

STICHTING
MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49
AMSTERDAM

SP 71 C

Rapport Delta Commissie, 3

Bijdragen van het Mathematisch Centrum



1960

BIJDRAGE II.3

TOETSING VAN DE ONAFHANKELIJKHEID
VAN HET HOOGWATER
TE HOEK VAN HOLLAND EN DE
WATERAFVOER VAN DE RIJN BIJ LOBITH

INHOUDSOVERZICHT

0.1 Inhoudsoverzicht	113
0.2 Overzicht van de tabellen	113
0.3 Opgenomen figuur	113
0.4 Summary	114
0.5 Résumé	114
1.0 Samenstelling van de bijdrage	115
1.1 Inleiding.	115
1.2 Auteur en medewerker.	115
2.0 Wiskundige formulering van het probleem	115
3.0 Gebruikte methoden	115
3.1 De periode 1900 t/m 1952 als geheel	115
3.2 De keuze van A en B	116
3.3 Combinaties van de resultaten van de jaren afzonderlijk	117
4.0 Resultaten van het onderzoek	117
5.0 Interpretatie der resultaten	118
0.2 OVERZICHT VAN DE TABELLEN	
3.1.1 Waarnemingen ingedeeld naar twee kenmerken	116
4.0.1 Periode 1900 t/m 1952 als geheel beschouwd	117
4.0.2 Jaren afzonderlijk en dan gecombineerd	118
0.3 OPGENOMEN FIGUUR	
3.1.1 Waarnemingen grafisch uitgezet (voorbeeld)	116

0.4 SUMMARY

Testing the independance of the levels of storm surges at Hook of Holland on the one hand and the discharge of the Rhine at Lobith on the other

As the water level of a lower river is determined by the sea level at the rivermouth and the discharge of the upper river, it is of importance to know whether any relation exists between these two quantities.

In this contribution due to J. Hemelrijk the independance is tested of the levels of storm surges at Hook of Holland on the one hand and the discharge of the Rhine at Lobith on the other.

The test is based on observations covering the period 1900–1952¹⁾, each observation consisting of one observed high tide and one discharge observed at the same time.

As we are especially interested in the occurrence of high water levels we only consider the “dangerous months” i.e. the periods: November–January, November–February and October–March and apply the test successively to each of these periods.

The observations are divided into four groups by dividing the high tides in “high” and “low” high tides, i.e. those higher respectively lower than the level that was exceeded on the average once in two years, and the discharges in “large” and “small” discharges, i.e. those larger respectively smaller than the average discharge of the period considered.

As the occurrence of combinations of high high tides and large discharges is of special interest, the exceedance probability of the observed number of those combinations is computed under the hypothesis that the high tides and discharges are independent. For the period 1900–1952 as a whole this is done by means of a binomial approximation of the method of the double dichotomy (2×2 -table), for the separate years by combining the double dichotomies of each year (using a normal approximation) into one test.

The results of the tests show that there is no indication of dependance of the high tides at Hook of Holland and the discharge of the Rhine at Lobith, neither for the period 1900–1952 as a whole nor for the separate years.

0.5 RÉSUMÉ

Examen de l'indépendance des marées hautes à Hoek van Holland et les débits du Rhin à Lobith

Le niveau de l'eau du cours inférieur d'un fleuve étant déterminé par le niveau de la mer à l'embouchure de ce fleuve et par le débit de son cours supérieur, il importe de savoir s'il y a un rapport entre ces deux quantités.

La présente contribution, due à J. Hemelrijk, traite du test de l'indépendance des marées hautes à Hoek van Holland et les débits du Rhin à Lobith.

Des observations embrassant une période allant de 1950–1952²⁾ nous ont fourni le matériel pour le test, chaque observation se constituant d'une marée haute et d'un débit observés au même instant.

Comme nous avons intérêt plus particulièrement à étudier les niveaux élevés, nous n'avons pris en considération que les „mois dangereux” c.à.d. les périodes de novembre–janvier, de novembre–février et d'octobre–mars et nous avons successivement appliqué le test à chacune de ces périodes.

On a obtenu quatre groupes d'observations en séparant les marées hautes en marées „fortes” et „faibles”, selon qu'elles sont respectivement supérieures ou inférieures au niveau dépassé en moyenne une fois tous les deux ans, et en séparant les débits en „grands” et „petits” débits selon qu'ils sont respectivement plus grands ou plus petits que le débit moyen ayant eu lieu dans la période en question.

Comme les combinaisons de marées hautes fortes et de grands débits présentent un intérêt particulier, on a calculé la probabilité de dépassement du nombre observé de ces combinaisons, en supposant que les marées hautes et les débits sont indépendants. Quant à la période de 1900–1952 prise dans son ensemble, le calcul s'est fait à l'aide d'une approximation binomiale de la méthode de la double dichotomie (la table 2×2), en ce qui concerne les années prises séparément, on l'a fait en réunissant les doubles dichotomies de chaque année en un seul test (en se servant d'une approximation normale).

Les résultats du test démontrent qu'il n'y a aucun indice de l'existence d'une interdépendance des marées hautes à Hoek van Holland et le débit du Rhin à Lobith, ni pour la période de 1900–1952 prise dans son ensemble, ni pour les années prises séparément.

¹⁾ The symbol – is meant as “upto and including”.

²⁾ Par le signe – on entend „jusqu'a . . . inclus”.

1.0 SAMENSTELLING VAN DE BIJDRAGE

1.1 Inleiding

Daar de waterstanden in plaatsen aan de benedenrivieren bepaald worden door zowel de waterstanden aan de kust als de waterafvoeren van de bovenrivieren, is het van belang, na te gaan of er tussen deze beide factoren enig verband bestaat. In het bijzonder interesseert ons het voorkomen van combinaties van hoge hoogwaterstanden aan de kust en grote rivierafvoeren. In deze bijdrage wordt de onafhankelijkheid nagegaan van de H.W.'s (hoogwaterstanden) te Hoek van Holland en de afvoeren van de Rijn bij Lobith.

1.2 Auteur en medewerker

Het onderzoek werd verricht en beschreven door J. Hemelrijk. De berekeningen, in deze bijdrage vermeld, zijn gebaseerd op gegevens van de Rijkswaterstaat en uitgevoerd door F. W. Steutel, assistent van de Statistische Afdeling van het Mathematisch Centrum, die ook bij de opstelling van deze bijdrage zijn assistentie heeft verleend.

2.0 WISKUNDIGE FORMULERING VAN HET PROBLEEM

H.W. te Hoek van Holland noemen we x^1), de afvoer van de Rijn bij Lobith y .

Deze grootheden bezitten kansverdelingen, die geschat kunnen worden op grond van waarnemingen en als x en y *onafhankelijk* zijn, d.w.z. als de kans op het optreden van een waarde van x niet wordt beïnvloed door de waarden, die y aanneemt, kan de *simultane verdeling* van x en y , d.i. de kansverdeling van paren (x, y) berekend worden uit de verdelingen van x en y ²⁾.

Op grond van de simultane verdeling kunnen dan uitspraken worden gedaan over de kansen op optreden van bepaalde combinaties van x en y .

Getoetst wordt in deze bijdrage de hypothese, dat x en y onafhankelijk zijn zowel in ieder jaar afzonderlijk als over een hele beschouwde periode van jaren.

We beschouwen hier alleen de „gevaarlijke” maanden, dat zijn die, waarin de hoge H.W.'s voornamelijk optreden, daar we ons speciaal voor de hoge waterstanden interesseren, t.w. de volgende perioden:

november t/m januari	} over de jaren 1900 t/m 1952.
november t/m februari	
oktober t/m maart	

Beperking van het onderzoek tot deze gedeelten van het jaar geeft een beter inzicht in de situatie en betekent bovendien een aanzienlijke werkbeparing.

3.0 GEBRUIKTE METHODEN

3.1 De periode 1900 t/m 1952 als geheel

Laat het waarnemingsmateriaal over een periode bestaan uit N paren (x_i, y_i) ($i = 1, 2, \dots, N$).

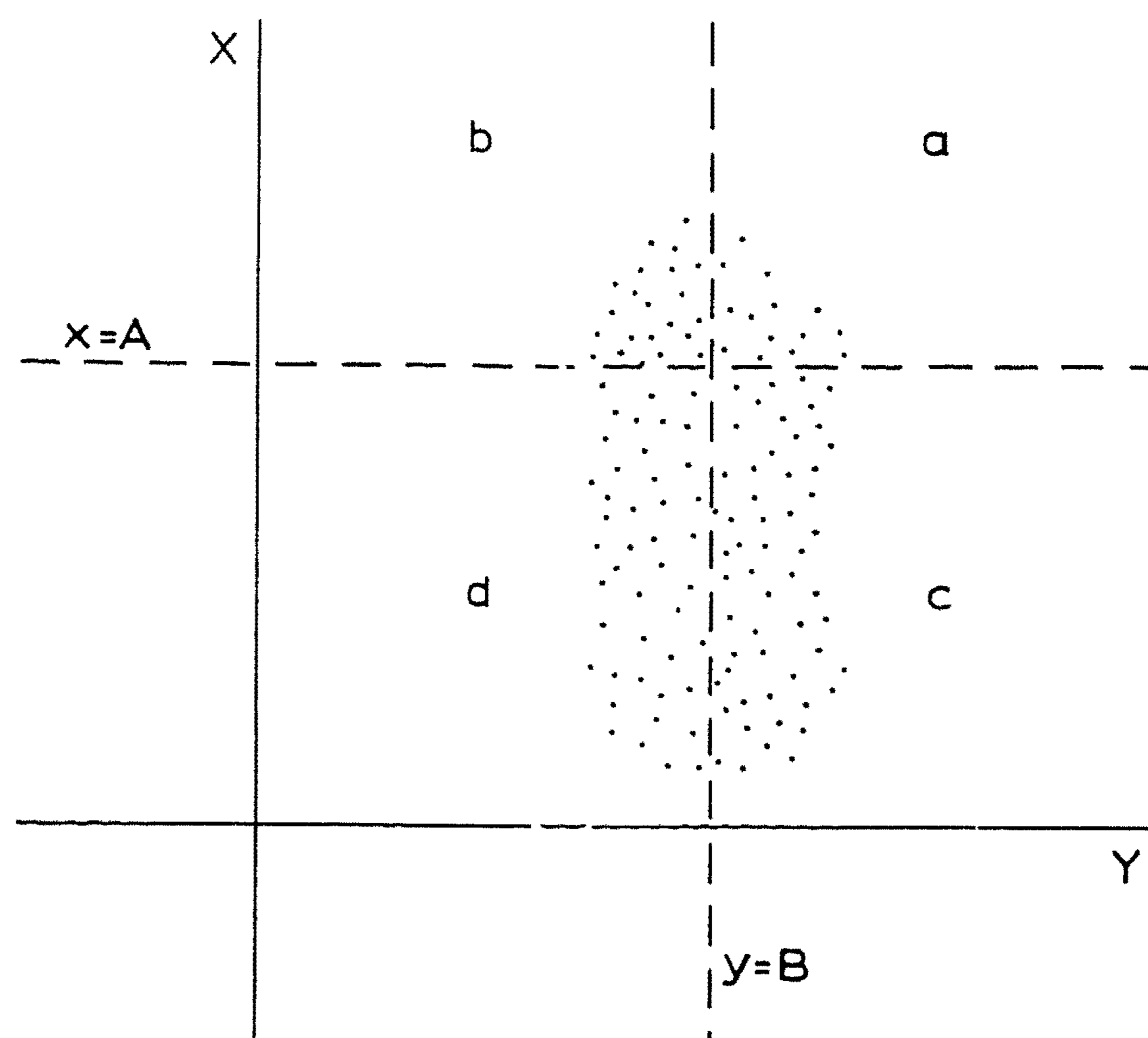
We kunnen deze waarnemingen als een „puntenwolk” uitzetten als in figuur 3.1.1.

¹⁾ Stochastische grootheden, d.w.z. grootheden, die een kansverdeling bezitten, worden onderscheiden van de waarden, die zij bij een experiment aannemen, door vet drukken van hun symbolen.

²⁾ In geval van afhankelijkheid van x en y is de simultane verdeling niet zonder meer uit de verdelingen van x en y af te leiden.

We verdelen nu deze puntenwolk door een horizontale lijn $x = A$ en een verticale lijn $y = B$ in vier delen:

$(x > A, y > B)$, $(x > A, y < B)$, $(x < A, y > B)$ en $(x < A, y < B)$, bestaande uit resp. a , b , c en d punten, zoals in tabel 3.1.1 aangegeven.



Figuur 3.1.1. Waarnemingen grafisch uitgezet (voorbeeld)

Zijn er m waarnemingen met $x > A$ en r met $y > B$, dan is onder de hypothese H_0 : „ x en y zijn onafhankelijk”, a een stochastische grootte met een bekende (hypergeometrische) verdeling, nl.:

$$P\{a \leq a \mid H_0\} = \sum_{i=0}^a \frac{\binom{r}{i} \binom{N-r}{m-i}}{\binom{N}{m}}. \quad (1)$$

Tabel 3.1.1

Waarnemingen ingedeeld naar twee kenmerken

	$y < B$	$y > B$	
$x > A$	b	a	m
$x < A$	d	c	$N - m$
	$N - r$	r	N

$a + b + c + d = N$
 $a + b = m$
 $a + c = r$

3.2 Keuze van A en B

In het onderhavige geval is de lijn $x = A$ gekozen met voor A het *grenspeil*, d.i. het peil, dat gemiddeld éénmaal in twee jaar wordt overschreden, hetgeen bij ca. 730 waarnemingen per jaar impliceert, dat $m \ll N$ is. Voor B is ongeveer het *gemiddelde* gekozen van de waterafvoeren in de beschouwde winterperioden, zodat $r \approx N - r \approx \frac{1}{2} N$ is en dus ook geldt: $m \ll r$ en $m \ll N - r$.

Voor grote N , $m \ll N$ en $m \ll N - r$ kan men (1) van 3.1 benaderen door de *binomiale verdeling*:

$$P\{a \leq a \mid H_0\} = \sum_{i=0}^a \binom{m}{i} \left(\frac{r}{N}\right)^i \left(\frac{N-r}{N}\right)^{m-i}, \quad (1)$$

met gemiddelde $\mathcal{E}(a \mid H_0) = \frac{mr}{N}$ en variantie $\sigma^2(a \mid H_0) = \frac{mr(N-r)}{N^2}$.

Op grond van deze verdeling kunnen we de *rechteroverschrijdingskans*, $k_r(a)$, behorende bij de gevonden waarde a van \mathbf{a} , berekenen. Deze is per definitie:

$$k_r(a) = P\{\mathbf{a} \geq a \mid H_0\} = 1 - P\{\mathbf{a} < a \mid H_0\}, \quad (2)$$

dus hier:

$$k_r(a) = 1 - \sum_{i=0}^{a-1} \binom{m}{i} \left(\frac{r}{N}\right)^i \left(\frac{N-r}{N}\right)^{m-i}, \quad (2')$$

welke kansen men kan vinden in een tabel van de binomiale verdeling.¹⁾

Is de zo gevonden overschrijdingskans kleiner dan een van tevoren gekozen onbetrouwbaarheidsdrempel α (dikwijls wordt gebruikt $\alpha = 0,05$), dan verwerpt men de hypothese H_0 en concludeert: combinaties van grote \mathbf{x} en grote \mathbf{y} komen vaker voor dan met de onafhankelijkheid van deze grootheden te rijmen is, m.a.w. \mathbf{x} en \mathbf{y} zijn niet onafhankelijk.

3.3 Combinaties van de resultaten der jaren afzonderlijk

Op de bovenbeschreven wijze wordt gezocht naar afhankelijkheid over de gehele periode van 52 jaar, doch de eventuele verschillen, tussen de jaren bijv. in aantal stormen, worden daarbij niet in de beschouwingen betrokken. Wenst men te onderzoeken of *binnen* de jaren, als afzonderlijke perioden beschouwd, onafhankelijkheid bestaat, dan dient men op andere wijze te werk te gaan. Men stelt dan tabel 3.1.1 op voor ieder der jaren afzonderlijk en uit de 52 op deze wijze verkregen tabelletjes – die ieder apart te weinig hoge H.W.'s bevatten om toetsing mogelijk te maken – wordt op hieronder te beschrijven wijze een gecombineerde toetsingsgrootte berekend.

De waarnemingen leveren van ieder der k jaren afzonderlijk:

een waarde a_i van \mathbf{a}_i ,

een waarde $\frac{m_i r_i}{N_i}$ van $\mathcal{E}(\mathbf{a}_i \mid H_0)$,

een waarde $\frac{m_i r_i (N_i - r_i)}{N_i^2}$ van $\sigma^2(\mathbf{a}_i \mid H_0)$.

Onder de hypothese H_0' , dat \mathbf{x}_i en \mathbf{y}_i onafhankelijk zijn voor *alle* i , dus in ieder der jaren afzonderlijk, is de grootte

$$z = \frac{\sum_{i=1}^k \left(a_i - \frac{m_i r_i}{N_i} \right)}{\sqrt{\sum_{i=1}^k \frac{m_i r_i (N_i - r_i)}{N_i^2}}}$$

bij benadering normaal verdeeld met gemiddelde 0 en spreiding 1.

Men berekent dus

$$z = \frac{\sum_{i=1}^k \left(a_i - \frac{m_i r_i}{N_i} \right)}{\sqrt{\sum_{i=1}^k \frac{m_i r_i (N_i - r_i)}{N_i^2}}}$$

zoekt $k_r(z)$ op in een tabel van de normale verdeling en verwerpt H_0' , indien $k_r(z) < \alpha$ is.

Opmerking. Daar $r \approx N-r \approx \frac{1}{2} N$ is, door de keuze van het deelpunt B , is de verdeling der \mathbf{a}_i ongeveer symmetrisch, zodat ook voor vrij kleine $m = \sum m_i$ de normale benadering bruikbaar is.

¹⁾ Bijv. Tables of the binomial probability distribution, National bureau of Standards, Appl. Math. Series 6, Washington, 1950.

4.0 RESULTATEN VAN HET ONDERZOEK

De resultaten van het onderzoek volgens 3.1, 3.2 en 3.3 zijn samengevat in de tabellen 4.0.1 en 4.0.2.

Tabel 4.0.1
Periode 1900 t/m 1952 als geheel beschouwd

	november t/m januari	november t/m februari	oktober t/m maart
N	9248	12085	18313
r	3348	4701	6567
m	15	16	20
r/n	0,3620	0,3890	0,3568
a	5	6	7
$k_r(a)$	0,68	0,64	0,62

Tabel 4.0.2
Jaren afzonderlijk en dan gecombineerd

	november t/m januari	november t/m februari	oktober t/m maart
Σa_i	5	6	7
$\Sigma \frac{m_i r_i}{N_i}$	6,2018	6,7614	7,9391
$\Sigma \frac{m_i r_i (N_i - r_i)}{N_i^2}$	2,4449	2,7153	3,5793
z	-0,77	-0,46	-0,50
$k_r(z)$	0,78	0,68	0,69

5.0 INTERPRETATIE DER RESULTATEN

Uit de gevonden overschrijdingskansen blijkt, dat er geen enkele aanleiding is om H_0 of H_0' te verwerpen, zodat we mogen concluderen, dat noch over de perioden als geheel, noch in de jaren afzonderlijk enige afhankelijkheid van grote x -waarden en grote y -waarden blijkt.

Men kan dus voor eventuele verdere berekeningen de H.W.'s te Hoek van Holland en de rivierafvoeren bij Lobith als onafhankelijk beschouwen.