MATHEMATISCH CENTRUM
AMSTERDAM
Statistische Afdeling.
Rapport S58
door Dr J.Hemelrijk Jr.

Februari 1951.

Vergelijking van een viertal waarnemingsreeksen met de mediaantoets.

1. De waarnemingen

Het in tabel I vermelde waarnemingsmateriaal werd ons voorgelegd met het verzoek deze 4 reeksen onderling te vergelijken; de bedoeling was in het byzonder hier niet de toets van Student toe te passen, daar de gegevens in vrij grote intervallen samengevat zijn en bovendien sommige der reeksen een nogal scheve verdeling vertonen.

Tabel I

Interval	Waarnemingsreeks I	II	III	IV
46-75	2	5	2	10
76-105	21	10	18	30
106-135	19	22	45	50
136-165	7	9	35	17
166-195	2	4	18	9
196-225	1	2	10	2
225-255	0	0	4	0

Voor de betekenis der horizontale dubbele streep vergelijke men de volgende paragraaf.

Ter verduidelijking van dit verslag is het memorandum S47(M6) over de toetsingstheorie als bijlage toegevoegd.

2. De gebruikte toetsingsmethode.

Als toets werd de mediaantoets gekozen[1]), die ook voor de discontinue gegevens gebruikt kan worden. Ter demonstratie van deze methode geven wij de volledige berekening voor de

1) Zie b.v. J.Hemelrijk, Symmetrietoetsen en andere toepassingen van de theorie van Neyman en Pearson, Diss. 1950, hoofdstuk 2 en, voor de gebruikte benaderingsmethode, opmerking 2 op blz 32. De mediaan van een waarschijnlijkheidsverdeling is die waarde, die dezelfde kans bezit om overschreden als om niet bereikt te worden.

vergelijking van de reeksen I en II.

A. Beide reeksen worden tezamen genomen en de zo verkregen groep waarnemingen wordt in twee groepen verdeeld, die uitsluitend grote, resp. kleine waarden bevatten. Daarbij wordt deze verdeling zo gekozen, dat het verschil van het aantal elementen in beide groepen zo klein mogelijk is. Deze verdeling is in tabel I door een horizontale dubbele streep aangegeven.

B. Vervolgens worden de gegevens, zoals in tabel II is aangegeven, in een 2 x 2 tabel samengevat.

Tabel II.
Waarn.-reeksen I en II in 2 x 2- tabel

	I	II	totaal
boven de dubbele streep:	23	15	38
onder de dubbele streep:	29	37	60
totaal	52	52	104

Noemen wij het aantal waarnemingen in het linksboven gelegen vakje a (dus hier is a = 23), dan is de mathematische verwachting (het "theoretisch gemiddelde") van a, onder de (te toetsen) hypothese, dat de waarnemingsreeksen uit dezelfde collectie (verdeling, populatie) komen, gelijk aan:

$$E\underline{a} = \frac{38 \times 52}{104} = 19.$$

Verder geldt voor de spreiding van $\underline{a}$:

$$\sigma_{\underline{a}} = \sqrt{\frac{38.60.52.52}{104^2.103}} = \frac{1}{0,425} .$$

Wij berekenen nu de grootheid

$$\frac{a - E\underline{a}}{\sigma_{\underline{a}}} = \frac{23 - 19}{0,425} = 0,09,$$

waarin a = 23 het bij het experiment gevonden aantal waarnemingen in het linkerbovenvakje aangeeft en kunnen dan, wegens de asymptotische normaliteit van deze grootheid (vgl. voetnoot 1)) de bijbehorende (tweezijdige) overschrijdingskans in een tabel van de normale verdeling opzoeken.

Deze bedraagt in dit geval 0,09, zodat wij niet kunnen concluderen tot een verschil tussen de waarnemingsreeksen

I en II; dat wil zeggen het gevonden verschil kan zonder bezwaar aan het toeval worden toegeschreven.

3. De resultaten.

Dezelfde toets werd op alle andere paren waarnemingsreeksen toegepast. Wij vonden daarbij de volgende overschrijdingskansen:

Tabel III.
resultaten van de mediaantoets.

Vergeleken waarn.-reeksen	Overschrijdingskans	Reeks met grootste mediaan
I,II	0,09	
I,III	0,0001	III
I,IV	0,27	
II,III	0,007	III
II,IV	0,54	
III,IV	0,0001	III

Alle overschrijdingskansen zijn tweezijdig.

De conclusie luidt dus: tussen de waarnemingsreeksen I,II, en IV is met deze toets geen verschil te ontdekken. De reeks III echter is ten opzichte van deze 3 reeksen sterk naar rechts verschoven.

4. Opmerkingen.

1. Het is niet onmogelijk, dat door de groepering der waarnemingen zoveel nuance verloren is gegaan, dat eventueel tussen de waarnemingsreeksen I,II, en IV bestaande niet-toevallige verschillen niet meer kunnen worden geconstateerd. Daarbij kan men dan in de eerste plaats denken aan de reeksen I en II, waarbij de overschrijdingskans vrij klein is.

2. Bij de boven beschreven methode wordt vaak een z.g. "continuïteitscorrectie" toegepast, die ten doel heeft onverantwoorde conclusies, die als gevolg van het gebruik van een benaderingsmethode kunnen optreden, te vermijden. Toepassing

van deze correctie vergroot alle overschrijdingskansen enigszins, maar deze verandering is gering en brengt in dit geval geen wijziging in de conclusies te weeg.

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.

S47(M6)

Algemene gang van zaken bij het toetsen van een hypothese.[1]

De toetsing van een hypothese H_0 berust steeds op een aantal waarnemingen $x_1, x_2, \dots, x_n$ van één of meer stochastische grootheden[2], of op enige groepen van waarnemingen (bv. twee steekproeven).

Bij een toets behoort een toetsingsgrootheid $\underline{u}$ (soms meer dan één), die een functie is van bovengenoemde stochastische grootheden en die, voor de waargenomen waarden $x_1, x_2, \dots, x_n$ een waarde aanneemt, die berekend kan worden (bv.: het gemiddelde der waarnemingen, of de spreiding, of het verschil van de gemiddelden van twee waarnemingen).

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat H_0 juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van $\underline{u}$, en wel op zodanige wijze, dat de kans, dat $\underline{u}$ een in Z gelegen waarde aanneemt, onder de hypothese H_0, gelijk is aan een gegeven getal α, zodat Z dus van α afhankelijk is[3]. Z heet de kritieke zône van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu H_0 op grond van de waarnemingen $x_1, x_2, \dots, x_n$, indien de bij deze waarnemingen behorende waarde van $\underline{u}$ in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien H_0 juist is, gelijk aan α. Derhalve is α de kans op ten onrechte verwerping van de juiste hypothese, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met α = 0,05 resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, waarin de hypothese die men toetst juist is, deze toch verwerpen.

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

2) Een stochastische grootheid is een grootheid, die een waarschijnlijkheidsverdeling bezit, of, anders gezegd, een grootheid, die voor de elementen van een collectie(universum,populatie) gedefinieerd is en daarop allerlei waarden aanneemt. Stochastische grootheden worden aangegeven door onderstreepte letters.

3) Soms kan men slechts bereiken, dat deze kans $\leq \alpha$ is.

De toetsingstheorie biedt in het algemeen <u>geen</u> mogelijkheid om tot <u>aanvaarding</u> van een hypothese te komen. Indien een bepaalde hypothese $\mathcal{H}_0$ niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

Wel zal men vaak in de loop van een statistische analyse bepaalde onderstellingen, die plausibel schijnen en voor de verdere analyse van nut zijn, toetsen, alvorens ze bij de verdere bewerking van het materiaal te gebruiken. Worden zij dan op grond van de toets niet verworpen, dan houdt dit in zo verre een rechtvaardiging van die onderstellingen in, dat een grote afwijking door de toets veelal wel zou zijn ontdekt. Indien men dan verder de onderstellingen gebruikt, verwaarloost men eventueel aanwezige afwijkingen van onbekende grootte, die echter niet zo groot zijn, dat zij door de toets zijn ontdekt.

Vele toetsen gelden zelf alleen onder bepaalde onderstellingen omtrent de waarschijnlijkheidsverdelingen der stochastische grootheden, waarvan waarnemingen zijn verricht. Deze <u>nevenvoorwaarden</u> dienen steeds uitdrukkelijk te worden vermeld en, zo mogelijk, zelf te worden getoetst.

In plaats van de onbetrouwbaarheidsdrempel α wordt vaak bij de uitslag van een toetsing de <u>overschrijdingskans</u> k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van $\mathcal{H}_0$ zou zijn overgegaan; anders gezegd: de kleinste α, waarvoor de gevonden waarde der toetsingsgrootheid nog juist in de (bij α behorende) kritieke zône Z ligt. <u>Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leq \alpha$ is.</u>

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijke men bv. de tweede hieronder gegeven litteratuurplaats. Wij moeten hier volstaan met de opmerking, dat éénzijdige toetsing veelal eerder tot verwerping van $\mathcal{H}_0$ leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

<u>Litteratuur:</u>

J.Neyman, <u>First</u> course in probability and statistics, New York, 1950, Chapter 5.

J.Hemelrijk en H.R. van der Vaart, Het gebruik van één- en tweezijdige overschrijdingskansen voor het toetsen van hypothesen, Statistica <u>4</u> (1950) p.54-66.