2 . Hetzelfde geldt voor de correlatie tussen deze percentages bij F_2 en F_3 .

Opmerking. Het bovenstaande betekent niet, dat het in de inleiding beschreven vereenvoudigde schema in het geheel niet aansluit bij de werkelijke situatie. Hoewel een verschil tussen de waarschijnlijkheden voor het optreden van A.A. bij de cultures van F_3 niet kan worden geconstateerd, evenmin als een positieve correlatie tussen de percentages A.A. in de cultures van F_2 en F_3 , is het zeer wel mogelijk, dat deze verschijnselen, in zwakke mate, toch aanwezig zijn. Grafiek 1 schijnt hierop wel enigszins te wijzen. Daar de kans op A.A. steeds vrij klein is, kunnen kleine verschillen hierin slechts door proeven van grote uitgebreidheid worden aangetoond. Wij geven, in dit verband, hieronder enkele aanwijzingen voor een proefopzet, die, bij een eventuele herhaling van de proef, de statistische verwerking zouden vergemakkelijken en de kans op het ontdekken van een eventueel aanwezige correlatie (resp. verschil in waarschijnlijkheden) zouden vergroten.

Aanwijzingen voor de proefopzet.

- a. Het aantal cultures F_1 kan beperkt worden tot ongeveer 15. Het is echter wenselijk de uitgebreidheden van de cultures, indien mogelijk, zo groot te maken, dat de aantallen exemplaren met Abnormaal Abdomen in de regel groter dan 5 zijn.
- b. Het verdient aanbeveling de cultures van F_2 en F_3 in twee groepen uiteen te laten vallen: één groep met ouderparen uit cultures van F_1 (resp. F_2) met een zeer gering percentage A.A. en één met ouderparen uit cultures met een hoog percentage A.A. Wat de grootte der cultures betreft geldt hetzelfde als onder a, echter is heir deze voorwaarde minder stringent.
- c. Het is wenselijk dat de percentages A.A. van de cultures van F_2 , waaruit F_3 wordt gekweekt, zo veel mogelijk gelijk genomen worden aan de percentages A.A. van de cultures van F_1 , waaruit F_2 verkregen is. In de grafieken 1 en 2

MATHEMATISCH CENTRUM
AMSTERDAM

zou dit ten gevolge hebben, dat de punten en kruisjes ongeveer voor de helft ver naar links en voor de andere helft ver naar rechts komen te liggen, terwijl de abscis-waarden in beide figuren ongeveer (en wel zo veel mogelijk) op dezelfde plaatsen liggen. Deze maatregelen maken een eenvoudiger en effectiever statistische verwerking mogelijk.

Bijlage 1.

Eerste generatie: F₁ (G g).

Nummer van de culture	Normaal Abdomen	Abnormaal Abdomen	Percentage Abn. Abd.
→ 1	39	2	4,8
2	35	2	5,4
→ 3	47	2	4,0
4	38	4	9,5
→ 5	81	10	11,0
6	37	5	11,9
7	38	2	5,0
8	39	3	7,1
9	49	1	2,0
10	29	1	3,3
11	35	3	7,9
12	29	4	12,2
13	34	1	2,8
14	17	3	15,0
→ 15	50	4	7,4
16	25	3	10,7
→ 17	32	0	0,0
18	30	2	6,2
19	32	1	3,0
21	41	5	10,8
22	23	4	15,8
23	16	7	30,2

→ betekent, dat uit de betreffende cultuur een serie ouderdieren genomen is voor de volgende generatie.
(Vervolg bijlage 1 op volgende pagina.)

Bijlage 1 (vervolg)
Tweede generatie: F₂

num- mer groep	nummer culture	Normaal Abdomen	Abnormaal Abdomen	Percentage Abn.Abd.
1	a	67	0	0,0
	→ b	61	1	1,6
	→ c	92	0	0,0
	d	71	1	1,4
	e	55	3	5,2
	f	38	2	5,0
	g	48	1	2,0
	h	23	3	11,6
	j	74	3	3,9
	3	a	64	2
b		50	1	1,9
c		35	4	10,0
→ d		61	3	4,7
e		54	1	1,8
f		64	7	9,7
g		60	3	4,8
j		64	2	3,0
h		11	1	8,3
k		35	0	0,0
5		→ a	83	2
	b	48	1	2,0
	c	70	1	1,4
	→ d	90	3	3,2
	e	83	3	3,5
	g	64	9	12,3
	h	92	6	6,1
	j	46	1	2,1
	n	50	2	3,8
	u	44	2	4,3
15	a	52	2	3,7
	b	76	3	3,8
	c	41	2	4,6
	→ d	59	6	9,2
	→ e	57	5	8,1
	f	63	0	0,0
	g	64	1	1,5
	j	69	3	4,2
	m	43	0	0,0
	n	39	2	4,9
	o	52	0	0,0
17	a	55	3	5,2
	b	69	5	6,7
	c	57	0	0,0
	d	37	0	0,0
	e	87	2	2,3
	f	55	0	0,0
	g	60	1	1,6
	h	62	2	3,1
	j	52	0	0,0

→ betekent, dat uit de betreffende cultuur een serie ouderdieren genomen is voor de volgende generatie.
(Vervolg van bijlage 1 op volgende pagina.)

Bijlage 1 (vervolg)

Derde generatie: F₃

num- mer groep	Nummer culture	Normaal Abdomen	Abnormaal Abdomen	Percentage Abn. Abd.
1b.	1	86	1	0,012
	2	80	3	0,036
	3	88	3	0,033
	4	51	0	0,000
1c.	1	61	1	0,016
	2	59	1	0,017
	3	35	0	0,000
	4	54	2	0,036
3d.	1	64	1	0,015
	2	58	1	0,017
	3	60	3	0,048
	4	49	2	0,039
5a.	1	32	0	0,000
	2	49	0	0,000
	3	56	4	0,067
	4	56	1	0,018
5d.	1	43	0	0,000
	2	84	1	0,012
	3	62	1	0,016
	4	80	1	0,012
15d.	1	-	-	-
	2	69	3	0,042
	3	72	0	0,000
	4	90	3	0,032
15e.	1	64	2	0,030
	2	53	5	0,086
	3	38	5	0,116
	4	64	1	0,015

Grafiek 1

Percentages Abnormaal Abdomen bij F_2 en F_3

Percentage AA bij F_3

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

0

1

2

3

4

5

6

7

8

9

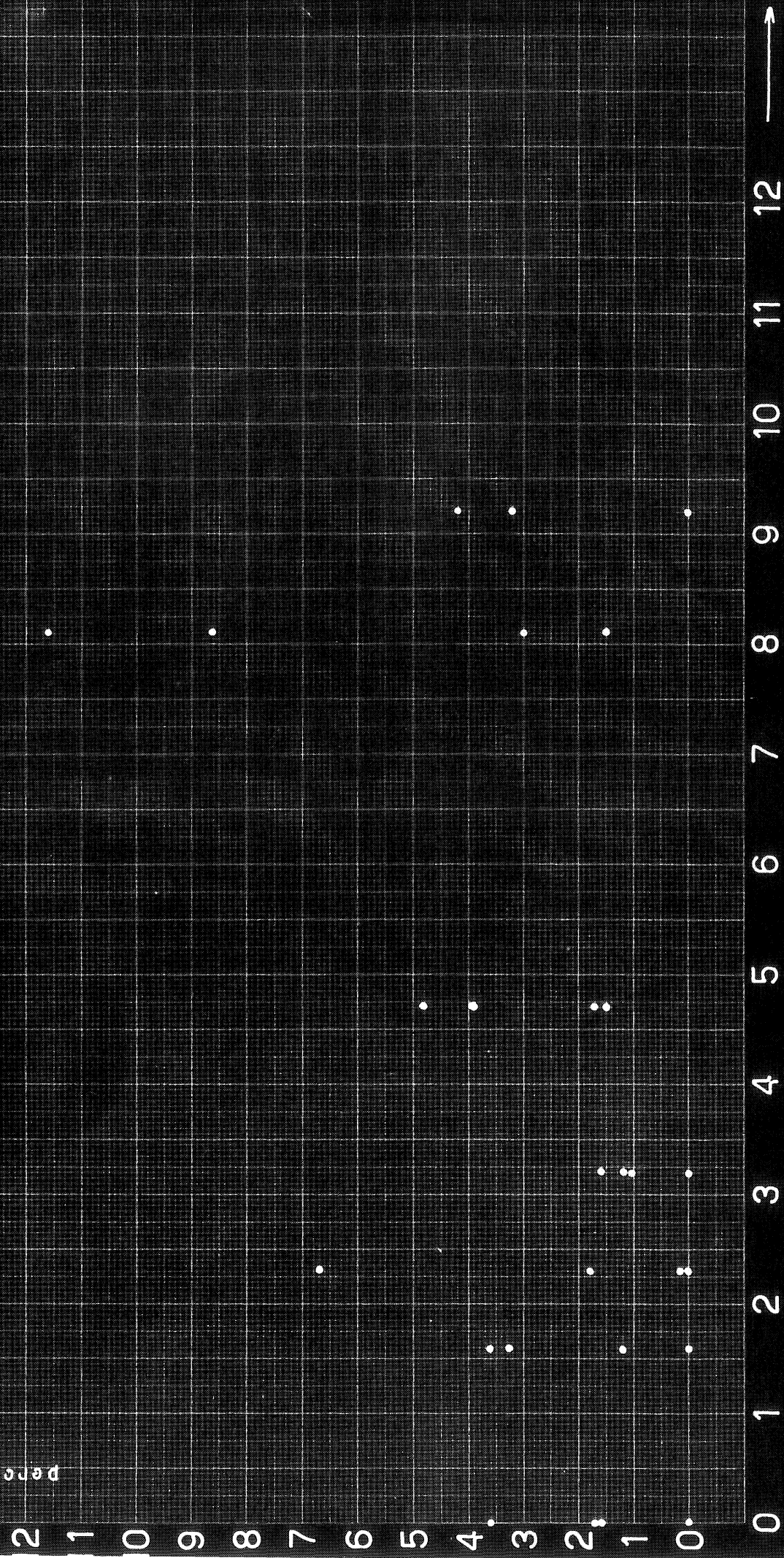
10

11

12

→

percentage AA bij F_2



Grafiek 2

Percentages Abnormaal Abdomen bij F_1 en F_2

Percentage AA bij F_2

0

1

2

3

4

5

6

7

8

9

10

11

12

1

2

3

4

5

6

7

8

9

10

11

12

percentage AA bij F_1

0

1

2

3

4

5

6

7

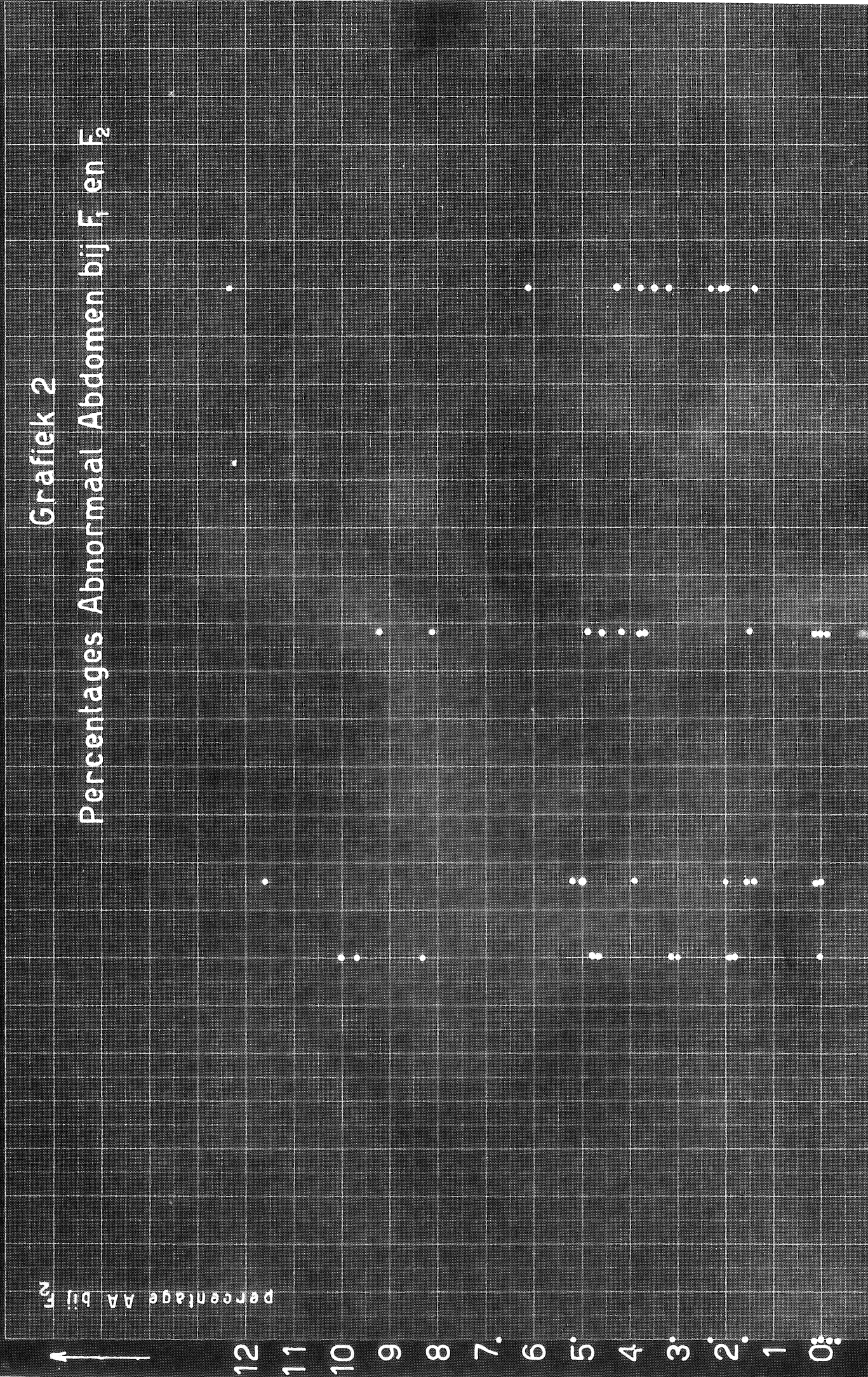
8

9

10

11

12



Algemene gang van zaken bij het toetsen van een ¹⁾
hypothese.

De toetsing van een hypothese H_0 berust steeds op een aantal waarnemingen x_1, x_2, \dots, x_n van één of meer stochastische grootheden ²⁾, of op enige groepen van waarnemingen (bv. twee steekproeven).

Bij een toets behoort een toetsingsgrootheid u (soms meer dan één), die een functie is van bovengenoemde stochastische grootheden en die, voor de waargenomen waarden x_1, x_2, \dots, x_n een waarde aanneemt, die berekend kan worden (bv.: het gemiddelde der waarnemingen, of de spreiding, of het verschil van de gemiddelden van twee waarnemingen).

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat H_0 juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van u , en wel op zodanige wijze, dat de kans, dat u een in Z gelegen waarde aanneemt, onder de hypothese H_0 , gelijk is aan een gegeven getal α , zodat Z dus van α afhankelijk is ³⁾. Z heet de kritieke zone van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance), Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu H_0 op grond van de waarnemingen x_1, x_2, \dots, x_n , indien de bij deze waarnemingen behorende waarde van u in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien H_0 juist is, gelijk aan α . Derhalve is α de kans op ten onrechte verwerping van de juiste hypothese, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met $\alpha = 0,05$ resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, waarin de hypothese die men toetst juist is, deze toch verwerpen.

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

2) Een stochastische grootheid is een grootheid, die een waarschijnlijkheidsverdeling bezit, of, anders gezegd, een grootheid, die voor de elementen van een collectie (universum, populatie) gedefinieerd is en daarop allerlei waarden aanneemt. Stochastische grootheden worden aangegeven door onderstreepte letters.

3) Soms kan men slechts bereiken, dat deze kans $\approx \alpha$ is.

De toetsingstheorie biedt in het algemeen geen mogelijkheid om tot aanvaarding van een hypothese te komen. Indien een bepaalde hypothese H_0 niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

Wel zal men vaak in de loop van een statistische analyse bepaalde onderstellingen, die plausibel schijnen en voor de verdere analyse van nut zijn, toetsen, alvorens ze bij de verdere bewerking van het materiaal te gebruiken. Worden zij dan op grond van de toets niet verworpen, dan houdt dit in zo verre een rechtvaardiging van die onderstellingen in, dat een grote afwijking door de toets veelal wel zou zijn ontdekt. Indien men dan verder de onderstellingen gebruikt, verwaarloost men eventueel aanwezige afwijkingen van onbekende grootte, die echter niet zo groot zijn, dat zij door de toets zijn ontdekt.

Vele toetsen gelden zelf alleen onder bepaalde onderstellingen omtrent de waarschijnlijkheidsverdelingen der stochastische grootheden, waarvan waarnemingen zijn verricht. Deze nevenvoorwaarden dienen steeds uitdrukkelijk te worden vermeld en, zo mogelijk, zelf te worden getoetst.

In plaats van de onbetrouwbaarheidsdrempel α wordt vaak bij de uitslag van een toetsing de overschrijdingskans k opgegeven; dit is de kleinste waarde van α , waarbij in het betrokken geval, nog tot verwerping van H_0 zou zijn overgegaan; anders gezegd: de kleinste α , waarvoor de gevonden waarde der toetsingsgrootte nog juist in de (bij α behorende) kritieke zone Z ligt. Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α , dan wordt verworpen, indien $k \leq \alpha$ is.

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijkte men bv. de tweede hieronder gegeven literatuurplaats. Wij moeten hier volstaan met de opmerking, dat éénzijdige toetsing veelal eerder tot verwerping van H_0 leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

Litteratuur:

J. Neyman, First course in probability and statistics, New York, 1950, Chapter 5.

J. Hemelrijk en H.R. van der Vaart, Het gebruik van één- en tweezijdige overschrijdingskansen voor het toetsen van hypothesen, Statistica 4 (1950) p.54-66.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A M S T E R D A M - O.

Statistische Afdeling.
S 73 (M 28)

Een parameter vrije toets tegen trend voor groepen waarnemingen¹⁾

door T.J. Terpstra.

Januari 1952.

Wij beschouwen het geval, dat h onafhankelijke steekproeven van h stochastische grootheden gegeven zijn:

$x_{1,1}$	$x_{1,2}$	\dots	x_{1,n_1}	van de	grootheid	\underline{x}_1 .
$x_{2,1}$	$x_{2,2}$	\dots	x_{2,n_2}	"	"	" \underline{x}_2 .
.
$x_{h,1}$	$x_{h,2}$	\dots	x_{h,n_h}	"	"	" \underline{x}_h .

De uitgebreidheden van de steekproeven zijn dus n_1, n_2, \dots, n_h en de eerste index van de waarneming $x_{i,j}$ geeft aan uit welke steekproef deze waarneming afkomstig is, terwijl de tweede index het nummer der waarneming binnen die steekproef aangeeft.

De hypothese H_0 , die wij wensen te toetsen luidt, dat al deze waarnemingen van dezelfde stochastische grootheid afkomstig zijn. Anders uitgedrukt: H_0 houdt in, dat de stochastische grootheden $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_h$ onderling onafhankelijk verdeeld zijn en alle dezelfde waarschijnlijkheidsverdeling bezitten.

De alternatieve hypothese, waartegen wij H_0 wensen te toetsen, is, dat de rij grootheden $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_h$ een (stijgende of dalende) trend vertonen. Een precieze definitie van "trend" zullen wij hier niet geven. Grofweg komt een stijgende trend hierop neer, dat van twee waarnemingen, één van \underline{x}_i en één van de meer naar rechts in de rij voorkomende grootheid \underline{x}_j (dus $j > i$), de laatste meer kans heeft om groter te zijn dan de eerste, dan andersom. Dit behoeft, strikt genomen, niet voor iedere i en j met $i < j$ te gelden, maar slechts voor het merendeel van dergelijke paren, maar daar gaan wij nu niet nader op in. Een analoge definitie geldt voor dalende trend.

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

De toetsingsgrootheid T wordt nu als volgt gedefiniëerd: Tel het aantal waarnemingen uit de 2^e , 3^e , ..., h^e steekproef, dat kleiner is dan $x_{1,1}$ (bij gelijkheid tellen wij $\frac{1}{2}$ in plaats van 1) en noem dit aantal $v_{1,1}$. Voer dezelfde telling uit voor $x_{1,2}$, $x_{1,3}$, ..., x_{1,n_1} en noem de gevonden aantallen $v_{1,2}$, ..., v_{1,n_1} . Noem de som van deze aantallen V_1 :

$$V_1 = v_{1,1} + v_{1,2} + \dots + v_{1,n_1}.$$

Tel vervolgens het aantal waarnemingen uit de 3^e , 4^e , ..., h^e steekproef, dat kleiner is dan $x_{2,1}$ (weer $\frac{1}{2}$ tellen bij gelijkheid) en noem dit $v_{2,1}$; analoog $v_{2,2}$, ..., v_{2,m_2} en noem

$$V_2 = v_{2,1} + \dots + v_{2,m_2}.$$

Bij deze tweede stap wordt dus de eerste steekproef buiten beschouwing gelaten. Bij de derde stap laten wij de eerste twee steekproeven buiten beschouwing en bepalen

$$V_3 = v_{3,1} + \dots + v_{3,n_3}$$

op analoge wijze. Dit wordt voortgezet tot en met

$$V_{h-1} = v_{h-1,1} + \dots + v_{h-1,n_{h-1}},$$

die uit de laatste twee steekproeven wordt bepaald. De toetsingsgrootheid is nu

$$T = V_1 + V_2 + \dots + V_{h-1}.$$

Deze grootheid kan dus kortweg gedefiniëerd worden als het aantal paren $(x_{i,a}, x_{j,b})$ met $i < j$ en a en b willekeurig²⁾, waarvoor $x_{i,a} > x_{j,b}$ is, vermeerderd met de helft van die paren, waarvoor $x_{i,a} = x_{j,b}$ is.

Het is duidelijk, dat T vooral grote waarden zal aannemen, indien er een dalende en kleine, indien er een stijgende trend is. Beschouwen wij de verzameling van alle bij het experiment mogelijke uitkomsten, dan bezit T op deze verzameling een waarschijnlijkheidsverdeling (vandaar de onderstreping van de letter T). Indien de hypothese H_0 , inhoudende, dat alle waarnemingen van dezelfde stochastische grootheid afkomstig zijn, juist is, geldt voor deze waarschijnlijkheidsverdeling het volgende:

1. T is bij benadering normaal verdeeld.
2. Het gemiddelde van deze verdeling is:

2) Met uiteraard $1 \leq i < h$, $1 < j \leq h$, $1 \leq a \leq n_i$ en $1 \leq b \leq n_j$.

$$\mu = E(\underline{T}|H_0) = \frac{1}{4} (N^2 - \sum_{i=1}^h n_i^2)$$

en de spreidingskwadraat is

$$\sigma^2 = \sigma^2(\underline{T}|H_0) = \frac{1}{72} \left\{ N(N+1)(2N+1) - \sum_{i=1}^h n_i(n_i+1)(2n_i+1) \right\}$$

met

$$N = \sum_{i=1}^h n_i.$$

De (tweezijdige) toets bestaat nu daarin, dat men H_0 verwerpt, indien de gevonden waarde T van \underline{T} te sterk van μ afwijkt, d.w.z. als

$$\frac{T - \mu}{\sigma} > \xi_{\alpha}, \quad 3)$$

waarin α de onbetrouwbaarheidsdrempel is en ξ_{α} volgt uit

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

en in een tabel van de normale verdeling kan worden opgezocht.

De (tweezijdige) overschrijdingskans k , behorende bij T , is gedefiniëerd als

$$k = \frac{2}{\sqrt{2\pi}} \int_{\left| \frac{T - \mu}{\sigma} \right|}^{\infty} e^{-\frac{1}{2}x^2} dx \quad 3)$$

en kan dus ook in een tabel van de normale verdeling worden gevonden. Bij ééNZijdige toetsing wordt α door 2α vervangen, resp. k gehalveerd.

Opmerkingen.

1. De boven voor σ^2 gegeven formule geldt eigenlijk alleen, indien er geen gelijke waarnemingen zijn. Als er wel gelijke waarnemingen zijn, kan σ^2 op de volgende wijze gecorrigeerd worden. Beschouw alle waarnemingen tezamen; zijn er k groepen van gelijke waarnemingen onder, die resp. m_1, m_2, \dots, m_k elementen

3) Deze formules berusten op de normale benadering van de verdeling van \underline{T} , als H_0 vervuld is. De exacte verdeling van \underline{T} onder H_0 is slechts voor enkele gevallen ($h = 2$ en $n_i \leq 10$, of $n_i = 1$ voor $i = 1, \dots, h$ en $h \leq 40$) bekend.

bevatten, dan wordt de reeds berekende σ^2 verminderd met het volgende bedrag:

$$\begin{aligned} & \frac{1}{72} \sum_{j=1}^k m_j (m_j - 1)(2m_j + 5) + \\ & - \frac{1}{36N(N-1)(N-2)} \left\{ \sum_{i=1}^h n_i (n_i - 1)(n_i - 2) \right\} \left\{ \sum_{j=1}^k m_j (m_j - 1)(m_j - 2) \right\} + \\ & - \frac{1}{8N(N-1)} \left\{ \sum_{i=1}^h n_i (n_i - 1) \right\} \cdot \left\{ \sum_{j=1}^k m_j (m_j - 1) \right\} . \end{aligned}$$

Deze correctie is steeds negatief en is, als de m_i niet te groot zijn, in het algemeen klein, zodat het dan van weinig belang is de berekening uit te voeren.

2. Strikt genomen is de geldigheid van de toets alleen bewezen voor het geval, dat er geen gelijke waarnemingen zijn. Zijn er wel gelijken, dan is het n.l. niet bekend, of de normale benadering nog goed is.

3. De toets is een toepassing van Kendall's theorie over de rangcorrelatiecoëfficiënt τ en kan anderzijds beschouwd worden als een generalisatie van Wilcoxon's toets voor het probleem van twee steekproeven.

MATHEMATISCH CENTRUM,
 2de Boerhaavestr. 49,
 A m s t e r d a m (0)

Statistische Afdeling
 S 73 (M 31)

Toets voor de hypothese, dat een aantal
 waarschijnlijkheden aan elkaar gelijk zijn. 1)

Wij beschouwen k reeksen van onafhankelijke waarnemingen, R_1, R_2, \dots, R_k , waarbij iedere waarneming als resultaat het kenmerk A of het kenmerk \bar{A} (niet $-A$) kan geven. De kans op A is, binnen ieder der reeksen, constant en wel gelijk aan p_i voor de waarnemingen van reeks R_i ($i=1, \dots, k$). Laat het aantal waarnemingen van reeks R_i ($i=1, \dots, k$) gelijk zijn aan n_i en laat hieronder m_i maal het kenmerk A optreden. Gevraagd wordt dan de hypothese

$$H_0: p_1 = p_2 = \dots = p_k$$

te toetsen op grond van deze gegevens. Dit kan geschieden met behulp van de volgende methode.

Bereken

$$m = m_1 + m_2 + \dots + m_k$$

$$n = n_1 + n_2 + \dots + n_k$$

en

$$\chi^2 = \frac{(m_1 - m \frac{n_1}{n})^2}{m \frac{n_1}{n}} + \frac{(m_2 - m \frac{n_2}{n})^2}{m \frac{n_2}{n}} + \dots + \frac{(m_k - m \frac{n_k}{n})^2}{m \frac{n_k}{n}}.$$

Onder de hypothese H_0 bezit deze grootte χ^2 bij benadering een χ^2 -verdeling met $k-1$ vrijheidsgraden (zie b.v. [1] p.445 e.v.). Deze benadering is goed, indien $m \frac{n_i}{n} \geq 5$ voor iedere i (zie [2]).

Indien H_0 onjuist is, dus als er verschillende kansen op A zijn, zal χ^2 gewoonlijk grotere waarden aannemen dan wanneer H_0 vervuld is. De kritieke zône wordt daarom van de vorm

$$\chi^2 \geq \chi_\alpha^2$$

genomen, waarbij α de gekozen onbetrouwbaarheidsdrempel voorstelt. De bij gegeven α behorende waarde van

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

χ^2 kan in tabellen of nomogrammen worden opgezocht (zie [3]), evenals de bij een gevonden waarde van χ^2 behorende overschrijdingskans (dit is de onbetrouwbaarheid van de kleinste kritieke zône van het bovenstaande type, die nog juist de gevonden waarde van χ^2 bevat).

Opmerking. Indien niet voldaan is aan de voorwaarde $m \frac{n_i}{n} \geq 5$ voor iedere i , kan men een (veel bewerkelijker) exacte toets baseren op de voorwaardelijke waarschijnlijkheidsverdeling van de grootheden \underline{m}_i ($i=1, \dots, k$), onder de voorwaarde, dat hun som de waarde m aanneemt ²⁾:

$$P[\underline{m}_1 = m_1, \underline{m}_2 = m_2, \dots, \underline{m}_k = m_k \mid \underline{m}_1 + \underline{m}_2 + \dots + \underline{m}_k = m; H_0] = \\ = \frac{\binom{n_1}{m_1} \binom{n_2}{m_2} \dots \binom{n_k}{m_k}}{\binom{n}{m}}.$$

De geldigheid van deze formule volgt direct uit de waarschijnlijkheidsverdelingen van de \underline{m}_i en van \underline{m} (onder H_0) en uit de definitie van een voorwaardelijke waarschijnlijkheid.

In dit geval definiëren wij de overschrijdingskans behorend bij een gevonden resultaat (m_1, m_2, \dots, m_k) met $m_1 + m_2 + \dots + m_k = m$ als de som van alle waarschijnlijkheden van bovengenoemde verdeling (met de gevonden waarde van m), die hoogstens gelijk zijn aan de waarschijnlijkheid van het gevonden resultaat.

Litteratuur

- [1]. H. Cramér. Mathematical Methods of Statistics, Princeton University Press, 1946.
- [2]. P.G. Hoel. On indices of dispersion. Ann. Math. Stat. 14(1943) p. 155-163.
- [3]. Tabellen en nomogrammen van de χ^2 -verdeling. M.G. Kendall. The advanced theory of Statistics, I, 1947, p. 444-446.
H. Cramér. Mathematical Methods of Statistics, Princeton University Press 1946, p. 559.
Statistica 1 (1946) p. 109.

2) Als wij de grootheden \underline{m}_i als stochastische grootheden (d.z. grootheden met een waarschijnlijkheidsverdeling) beschouwen geven wij dit door onderstreping aan. Niet onderstreepte letters geven waarden aan, die door de stochastische grootheden worden aangenomen.