

Onderzoek naar de Razzia in Rotterdam
op 10 en 11 November 1944:

1. Inleiding:

Door het Rijksinstituut voor Oorlogsdocumentatie werden ons gegevens verstrekt over de Razzia, die op 10 en 11 November 1944 in Rotterdam en Schiedam werd gehouden, met het verzoek te trachten de volgende vragen te beantwoorden:

- 1) Zijn er inderdaad 50.000 mannen weggevoerd?
- 2) Welke zijn de middelen waarmee we dit aantal zouden kunnen schatten?
- 3) Hoeveel mannen vertrokken uit Rotterdam lopend, per trein of per boot?

De ons verstrekte gegevens zijn:

- [1] De resultaten van een onder ongeveer 100 teruggekeerde personen gehouden enquête.
- [2] De gegevens van de volkstelling van 31 Mei 1947.
- [3] Gegevens over het aantal rijnen waarmee de mannen werden weggevoerd en tevens de per stuk voor dit vervoer gebruikte oppervlakte.

Gegevens van ooggetuigen over:

- [4] Het aantal treinen waarmee de mannen werden weggevoerd.
- [5] Het aantal mannen dat naar schatting per trein werd vervoerd.

We hebben om van de gegevens van de volkstelling gebruik te maken, moeten onderstellen, dat de volgende voorwaarden bij benadering vervuld zijn:

- a. De mannelijke bevolking tussen 17 en 40 jaar in 1944 was even groot als die bij de volkstelling van 31 Mei 1947.
- b. De totale mannelijke bevolking in 1944 was even groot als die op 31 Mei 1947.
- c. De fractie van het aantal mannen tussen 17 en 40 jaar was in Rotterdam en Schiedam

Schiedam even groot.

Dit is zeker niet geheel juist, maar daar ons geen gegevens ter beschikking staan over 1944, zijn wij wel gedwongen hiervan uit te gaan. De mannelijke bevolking tussen 17 en 40 jaar zal niet veel veranderd zijn, daar de toename der bevolking tussen 1944 en 1947 voor het grootste deel uit geboortes bestond. De totale mannelijke bevolking is hierdoor echter wel toegenomen, zodat we voor de fractie tussen 17 en 40 jaar een te kleine waarde vinden. De fout, die hierdoor ontstaat zal echter wellicht ongeveer worden gecompenseerd doordat er in 1944 een aantal mannen tussen 17 en 40 jaar was ondergedoken of reeds naar Duitsland was weggevoerd.

2. Beantwoording vragen 1 en 2:

Bij de razzia werden Rotterdam en Schiedam in elf wijken ¹⁾ verdeeld. Ieder van deze wijken had een verzamelplaats. Op Bijlage I vindt men het aantal mannen per wijk zoals dit gevonden (en gedeeltelijk geschat ²⁾) is uit de gegevens der volkstelling. Dit geeft een totaal van 348.600.

De fractie hiervan, die tussen 17 en 40 jaar ligt, noemen wij α . Wij vinden deze uit:

Aantal mannen in Gemeente Rotterdam:

318.036. (zie pag. 18 van het boekje "Jaarcijfers 1948, Rotterdam", Rotterdam, Gemeentelijk Bureau voor de Statistiek).

Aantal mannen tussen 17 en 40 jaar: 119.668.
(zie pag. 16 en 18 van "Jaarcijfers 1948, Rotterdam").

¹⁾ Deze wijken zijn niet precies dezelfde als die bij de volkstelling van 31 Mei 1947. Wij merken verder met de wijkindeling van de razzia.

²⁾ Daar de wijkindeling bij de razzia niet precies dezelfde was als die bij de volkstelling hebben we voor sommige wijken het aantal mannen moeten schatten m.b.v. de oppervlakte der wijk.

$$\text{Dit geeft: } \alpha = \frac{119.668}{318.036} = 0,376272.$$

Voor het aantal mannen tussen 17 en 40 jaar in die delen van de Gemeente Rotterdam waar razzia's zijn gehouden, vinden we dus:

$$\alpha \times 348.600 \approx 131.000.$$

Uit de gegevens [4] van §.1 blijkt dat deze mannen zijn weggevoerd:

<u>vanuit:</u>	<u>met:</u>
Amersfoort	3 à 5 treinen
Rotterdam	6 "
Zwolle-Kampen	11 à 13 "
Wexep	-
Delft	4 à 7 "
Totaal	24 à 31 treinen.

Per trein zouden dus, indien alle mannen tussen 17 en 40 jaar weggevoerd waren, ongeveer 5000 mannen weggevoerd zijn.

In de gegevens [5] van §1 wordt gesproken van de volgende schattingen:

1350 per trein
 1500 à 2000 " "
 1500 " "
 2000 à 3000 " " (2x).

Een gemiddelde schatting is dus ongeveer 2000 per trein, zodat volgens deze schatting ongeveer 2/5 van de daarvoor in aanmerking komende mannen werd weggevoerd.

Uitgaande van deze schatting vindt men als schatting voor het aantal weggevoerden

$$\frac{2000}{5000} \times 131.000 \approx 52.500$$

Dit aantal klopt zeer goed met de reeds bestaande schatting van 50.000.

Opmerking: Daar wij niet beschikken over de aantallen mannen tussen 17 en 40 jaar in ieder der wijken apart, nemen we noodgedwongen voor alle wijken gelijk. Ook de factor 2/5, die (naar schatting) het aantal mannen tussen 17 en 40 jaar in 1947 reduceert t

het aantal in 1944 weggevoerde mannen, moeten wij voor alle wijken gelijk nemen.

3. Beantwoording van vraag 3 :

Bij de beantwoording van deze vraag is gebruik gemaakt van de in §1 onder [1] vermelde enquêteresultaten.

Van de binnengekomen antwoorden werd ons opgegeven dat er 867 als betrouwbaar beschouwd kunnen worden, terwijl de overige 90 minder betrouwbaar zijn. Hierbij doet zich nu de vraag voor of we deze laatste 90 gevallen misschien toch mogen gebruiken en om hierop een antwoord te vinden werd onderzocht of ze in overeenstemming zijn met de eerste 867 antwoorden. Het enquêteformulier bevatte vele vragen. Eén van de eenvoudigste hieronder was wel de vraag of men per trein, te voet of per schip uit Rotterdam was weggevoerd. Het is dus niet onredelijk te verwachten dat deze vraag ook door hen die de overige vragen slecht beantwoord hebben, correct beantwoord is.

De overeenstemming tussen de 867 goed beantwoorde en de 90 slecht beantwoorde vragenlijsten, voor zover ^{het} dit genoemd gegeven betreft, werd getoetst met behulp van het χ^2 -criterium (zie bijlage S47 (M6) en S53 (M20)), waarbij getoetst werd of de 90 onbetrouwbare gegevens een steekproef kunnen vormen uit een collectie van de samenstelling, die door de 867 betrouwbare worden ^t gevormd.

Daar de aantallen per versamelplaats per vervoermiddel te klein zijn, kunnen we alleen de overeenstemming van de totalen per vervoermiddel (en even zo die per wijk) tussen de 867 en de 90 gegevens toetsen.

Op Bijlage II vindt men deze totalen en hieruit krijgen we:

Tabel I: ')

Wijze van weg- voeren	n_1	np_1	$\frac{(n_1 - np_1)^2}{np_1}$
Te voet	21	18,1	0,5
Per trein	22	19,1	0,4
Per boot	30	35,0	0,7
Naar Delft	17	17,8	0,0
	90	90,0	$\chi^2 = 1,6$ $v = 3$ $k \approx 0,6$

Bij de totalen per wijk zijn verschillen-
de aantallen nog te klein; vandaar dat enige
wijken samengenomen zijn:

Tabel II:

Verzamel- plaats	n_1	np_1	$\frac{(n_1 - np_1)^2}{np_1}$
I	15	14,9	90,0
II	10	9,3	0,0
III + VIII	10	8,6	0,2
IV	17	20,8	0,7
V + VI + X	11	7,5	1,6
VII + IX	14	16,8	0,5
XI	13	12,1	0,1
	90	90,0	$\chi^2 = 3,1$ $v = 6$ $k \approx 98$

In beide gevallen vinden we een grote o-
verschrijdingskans k . De conclusie is dus:
geen verwerping van de hypothese dat de twee
groepen antwoorden met elkaar in overeenstem-
ming zijn.

') In deze tabellen hebben n_1 en m_1 dezelfde
betekenis als in bijlage II, terwijl $p_1 =$
 $= \frac{N}{867}$ en $n = 90$ is. Verder is $v =$ aantal
vrijheidsgraden (= aantal groepen, vermin-
derd met 1.) en $k =$ de overschrijdingskans.

In verband met deze uitkomst hebben wij verder de genoemde 90 antwoorden bij de groep van 867 gevoegd en met al deze antwoorden tezamen verder gewerkt.

Er doet zich hier nog een tweede vraag voor n.l. of de antwoorden van de enquête per wijk wel representatief zijn voor die wijken. Dit toetsen we weer met het χ^2 -criterium, door de aantallen antwoorden per wijk (nu van beide groepen tezamen; zie bijlage III) te vergelijken met het totaal aantal mannen per wijk (zie bijlage I).

We vinden dan:

Tabel III')

Versamel- plaats.	n_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
I	159	187,1	4,2
II	100	69,3	13,6
III	51	65,1	3,1
IV	216	194,2	2,4
V	46	59,9	3,2
VI	10	19,3	4,5
VII	98	122,3	4,8
VIII	42	51,2	1,7
IX	78	59,6	5,7
X	27	33,8	1,4
XI	130	95,2	12,7
	957	957,0	$\chi^2 = 57,3$ $v = 10$ $k < 0,001.$

Daar in dit geval $k < 0,001$ is, is de con-
clusie: verwerping van de hypothese dat de ant-
woorden representatief zijn voor de wijken.

*) De betekenis van n_i , n , p_i , np_i en v is ana-
loog aan de betekenis van de overeenkomstig-
symbolen in tabel II.

Dit betekent dat we bij de verdere berekeningen de wijken niet tesamen kunnen nemen.

We beschikken nu voor iedere wijk over een schatting voor het totaal aantal mannen (Bijlage I) en voor de fracties die te voet, per trein, per boot of naar Delft zijn weggevoerd (Bijlage III).

Met behulp hiervan en door vermenigvuldiging met de factor $\frac{2}{3} \times = 0,1505$ (zie opmerking aan het einde van § 2) verkrijgen we de volgende tabel:

Tabel IV:')

Het aantal weggevoerde mannen, gesplitst naar vervoersmiddel en naar versamelplaats.

Versamel- plaats	Te voet	Per trein	Per boot	Naar Delft	Totaal
I	-	1800	8450	-	10.250
II	-	1450	2350	-	3800
III	-	300	3300	-	3600
IV	400	-	5050	5250	10700
V	3300	-	-	-	3300
VI	100	-	-	950	1050
VII	5600	-	1100	-	6700
VIII	1850	950	-	-	2800
IX	1250	2000	-	-	3250
X	-	-	600	1250	1850
XI	-	2950	-	2200	5150
	12.500	9450	20.850	9650	52.450 ²⁾

4. Een controleberekening.

Op het aantal mannen dat per boot is weggevoerd hebben we nog een controle. Het aantal mannen dat per boot vervoerd is, is volgens

1) De aantallen in deze tabel zijn afgerond op een veelvoud van 50.

2) Het verschil, dat hier optreedt met het aantal 52.500 van §1, is een gevolg van de afronding.

tabel IV ongeveer 21.000. Hiervan zijn er volgens aan ons verstrekte gegevens ongeveer 1000 vervoerd op begeleidende Duitse schepen, dus ongeveer 20.000 per aak.

Uit de gegevens[3] van §1 vindt men voor het gemiddelde tonnage der aken 1330 ton. Het gemiddeld oppervlak per aak berekenen we uit de gegevens:

1300 ton	oppervlakte 760 m^2
1400 ton	oppervlakte 810 m^2

Lineaire interpolatie geeft: 1330 ton: oppervlakte 775 m^2 .

Hiervan moet, volgens de ons verstrekte gegevens, $150 \text{ à } 200 \text{ m}^2$ afgetrokken worden voor woonruimte, die niet gebruikt werd voor vervoer.

De gebruikte oppervlakte per rijnaak is dus $575 \text{ à } 625 \text{ m}^2$.

Per aak werden gemiddeld vervoerd:

$\frac{20.000}{13} \approx 1550$ mannen. Per m^2 dus ongeveer $2,5 \text{ à } 3$.

Dit klopt goed met het door ooggetuigen vermelde feit, dat men tijdens het vervoer kon zitten, maar slechts met opgetrokken knieën.

6. Splitsing van het aantal weggevoerden naar de datum van wegvoering.

Op 10 November werden, volgens de ons verstrekte gegevens, razzia's gehouden in die delen van Rotterdam, die corresponderen met de verzamelplaatsen I t/m V en een deel van VI en VII.

Als we bij VI en VII de aantallen weggevoerden over de twee delen verdelen in verhouding tot de aantallen inwoners van die delen (geschat naar aanleiding van het oppervlak der delen) dan krijgen we (zie tabel IV):

Tabel V :
10 November:

Verzamel- plaats	te voet	per trein	per boot	naar Delft	Totaal
I	-	1800	8450	-	10.250
II	-	1450	2350	-	3800
III	-	300	3300	-	3600
IV	400	-	5050	5250	10.700
V	3300	-	-	-	3300
VI	100	-	-	950	1050
VII	1850	-	400	-	2250
	5650	3550	19550	6200	34.950

11 November:

Verzamel- plaats	te voet	per trein	per boot	naar Delft	Totaal
VI	-	-	-	-	-
VII	3750	-	700	-	4450
VIII	1850	950	-	-	2800
IX	1250	2000	-	-	3250
X	-	-	600	1250	1850
XI	-	2950	-	2200	5150
	6850	5900	1300	3450	17.500

7. Samenvatting:

Uit het bovenstaande blijkt de schatting van ongeveer 50.000 (van de ongeveer 130.000) weggevoerde mannen in overeenstemming is met de overige gegevens omtrent de razzia.

Liese spreekt in een aan ons verstrekt rapport d.d. 16 November 1944 over ongeveer 54.000 weggevoerden. Hij vermeldt hierin tevens dat het aantal mannelijke inwoners tussen 17 en 40 jaar 60 à 70.000 zou zijn, hetgeen veel te laag is. Wellicht heeft hij dit te lage aantal opgegeven, om het te doen voorkomen alsof vrijwel alle mannen tussen 17 en 40 jaar weggevoerd waren.

Verder is het aantal weggevoerden gesplitst naar vervoermiddel, hetgeen de volgende schat-

tingen gaf:

	Aantallen	in %
Te voet	12.500	24
Per trein	9450	18
Per boot	20.850	40
Naar Delft	9650	18
Totaal	52.450	100

XXX

Bijlage I:

MATHEMATISCH CENTRUM
AMSTERDAM

Aantal mannen per wijk volgens de
volkstelling van 31 Mei 1947:

Verzamel-
plaats

Aantal mannen
in wijk $i=m_i$

I	$\frac{3}{4} x(80) + \frac{2}{3} x(81) + (82) t/m (89) + (00) + (17)$	= 68.200
II	$\frac{1}{4} x(80) + \frac{1}{3} x(81) + \frac{4}{5} x(71) + \frac{1}{2} x(72) + \frac{1}{3} x(73) + (76)$	= 25.250
III	$\frac{1}{5} x(71) + \frac{1}{2} x(72) + \frac{2}{3} x(73) + (74) + (75) + (91) t/m (96)$	= 23.700
IV	Schiedam + (20) t/m (23) + (26) t/m (29)	= 70.750
V	(41) + (42) + $\frac{1}{3} x(44) + (45) t/m (48)$	= 21.800
VI	(33) + (51) + (54) t/m (58) + (66)	= 7.050
VII	(60) t/m (62) + (65) + (34) + (35) + (38) + (39) + $x(15)$	= 44.550
VIII	(14) + (36) + (37) + $\frac{1}{2} x(10)$	= 18.650
IX	(16) + (31) + (32) + $\frac{1}{3} x(15)$	= 21.700
X	$\frac{1}{2} x(10) + (12) + \frac{1}{2} x(13) + \frac{2}{3} x(11) + \frac{3}{5} x(18) + \frac{2}{5} x(19)$	= 12.300
XI	$\frac{2}{3} x(18) + \frac{3}{5} x(19) + (24) + (25) + \frac{1}{3} x(11) + \frac{1}{2} x(13)$	= 34.650

Totaal $m = 348.600$

De getallen tussen de haakjes verwijzen naar de code-nummers van de wijken waarin Rotterdam bij de volkstelling van 31 Mei 1947 was verdeeld.

Alle aantallen zijn afgerond op een veelvoud van 50.

Wijze en plaats van wegvoering der
geenquesteerden:

Wijze van wegvoeren	n_1	n_2
Te voet	174	21
Per trein	124	22
Per boot	338	30
naar Delft	171	17
	$n = 867$	$n = 90$

Verzamel- plaats	n_1	n_2
I	144	15
II	90	10
III + VIII	83	10
IV	199	17
V + IV + X	72	11
VII + IX	162	14
XI	117	13
	$n = 867$	$n = 90$

n_1 resp. n_2 is in deze tabellen het aantal per vervoermiddel en per verzamelplaats weggevoerden onder de groep van 90 resp. 867 geenquesteerden

Totaal aantal antwoorden op de enquête,
gesplitst naar vervoersmiddel en naar
verzamelplaats:

Verzamel- plaats	Te voet	Per trein	Per boot	Naar Belft	Totaal n_i
I	-	28	131	-	159
II	-	38	62	-	100
III	-	4	47	-	51
IV	8	-	102	106	216
V	46	-	-	-	46
VI	1	-	-	9	10
VII	82	-	16	-	98
VIII	28	14	-	-	42
IX	30	48	-	-	78
X	-	-	9	18	27
XI	-	74	1	55	130
	195	206	368	188	$n = 957$

In deze tabel is n_i het totaal aantal antwoorden
van wijk i.

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 grootheden²⁾, 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 is³⁾. 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 verworpt 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 $\leq \alpha$ 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 vergelijk men bv. de tweede hieronder gegeven litteratuurplaats. 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.

De χ^2 -toets (Algemeen).

Wij beschouwen een experiment, dat als resultaat één van de, elkaar uitsluitende, uitkomsten

$$(1) \quad A_1, A_2, \dots, A_k$$

heeft. De χ^2 -toets stelt ons in staat om, op grond van de resultaten van een aantal onafhankelijke uitvoeringen van het experiment, de hypothese H_0 te toetsen, dat de kansen op de uitkomsten $A_1, \dots, A_k$ gelijk zijn aan gegeven getallen

$$(2) \quad p_1, p_2, \dots, p_k,$$

waarbij $p_1 + p_2 + \dots + p_k = 1$ is. Deze voorwaarde is noodzakelijk, daar de uitkomsten elkaar uitsluiten, maar één van alle moeten optreden.

Indien H_0 juist is en er n waarnemingen worden verricht, is de mathematische verwachting van de aantallen malen, dat $A_1, \dots, A_k$ hierbij zullen optreden:

$$(3) \quad np_1, np_2, \dots, np_k.$$

Worden in werkelijkheid de frequenties

$$(4) \quad n_1, n_2, \dots, n_k$$

gevonden, dan zal men, indien de rijen (3) en (4) goed overeenstemmen de hypothese H_0 niet verwerpen, maar indien zij slecht overeenstemmen wel. Als maat voor de overeenstemming van de twee rijen neemt men nu de grootheid χ^2 , die gedefinieerd is als:

$$(5) \quad \chi^2 = \frac{(n_1 - np_1)^2}{np_1} + \frac{(n_2 - np_2)^2}{np_2} + \dots + \frac{(n_k - np_k)^2}{np_k}$$

Bij goede overeenstemming van (3) en (4) is χ^2 klein (bij volledige overeenstemming zelfs 0) en bij slechte overeenstemming is χ^2 groot.

De waarschijnlijkheidsverdeling van χ^2 , onder aanname van de hypothese H_0 , is voor voldoende grote n bij benadering bekend en getabelleerd.

Deze waarschijnlijkheidsverdeling hangt van k af: $v = k - 1$ wordt het aantal graden van vrijheid genoemd. De benadering

is voldoende nauwkeurig, indien voor iedere i geldt:

$$(6) \quad np_i > 10. \quad i = 1, \dots, k.$$

Indien hieraan niet voldaan is, neemt men kenmerken, waarvoor $np_i < 10$ is, tezamen tot nieuwe kenmerken, totdat wel aan (6) voldaan is. (NB: hierdoor vermindert dan de waarde van k , dus het aantal vrijheidsgraden wordt kleiner).

Als kritieke zône wordt, bij gebruik van de onbetrouwbaarheidsdrempel α , het gebied

$$(7) \quad \chi^2 \geq \{\chi_0(\alpha, \nu)\}^2 \text{ genomen,}$$

waarbij $\chi_0(\alpha, \nu)$ zo gekozen is, dat, onder hypothese H_0 , de kans, dat (7) vervuld is, gelijk aan α is.

Litteratuur:

M.G. Kendall, The advanced Theory of Statistics, I, 1947,
Chapter 12.

Tabellen en nomogrammen:

M.G. Kendall, ibidem, p. 444 - 446,

H. Cramér, Mathematical Methods of Statistics, 1946, p.559,
Statistica 1 (1946), p. 109.