STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 229 (M 79)

Het combineren van een aantal onafhankelijke toetsen voor de gelijkheid van twee kansen.

MATHEMATISCH CENTRUM,
2e Boerhaavestraat 49,
A M S T E R D A M - O.
Statistische Afdeling.

Rapport S 229 (M 79).

S M 79

Het combineren van een aantal onafhankelijke toetsen voor de gelijkheid van twee kansen 1)

Met behulp van de methode der 2x2-tabel (zie memorandum S 53 (M 23)) kan men de hypothese toetsen dat twee kansen, p en p', aan elkaar gelijk zijn. Om deze toets te kunnen toepassen moet men beschikken over twee onafhankelijke reeksen van onafhankelijke experimenten, waarbij ieder experiment twee mogelijke uitkomsten: "succes" (S) of "mislukking" (M) heeft. Laat nu de eerste (resp. tweede) reeks uit m (resp. n) experimenten met a (resp. b) maal S en $c = m - a$ (resp. $d = n - b$) maal M bestaan. Deze gegevens kan men als volgt samenvatten:

Tabel 1

	Aantal malen		
	S	M	totaal
eerste reeks	a	c	m
tweede reeks	b	d	n
totaal	r	s	N.

Laat verder p (resp. p') de kans op succes voorstellen bij ieder experiment van de eerste (resp. tweede) reeks dan kan men met behulp van de methode der 2x2-tabel de hypothese toetsen dat $p = p'$ is. De toets kan echter alleen dan worden toegepast als, zowel binnen de eerste als binnen de tweede reeks, de kans op S voor ieder experiment dezelfde is. Is dit niet het geval dan is het vaak mogelijk het waarnemingsmateriaal te splitsen in een aantal (stel h) groepen, zodanig dat binnen ieder van deze groepen wel aan de genoemde voorwaarde voldaan is. Men krijgt dan h tabellen van de vorm van tabel 1, dus

---

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

	aantal malen		totaal	
	S	M		
eerste reeks	a_i	c_i	m_i	voor $i = 1, \ldots, h$,
tweede reeks	b_i	d_i	n_i	
totaal	r_i	s_i	N_i	

waarbij, in de i^e tabel, p_i (resp. p'_i) de kans op succes voor ieder experiment van de eerste (resp. tweede) reeks voorstelt.

Op grond van deze h tabellen kan men dan de hypothese H_0 toetsen dat $p_i = p'_i$ voor iedere i. Hiervoor kan men een van de volgende twee methoden gebruiken:

1. Bereken voor iedere i de grootheid

$$A_i = \frac{a_i d_i - b_i c_i}{m_i n_i}.$$

De toetsingsgrootheid is dan

$$A = A_1 + \ldots + A_h.$$

Bereken verder

$$B_i = \frac{r_i s_i}{n_i m_i (N_i - 1)} \qquad \text{en} \quad B = B_1 + \ldots + B_h$$

dan bezit de toetsingsgrootheid $\underline{A}$ (onder de hypothese H_0) voor voldoende grote $N = N_1 + \ldots + N_h$ bij benadering een normale verdeling met gemiddelde 0 en variantie B. De grootheid $\underline{u} = \frac{A}{\sqrt{B}}$ bezit dus bij benadering een $N(0,1)$-verdeling zodat de overschrijdingskans in een tabel van de normale verdeling kan worden opgezocht.

Past men deze toets tweezijdig toe, d.w.z. bepaalt men de tweezijdige overschrijdingskans van u, dan leidt de toets speciaal tot verwerping van de getoetste hypothese als

$$\sum_{i=1}^{h} (p_i - p'_i) \neq 0$$

is. Bij linkseenzijdige toetsing is de toets speciaal gevoelig voor de alternatieve hypothesen

$$\sum_{i=1}^{h} (p_i - p'_i) < 0$$

en bij rechtseenzijdige toetsing voor de alternatieven

$$\sum_{i=1}^{h} (p_i - p_i') > 0.$$

2. Bij de tweede methode berekent men voor iedere i

$$C_i = \frac{A_i^2}{B_i} = \frac{(a_i d_i - b_i c_i)^2}{r_i s_i m_i n_i} (N_i - 1).$$

De toetsingsgrootheid is dan

$$C = C_1 + \ldots + C_h$$

en deze grootheid bezit onder de hypothese H_0 bij benadering een χ^2-verdeling met h vrijheidsgraden.
De (rechtseenzijdige) overschrijdingskans kan dus in een tabel van de χ^2-verdeling worden opgezocht.

Deze toets is speciaal gevoelig voor alternatieve hypothesen, waarbij $p_i - p_i' \neq 0$ voor minstens één waarde van i.

Opmerking

De eerste hierbovenbeschreven methode komt overeen met de eerste in memorandum S 102 (M 17b) beschreven toets, als we daarin $c_i = \frac{N_i}{m_i n_i}$ nemen. De tweede methode uit dit memorandum is identiek met methode 4 uit S102 (M 17b).