MATHEMATISCH CENTRUM
AMSTERDAM

Statistische Afdeling. December 1951.
Rapport S 72,

door
Prof.Dr D.van Dantzig,
Mej.C.van Eeden en
Dr J.Hemelrijk Jr.

Verslag van een oriënterende statistische analyse van een onderzoek van cosmische straling.

1. Inleiding.

Door Prof.Dr J.Clay werden ons waarnemingen van de cosmische straling verstrekt, die ieder uur, gedurende ongeveer 2 maanden (10 Mei tot 15 Juli 1948) zijn verricht.

De gegevens bestonden uit:

a. Uurwaarnemingen van de intensiteit der cosmische straling, verricht met tellers en met een ionisatievat.

b. Voor ieder uur één waarneming van de barometerstand en één van de temperatuur.

De bedoeling van de statistische analyse van deze gegevens is, na te gaan, of er uit deze gegevens een conclusie gevormd kan worden over de aanwezigheid van een dagperiodiciteit van de intensiteit der cosmische straling, die niet door de invloed van de barometerstand of van de temperatuur ontstaat.

Hoewel het mogelijk zou zijn de gegevens aan een Fourier-analyse te onderwerpen en op die wijze een schatting te verkrijgen van de amplitude van de term, die de periode van 1 dag vertegenwoordigt, waarbij een formele eliminatie van de invloed van de barometerstand wel op een of andere manier kan worden uitgevoerd, werd in deze richting om twee redenen geen poging gewaagd. De eerste was, dat wij vooralsnog niet beschikken over een bij deze analyse aansluitende toetsingsmethode. Wij zouden dus een schatting van de amplitude van de dagperiodiciteit vinden zonder dat wij er ons van zouden kunnen vergewissen of deze amplitude werkelijk van nul verschilt dan wel of de afwijking van nul aan het toeval toegeschreven zou kunnen worden. Het gevonden getal zou dus geen physische interpretatie toelaten.

Weliswaar is het in principe wel mogelijk, hieraan tegemoet te komen, door de Fourier componenten te bepalen behorend bij perioden van b.v. 22, 23, 24, 25, 26 uren, en na te gaan of er bij 24 uren een piek optreedt, maar dit zou zeer veel en zeer kostbaar rekenwerk vereisen, terwijl de uitslag zeer twijfelachtig is. We hebben echter een andere, minder kostbare methode in gedachte, waartoe een aantal voorbereidende onderzoekingen zijn geschied, die hierna vermeld zullen worden.

De tweede reden bestaat daarin, dat we niet over een duidelijk fysisch model beschikken, dat ondubbelzinnig vastlegt, van welke grootheid de periodiciteit bepaald moet worden. Is $\underline{x}$ de waargenomen barometerstand, en $\underline{y}$ de waargenomen intensiteit der straling, dan kan men vragen naar een periodiciteit van een grootheid van de vorm $\underline{y}+\beta\underline{x}$, of ook in een van de vorm $\underline{y}.\underline{x}^{\beta}$, (of van log $\underline{y}$+βlog $\underline{x}$) Ook indien $\underline{x}$ niet al te grote schommelingen om een gemiddelde waarde uitvoert, waardoor de bepaling der <u>gemiddelden</u> in beide gevallen ongeveer tot elkaar herleidbare resultaten geeft, kan echter juist een Fourier component in beide gevallen wezenlijk verschillen. (B.v. is de bij sin3x behorende Fourier component in de ontwikkeling van sin x nul, maar in die van log(a+sin x) ongelijk nul). De vraag, welke keuze hier gedaan moet worden, kan niet op statistische gronden beantwoord worden, maar alleen op een fysische overweging, die beslist, of, als $\underline{y}_0$ de intensiteit der in de atmosfeer binnenvallende straling is, de waargenomen intensiteit daaruit ontstaat door vermindering met een term $\beta\,\underline{x}$, dan wel door vermenigvuldiging met een van $\underline{x}$ afhankelijke <u>factor</u>. Overigens zou ook dit bezwaar bij de bovenbedoelde statistische methode minder zwaar wegen.

Verder lijkt het niet uitgesloten, dat een hoge barometerstand zou kunnen ontstaan door verschillende oorzaken (een dikke luchtlaag of neerwaartse luchtstromingen b.v.), die verschillende invloeden op de intensiteit der cosmische straling uitoefenen.

Het leek daarom gewenst, eerst de invloed van de barometerstand op de cosmische straling te onderzoeken. De invloed van de temperatuur, die veel geringer is, werd tot nu toe niet systematisch in het onderzoek betrokken.

Ter oriëntering werd eerst van de verschillende waarnemingsreeksen een voortschrijdend gemiddelde bepaald. De zo verkregen tijdreeksen werden vervolgens onderzocht op de in § 2 besproken wijze.

Daarna werd de methode der klassieke regressieanalyse voor een preciezer onderzoek ingeschakeld. Dit gedeelte van het onderzoek is vervat in § 3.

Bij dit onderzoek werden een aantal ter plaatse te bespreken merkwaardige conclusies bereikt, die het verdere onderzoek enigszins bemoeilijken en de aanleiding waren tot het opstellen van dit voorlopige rapport.

Ter oriëntering van de lezer is aan het eind van dit rapport een memorandum S 47 (M 6) toegevoegd over de algemene gang van zaken bij het toetsen van hypothesen. De in het rapport genoemde grafieken zijn groot en moeilijk te reproduceren. Zij zijn daarom niet aan dit rapport toegevoegd.

2. Methode der voortschrijdende gemiddelden.

Door de Rekenafdeling van het Mathematisch Centrum werden de sommen van telkens 12 opvolgende waarnemingen bepaald, die dus, op een factor $\frac{1}{12}$ na, de voortschrijdende gemiddelden vormen. Dit werd uitgevoerd voor alle waarnemingsreeksen. Bovendien werden van de zo verkregen reeksen de verschillen van opeenvolgende waarden bepaald, hetgeen, in de oorspronkelijke waarnemingen uitgedrukt, neer komt op het verschil van de $n+12^e$ en de n^e waarneming, voor $n=1,2,3,\ldots$

De reeksen gaven, grafisch uitgezet, een vrij regelmatig verloop te zien, waaruit de invloed van de barometerstand duidelijk blijkt. Bij stijgende barometer treedt daling van de intensiteit der cosmische stralen op en omgekeerd. Opvallend is echter, dat de verandering in de intensiteit der cosmische straling soms onevenredig groot is in vergelijking met de verandering van de barometerstand. Dit verschijnsel treedt b.v. zeer duidelijk aan de dag op 14 Juni.

Om het verband tussen de intensiteit der cosmische stralen en de barometerstand nader te beschouwen, werden die perioden uitgekozen, waarin de stijging van het voortschrijdend gemiddelde van de barometerstand ongeveer constant was (dus waarin de verschillen tussen $n+12^e$ en n^e waarneming constant waren). Voor die perioden werd de stijging per uur van de voortschrijdende gemiddelden van de waarnemingen der cosmische stralen geschat en tegen de stijging van barometerstand uitgezet. Deze grafiek geeft een duidelijke daling van de intensiteit der cosmische stralen te zien bij stijgende barometerstand.

Uit deze grafiek hebben wij, apart voor stijgende en dalende barometerstanden, een schatting afgeleid voor de helling van de rechte, die gemiddeld het verband tussen cosmische stralen en barometerstand geeft. Hierbij was geen enkel verschil te zien tussen het verband bij dalende en dat bij stijgende barometer.

De schattingen werden verkregen door in de genoemde grafiek een lijn door 0 te trekken, zo, dat aan beide zijden van deze lijn een gelijk aantal waarnemingspunten lag. Daarbij werd rekening gehouden met het aantal der oorspronkelijke waarnemingen, dat in ieder punt van deze hellingen-grafiek verwerkt was. De helling van deze lijn is een schatting van de regressiecoëfficiënt van het verband tussen de flucties van barometerstand en teller-waarnemingen, maar tevens van de regressiecoëfficiënt tussen deze twee grootheden zelf. Dit alles geldt uiteraard slechts onder de aanname van een lineair verband tussen de betrokken grootheden.

Bovenbeschreven methode werd zowel op de teller- als op de ionisatievatwaarnemingen toegepast, en uit het quotiënt van beide hellingen werd een schatting gevonden voor de helling van de rechte die het verband aangeeft tussen de waarnemingen van de tellers- en van het ionisatievat.

In de grafiek van de flucties van barometerstand en teller-waarnemingen komen twee punten voor, die opvallend ver van de andere afwijken. Deze punten zijn afkomstig van waarnemingen verricht op 23 Mei en op 10 Juni. Uit de aantekeningen omtrent magnetische storingen e.d., die bij het waarnemingsmateriaal waren verstrekt, bleek niet, dat er op die dagen iets bijzonders aan de hand was. In i iets mindere mate valt een punt, behorende bij de waarnemingen van 4 Juni buiten de wolk, Deze drie punten vertoonden eveneens een afwijkend gedrag in de overeenkomstige grafiek behorend bij de waarnemingen van het ionisatievat. Deze overeenstemming wijst erop, dat het afwijkende gedrag niet aan één van beide apparaturen geweten kan worden tenzij er een gemeenschappelijke oorzaak, b.v.een sterke daling van de spanning van het lichtnet, in het spel is. Een fout in de registratie van de barometerstand zou natuurlijk wel de oorzaak kunnen zijn, echter ook een atmospherische storing of een cosmischeverschijnsel. Wij beschikken niet over de gegevens, die nodig zijn, om hier een conclusie te formuleren.

Er waren verschillende redenen, die ons ertoe brachten, om behalve de hier beschreven methode ook nog de klassieke regressieanalyse in het onderzoek te betrekken. Deze regressieanalyse, toegepast op de oorspronkelijke waarnemingen, stelt ons nl. in staat de gelijkheid van hellingen te toetsen, hetgeen met de bovenbeschreven oriënterende methode niet mogelijk is. Wij kunnen dus nagaan, of de barometerinvloed constant was gedurende de periode van de proef. Tevens is het interessant, om de hellingen, berekend volgens verschillende methoden met elkaar te vergelijken. De onderstelling, dat de stochastische afwijkingen een normale waarschijnlijkheidsverdeling bezitten, noodzakelijk voor het toepassen van de hier gebruikelijke statistische toetsen, lijkt bij het onderhavige materiaal niet zo onredelijk, dat aan de realiteit van verschijnselen, die bij deze analyse duidelijk naar voren komen, getwijfeld behoeft te worden.

3. Methode der regressieanalyse:

3.1. Methode van onderzoek:

Bij de toepassing van deze methode werd de gehele waarnemingsperiode oorspronkelijk in twee gelijke perioden verdeeld n.l.

1^e periode: 10 Mei tot 12 Juni
2^e periode: 13 Juni tot 15 Juli

Boverdien werden de dag- en nachtwaarnemingen gescheiden zowel als tezamen onderzocht. De scheidingslijn tussen dag en nacht werd daarbij bepaald op een in § 3.2.1 beschreven wijze naar aanleiding van de temperatuur. Een eventuele invloed van de temperatuur of van de stand van de zon zou op die wijze kunnen blijken.

Voor beide perioden, en voor dag en nacht apart, werd nu een grafiek gemaakt van de waarnemingen (in de oorspronkelijke vorm) van tellers resp. ionisatievat, tegen de barometerstand. De ligging der punten verschilde duidelijk voor de beide perioden, zowel bij tellers als bij ionisatievat , overdag en 's nachts, en voor al deze 4 grafieken op gelijksoortige wijze. De regressielijn van de tweede periode is ten opzichte van die van de eerste naar rechts gedraaid om een nogal ver weg gelegen punt, behorend bij lage barometerstand en hoge intensiteit der cosmische straling. Het verschil tussen beide perioden is dus het grootst bij de hoge barometerstanden, en bij overgang van de eerste naar de tweede periode is er een toeneming van de intensiteit der straling. Deze toename is vermoedelijk een bekend verschijnsel (seizoeninvloed of iets dergelijks). De verandering der helling is echter vermoedelijk minder bekend. De details van het onderzoek zijn verderop beschreven.

De veranderlijkheid van de barometer-invloed gaf ons aanleiding, ook een grafiek te maken van de ionisatievatwaarnemingen tegen de teller-waarnemingen. Ook hier bleken de regressielijnen der beide perioden duidelijk te verschillen. Zij waren weliswaar zo goed als evenwijdig, maar een verschuiving was duidelijk te constateren. In de tweede periode lagen de tellerwaarnemingen gemiddeld hoger in vergelijking met de bijbehorende ionisatievatwaarnemingen dan in de eerste. Details worden in § 3.3 besproken. Deze opschuiving van de regressielijn kan een gevolg zijn van een verloop van twee apparaturen ten opzichte van elkaar. Wij kunnen er niet over oordelen, of er misschien ook andere verklaringen mogelijk zijn.

In sommige der hier besproken grafieken komen weer punten voor, die sterk van de overige afwijken. Ook hiervoor konden wij uit de aantekeningen omtrent magnetische storingen e.d. geen verklaring vinden.

3.2. Onderzoek naar het verschil in helling:

3.2.1 De splitsing naar dag en nacht:

De splitsing naar dag en nacht is gemaakt aan de hand van de temperatuur. Hiertoe werd op de temperatuurwaarnemingen de methode der m rangschikkingen toegepast (Vgl. S 47 (M 14) aan het eind van dit rapport.) Elke rangschikking is hierbij afkomstig van de 24 temperatuurwaarnemingen van één dag. De uren, waarop wij de hoogste twaalf kolomtotalen vonden werden tot de dag gerekend; de andere tot de nacht.

Op deze wijze vonden wij:

dag: 10-21 uur,
nacht: 22-9 uur.

3.2.2. De hellingen der regressielijnen.

De schattingen der hellingen, berekend uit de oorspronkelijke waarnemingen, waarover in § 3.1 gesproken is, zijn in onderstaande tabellen samengevat. Daarbij is steeds de barometerstand als onafhankelijke variabele genomen.

Tabel I:

Schattingen der hellingen van tellers tegen barometerstand.

periode \ temp.	dag 10-21 uur	nacht 22-9 uur	dag + nacht
1	- 4,75	- 5,03	-4,90
2	- 3,44	- 3,64	-3,54
1+2	- 4,46	- 4,72	-4,59

Tabel II:

Schattingen der hellingen van ionisatievat tegen barometerstand.

periode \ temp.	dag 10-21 uur	nacht 22-9 uur	dag + nacht
1	-4,47	- 4,33	-4,41
2	-3,01	- 3,08	-3,04
1+2	-3,96	- 3,89	-3,93

Hieruit zien wij direct, dat er voor de beide perioden grote verschillen zijn, waarvan in de volgende paragraaf dan ook zal blijken, dat zij niet aan het toeval geweten kunnen worden. Er is echter betrekkelijk weinig verschil tussen dag en nacht. Men kan dus wel onderstellen dat de invloed van de barometerstand op de cosmische straling in eerste benadering overdag en 's nachts dezelfde is.

In tabel III zijn de schattingen der hellingen van de ionisatievat-waarnemingen tegen de teller-waarnemingen opgenomen, berekend als quotiënten van de corresponderende hellingen uit de vorige twee tabellen.

Tabel III:

Schattingen der hellingen van ionisatievat tegen tellers.

periode \ temp.	dag 10-21 uur	nacht 22-9 uur	dag + nacht.
1	0,94	0,86	0,90
2	0,88	0,85	0,86
1+2	0,89	0,82	0,86

Hier is het verschil tussen de hellingen voor de twee perioden veel kleiner dan boven. De verschuiving van de regressielijnen ten opzichte van elkaar wordt in § 3.3 besproken.

3.2.3. Toetsing van het verschil tussen de hellingen.

Hierbij is de in memorandum S 53 (M27) beschreven methode toegepast (zie eind van dit rapport), die op de toets van Student berust.

De resultaten van deze toets vinden we voor de vergelijkingen van 1e en 2e periode in tabel IV en voor de vergelijking van dag en nacht in tabel V.
De grootheid t stelt de voor de toetsingsgrootheid gevonden waarde voor.

Tabel IV:

Toepassing van de toets van Student op de hellingen uit de 1e en 2e periode.

	Periode	tijd	hellingen	t	Overschrijdingskans
Tellers tegen barometer	1e 2e	10-21	-4,75 -3,44	5,0	<0.00001
	1e 2e	22-9	-5,033 -3,64	5,6	<0.00001
	1e 2e	dag + nacht	-4,90 -3,54	7,4	<<0.00001
Ionisatievat tegen barometer	1e 2e	10-21	-4,47 -3,01	6,2	<0.00001
	1e 2e	22-9	-4,33 -3,08	6,2	<0.00001
	1e 2e	dag + nacht	-4,41 -3,04	8,8	<<0.00001

De overschrijdingskansen zijn alle zeer klein, zodat de toets steeds tot verwerping van de hypothese van gelijkheid der regressiecoëfficiënten leidt. Hieruit zien we dus dat er een duidelijk verschil is tussen de hellingen uit de eerste en die uit de tweede periode, de helling is in de tweede periode groter, d.w.z. minder sterk negatief dan in de eerste.

Tabel V:

Toepassing van de toets van Student op de hellingen overdag en 's nachts.

	Periode	tijd	hellingen	t	overschrijdingskans
Tellers tegen barometerstand	1	10-21 22-9	-4,75 -5,03	0,92	0,36
	2	10-21 22-9	-3,44 -3,64	0,99	0,32
	1+2	10-21 22-9	-4,46 -4,72	1,10	0,27
Ionisatievat tegen barometer	1	10-21 22-9	-4,47 -4,33	0,70	0,48
	2	10-21 22-9	-3,01 -3,08	0,30	0,76
	1+2	10-21 22-9	-3,96 -3,89	0,40	0,69

Uit tabel V zien we dus dat de boven, naar aanleiding van tabel II getrokken voorlopige conclusie, t.w. dat er geen reden is om aan te nemen dat er een verschil is tussen de hellingen overdag en 's nachts, bevestigd wordt.

Of de hellingen van ionisatievat tegen tellers uit 1e en 2e periode en overdag en 's nachts significant verschillen hebben wij niet onderzocht. Dit zou nl. zeer veel rekenwerk eisen en waarschijnlijk niets opleveren, daar deze verschillen ook klein zijn (vgl.tabel III). Ook het verschil voor de twee perioden is voor die hellingen niet onderzocht.

3.3 Onderzoek naar de verschuiving der ionisatievat waarnemingen t.o.v. de teller waarnemingen.

Zoals reeds in § 3.1 werd opgemerkt, vonden wij in een grafiek, waarin de ionisatievat waarnemingen uitgezet waren tegen de teller waarnemingen, dat de regressielijnen van de twee perioden ten opzichte van elkaar verschoven zijn.

Om dit nader te onderzoeken hebben wij de methode der dubbele dichotomie toegepast (vgl.memorandum S 53(M 23)). Hiertoe werd de gehele puntenwolk in twee delen gedeeld, door er een rechte door te trekken met als helling de in § 2 gevonden helling van ionisatievat tegen tellers en wel zodanig dat het aantal waarnemingen dat boven de rechte ligt ongeveer gelijk is aan het aantal dat er onder ligt.

We vinden dan voor de aantallen waarnemingen:

	1^e periode	2^e periode	totaal
boven de rechte	125	428	553
onder de rechte	338	271	609
Totaal	463	699	1162

Onder de hypothese, dat de regressielijn in beide perioden dezelfde is, is de kans, dat een willekeurig punt, dat boven de rechte ligt, tot de eerste periode behoort gelijk aan de kans, dat dit met een willekeurig punt van onder de streep het geval is. Anders uitgedrukt: de kenmerken "boven" resp. "onder de streep" zijn dan stochastisch onafhankelijk van "tot de eerste" resp. "tot de tweede periode behorend". Dit kan getoetst worden met behulp van de in bovengenoemd memorandum beschreven methode.

We vinden daarbij een overschrijdingskans $\ll 10^{-6}$, hetgeen dus de verschuiving overtuigend aantoont.

3.4. Verdeling der gehele waarnemingsperiode in vijf gelijke perioden:

Wij hebben de gehele waarnemingsperiode ook nog in vijf gelijke perioden verdeeld en voor ieder van deze perioden voor dag en nacht apart de schattingen van de hellingen van tellers (resp. ionisatievat) tegen barometer berekend.

Als perioden hebben we genomen:

I 10 Mei - 23 Mei
II 24 Mei - 6 Juni
III 7 Juni - 19 Juni
IV 20 Juni - 2 Juli
V 3 Juli - 15 Juli

De voor deze perioden gevonden hellingen vinden we in de tabellen VI en VII.

Tabel VI: (zie volgende blz.)

Tabel VI:

Schattingen van de hellingen van tellers tegen barometer voor ieder der vijf perioden:

periode \ temp.	dag 10-21 uur	nacht 22-9 uur
I	-4,32	-3,58
II	-2,44	-3,79
III	-7,20	-7,42
IV	-4,27	-4,48
V	-3,04	-3,14

Tabel VII:

Schattingen van de hellingen van ionisatievat tegen barometer voor ieder der vijf perioden:

periode \ temp.	dag 10-21 uur	nacht 22-9 uur
I	-4,17	-4,13
II	-3,85	-3,60
III	-6,17	-4,94
IV	-3,69	-3,85
V	-3,01	-2,80

Uit de tabellen VI en VII zien we dus dat, zowel voor tellers als voor ionisatievat, de hellingen zeer sterk wisselen, en speciaal in de derde periode zeer hoog zijn.

4. Samenvatting der resultaten en conclusies:

In het bovenstaande hebben wij dus gevonden;

1. dat de helling van tellers (resp. ionisatievat) tegen barometer niet constant is.
2. dat, bij eenzelfde tellerwaarneming, deionisatievat-waarnemingen in de tweede periode gemideld hoger zijn dan in de eerste periode.
3. dat het verschil van de invloed van de barometerstand op de intensiteit der cosmische straling overdag en 's nahhts in eerste benadering te verwaarlozen is.
4. dat er onverwacht grote incidentele afwijkingen optreden bij tellers en ionisatievat tegelijk, waarvan (op grond van de beschikbare gegevens) niet uit te maken is, of deze aan fouten in de waarneming geweten moeten worden, of dat men hier met wezenlijke veranderingen in de cosmische straling te maken heeft.

Opmerking: Conclusie 1 maakt het onmogelijk één barometercorrectie voor het gehele materiaal te berekenen en na toepassing daarvan naar een dagperiodiciteit te gaan zoeken. Deze periodiciteit kan wel worden onderzocht door ieder der dagen apart te beschouwen en de uitkomsten der verschillende dagen te combineren. Daarbij zou men dan moeten onderstellen, dat de barometerinvloed gedurende iedere periode van één dag constant is. Deze onderstelling wordt ondersteund door conclusie 3. Tevens zou men dan de invloed van de temperatuur op de cosmische straling dienen te onderzoeken, daar de temperatuur een duidelijke dagperiodiciteit vertoont (vgl. § 3.2.1).

Een andere mogelijkheid is, voor ieder der verschillende tijdstippen 0 uur, 1 uur, 2 uur, etc. de waarnemingen tezamen nemen en deze groepen apart te analyseren. Een verschillende invloed van de barometerstand op de intensiteit der cosmische stralen op verschillende tijdstippen van de dag zou een interessant resultaat zijn, terwijl een systematisch verschil in intensiteit de mogelijkheid vah een onderzoek naar de dagperiodiciteit van deze intensiteit zou openen. Daarbij zou gebruik gemaakt kunnen worden van een voor ieder der genoemde tijdstippen afzonderlijk bepaalde langzaam met de tijd veranderende barometercorrectie.

Het leek echter gewenst het uitvoeren van deze uitgebreide bewerkingen te laten voorafgaan door een bespreking van de tot nu toe behaalde resultaten.
Daartoe werd dit rapport geschreven. Een vraag, die voor het voortzetten van het onderzoek van veel belang is, betreft de relatieve betrouwbaarheid van de waarnemingen verricht met de tellers en met het ionisatievat en de betrouwbaarheid der waarnemingen in het algemeen. Dit laatste is van belang in verband met conclusie 4. Ook een bespreking van de in de inleiding genoemde moeilijkheden kan voor een verder onderzoek van belang zijn.

Een gedeelte van de voor dit rapport noodzakelijke berekeningen is uitgevoerd door de Rekenafdeling van het Mathematisch Centrum.

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.
S47(M6)

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

De toetsing van een hypothese $\mathcal{H}_0$ berust steeds op een aantal waarnemingen $x_1, x_2, \ldots, x_n$ 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 $x_1, x_2, \ldots, x_n$ 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 $\mathcal{H}_0$ 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 $\mathcal{H}_0$, 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 $\mathcal{H}_0$ op grond van de waarnemingen $x_1, x_2, \ldots, x_n$, 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 $\mathcal{H}_0$ 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 $\alpha = 0{,}05$ 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 *geen* mogelijkheid om tot *aanvaarding* van een hypothese te komen. Indien een bepaalde hypothese 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 *nevenvoorwaarden* 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 *overschrijdingskans* k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van 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. *Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leq \alpha$ is.*

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 H_0 leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

Litteratuur:

J.Neyman, First 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 4 (1950) p.54-66.

THEMATISCH CENTRUM
le Boerhaavestr. 49
M S T E R D A M O.
tatistische Afdeling
17 (M14)

Methode der m rangschikkingen [1])

Een duidelijk voorstelling van deze toetsingsmethode verkrijgt men door n elementen te beschouwen, die een bepaald kenmerk, eventueel in verschillende mate, bezitten. dit kenmerk wordt door m waarnemers beoordeeld en ieder van deze waarnemers rangschikt deze n elementen volgens zijn beoordeling naar opklimmende waardering. Op deze wijze ontstaan m rijen van rangschikkingen. We willen nu een maat aangeven voor de overeenstemming tussen deze rangschikkingen, m.a.w. een maat voor de overeenstemming tussen de m beoordelingen. De hypothese H_0 , die met deze methode getoetst kan worden, houdt in dat er geen overeenstemming tussen de waarnemers bestaat; precieser gezegd, dat alle rangschikkingen onafhankelijk van elkaar op toevallige wijze zijn ontstaan. Dit is b.v. het geval, als het betrokken kenmerk in werkelijkheid voor alle elementen dezelfde waarde bezit.

We kunnen de afleiding voor de maat van overeenstemming het eenvoudigst geven aan de hand van een voorbeeld.

elementen		A	B	C	D	E	F
rangnummers toegekend door waarnemer	a	5	4	1	6	3	2
	b	2	3	1	5	6	4
	c	4	1	6	3	2	5
	d	4	3	2	5	1	6
		15	11	10	19	12	17

De som van alle rangnummers is $\frac{1}{2} n\, m(n+1)$. Onder de hypothese H_0 is het theoretische gemiddelde van iedere kolom: $\frac{1}{2} m(n+1)$

We beschouwen nu de afwijkingen van dit gemiddelde. In ons voorbeeld is het theoretisch kolomgemiddelde gelijk aan 14. De afwijkingen daarvan zijn

$$1 \quad -3 \quad -4 \quad 5 \quad -2 \quad 3$$

---

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

De som der kwadraten van deze afwijkingen noemen we S
In ons voorbeeld is $S = 64$
Als alle m rangschikkingen gelijk zijn wordt het maximum van S bereikt.
Dit maximum is $\frac{1}{12} m^2(n^3-n)$
We definiëren nu als <u>coëfficiënt van overeenstemming</u>

$$W = \frac{12\,S}{m^2(n^3-n)}$$

In ons voorbeeld is $W = \frac{12 \times 64}{16 \times 210} = 0{,}229$

W varieert dus tussen 0 en 1
De verdeling van W onder de hypothese H_0 is exact berekend voor een aantal waarden van n en m [1] terwijl voor grote n en m benaderingen bekend zijn. [1] Hiermee kan H_0 dus getoetst worden, waarbij H_0 verworpen wordt, als W waarden dichtbij 1 aanneemt. De kritieke zône is dus van de vorm $W \geqq W_0$
Het kan voorkomen dat de waarnemers geen onderscheid ontdekken in de mate waarin verschillende elementen het kenmerk bezitten. Ze geven deze elementen dan gelijke rangnummers.
Veronderstel, dat door een waarnemer geen onderscheid wordt gemaakt tussen de elementen, die de rangnummers 3 t/m 6 moeten dragen. Dan wordt als rangnummer van iedere van deze elementen het gemiddelde van de rangnummers $\frac{1}{4}(3+4+5+6) = 4\frac{1}{2}$
gebruikt.
Daar het maximum van S nu verandert, moeten we een correctie op de formule van W toepassen; deze vindt men in het boek van Kendall behandeld.

Litteratuur:

[1] M.G.Kendall, Rank correlation methods, London 1948
hoofdstuk 6 p.80.

tabel verdelingsfunctie van W voor

$n=3 \quad m = 2$ t/m 10
$n=4 \quad m = 2$ t/m 6
$n=5 \quad m = 3$

op p. 146-149

THEMATISCH CENTRUM,
le Boerhaavestr. 49,
m s t e r d a m -O.

atistische Afdeling.
S 53 (M 23).

Toetsing van de hypothese $p_1 = p_2$ met behulp van een 2 x 2-tabel[1].

Wij beschouwen twee reeksen van onafhankelijke experimenten, waarbij ieder experiment van de ene reeks één van de twee resultaten A of $\bar{A}$ (non-A) heeft en ieder experiment van de tweede reeks één van de beide resultaten B of $\bar{B}$ (hierbij kan A=B zijn). Daarbij wordt ondersteld, dat bij ieder der experimenten van de ene reeks de kans op A gelijk aan p_1 (en dus de kans op $\bar{A}$ gelijk aan $1-p_1$) is en bij ieder der experimenten van de tweede reeks de kans op B gelijk aan p_2 (en dus de kans op $\bar{B}$ gelijk aan $1-p_2$). De te toetsen hypothese luidt nu:

$$H_0: \quad p_1 = p_2.$$

Indien de eerste reeks uit n en de tweede reeks uit m waarnemingen bestaat, waaronder n_1 (resp. m_1) maal A (resp. B) voorkomt, kunnen deze gegevens in de volgende 2 x 2-tabel worden samengevat:

	A resp. B	$\bar{A}$ resp. $\bar{B}$	totaal
eerste reeks	n_1	$n-n_1$	n
tweede reeks	m_1	$m-m_1$	m
totaal	r	n+m-r	n+m

Als toetsingsgrootheid wordt n_1, het aantal malen A in de eerste reeks waarnemingen, gebruikt. Indien H_0 juist is bezit deze grootheid onder de voorwaarde, dat r de bij het experiment gevonden waarde aanneemt, de volgende waarschijnlijkheidsverdeling: de kans, dat een bepaalde waarde n_1 aangenomen wordt, is gelijk aan:

$$\frac{\binom{n}{n_1}\binom{m}{m_1}}{\binom{n+m}{r}}$$

Als kritieke zône worden de waarden van n_1 met de kleinste waarschijnlijkheden bijeengezocht, tot de gekozen betrouwbaarheidsdrempel het toevoegen van een nieuwe waarde verhindert (bij éénzijdige toetsing bestaat de kritieke zône uitsluitend uit grote of uitsluitend uit kleine waarden van n_1).

---

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

De overschrijdingskans, behorende bij de gevonden waarde van n_1, is gedefiniëerd als de som van alle waarschijnlijkheden van bovenstaande verdeling, die hoogstens gelijk aan de waarschijnlijkheid van de gevonden waarde zijn (bij éénzijdige toetsing echter gelijk aan de som van de waarschijnlijkheden van alle waarden die groter of gelijk aan de gevondene, of van alle waarden, die kleiner of gelijk aan de gevondene zijn). Deze exacte toetsingsmethode voor H_0 is afkomstig van R.A. Fisher.

Indien n en m zo groot zijn, dat deze exacte berekening te omslachtig wordt, maakt men gebruik van de volgende benadering:

Gemiddelde en spreiding van de grootheid n_1 zijn (indien H_0 juist is):

$$\frac{nr}{n+m} \quad \text{resp.} \quad \sqrt{\frac{nmrs}{(n+m)^2(n+m-1)}}$$

Men gebruikt dan in plaats van de exacte waarschijnlijkheidsverdeling van n_1 de normale verdeling met hetzelfde gemiddelde en dezelfde spreiding en in plaats van de gevonden waarde van n_1 neemt men het getal, dat $\frac{1}{2}$ dichter bij het gemiddelde ligt dan deze gevonden waarde (dit laatste is de z.g. "continuïteitscorrectie", die bij toenemende n en m weldra verwaarloosd kan worden). Met behulp van de benadering gaat men dan verder te werk als boven beschreven, daarbij gebruik makende van een tabel van de normale verdeling.

Litteratuur:

R.A.Fisher, Statistical Methods for Research Workers, London 1948, p. 96. Opmerking: Fisher gebruikt hier de éénzijdige overschrijdingskans.

J.Hemelrijk, Waarschijnlijkheidsrekening en Statistiek, Vacantiecursus Mathematisch Centrum, Amsterdam 1950, § 4.

MATHEMATISCH CENTRUM,
2de Boerhaavestraat 49,
A m s t e r d a m (O).

Statistische Afdeling.
S 53 (M 27)

Toetsing van de gelijkheid van twee regressiecoëfficiënten indien de afwijkingen normaal verdeeld zijn met gelijke spreidingen.[1]

Zij gegeven, dat van vier grootheden ξ_1, η_1, ξ_2 en η_2 de grootheden ξ_1 en ξ_2 foutloos kunnen worden waargenomen, terwijl bij de waarnemingen van η_1 en η_2 onderling onafhankelijk verdeelde fouten[2] worden gemaakt, die alle een normale verdeling bezitten[3] met gemiddelde 0 en onbekende, maar steeds dezelfde, spreiding σ.

Verder zij gegeven, dat η_1 en η_2 lineair van ξ_1 resp. ξ_2 afhankelijk zijn, dus dat geldt

$$\eta_1 = \beta_1 \xi_1 + \alpha_1$$
$$\eta_2 = \beta_2 \xi_2 + \alpha_2,$$

waarin de parameters α_1, β_1, α_2 en β_2 onbekend zijn.

Gevraagd wordt nu, op grond van een aantal waarnemingsparen

$$(x_1',y_1'),\ldots,(x_n',y_n') \quad \text{van } (\xi_1,\eta_1)$$
en
$$(x_1'',y_1''),\ldots,(x_m'',y_m'') \quad \text{van } (\xi_2,\eta_2)$$

de hypothese te toetsen, dat de twee regressiecoëfficiënten β_1 en β_2 gelijk zijn:

---

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

[2] Onder "fouten" vallen in dit geval ook toevallige afwijkingen van ander karakter, bv. physiologische afwijkingen.

[3] De stochastische grootheid $\underline{x}$ bezit een normale verdeling met gemiddelde μ en spreiding σ, indien voor iedere a geldt:

$$P[\underline{x} \leq a] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}\frac{(u-\mu)^2}{\sigma^2}}\, du.$$

H_0: $\beta_1 = \beta_2$.

Om deze hypothese te toetsen, worden schattingen b_1 en b_2 van β_1 en β_2 berekend volgens de formules:

$$b_1 = \frac{\sum_{i=1}^{n}(y_i'-\bar{y}')(x_i'-\bar{x}')}{\sum_{i=1}^{n}(x_i'-\bar{x}')^2}$$

en

$$b_2 = \frac{\sum_{j=1}^{m}(y_j''-\bar{y}'')(x_j''-\bar{x}'')}{\sum_{j=1}^{m}(x_j''-\bar{x}'')^2},$$

waarin

$$\bar{y}' = \frac{1}{n}\sum_{i=1}^{n} y_i', \quad \bar{x}' = \frac{1}{n}\sum_{i=1}^{n} x_i', \text{ enz.}$$

Deze schattingen worden verkregen door toepassing van de methode der kleinste kwadraten, die in dit geval overeenkomt met de methode der meest aannemelijke schattingen (Engels: method of maximum likelihood). Op dezelfde wijze kan men als schatting voor σ^2 afleiden:

$$s^2 = \frac{\sum_{i=1}^{n}\left\{y_i'-b_1(x_i'-\bar{x}')-\bar{y}'\right\}^2 + \sum_{j=1}^{m}\left\{y_j''-b_2(x_j''-\bar{x}'')-\bar{y}''\right\}^2}{(n-2)+(m-2)} =$$

$$= \frac{\sum_{i=1}^{n}(y_i'-\bar{y}')^2 - b_1^2\sum_{i=1}^{n}(x_i'-\bar{x}')^2 + \sum_{j=1}^{m}(y_j''-\bar{y}'')^2 - b_2^2\sum_{j=1}^{m}(x_j''-x'')^2}{n+m-4}$$

Beschouwen wij niet alleen de gevonden waarden $y_1',\ldots,y_n'$, $y_1'',\ldots,y_m''$, maar de verzameling van alle mogelijke waarden, die deze grootheden kunnen aannemen, dan bezitten zij op deze verzameling een waarschijnlijkheidsverdeling, die afhankelijk is van de (exact waargenomen) waarden $x_1',\ldots,x_n'$, $x_1'',\ldots,x_m''$. De schattingen b_1, b_2 en s^2 zijn dan ook stochastisch, hetgeen door onderstreping aangegeven wordt. Zij zijn verder onderling onafhankelijk verdeeld, zoals bv. door Mood [1] bewezen wordt, en $\underline{b}_1$ en $\underline{b}_2$ bezitten beide een normale verdeling met β_1 resp. β_2 als gemiddelden en spreidingen

$$\frac{\sigma}{\sqrt{\sum_i (x_i'-\bar{x}')^2}} \text{ resp. } \frac{\sigma}{\sqrt{\sum_j (x_i''-\bar{x}'')^2}}.$$

Verder bezit $(n+m-4)\underline{s}^2/\sigma^2$ een χ^2-verdeling met (n+m-4) vrijheidsgraden. Hieruit volgt, dat de stochastische grootheid

$$\underline{t} = \frac{\underline{b}_1 - \underline{b}_2 - (\beta_1 - \beta_2)}{\underline{s}\sqrt{\frac{1}{\sum\limits_i (x_i' - \bar{x}')^2} + \frac{1}{\sum\limits_j (x_j'' - \bar{x}'')^2}}}$$

verdeeld is volgens de verdeling van Student met (n+m-4) vrijheidsgraden (zie bv. Cramèr [2] p.237) Indien $\beta_1 - \beta_2 = 0$ is, gaat deze grootheid over in

$$\underline{t} = \frac{\underline{b}_1 - \underline{b}_2}{\underline{s}\sqrt{\frac{1}{\sum\limits_i (x_i' - \bar{x}')^2} + \frac{1}{\sum\limits_j (x_j'' - \bar{x}'')^2}}},$$

die dus, onder de hypothese $\beta_1 = \beta_2$, de bovengenoemde Student-verdeling bezit. Is $\beta_1 \neq \beta_2$, dan zullen bepaalde van nul verwijderd liggende waarden van t een grotere waarschijnlijkheidsdichtheid bezitten dan wanneer $\beta_1 = \beta_2$ is. Voor de toetsing van deze hypothese gebruikt men daarom in het tweezijdige geval (als zowel $\beta_1 > \beta_2$ als $\beta_1 < \beta_2$ als alternatieve mogelijkheid wordt toegelaten) een <u>tweezijdige kritieke zône</u> van de vorm

$$|\underline{t}| \geq t_0,$$

waarbij t_0, behorend bij een bepaalde onbetrouwbaarheidsdrempel α, opgezocht kan worden in tabellen, vermeld in onderstaande litteratuurlijst. Indien slechts alternatieve mogelijkheden van de vorm $\beta_1 < \beta_2$ resp. $\beta_1 > \beta_2$ worden toegelaten gebruikt men de éénzijdige kritieke zônes

$$\underline{t} > t_1, \quad \text{resp.} \quad \underline{t} < -t_1,$$

waarin t_1 in dezelfde tabellen kan worden gevonden als t_0 door te zoeken in een tabel voor de tweezijdige toets met onbetrouwbaarheidsdrempel 2α. Het bij de tabellen vermelde aantal vrijheidsgraden (vaak aangegeven door ν of n' of n-1) is in dit geval gelijk aan n+m-4.

Litteratuur:

[1] A.M. Mood, Introduction to the theory of statistics, Mc Graw-Hill, New York-Toronto-London, 1950, hoofdstuk 13 (regressietheorie) en p. 425 (tabel).

[2] H. Cramer, Mathematical methods of statistics, Princeton Un. Press, Princeton 1946.

[3] M.G. Kendall, The advanced theory of statistics I, Griffin and Co., London 1947, p. 440-441 (tabel).