

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 53 (M 20)

De χ^2 -toets (Algemeen)



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 grootte 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: $\nu = k - 1$ wordt het aantal graden van vrijheid genoemd. De benadering

dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid

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.