MATHEMATISCH CENTRUM
2de Boerhaavestr.49
A M S T E R D A M 0.
Statistische Afdeling
S53 (M20)

De χ^2-toets (Algemeen).[1)]

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

(1) $A_1, A_2, \ldots, A_k$

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_o te toetsen, dat de kansen op de uitkomsten $A_1, \ldots, A_k$ gelijk zijn aan gegeven getallen

(2) $p_1, p_2, \ldots, p_k$,

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

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

(3) $np_1, np_2, \ldots, np_k$.

Worden in werkelijkheid de frequenties

(4) $n_1, n_2, \ldots, n_k$

gevonden, dan zal men, indien de rijen (3) en (4) goed overeenstemmen de hypothese H_o 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) $$\chi^2 = \frac{(n_1-np_1)^2}{np_1} + \frac{(n_2-np_2)^2}{np_2} + \ldots + \frac{(n_k-np_k)^2}{np_k}\ .$$

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

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

De waarschijnlijkheidsverdeling van χ^2, onder aanname van de hypothese H_o, is voor voldoende grote n bij benadering bekend en getabelleerd.
Deze waarschijnlijkheidsverdeling hangt van k af: $\nu = k - 1$ wordt het <u>aantal graden van vrijheid</u> genoemd. De benadering is voldoende nauwkeurig, indien voor iedere i geldt:

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

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) $\quad \chi^2 \geqq \left\{ \chi_o(\alpha,\nu) \right\}^2$ genomen,

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

<u>Litteratuur:</u>
M.G. Kendall, The advanced Theory of Statistics, I, 1947, Chapter 12.

<u>Tabellen en nomogrammen:</u>
M.G. Kendall, ibidem, p.444 - 446,
H.Cramér, Mathematical Methods of Statistics, 1946, p.559,
Statistica <u>1</u> (1946), p. 109.