STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 73 (M 31)

Toets voor de hypothese, dat een aantal waarschijnlijkheden aan elkaar gelijk zijn.

1952

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
Amsterdam (O)

Statistische Afdeling
S 73 (M 31)

Toets voor de hypothese, dat een aantal waarschijnlijkheden aan elkaar gelijk zijn. 1)

Wij beschouwen k reeksen van onafhankelijke waarnemingen, $R_1, R_2, \ldots, R_k$, waarbij iedere waarneming als resultaat het kenmerk A of het kenmerk $\overline{A}$ (niet -A) kan geven. De kans op A is, binnen ieder der reeksen, constant en wel gelijk aan p_i voor de waarnemingen van reeks $R_i (i=1,\ldots,k)$. Laat het aantal waarnemingen van reeks $R_i (i=1,\ldots,k)$ gelijk zijn aan n_i en laat hieronder m_i maal het kenmerk A optreden. Gevraagd wordt dan de hypothese

$$H_o: \quad p_1 = p_2 = \ldots = p_k$$

te toetsen op grond van deze gegevens. Dit kan geschieden met behulp van de volgende methode.

Bereken

$$m = m_1 + m_2 + \ldots + m_k$$
$$n = n_1 + n_2 + \ldots + n_k$$

en

$$\chi^2 = \frac{(m_1 - m\frac{n_1}{n})^2}{m\frac{n_1}{n}} + \frac{(m_2 - m\frac{n_2}{n})^2}{m.\frac{n_2}{n}} + \ldots + \frac{(m_k - m\frac{n_k}{n})^2}{m\frac{n_k}{n}} .$$

Onder de hypothese H_o bezit deze grootheid χ^2 bij benadering een χ^2-verdeling met k-1 vrijheidsgraden (zie b.v. [1] p.445 e.v.). Deze benadering is goed, indien $m\frac{n_i}{n} \geqq 5$ voor iedere i (zie [2]).

Indien H_o onjuist is, dus als er verschillende kansen op A zijn, zal χ^2 gewoonlijk grotere waarden aannemen dan wanneer H_o vervuld is. De <u>kritieke zône</u> wordt daarom van de vorm

$$\chi^2 \geqq \chi_\alpha^2$$

genomen, waarbij α de gekozen <u>onbetrouwbaarheidsdrempel</u> voorstelt. De bij gegeven α behorende waarde van

---

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

χ^2_α kan in tabellen of nomogrammen worden opgezocht (zie [3]), evenals de bij een gevonden waarde van χ^2 behorende <u>overschrijdingskans</u> (dit is de onbetrouwbaarheid van de kleinste kritieke zône van het bovenstaande type, die nog juist de gevonden waarde van χ^2 bevat).

<u>Opmerking</u>. Indien niet voldaan is aan de voorwaarde $m\frac{n_i}{n} \geqq 5$ voor iedere i, kan men een (veel bewerkelijker) exacte toets baseren op de voorwaardelijke waarschijnlijkheids-verdeling van de grootheden $\underline{m}_i (i=1,\dots,k)$, onder de voorwaarde, dat hun som de waarde m aanneemt [2]:

$$P\left[\underline{m}_1=m_1,\underline{m}_2=m_2,\dots,\underline{m}_k=m_k \mid \underline{m}_1+\underline{m}_2+\dots+\underline{m}_k=m;\ H_o\right] =$$
$$= \binom{n_1}{m_1}\binom{n_2}{m_2}\cdots\binom{n_k}{m_k}\Big/\binom{n}{m}.$$

De geldigheid van deze formule volgt direct uit de waarschijnlijkheidsverdelingen van de $\underline{m}_i$ en van $\underline{m}$ (onder H_o) en uit de definitie van een voorwaardelijke waarschijnlijkheid.

In dit geval definiëren wij de overschrijdingskans behorend bij een gevonden resultaat $(m_1,m_2,\dots,m_k)$ met $m_1+m_2+\dots+m_k=m$ als de som van alle waarschijnlijkheden van bovengenoemde verdeling (met de gevonden waarde van m), die hoogstens gelijk zijn aan de waarschijnlijkheid van het gevonden resultaat .

<u>Litteratuur</u>

[1]. <u>H. Cramér</u>. Mathematical Methods of Statistics, Princeton University Press, 1946.

[2]. <u>P.G. Hoel</u>. On indices of dispersion. Ann. Math. Stat. <u>14</u>(1943) p. 155-163.

[3]. <u>Tabellen en nomogrammen</u> van de χ^2-verdeling.
<u>M.G. Kendall</u>.The advanced theory of Statistics, I, 1947, p. 444-446.
<u>H. Cramér</u>. Mathematical Methods of Statistics, Princeton University Press 1946,p. 559.
<u>Statistica 1</u> (1946) p. 109.

---

2) Als wij de grootheden $\underline{m}_i$ als stochastische grootheden (d.z.grootheden met een waarschijnlijkheidsverdeling) beschouwen geven wij dit door onderstreping aan. Niet onderstreepte letters geven waarden aan, die door de stochastische grootheden worden aangenomen.