MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 148

(Vertrouwelijk)

Dienstregelmaat van vliegtuigen

door

A.Benard
en
Emily C. Bos-Levenbach

1954

1. Inleiding.

Het onderhavige statistische onderzoek heeft betrekking op het verschil ϑ tussen de werkelijke aankomsttijd en de aankomsttijd volgens de dienstregeling van schroefvliegtuigen op de eindpunten der voornaamste intercontinentale lijnen der K.L.M.

Op grond van practische overwegingen van de onderzoeker was het materiaal als volgt in klassen ingedeeld:

a naar het verschil ϑ,
0-45, 45-90, 90-180, 180-360 of meer dan 360 min. te vroeg en 0-45, 45-90, 90-180, 180-360, 360-720, 720-1440, 1440-2880 of meer dan 2880 min. te laat,

b naar uit- of thuisreis,

c naar de maand van waarneming,
Januari, Februari,..., t/m Juni 1953,

d naar de luchtlijnen,

1.	Amsterdam - Dahran	v.v.
2.	Amsterdam - Bangkok	"
3.	Amsterdam - Johannesburg	"
4.	Amsterdam - Dakar-Curaçao	"
5.	Amsterdam - Buenos Aires	"
6.	Amsterdam - Montreal-Mexico City	"
7.	Amsterdam - New York	" .

Een aankomst tussen 90 min. te vroeg en 45 min. te laat ($-90 < \vartheta < +45$) werd, op grond van diverse overwegingen, door de onderzoeker als punctueel beschouwd. Als maat voor de punctualiteit van een groep aankomsten werd hier de verhouding tussen het aantal punctuele en het totale aantal aankomsten van die groep gebruikt.

2. Resultaten van de onderzoeker.

We zullen nu eerst een kort overzicht geven van de door de onderzoeker getrokken conclusies, waarbij we tevens vermelden in welke paragrafen van dit rapport deze conclusies, voor zover dat mogelijk was, nader zijn onderzocht.

2.1. "Het is niet mogelijk de punctualiteit van aankomst op de eindstations van alle intercontinentale lijnen der K.L.M. vanuit één gezichtspunt te beschouwen".

De punctualiteit van de aankomsten op de diverse intercontinentale lijnen wordt vergeleken in par. 3.1 (zie ook par. 4.3).

2.2. "Er schijnt enige betekenis gehecht te kunnen worden aan het percentage vluchten, dat tussen 90 min. te vroeg en 45 min. te laat aan een eindpunt aankomt".

In zoverre hiermee bedoeld is, dat genoemd percentage een practisch bruikbare maat voor de punctualiteit van de vluchten is, hangt de juistheid van deze uitspraak in hoofdzaak van practische overwegingen af, die nader gepreciseerd zouden moeten worden om een statistisch onderzoek mogelijk te maken.

2.3. "Enige betekenis is ook te hechten aan de centrale groepering vluchten om het modale percentage".

In verband hiermee zal in par. 4 worden nagegaan in hoeverre een wijziging van de punctualiteitsdefinitie in die zin, dat we onder een punctuele aankomst een aankomst in die periode van 135 min., waarin de meeste aankomsten vallen, verstaan, invloed heeft op de verdere conclusies.

2.4. "Bepaalde betekenis kan gehecht worden aan een vergelijking van maandelijkse en meermaandelijkse met zesmaandelijke gemiddelden. Vervlakking wijst op seizoensinvloed, zie b.v. route 7a en b" (New York).

In par. 3.2 wordt de seizoensinvloed onderzocht; wij hebben ons hierbij voorlopig beperkt tot een vergelijking van eerste en tweede kwartaal (zie ook par. 4.4).

2.5. "Bepaalde betekenis kan ook gehecht worden aan het voorkomen van een tweede piek bij vluchten, die in de orde van grootte van een dag te laat zijn, zie b.v. route 4b" (Curaçao-Dakar-Amsterdam).

De betekenis van de tweede pieken is niet zonder meer aan de hand van het gegeven materiaal te onderzoeken (zie ook par. 2.9).

2.6. "Bij toenemende snelheid der vliegtuigen kan, waar de weersinvloed overheerst, de punctualiteit in de toekomst beïnvloed worden. Bij een constante weersinvloed, zoals wind, in gunstige zin. Bij een onregelmatige invloed, zoals die van de temperatuur (op hoogte), eventueel in ongunstige zin".

Dit punt kon, wegens gebrek aan gegevens, nog niet onderzocht worden (zie ook par. 5).

2.7. "Er zijn aanwijzingen, dat uitvluchten meer punctueel verlopen dan thuisvluchten".

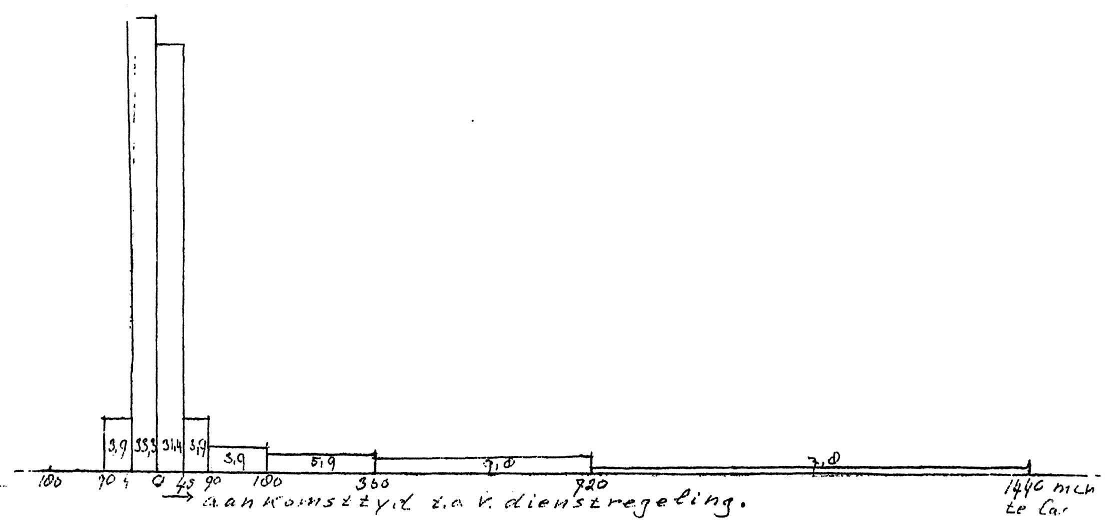
Het verschil tussen uit- en thuisvluchten is onderzocht in par. 3.3 (zie ook par. 4.5).

2.8. "Er is een aanwijzing dat betere punctualiteit betekent: een meest voorkomend percentage vluchten in een periode van <u>te vroeg</u> aankomen. D.w.z. naar mate men blijkbaar meer streeft om op tijd te komen, zal men meer te vroeg aankomen".

Wij verwijzen hiervoor naar par. 3.4.

2.9. <u>De grafieken</u>.

Bij het materiaal bevonden zich een aantal grafieken, waarin van iedere luchtlijn, voor uit- en thuisreis apart, het percentage binnengekomen vliegtuigen uitgezet was tegen het verschil ν (zie par. 1). De tijdas was echter verdeeld volgens de op pag. 1 gegeven intervallen, dus niet lineair. Hierdoor werd een minder juiste afbeelding verkregen. Wanneer men in plaats hiervan een histogram tekent, waarin het oppervlak der blokjes een maat is voor het percentage vluchten, terwijl de breedte correspondeert met het aantal minuten te vroeg of te laat, verkrijgt men een juister beeld. Wij voerden dit als voorbeeld uit voor de luchtlijn 5b, zie figuur 1 (Buenos Aires-Amsterdam).



Figuur 1. <u>Dienstregelmaat route 5b (Buenos Aires-Amsterdam)</u>.

De tweede piek, die in de oorspronkelijke tekening aanwezig was, is nu verdwenen; deze was dus alleen ontstaan doordat intervallen van gelijke lengte op de tijdas niet overeenkwamen met gelijke tijdsintervallen.

Het zou de figuren nog ten goede komen, wanneer de aankomsttijden nauwkeuriger bekend waren, zodat een fijnere intervalverdeling gemaakt zou kunnen worden.

Opmerking.

Bij het onderzoek van het materiaal wordt gebruik gemaakt van statistische methoden, die berusten op het toetsen van bepaalde hypothesen (zie bijlage: memorandum S 47 (M 6)). Voor de speciale methoden, die werden gebruikt, wordt verwezen naar des betreffende memoranda, die als bijlage aan dit rapport zijn toegevoegd.

3. Onderzoek van het materiaal met gebruik van de oorspronkelijke punctualiteitsdefinitie.

3.1. Vergelijking van de luchtlijnen wat betreft hun punctualiteit (ad par. 2.1)

We gebruiken hiervoor de toets voor de hypothese $p_1 = p_2 = \dots = p_k$ met behulp van een 2xk-tabel, met $k = 7$ (zie memorandum S 73 (M 23a)).

We vinden dan, gedifferentiëerd naar eerste en tweede kwartaal, uit- en thuisreis, de resultaten, weergegeven in tabel I.

Tabel I

Vergelijking van de luchtlijnen wat betreft hun punctualiteit

		$\chi^2_{(6)}$	overschrijdingskans
uitreis,	eerste kwartaal	30,5	$< 10^{-4}$
" ,	tweede "	22,0	0,001
thuisreis,	eerste "	34,3	$< 10^{-4}$
" ,	tweede "	10,3	0,043

Conclusie.

De punctualiteit van de vliegdiensten is voor de diverse routes duidelijk verschillend, zowel in het eerste als in het tweede kwartaal, zowel bij de uit- als bij de thuisreizen.

Om een overzicht te krijgen van deze verschillen, vermelden we in tabel II het percentage punctuele vluchten voor iedere route, weer gesplitst naar uit- en thuisreis, eerste en tweede kwartaal.

Tabel II

Punctualiteit van de routes in procenten

route		1	2	3	4	5	6	7
uitreis,	eerste kwartaal	56	<u>84</u>	80	73	65	51	50
" ,	tweede "	78	<u>84</u>	58	81	54	60	59
thuisreis,	eerste "	62	43	50	58	72	<u>18</u>	<u>27</u>
" ,	tweede "	54	55	77	58	65	60	72

In deze tabel valt de grote punctualiteit van de route 2 (Bangkok) bij de uitreis op, benevens de geringe punctualiteit van de routes 6 (Montreal-Mexico City) en 7 (New York) bij de thuisreis in het eerste kwartaal.

3.2. Onderzoek seizoensinvloed (ad par. 2.4).

Hiertoe maakten we gebruik van de toets voor een 2x2-tabel (memorandum S 53 (M 23)).

Gezien de mogelijke heterogeniteit in het materiaal werd dit onderzoek voor de diverse routes èn voor uit- en thuisreis afzonderlijk verricht en bovendien zijn de resultaten gecombineerd door optelling der toetsingsgrootheden α volgens de methode 1 beschreven in memorandum S 102 (M 17b).

De resultaten staan vermeld in tabel III.

Tabel III

Onderzoek naar verschil in punctualiteit tussen eerste en tweede kwartaal

route	overschrijdingskans 1)	
	uitreis	thuisreis
1	0,10 (-)	0,58 (+)
2	0,92 (+)	0,11 (-)
3	0,09 (+)	0,05 (-)
4	0,52 (-)	1
5	0,40 (-)	0,62 (+)
6	0,41 (-)	$< 10^{-4}$ (-)
7	0,19 (-)	$< 10^{-6}$ (-)
gecom.	0,14 (-)	$< 10^{-6}$ (-)

Conclusie.

Er werd geen systematisch verschil in punctualiteit tussen de uitreizen in het eerste en die in het tweede kwartaal gevonden. De thuisreizen verlopen in het tweede kwartaal systematisch vaker punctueel dan in het eerste kwartaal; bij nadere beschouwing blijken alleen de luchtlijnen 7 (New York), 6 (Montreal-Mexico City) en in mindere mate 3 (Johannesburg) tot dit resultaat bij te dragen.

---

1) Het teken (-) resp. (+) achter een overschrijdingskans betekent, dat in de steekproef in het tweede kwartaal relatief vaker (resp. minder vaak) punctuele aankomsten voorkomen dan in het eerste kwartaal. De overschrijdingskansen zijn tweezijdig.

3.3. Vergelijking tussen de punctualiteit van uit- en thuisreizen (ad par. 2.7).

We onderzochten met behulp van de toets voor een 2x2-tabel of er een verschil bestaat tussen de punctualiteit van uit- en thuisreizen, afzonderlijk voor de verschillende routes en voor eerste en tweede kwartaal. De resultaten voor de verschillende routes zijn wederom gecombineerd als aangegeven in par. 3.2.

Tabel IV

Onderzoek naar verschil in punctualiteit tussen uit- en thuisreis

route	overschrijdingskans 2)	
	eerste kwartaal	tweede kwartaal
1	0,69 (-)	0,07 (+)
2	$< 10^{-6}$ (+)	$< 10^{-4}$ (+)
3	0,03 (+)	0,14 (-)
4	0,25 (+)	0,08 (+)
5	0,62 (-)	0,40 (-)
6	0,003 (+)	0,97 (-)
7	0,001 (+)	0,03 (-)
gecombineerd	$< 10^{-6}$ (+)	0,36 (+)

Conclusie.

In het eerste kwartaal verlopen de uitreizen systematisch vaker punctueel dan de thuisreizen; dit resultaat is hoofdzakelijk te wijten aan hetgeen gevonden werd bij de routes 2 (Bangkok), 7 (New York), 6 (Montreal-Mexico City) en 3 (Johannesburg) In het tweede kwartaal verloopt bij de route 2 (Bangkok) de uitreis systematisch vaker punctueel dan de thuisreis; zwakke aanwijzingen in dezelfde richting vindt men bij route 1 (Dahran) en 4 (Dakar-Curaçao); bij route 7 (New York) constateert men echter het tegengestelde.

Mede tengevolge hiervan vindt men bij combinatie van de toetsen in het tweede kwartaal geen systematisch verschil meer.

3.4. Onderzoek naar het verband tussen de aantallen vliegtuigen die punctueel arriveren en de aantallen die te vroeg arriveren (ad par. 2.8).

De opmerking van de onderzoeker, vermeld in par. 2.8, hebben wij als volgt geïnterpreteerd: Naarmate bij een bepaalde

----

2) Het teken (-) resp. (+) achter een overschrijdingskans betekent, dat in de steekproef de thuisreis vaker resp. minder vaak punctueel verlopen is dan de uitreis.

route een groter percentage punctuele vliegtuigen optreedt, is er onder de punctuele vluchten een grotere fractie te vroeg. Wij hebben dit voor uit- en thuisreizen en voor eerste en tweede kwartaal afzonderlijk onderzocht. Dit is verder als volgt gedaan: We rangschikken de routes in volgorde van opklimmende punctualiteit. Wij toetsen nu of er in deze volgorde een verloop bestaat in de fractie te vroege vluchten ($-90 < \vartheta < 0$) onder de punctuele vluchten ($-90 < \vartheta < +45$) met behulp van de toets beschreven in memorandum S 139 (M 48) (zie tevens S 47 (M 7) ,voor de definitie van ϑ zie par. 1). De resultaten zijn vermeld in tabel V.

Tabel V

Onderzoek naar een verloop in de fractie te vroege van de punctuele aankomsten

	overschrijdingskans [3]
uitreis, eerste kwartaal	0,15 (-)
" , tweede "	0,61 (-)
thuisreis, eerste "	0,32 (-)
" , tweede "	0,008 (+)

Conclusie.

Alleen bij de thuisreizen van het tweede kwartaal vinden wij een systematisch verloop van de fractie te vroege aankomsten onder de punctuele aankomsten met de punctualiteit van de luchtlijnen. In de andere gevallen is de correlatie negatief, maar zo zwak dat niet van een systematische toename met de punctualiteit gesproken mag worden.

4. Onderzoek van het materiaal bij gebruik van een gewijzigde punctualiteitsdefinitie.

4.1. Inleiding.

Het is bij bepaalde groepen vluchten mogelijk, de gevonden waarde voor de punctualiteit te verhogen, door deze te definiëren als de fractie van het totale aantal vluchten, gevormd door de vluchten met een waarde ϑ gelegen in dat interval van 135 minuten, waarin de meeste ϑ-waarden van de groep vallen, hetgeen dus overeenkomt met het gebruik van het "modale percentage"

---

3) Een (+) resp. (-) teken achter de overschrijdingskans betekent, dat in de steekproef de fractie te vroege aankomsten onder de punctuele aankomst positief resp. negatief gecorreleerd is met de fractie punctuele aankomsten van het totale aantal.

(zie par. 2.3).

In vele gevallen is deze periode gelijk aan die van $-90 < t_a < +45$, waarop onze oorspronkelijke punctualiteitsdefinitie gebaseerd was. Voor zover dit niet het geval is, bestaat de mogelijkheid dat de dienstregeling niet voldoende aangepast was aan de op de beschouwde luchtlijn heersende weersomstandigheden, accommodatie op de vliegvelden etc. In verband met deze mogelijkheid willen we hier onderzoeken in hoeverre genoemde wijziging in de punctualiteitsdefinitie onze conclusies zouden kunnen beïnvloeden.

Hierbij moeten we nog twee opmerkingen maken. Ten eerste is het niet nodig, dat de dienstregeling optimaal is, als de beide punctualiteitsperioden samenvallen, immers, ook bij een te strakke of te ruime dienstregeling zal het streven bestaan om de officiële aankomsttijden zo dicht mogelijk te benaderen; ten tweede is het mogelijk dat een dienstregeling goed is, ook als de beide punctualiteitsperioden niet samenvallen. Dat laatste kan dan te wijten zijn aan b.v. een abnormaal ongunstige weersperiode, of andere oorzaken.

Het hier gegeven materiaal is onvoldoende om het vraagstuk van de optimale dienstregeling op te lossen.

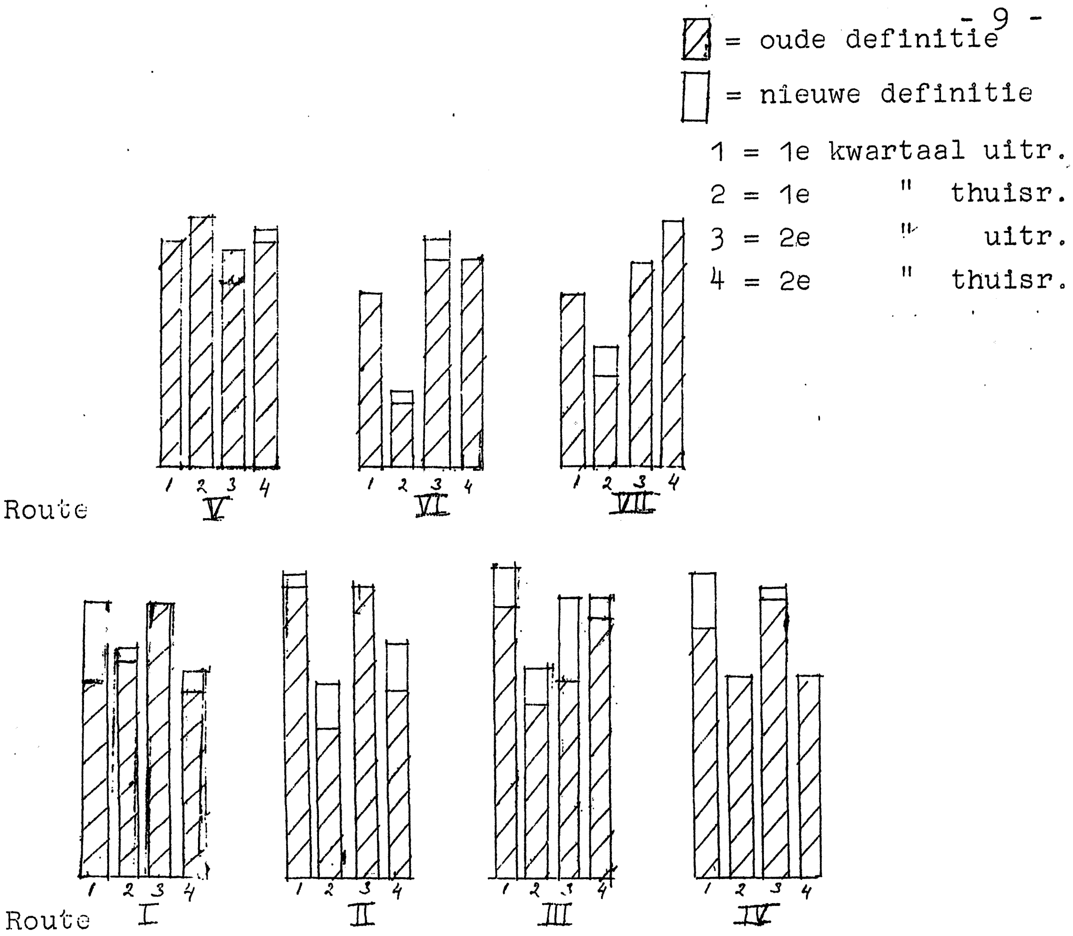
4.2. Vergelijking van de punctualiteitspercentages.

In tabel VI hebben wij de cijfers gegeven van de punctualiteit volgens de oude en de nieuwe definitie bij de diverse luchtlijnen, onderverdeeld naar eerste en tweede kwartaal, uit- en thuisreis; in figuur 2 zijn deze getallen in een grafiek weergegeven.

Tabel VI

Vergelijking tussen de punctualiteiten in procenten volgens oude en nieuwe definitie

route	eerste kwartaal				tweede kwartaal			
	uitreis		thuisreis		uitreis		thuisreis	
	oude def.	nwe def.	oude def.	nwe def.	oude def.	nwe def.	oude def.	nwe def.
1	56	78	62	65	76	78	54	58
2	84	88	43	56	84	84	55	68
3	80	92	50	62	58	81	77	81
4	73	89	58	58	81	85	58	58
5	65	65	72	72	54	62	65	69
6	51	51	18	21	60	66	60	60
7	50	50	27	35	59	59	72	72



Figuur 2. <u>Punctualiteit vergeleken volgens oude en nieuwe definitie</u>

In tabel en grafiek is duidelijk te zien, dat de punctualiteit volgens de nieuwe definitie minstens even groot of groter is dan volgens de oude, hetgeen ook uit de definities zelf volgt. In 11 van de 28 gevallen blijft het percentage ongewijzigd, in 21 van de 28 gevallen is het verschil minder dan 10%.

4.3. <u>Vergelijking van de luchtlijnen wat betreft hun punctualiteit</u> (vgl. par. 3.1).

We gebruiken hiervoor weer de toets voor de hypothese $p_1 = p_2 = \ldots = p_k$ met behulp van een 2xk-tabel met $k = 7$. We splitsen uiteraard weer naar eerste en tweede kwartaal, uit- en thuisreis; de resultaten staan vermeld in tabel VII.

<u>Tabel VII</u>

<u>Vergelijking van de luchtlijnen wat betreft hun punctualiteit volgens de nieuwe definitie</u>

		χ^2_6	overschrijdingskans
uitreis,	eerste kwartaal	48,7	$\ll 10^{-4}$
" ,	tweede "	20,9	0,002
thuisreis,	eerste "	31,4	$< 10^{-4}$
"	tweede "	6,6	0,36

Conclusie.

Het verschil in punctualiteit op de diverse routes blijft bestaan, behalve op de thuisreis in het tweede kwartaal, in tegenstelling tot par. 3.1, waar verschil optrad in alle vier de gevallen.

4.4. Onderzoek seizoensinvloed (vgl. par. 3.2).

We gebruikten hiervoor weer de toets van de 2x2-tabel; de combinatie geschiedde op dezelfde wijze als in par. 3.2. Het resultaat staat vermeld in tabel VIII.

Tabel VIII

Onderzoek naar verschil in punctualiteit tussen eerste en tweede kwartaal volgens de nieuwe definitie

route	overschrijdingskans [4]	
	uitreis	thuisreis
1	0,88 (-)	0,57 (+)
2	0,46 (+)	0,10 (-)
3	0,25 (+)	0,13 (-)
4	0,69 (+)	1,-- (+)
5	0,77 (+)	0,83 (+)
6	0,16 (-)	3×10^{-4} (-)
7	0,19 (-)	$< 10^{-6}$ (-)
gecombineerd	0,47 (-)	$< 10^{-6}$ (-)

Conclusie.

De conclusie van par. 3.2 blijft bijna onveranderd gelden bij invoering van de nieuwe punctualiteitsdefinitie. De luchtlijn 3 (Johannesburg) levert echter geen belangrijke bijdrage meer tot het systematische verschil in de punctualiteit tussen de thuisreizen van het eerste en van het tweede kwartaal.

4.5. Vergelijking tussen de punctualiteit volgens de nieuwe definitie van uit- en thuisreizen (vgl. par. 3.3).

We gebruiken de toets van 2x2-tabel en de combinatiemethode vermeld in par. 3.3. Het resultaat wordt vermeld in tabel [illegible]

----

4) zie voetnoot 1) blz. 5

Tabel IX

Onderzoek naar verschil in punctualiteit tussen uit- en thuisreis volgens de nieuwe definitie

route	overschrijdingskans [5]	
	eerste kwartaal	tweede kwartaal
1	0,41 (+)	0,12 (+)
2	$< 10^{-4}$ (+)	0,02 (+)
3	0,01 (+)	1
4	0,01 (+)	0,03 (+)
5	0,62 (-)	0,56 (-)
6	0,006 (+)	0,57 (+)
7	0,04 (+)	0,03 (-)
gecombineerd	$< 10^{-6}$ (+)	0,42 (+)

Conclusie.

We vinden hier dezelfde resultaten als in par. 3.3, met uitzondering van route 4 (Amsterdam-Dakar-Curaçao). Hierbij verloopt nu de uitreis ook systematisch vaker punctueel dan de thuisreis, zowel voor eerste als tweede kwartaal.

De wijziging van de punctualiteitsdefinitie brengt bij dit materiaal overwegend geen of slechts een geringe verandering te weeg in de punctualiteitspercentages van de verschillende routes. Evenmin heeft zij een invloed van betekenis op de conclusies geformuleerd in par. 3. Dit wordt nog nader geïllustreerd in figuur 3 (zie blz. 12). Hierin zijn nl. uitgezet de verschillen in punctualiteit tussen uit- en thuisreis voor het eerste en tweede kwartaal, oude en nieuwe definitie, benevens tussen eerste en tweede kwartaal voor uit- en thuisreis, oude en nieuwe definitie.

5. Samenvatting der conclusies en aanwijzingen voor een eventueel verder onderzoek.

Wij kunnen onze conclusies als volgt samenvatten:

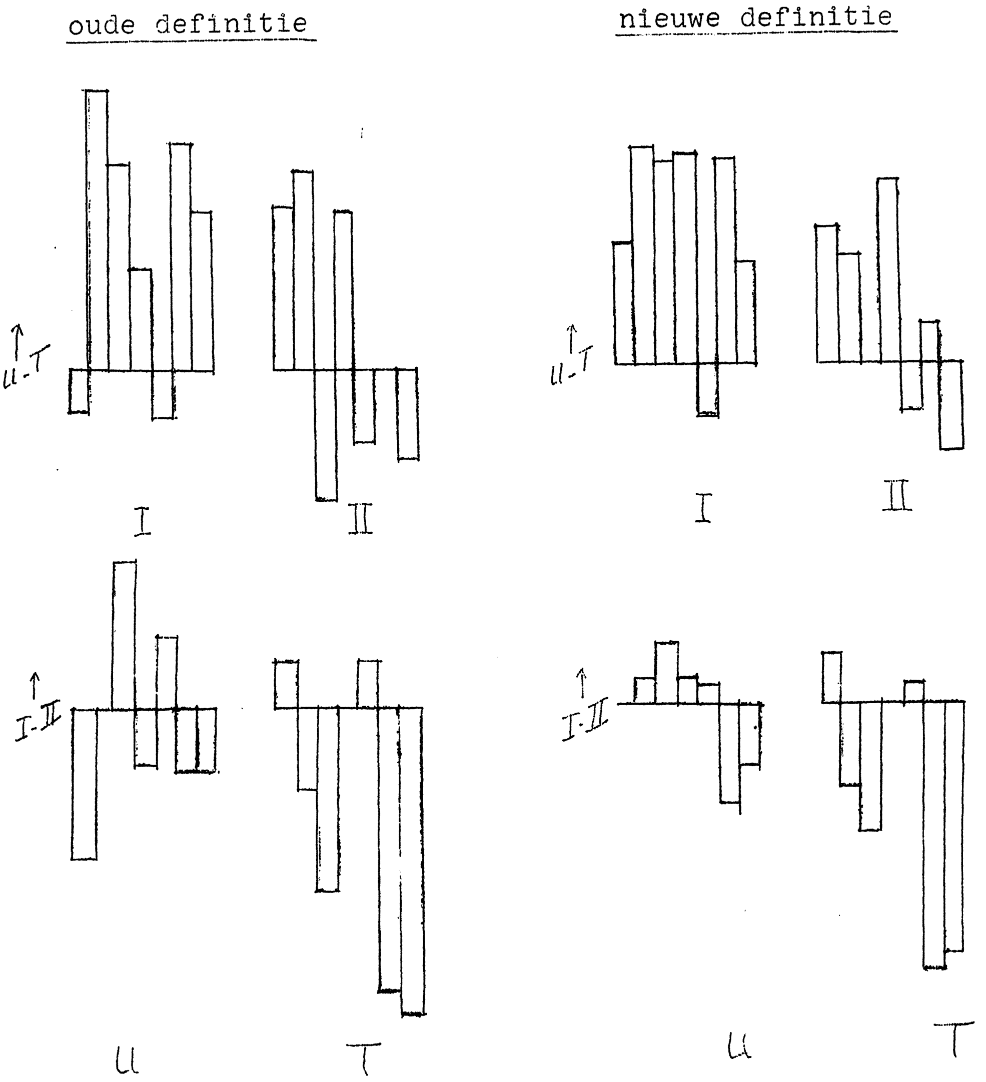
5.1. Er is een systematisch verschil gevonden tussen de punctualiteit van de aankomsten op de verschillende luchtlijnen.

5.2. De thuisreizen in het tweede kwartaal verlopen bij een aantal luchtlijnen (3, 6, 7) systematisch punctueler dan in het eerste kwartaal.

5.3. In het eerste kwartaal verlopen de uitreizen bij een aantal luchtlijnen (2, 3, 6, 7) systematisch punctueler dan de thuis

---

5) Zie voetnoot 2) blz. 6.



Figuur 3[6]. Verschillen tussen punctualiteiten van uit- en thuisreizen en van eerste en tweede kwartaal, volgens de oude en volgens de nieuwe definitie
U = uitreis   T = thuisreis
I = 1e kwartaal   II = 2e kwartaal.

reizen. (In het tweede kwartaal is dit het geval bij één luchtlijn (2)). Bij een andere luchtlijn constateert men het tegengestelde.

5.4. Bij de thuisreizen in het tweede kwartaal vinden wij dat de fractie te vroege aankomsten onder de punctuele van de verschillende luchtlijnen systematisch positief gecorreleerd is met de puntualiteit van de aankomsten. Bij de uitreis in het eerste en tweede en bij de thuisreis in het eerste kwartaal wordt een dergelijk effeqt in het geheel niet geconstateerd.

5.5. Een wijziging van de punctualiteitsdefinitie in die zin, dat men een aankomst punctueel noemt, als zij valt in die periode van 135 minuten, die een maximaal aantal aankomsten bevat,

6) De grafieken zijn voor ieder der zeven luchtlijnen apart gegeven. Ter vergelijking zijn deze naast elkaar geplaatst in de volgorde 1 t/m 7.

heeft betrekkelijk weinig invloed en verandert de conclusies 2 en 3 slechts op onbelangrijke punten.

5.6. Men dient goed te bedenken, dat de conclusies 5.2 en 5.3 niet onderling onafhankelijk zijn; de zeer geringe punctualiteit van de thuisreizen van de routes 6 en 7 en de betrekkelijk geringe punctualiteit van vluchten van de route 3 in het eerste kwartaal zijn op beide conclusies van invloed.

5.7. Een groot aantal factoren, die de punctualiteit gunstig of ongunstig kunnen beïnvloeden, worden in het materiaal in het geheel niet genoemd.

We denken hierbij b.v. aan

1e: De invloed van de samenstelling van de bemanning;

2e: Het vliegtuigtype;

3e: De weersomstandigheden;

4e: De vertrektijden van de vliegtuigen van het beginpunt der route en de invloed van de leiding van de vliegvelden en andere bijkomende omstandigheden op het tijdstip van vertrek.

5.8. Voor een verder onderzoek komt het ons daarom gewenst voor na te gaan in hoeverre factoren genoemd in par. 5.7 van invloed geweest kunnen zijn op het ontstaan van de in 5.1 t/m 5.3 vermelde verschillen in punctualiteit.

5.9. Onderzoekingen welke ten doel hebben de definitie van punctualiteit te verbeteren en eventueel ook de dienstregeling kunnen in eerste instantie beter beperkt worden tot een bepaald vliegtuigtype en niet meer dan één lijn tegelijk, waarbij dan gegevens betreffende de in 5.7 genoemde nevenfactoren zeer gewenst zijn.

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 $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ 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 <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 $k \leqq \alpha$ 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 course in probability and statistics</u>, 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.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - O.
Statistische Afdeling
S 145 (M 23a)

Toets voor de hypothese $p_1 = p_2 = \dots = p_k$ met behulp van een $2 \times k$ tabel [1)]

Wij beschouwen k reeksen $R_1, R_2, \dots, R_k$ van onafhankelijke waarnemingen, waarbij iedere waarneming als resultaat het kenmerk A of het kenmerk $\bar{A}$ (non 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 ($i = 1,2,\dots,k$).
Laat het aantal waarnemingen van reeks R_i ($i = 1,2,\dots,k$) gelijk zijn aan n_i en laat hieronder het aantal met kenmerk A, m_i zijn.
Gevraagd wordt dan de hypothese H_0: $p_1=p_2= \dots =p_k$ te toetsen op grond van deze gegevens.
De gegevens kunnen in een $2 \times k$-tabel worden samengevat;

	R_1	R_2		R_i		R_k	totaal
A	m_1	m_2	...	m_i	...	m_k	m
$\bar{A}$	n_1-m_1	n_2-m_2	...	n_i-m_i	...	n_k-m_k	$n-m$
totaal	n_1	n_2		n_i		n_k	n

waarin dus $m_1 + m_2 + \dots + m_k = m$
en $n_1 + n_2 + \dots + n_k = n$.
De hypothese H_0 wordt getoetst met de grootheid

$$\chi_c^2 = \sum_i \frac{\left(m_i - \frac{m \cdot n_i}{n}\right)^2}{\frac{m\, n_i}{n}} + \sum_i \frac{\left(n_i - m_i - \frac{(n-m)n_i}{n}\right)^2}{\frac{(n-m)\,n_i}{n}} =$$

$$= \sum_i \frac{(n m_i - m n_i)^2}{m(n-m)n_i} = \frac{n^2}{m(n-m)} \sum_i \frac{m_i^2}{n_i} - \frac{nm}{n-m}$$

Deze grootheid $\underline{\chi}_c^2$ [2)] heeft onder de hypothese H_0 bij benadering een χ^2-verdeling met $k-1$ vrijheidsgraden (zie b.v. [1] p. 445 e.v.).

---

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

2) Als wij grootheden als stochastische grootheden (dit zijn grootheden met een waarschijnlijkheidsverdeling) beschouwen geven wij dit door onderstreping aan. Niet onderstreepte letters geven waarden aan, die door de stochastische grootheden worden aangenomen.

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 bij verschillende reeksen verschillende kansen op A zijn, zal $\underline{\chi}_c^2$ gewoonlijk grotere waarden aannemen, dan wanneer H_o juist is.
De kritieke zone bestaat uit die waarden van $\underline{\chi}_c^2$, waarvoor geldt $\underline{\chi}_c^2 \geqq \chi_\alpha^2$. Hierin is χ_α^2 die waarde van $\underline{\chi}^2$, die voldoet aan

$$P[\underline{\chi}^2 \geqq \chi_\alpha^2] = \alpha$$

met α als van te voren vastgelegde onbetrouwbaarheid.
De overschrijdingskans behorende bij een bepaalde gevonden waarde χ_c^2 van $\underline{\chi}_c^2$ is gedefinieerd als

$$P[\underline{\chi}_c^2 \geqq \chi_c^2 | H_o]$$

waarin "$|H_o$" aangeeft, dat deze kans berekend wordt op grond van H_o.
χ_α^2 en de overschrijdingskans kunnen in tabellen of nomogrammen worden opgezicht (zie [3]).

Opmerking. Indien niet voldaan is aan de voorwaarde $m\frac{n_i}{n} \geqq 5$ voor iedere i, kan men een (meer bewerkelijke) exacte toets baseren op de voorwaardelijke waarschijnlijkheidsverdeling van de grootheden $\underline{m}_i$ ($i = 1,\ldots,k$), onder de voorwaarde, dat hun som de waarde m aanneemt:

$$P[\underline{m}_1 = m_1, \underline{m}_2 = m_2, \ldots, \underline{m}_k = m_k | \underline{m}_1 + \underline{m}_2 + \cdots + \underline{m}_k = m ; H_o] =$$

$$= \binom{n_1}{m_1}\binom{n_2}{m_2}\cdots\binom{n_k}{m_k} \Big/ \binom{n}{m}$$

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 $(m_1, m_2, \ldots, m_k)$ met $m_1 + m_2 + \ldots + m_k = m$ 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.

Literatuur.

[1] H.Cramér, Mathematical methods of statistics, Princeton University Press, 1946.

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

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

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

Statistische 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:

$$H_0 : p_1 = p_2.$$

Indien de eerste reeks uit n en de tweede reeks uit m waarnemingen bestaat, waaronder a (resp. b) 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	a	c	n
tweede reeks	b	d	m
totaal	r	s	N

Als toetsingsgrootheid wordt a, 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 a aangenomen wordt, is gelijk aan:

$$\frac{\binom{n}{a}\binom{m}{b}}{\binom{N}{r}}$$

Als kritieke zone worden de waarden van a met de kleinste waarschijnlijkheden bijeengezocht, tot de gekozen onbetrouwbaarheidsdrempel het toevoegen van een nieuwe waarde verhindert (bij éénzijdige toetsing bestaat de kritieke zone uitsluitend uit grote of uitsluitend uit kleine waarden van a). De overschrijdingskans, behorende bij de gevonden waarde van a, is gedefinieerd als de som van alle waarschijnlijkheden van bovenstaande verdeling, die hoogstens gelijk aan de waarschijn-

---

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

lijkheid 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 a zijn (indien H_0 juist is):

$$\frac{nr}{N} \text{ resp. } \sqrt{\frac{nmrs}{N^2(N-1)}} .$$

Men gebruikt dan in plaats van de exacte waarschijnlijkheidsverdeling van a de normale verdeling met hetzelfde gemiddelde en dezelfde spreiding en in plaats van de gevonden waarde van a 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). Bij positieve $a - \frac{nr}{N}$ berekent men dus:

$$a^* = \frac{a - \frac{nr}{N} - \frac{1}{2}}{\sqrt{\frac{nmrs}{N^2(N-1)}}} = \frac{aN - nr - \frac{1}{2}N}{\sqrt{\frac{nmrs}{N-1}}} = \frac{ad - bc - \frac{1}{2}N}{\sqrt{\frac{nmrs}{N-1}}}$$

en bij negatieve $a - \frac{nr}{N}$ berekent men:

$$a^* = \frac{a - \frac{nr}{N} + \frac{1}{2}}{\sqrt{\frac{nmrs}{N^2(N-1)}}} = \frac{ad - bc + \frac{1}{2}N}{\sqrt{\frac{nmrs}{N-1}}} .$$

De overschrijdingskans wordt nu opgezocht in een tabel der normale verdeling met gemiddelde 0 en spreiding 1. De rechts-éénzijdige (resp. links-éénzijdige) overschrijdingskans is het oppervlak rechts (resp. links) gelegen van a^*. De tweezijdige overschrijdingskans is twee maal het oppervlak der normale verdeling dat rechts van $\frac{|ad - bc| - \frac{1}{2}N}{\sqrt{\frac{nmrs}{N-1}}}$ ligt.

<u>Literatuur</u>.

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

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

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - O.

Statistische Afdeling
Rapport S 102 (M 17b)

Het combineren van onafhankelijke toetsen (aanvulling) [1].

In memorandum S 73 (M 17a) wordt een methode voor combinatie van onafhankelijke toetsen behandeld, waarbij het nodig is de overschrijdingskans van iedere toets te bepalen. In vele gevallen kan men de combinatie ook direct op de afzonderlijke toetsingsgrootheden baseren en dit verdient zelfs de voorkeur.

Wij beschouwen hier het geval, dat een bepaalde toets moet worden toegepast op een heterogeen materiaal. Dit materiaal wordt dan eerst verdeeld in h homogeen geachte groepen. Het aantal waarnemingen van de i^e groep zij n_i en de toetsingsgrootheid $\underline{t}_i$ [2]. Laat verder gegeven zijn, dat de verdeling van $\underline{t}_i$ onder de getoetste hypothese (voor de i^e groep aangeduid door H_i) voor grote n_i asymptotisch normaal [3] is, met bekende verwachting μ_i en bekende spreiding σ_i. Aan deze voorwaarden is o.a. voldaan, indien wij te doen hebben met toetsen van WILCOXON, rangcorrelatietoetsen van KENDALL of SPEARMAN, tekentoetsen enz.

Wij toetsen met al de hier te behandelen gecombineerde methoden de hypothese H, dat voor iedere groep de desbetreffende hypothese H_i geldt, terwijl de groepen onderling onafhankelijk zijn. De toetsen verschillen echter ten aanzien van de alternatieve (van H afwijkende) hypothesen waarvoor zij gevoelig [4]) zijn.

De meest gebruikelijke toetsingsgrootheden van gecombineerde toetsen zijn van de gedaante:

$$\underline{T} = \sum_{i=1}^{h} c_i(\underline{t}_i - \mu_i)$$

waarin de letters c_i ($i = 1,2,\ldots\ldots,h$) constanten voorstellen, die voor ieder van de combinatiemethoden op een bepaalde wijze

---

1) Dit memorandum dient slechts ter oriëntatie en streeft niet naar volledigheid of volledige exactheid. Het is bedoeld als een aanvulling op Rapport S 73 (M 17a).

2) De onderstreping geeft aan dat een toetsingsgrootheid stochastisch is, d.w.z. een waarschijnlijkheidsverdeling bezit.

3) Dit houdt in dat $\underline{t}_i$ een waarschijnlijkheidsverdeling heeft, die als n_i toeneemt, steeds minder van een normale verdeling (verdeling van Gauss) afwijkt.

4) Een toets van hypothese H is gevoelig ten opzichte van een alternatieve hypothese H', als de kans dat H verworpen wordt, indien H' juist is, groot is

gekozen worden. Onder de hypothese H zal $\underline{T}$ asymptotisch (voor grote h en/of grote n_i) normaal verdeeld zijn met verwachting 0 en spreiding $\sigma = \sqrt{\sum_{i=1}^{h} c_i^2 \sigma_i^2}$. De dubbele overschrijdingskans van een gevonden waarde T van $\underline{T}$ is dus bij benadering gelijk aan:

$$\frac{2}{\sqrt{2\pi}} \int_{\left|\frac{T}{\sigma}\right|}^{\infty} e^{-\frac{1}{2}x^2} \, dx$$

en kan bepaald worden met behulp van een tabel van de normale verdeling. Indien de dubbele overschrijdingskans kleiner is dan de onbetrouwbaarheidsdrempel α , zal men H verwerpen.

Wij geven hier 3 combinatiemethoden van dit type:

<u>Methode 1</u>: $c_1 = c_2 = \cdots\cdots = c_h = 1$

dus: $\underline{T} = \sum_{i=1}^{h} \underline{t}_i - \sum_{i=1}^{h} \mu_i$; $\sigma = \sqrt{\sum_{i=1}^{h} \sigma_i^2}$

<u>Methode 2</u>: $c_1 = \frac{1}{\sigma_1}$, $c_2 = \frac{1}{\sigma_2}$, ... , $c_h = \frac{1}{\sigma_h}$

dus: $\underline{T} = \sum_{i=1}^{h} \frac{\underline{t}_i - \mu_i}{\sigma_i}$; $\sigma = \sqrt{h}$

<u>Methode 3</u>: $c_1 = \frac{1}{n_1}$, $c_2 = \frac{1}{n_2}$, ... , $c_h = \frac{1}{n_h}$

dus: $\underline{T} = \sum_{i=1}^{h} \frac{\underline{t}_i - \mu_i}{n_i}$; $\sigma = \sqrt{\sum_{i=1}^{h} \left(\frac{\sigma_i}{n_i}\right)^2}$

Deze methoden zijn alleen gevoelig ten aanzien van alternatieve hypothesen volgens welke de grootheden $\underline{t}_i$ verdelingen hebben die, voor zover zij afwijken van de verdelingen onder de corresponderende hypothesen H_i, dit over het algemeen in dezelfde richting doen. Men zal dan <u>methode 1</u> bij voorkeur toepassen als men aan de $\underline{t}_i$ met een kleine spreiding (in de regel zullen dat de $\underline{t}_i$ van kleine groepen zijn) een geringer gewicht wil toekennen dan aan de $\underline{t}_i$ met een grote spreiding. De <u>methoden 2 en 3</u> zijn te gebruiken als men aan de verschillende groepen waarnemingen, ongeacht hun grootte, een ongeveer gelijke invloed op het resultaat wil toekennen. De keuze tussen deze twee methoden hangt verder van hier niet te behandelen theoretische overwegingen af (zie literatuur [1]).

Indien men verwacht dat mogelijke verschuivingen van de verdelingen der $\underline{t}_i$ in beide richtingen kunnen liggen, verdient het de voorkeur om gebruik te maken van de volgende toetsingsgrootheid:

$$\sum_{i=1}^{h} \left(\frac{\underline{t}_i - \mu_i}{\sigma_i} \right)^2 \qquad \text{(\underline{methode 4})}$$

Deze grootheid is onder de hypothese H asymptotisch verdeeld volgens een χ^2-verdeling met h vrijheidsgraden. De overschrijdingskans van een gevonden waarde van deze grootheid kan dus met behulp van een tabel van de χ^2-verdeling bepaald worden.

De toets, behandeld in memorandum S 73 (M 17a) par. 1, waarbij men het product van linkszijdige en product van alle rechtszijdige overschrijdingskansen bepaalt en het kleinste van deze twee producten gebruikt, heeft betrekking op dezelfde gevallen als de hier behandelde methoden 2 of 3, terwijl de methode, behandeld in S 73 (M 17a) par. 2, berustend op het product van de tweezijdige overschrijdingskansen, meer overeenkomt met methode 4. Men mag echter verwachten, dat, zo aan de asymptotische normaliteit der t_i voldaan is, de in dit memorandum behandelde methodenscherper zijn dan de toetsen behandeld in S 73 (M 17a).

Literatuur:

1 C.van Eeden, Methoden voor het vergelijken, toetsen en schatten van onbekende kansen, Rapport S 115 (M 45) van het Mathematisch Centrum (1953).

2 -----------, Trendtoets met behulp van rangcorrelatie, Memorandum S 73 (M 13a). (Voorbeeld van toepassing van methode 1.)

3 Dr J.Hemelrijk, Het combineren van onafhankelijke toetsen, Memorandum S 73 (M 17a).

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - O.

Statistische Afdeling
Rapport S 139 (M 48)
door
A.Benard
en
Constance van Eeden.

Toets tegen verloop voor een aantal kansen.[1]

Deze toets kan worden toegepast in het volgende geval: $R_i (i = 1,2,\ldots,k)$ zijn k onafhankelijke reeksen van onafhankelijke experimenten, waarbij ieder experiment één der mogelijke uitkomsten: succes of mislukking heeft. Bestaat de i^e reeks uit t_i experimenten, treedt hierbij $\underline{n}_i$ maal de uitkomst: succes op en stellen wij $\underline{m}_i = t_i - \underline{n}_i$, dan kunnen wij de resultaten als volgt samenvatten:

reeks no.	aantal malen		aantal experimenten
	succes	mislukking	
1	$\underline{n}_1$	$\underline{m}_1$	t_1
2	$\underline{n}_2$	$\underline{m}_2$	t_2
.	.	.	.
.	.	.	.
.	.	.	.
k	$\underline{n}_k$	$\underline{m}_k$	t_k
totaal	$\underline{n}$	$\underline{m}$	N

Hierin zijn $\underline{n}_i$ en $\underline{m}_i$ $(i = 1,2,\ldots,k)$, $\underline{n}$ en $\underline{m}$ stochastisch, terwijl t_i $(i = 1,2,\ldots,k)$ en N gegeven getallen zijn.

Is nu bij ieder experiment van de i^e reeks de kans op succes p_i (en dus de kans op mislukking $q_i = 1 - p_i$), dan luidt de hypothese H_0, die we willen toetsen:

$$H_0 : p_1 = p_2 = \ldots\ldots = p_k ,$$

terwijl de alternatieve hypothesen inhouden, dat $p_1, p_2, \ldots, p_k$ in deze volgorde een stijgend of dalend verloop vertonen.

De toets wordt nu voorwaardelijk uitgevoerd onder de voorwaarde dat $\underline{n}$ en $\underline{m}$ de bij het experiment gevonden waarden n en m aannemen.

Wij kunnen nu de N waarnemingen opvatten als n waarnemingen van een stochastische grootheid $\underline{x}$ en m van een stochastische

---

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

grootheid y , waarbij x en y beide een discrete verdeling bezitten en de waarden $1, 2, \ldots, k$ aannemen. In de twee steekproeven heeft x n_i maal de waarde i aangenomen, terwijl bij y deze waarde m_i maal is opgetreden.

Stellen we nu de kans dat x de waarde i aanneemt p_i' en de kans dat y deze waarde aanneemt p_i'' dan geldt, als H_0 juist is

$$p_i' = \frac{t_i}{N} \qquad\qquad p_i'' = \frac{t_i}{N} \, , \qquad (i = 1, 2, \ldots\ldots, k)$$

dus:

$$p_i' = p_i'' \qquad (i = 1, 2, \ldots\ldots, k)$$

en dit kunnen we toetsen door de toets van WILCOXON (zie memorandum S 47 (M 7)) toe te passen op de twee steekproeven van x en y . Omgekeerd geldt ook dat, als $p_i' = p_i''$ voor iedere i , de hypothese H_0 vervuld is.

Indien de kansen $p_1, p_2, \ldots, p_k$ een stijgend verloop vertonen, dan zullen wij bij de waarnemingen van x weinig kleine en veel grote vinden, terwijl bij die van y veel kleine en weinig grote zullen optreden, zodat dus de twee steekproeven van x en y systematisch zullen gaan verschillen. Hieruit zien we dat de alternatieve hypothesen waartegen wij H_0 willen toetsen overeenstemmen met de alternatieven van de toets van WILCOXON, zodat dus een verloop in de p's inderdaad door toepassing van de toets van WILCOXON aangetoond zal kunnen worden.

Opmerking.

Men kan gemakkelijk bewijzen dat de bovenbeschreven toets voor de hypothese

$$H_0 \; : \; p_1 = p_2 = \ldots\ldots = p_k$$

identiek is met de door T.J.TERPSTRA gegeven toets tegen verloop voor groepen waarnemingen (zie memorandum S 73 (M 28) en de opgegeven literatuur), waarbij dan iedere reeks R_i een groep van waarnemingen is, die ieder de waarde 0 of 1 bezitten.

Literatuur.

Terpstra, T.J., The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking, Proc. Kon. Ned. Akad. v. Wet., A 55 (1952).

Mathematisch Centrum,
2de Boerhaavestraat 49,
Amsterdam O.
Statistische Afdeling,
S47 (M7).

Maart, 1952.

De toets van Wilcoxon.[1]

Deze methode dient tot het toetsen van de hypothese H_o, inhoudende, dat twee steekproeven $x_1,\ldots,x_n$ en $y_1,\ldots,y_m$ afkomstig zijn uit één collectie (ook wel populatie of universum genaamd).

Voor het toetsen van de hypothese H_o wordt gebruik gemaakt van een toetsingsgrootheid $\underline{U}$ [2]), die als volgt uit de waarnemingen berekend wordt. Onderstellen we, dat de waarnemingen $x_1,\ldots,x_n$ en $y_1,\ldots,y_m$ naar opklimmende grootte gerangschikt zijn, dan bepalen we eerst het aantal waarnemingen uit de tweede steekproef, dat kleiner is dan de kleinste waarneming x_1 uit de eerste steekproef (bij gelijkheid tellen wij $\frac{1}{2}$ in plaats van 1). Noem dit aantal V_1. Vervolgens wordt het aantal waarnemingen uit de tweede steekproef bepaald, dat kleiner is dan de op één na kleinste waarneming x_2 uit de eerste steekproef (bij gelijkheid wordt weer een $\frac{1}{2}$ in plaats van 1 geteld). Dit aantal noemen we V_2. Evenzo worden met betrekking tot $x_3,x_4,\ldots,x_n$ de aantallen $V_3,V_4,\ldots,V_n$ bepaald. De waarde U van de toetsingsgrootheid $\underline{U}$ wordt voor de twee steekproeven dan gegeven door

$$U=V_1+V_2+\ldots+V_n.$$

Wanneer onder de waarnemingen niet te veel gelijken voorkomen, kan bewezen worden, dat de toetsingsgrootheid $\underline{U}$ onder de hypothese H_o voor grote waarden van n en m (beide $\geq$ 10) bij benadering een normale verdeling bezit. De waarnemingen $x_1,\ldots,x_n$ en $y_1,\ldots,y_m$ tezamen genomen vallen uiteen in een aantal groepen van gelijke waarnemingen. Noem het aantal van deze groepen k, dan is k minstens 1 (als alle waarnemingen gelijk zijn) en hoogstens m+n (als alle waarnemingen verschillend zijn).

---

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

2) Stochastische grootheden worden door onderstreping aangeduid.

Zijn $t_1, \ldots, t_k$ de aantallen waarnemingen in deze groepen van gelijken, dan worden het gemiddelde μ en de variantie σ^2 van de toetsingsgrootheid $\underline{U}$ gegeven door

$$\mu(\underline{U}) = \tfrac{1}{2}nm,$$

en

$$\sigma^2 = \mathrm{Var}\,(\underline{U}) = \frac{1}{12}\,\frac{nm}{(n+m)(n+m-1)}\left\{(n+m)^3 + \right.$$
$$\left. - (t_1^3 + t_2^3 + \ldots + t_k^3)\right\} \qquad {}^{1})$$

De grootheid $\mu(\underline{U})$ is dus onafhankelijk van de waarden vast. Indien de hypothese H_0 niet vervuld is, zal de grootheid $\underline{U}$ grote of kleine waarden bezitten, al naar gelang $\underline{y}$ systematisch kleiner of groter is dan $\underline{x}$.
De (tweezijdige) toets bestaat nu daarin, dat men H_0 verwerpt indien de gevonden waarde U van $\underline{U}$ te sterk van μ afwijkt, d.w.z. als

$$\frac{|U-\mu|}{\sigma} > \xi_\alpha\,, \qquad {}^{2})$$

waarin α de onbetrouwbaarheidsdrempel is en ξ_α volgt uit

$$\frac{1}{\sqrt{2\pi}}\int_{\xi_\alpha}^{\infty} e^{-\frac{1}{2}x^2}\,dx = \tfrac{1}{2}\alpha\,,$$

en in een tabel van de normale verdeling kan worden opgezocht.
De (tweezijdige) overschrijdingskans k , behorende bij T, is gedefiniëerd als

$$k = \frac{2}{\sqrt{2\pi}}\int_{\left|\frac{U-\mu}{\sigma}\right|}^{\infty} e^{-\frac{1}{2}x^2}\,dx \qquad {}^{2})$$

en kan ook in eenttabel van de normale verdeling worden gevonden.
Bij eenzijdige toetsing wordt α door 2α vervangen, resp. k gehalveerd.
Een bijzonder geval van het bovenstaande is, dat onder de waarnemingen voor $\underline{x}$ en $\underline{y}$ in 't geheel geen gelijken voorkomen. In dat geval kan de uitdrukking voor de variantie herleid worden tot

$$\sigma^2 = \frac{1}{12}\,nm(n+m+1).$$

---

1) Deze formule is een door T.J.Terpstra gegeven vereenvoudiging van de door J.Hemelrijk([5] en [7]) afgeleide formule. De afleiding van deze vereenvoudigde formule zal nog gepubliceerd worden.

2) Deze formules berusten op de normale benadering van de verdeling van $\underline{U}$.

Indien n en m kleiner zijn dan 10, zijn tabellen beschikbaar voor het berekenen van de overschrijdingskans k voor de uit de steekproef bepaalde waarde U van U (zie [2] en [4]). Dergelijke tabellen bestaan echter niet voor het geval, dat gelijke waarnemingen optreden.

Opmerking. Men kan gemakkelijk bewijzen, dat de variantie van U door het optreden van gelijke waarnemingen vermindert. Het verschil, dat door deze gelijken optreedt, is echter in het algemeen gering. Men kan daarom in eerste instantie deze correctie op σ^2 verwaarlozen. De overschrijdingskansen, die men dan vindt, zijn iets te groot.

Litteratuur:

1. F.Wilcoxon, — Individual comparisons by ranking methods, Biometrics 1 (1945), p.80-83.
2. H.B.Mann and D.R.Whitney — On a test of whether one of two random variables is stochastically larger than the other, Amer.Math.Stat. 18 (1947),p. 50-60.
3. H.R.van der Vaart — Some remarks on the power function of Wilcoxon's test for the problem of two samples, Proceedings van de Kon. Ned.Ak.v.Wet., 53 (1950),p. 494-520.
4. H.R.van der Vaart — Gebruiksaanwijzing voor de toets van Wilcoxon, met tabellen voor n en m≤10, Rapport S32 (M4) (1950).
5. H.R.van der Vaart — De toets van Wilcoxon voor het probleem van twee steekproeven. (Cursus "Parametervrije Methoden", 1951-'52).
6. D.van Dantzig — Kadercursus Mathematische Statistiek, Math. Centrum, Amsterdam (1947-'50), hoofdst. 6, § 3.
7. J.Hemelrijk — Note on Wilcoxon's two sample test, when ties are present, Ann.Math.Stat. 23 (1952) no. 2.