STICHTING

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

S 53 (M 15a)

De χ^2-toets voor onafhankelijkheid van 2 grootheden.

MATHEMATISCH CENTRUM
2de Boerhaavestr. 49
A M S T E R D A M O.
Statistische Afdeling.

Rapport S53 (M 15a)

De χ^2-toets voor onafhankelijkheid van 2 grootheden.[1]

De χ^2-toets voor de hypothese H_o van onafhankelijkheid van 2 grootheden $\underline{x}$ en $\underline{y}$ kan beschouwd worden als een uitbreiding van de in het Memorandum S47 (M15) beschreven methode der Dubbele Dichotomie voor het toetsen van dezelfde hypothese.

Bij deze laatste toetsingsmethode wordt de puntenwolk van de waarnemingen $(x_1,y_1),\ldots,(x_N,y_N)$ door één horizontale en één verticale lijn in 4 groepen verdeeld. Deze twee lijnen worden zodanig getrokken, dat de verschillen tussen de aantallen punten boven en beneden de horizontale lijn, resp. links en rechts van de verticale lijn, zo klein mogelijk zijn.

De χ^2-toets voor onafhankelijkheid is een uitbreiding van de toetsingsmethode der Dubbele Dichotomie, in die zin, dat i.p.v. één horizontale en één verticale lijn, verschillende horizontale en verticale lijnen getrokken worden, zodanig dat in alle verticale stroken, voor zover mogelijk, evenveel punten liggen en eveneens in alle horizontale stroken.

Zijn de aantallen horizontale lijnen en verticale lijnen, die getrokken worden, resp. r-1 en s-1, dan kiezen wij r en s zodanig, dat $\frac{N}{r.s} \geqq 10$ is en zo mogelijk in de buurt van 20 ligt. Hierbij is N het totale aantal waarnemingen. Wanneer n.l. r-1 horizontale en s-1 verticale lijnen getrokken worden op boven beschreven wijze, is, onder de hypothese H_o, de verwachting van de aantallen punten in ieder van de r s door de verticale en horizontale lijnen bepaalde vakken, ongeveer gelijk aan $\frac{N}{r\,s}$. Voor toepassing van de χ^2-toets is het noodzakelijk, dat deze aantallen $\geqq 10$ zijn.

In het algemeen worden r en s ongeveer even groot gekozen, terwijl de keuze soms, door de aard van het speciale onderzoek bepaald wordt.

---

[1] Dit memorandum is slechts bedoeld ter oriëntering en streeft niet naar volledigheid of volledige exactheid.

Uit het bovenstaande blijkt, dat de χ^2-toets voor onafhankelijkheid alleen gebruikt kan worden voor een groot aantal waarnemingen. Wanneer het aantal klein is, kan men beter de exacte toetsingsmethode der Dubbele Dichotomie toepassen. (Zie hiertoe Memorandum S47 (M15)).

Nadat de horizontale en verticale lijnen aangebracht zijn, worden de aantallen punten in de rs vakken geteld. Stel deze aantallen zijn $f_{1,1}$, $f_{1,2}, \dots, f_{r,s}$.
Het volgende schema wordt nu opgesteld:

$f_{1,1}$ $w_{1,1}$ $t_{1,1}$	$f_{1,2}$ $w_{1,2}$ $t_{1,2}$			$f_{1,s-1}$ $w_{1,s-1}$ $t_{1,s-1}$	$f_{1,s}$ $w_{1,s}$ $t_{1,s}$	V_1
$f_{2,1}$ $w_{2,1}$ $t_{2,1}$	$f_{2,2}$ $w_{2,2}$ $t_{2,2}$			$f_{2,s-1}$ $w_{2,s-1}$ $t_{2,s-1}$	$f_{2,s}$ $w_{2,s}$ $t_{2,s}$	V_2
$f_{r-1,1}$ $w_{r-1,1}$ $t_{r-1,1}$	$f_{r-1,2}$ $w_{r-1,2}$ $t_{r-1,2}$			$f_{r-1,s-1}$ $w_{r-1,s-1}$ $t_{r-1,s-1}$	$f_{r-1,s}$ $w_{r-1,s}$ $t_{r-1,s}$	V_{r-1}
$f_{r,1}$ $w_{r,1}$ $t_{r,1}$	$f_{r,2}$ $w_{r,2}$ $t_{r,2}$			$f_{r,s-1}$ $w_{r,s-1}$ $t_{r,s-1}$	$f_{r,s}$ $w_{r,s}$ $t_{r,s}$	V_r
H_1	H_2			H_{s-1}	H_s	N

In bovenstaand schema zijn H_1, $H_2, \dots, H_s$, V_1, $V_2, \dots, V_r$ de zgn. "randtotalen", N is het totale aantal waarnemingen.

Wanneer x en y werkelijk twee onafhankelijke grootheden zijn, dan bezitten bij gegeven randtotalen de waarden f_{ij} als verwachting de waarden:

$$w_{ij} = \frac{H_i V_j}{N} \ , \quad 1 \leqq i \leqq r \ , \quad 1 \leqq j \leqq s .$$

Deze waarden $w_{i,j}$ zijn dus uit de gegeven randtotalen te berekenen en worden in het schema ingevuld. Daarna wordt voor ieder der r s vlakken de waarde

$t_{ij} = \dfrac{(f_{i,j} - w_{i,j})^2}{w_{i,j}}$ bepaald en weer in het schema ingevuld.

Als toetsingsgrootheid bij deze toets wordt de grootheid χ^2 gebruikt, met

$$\chi^2 = (t_{1,1} + \dots + t_{1,s}) + (t_{2,1} + \dots + t_{2,s}) + \dots + (t_{r,1} + \dots + t_{r,s}) = \sum_{i,j} t_{i,j}$$

welke onder de hypothese H_0 een χ^2-verdeling bezit, met (r-1)(s-1) vrijheidsgraden.(d.i. het aantal f's dat vrij gevariëerd kan worden, zonder de randtotalen te wijzigen).

Voor deze χ^2-verdeling bestaan tabellen en nomogrammen, die men in vrijwel ieder statistisch leerboek kan vinden.

Grote waarden van χ^2 wijzen op grote verschillen tussen de gevonden aantallen $f_{i,j}$ en hun verwachtingen $w_{i,j}$, berekend in de onderstelling, dat H_o juist is. We verwerpen daarom de hypothese H_o indien

$$\underline{\chi}^2 \geqq \chi_\alpha^2$$

is, waarbij χ_α^2 de waarde van χ^2 is, waarvoor

$$P\left[\underline{\chi}^2 \geqq \chi_\alpha^2\right] = \alpha$$

Hierin is α de onbetrouwbaarheidsdrempel. De hypothese H_o zal dan gemiddeld in een fractie α van die gevallen, waarin H_o juist is, ten onrechte verworpen worden.

De waarde χ_α^2, waarvoor bovenstaande betrekking geldt, kan gemakkelijk uit de tabellen of nomogrammen van de χ^2-verdeling bepaald worden.

<u>Litteratuur:</u>

M.G.Kendall, The advanced Theory of Statistics, Volume I (London 1947).

A.M.Mood, Introduction to the Theory of Statistics (London 1950).

R.A.Fisher and F.Yates, Statistical Tables, (London 1949).