

Dr W. VAN DER BIJL

**TOEPASSING
VAN STATISTISCHE METHODEN
IN DE KLIMATOLOGIE**

*

**APPLICATIONS OF STATISTICAL METHODS
IN CLIMATOLOGY**

1952



F 7.60

Ned. Meteor. Inst.

De Bilt

02 - 58

STAATSDRUKKERIJ- EN UITGEVERIJBEDRIJF 'S-GRAVENHAGE

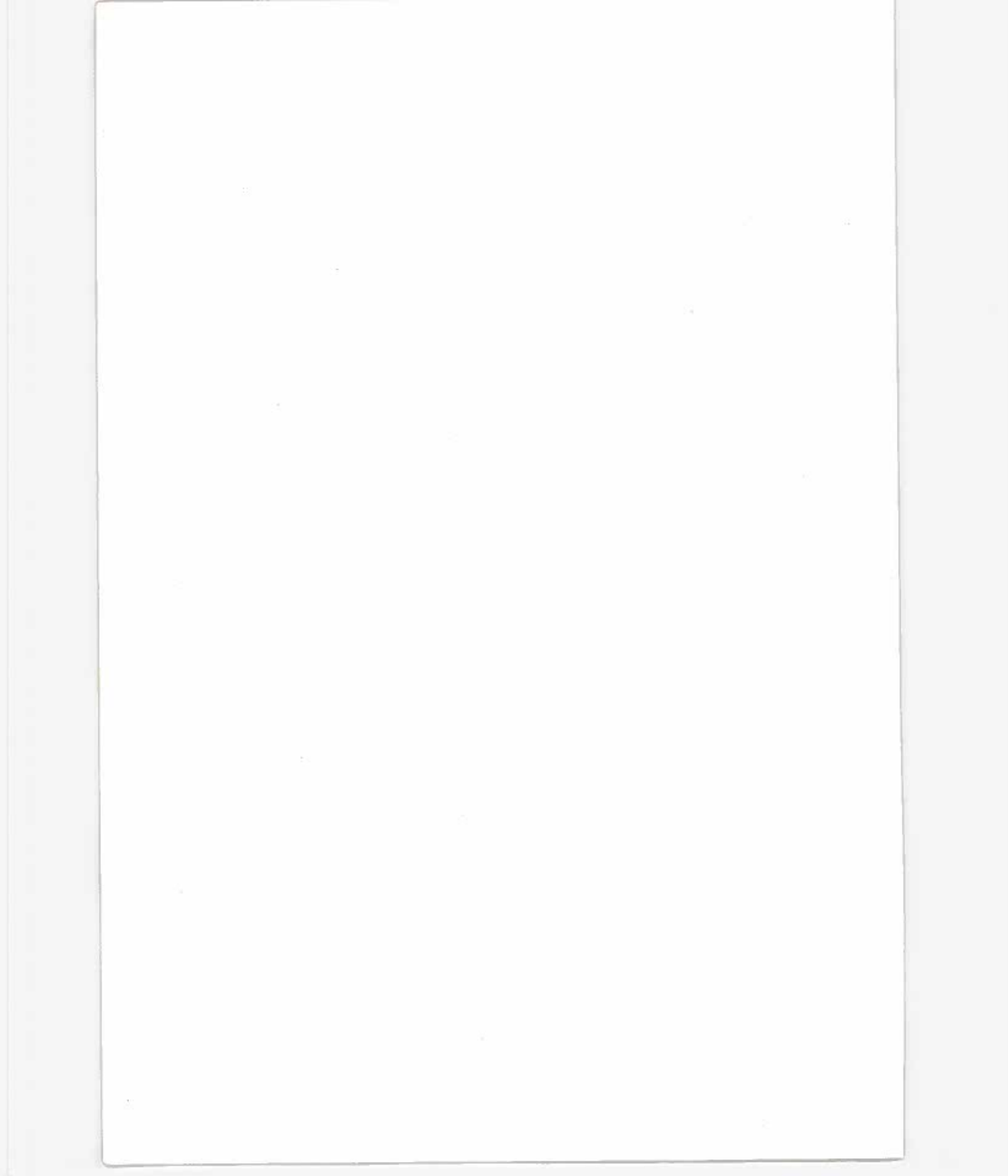
TOEPASSING
VAN STATISTISCHE METHODEN
IN DE KLIMATOLOGIE

DR W. VAN DER BIJL

1952



STAATSDRUKKERIJ- EN UITGEVERIJBEDRIJF / 'S-GRAVENHAGE



INDEX

ALPHABETICALLY	1
ALPHABETICALLY BY SUBJECT	2
ALPHABETICALLY BY AUTHOR	3
ALPHABETICALLY BY TITLE	4
ALPHABETICALLY BY DATE	5
ALPHABETICALLY BY PERIOD	6
ALPHABETICALLY BY COUNTRY	7
ALPHABETICALLY BY LANGUAGE	8
ALPHABETICALLY BY FORM	9
ALPHABETICALLY BY METHOD	10
ALPHABETICALLY BY MATERIAL	11
ALPHABETICALLY BY INSTRUMENT	12
ALPHABETICALLY BY MEASURE	13
ALPHABETICALLY BY UNIT	14
ALPHABETICALLY BY FORCE	15
ALPHABETICALLY BY EFFECT	16
ALPHABETICALLY BY CAUSE	17
ALPHABETICALLY BY RESULT	18
ALPHABETICALLY BY ACTION	19
ALPHABETICALLY BY PASSION	20
ALPHABETICALLY BY AFFECTION	21
ALPHABETICALLY BY SENSATION	22
ALPHABETICALLY BY FACULTY	23
ALPHABETICALLY BY POWER	24
ALPHABETICALLY BY FACULTY	25
ALPHABETICALLY BY POWER	26
ALPHABETICALLY BY FACULTY	27
ALPHABETICALLY BY POWER	28
ALPHABETICALLY BY FACULTY	29
ALPHABETICALLY BY POWER	30
ALPHABETICALLY BY FACULTY	31
ALPHABETICALLY BY POWER	32
ALPHABETICALLY BY FACULTY	33
ALPHABETICALLY BY POWER	34
ALPHABETICALLY BY FACULTY	35
ALPHABETICALLY BY POWER	36
ALPHABETICALLY BY FACULTY	37
ALPHABETICALLY BY POWER	38
ALPHABETICALLY BY FACULTY	39
ALPHABETICALLY BY POWER	40
ALPHABETICALLY BY FACULTY	41
ALPHABETICALLY BY POWER	42
ALPHABETICALLY BY FACULTY	43
ALPHABETICALLY BY POWER	44
ALPHABETICALLY BY FACULTY	45
ALPHABETICALLY BY POWER	46
ALPHABETICALLY BY FACULTY	47
ALPHABETICALLY BY POWER	48
ALPHABETICALLY BY FACULTY	49
ALPHABETICALLY BY POWER	50
ALPHABETICALLY BY FACULTY	51
ALPHABETICALLY BY POWER	52
ALPHABETICALLY BY FACULTY	53
ALPHABETICALLY BY POWER	54
ALPHABETICALLY BY FACULTY	55
ALPHABETICALLY BY POWER	56
ALPHABETICALLY BY FACULTY	57
ALPHABETICALLY BY POWER	58
ALPHABETICALLY BY FACULTY	59
ALPHABETICALLY BY POWER	60
ALPHABETICALLY BY FACULTY	61
ALPHABETICALLY BY POWER	62
ALPHABETICALLY BY FACULTY	63
ALPHABETICALLY BY POWER	64
ALPHABETICALLY BY FACULTY	65
ALPHABETICALLY BY POWER	66
ALPHABETICALLY BY FACULTY	67
ALPHABETICALLY BY POWER	68
ALPHABETICALLY BY FACULTY	69
ALPHABETICALLY BY POWER	70
ALPHABETICALLY BY FACULTY	71
ALPHABETICALLY BY POWER	72
ALPHABETICALLY BY FACULTY	73
ALPHABETICALLY BY POWER	74
ALPHABETICALLY BY FACULTY	75
ALPHABETICALLY BY POWER	76
ALPHABETICALLY BY FACULTY	77
ALPHABETICALLY BY POWER	78
ALPHABETICALLY BY FACULTY	79
ALPHABETICALLY BY POWER	80
ALPHABETICALLY BY FACULTY	81
ALPHABETICALLY BY POWER	82
ALPHABETICALLY BY FACULTY	83
ALPHABETICALLY BY POWER	84
ALPHABETICALLY BY FACULTY	85
ALPHABETICALLY BY POWER	86
ALPHABETICALLY BY FACULTY	87
ALPHABETICALLY BY POWER	88
ALPHABETICALLY BY FACULTY	89
ALPHABETICALLY BY POWER	90
ALPHABETICALLY BY FACULTY	91
ALPHABETICALLY BY POWER	92
ALPHABETICALLY BY FACULTY	93
ALPHABETICALLY BY POWER	94
ALPHABETICALLY BY FACULTY	95
ALPHABETICALLY BY POWER	96
ALPHABETICALLY BY FACULTY	97
ALPHABETICALLY BY POWER	98
ALPHABETICALLY BY FACULTY	99
ALPHABETICALLY BY POWER	100

If a man will begin with certainties he shall end in doubts; but if he will be content to begin with doubts he shall end in certainties.

FRANCIS BACON

INHOUD

blz.

1	0	INLEIDING
1	0. 0	Opmerkingen van algemene aard
1	0. 0.1	Ontstaan
1	0. 0.2	Gestelde doel
2	0. 0.3	Bereikte doel
3	0. 0.4	Korte inhoud der hoofdstukken
3	0. 1	Klimatologie en statistica
4	0. 2	Benamingen voor het bestudeerde gebied
5	0. 3	Klimatologische publicaties in Nederland
5	0. 4	Klassieke klimatologie
5	0. 5	Luchtsoorten-klimatologie
6	0. 6	Statistische klimatologie
7	1	STATISTISCHE METHODEN
7	1. 0	Samenvatting
7	1. 1	Universum en steekproef (population and sample)
13	1. 1.1	Voorwaarden, waaraan een steekproef behoort te voldoen
15	1. 2	Verschillen tussen steekproef- en universumwaarden
17	1. 2.1	Het begrip overschrijdingskans
18	1. 2.2	De begrippen significantie en betrouwbaarheidsdrempel
19	1. 2.3	Selectieeffect
20	1. 3	Enkele veel voorkomende verdelingen
20	1. 3.1	De verdeling van GAUSS
21	1. 3.2	De verdeling van BERNOULLI
22	1. 3.3	De verdeling van POISSON
22	1. 3.4	De U-verdeling
23	1. 4	Centrum en gemiddelde (universum- en steekproefgemiddelde)
23	1. 5	Het aantal vrijheidsgraden
24	1. 6	Standaarddeviatie en variantie
26	1. 6.1	De frequentiefunctie van de standaarddeviatie van de enkelvoudige waarneming
26	1. 6.2	De frequentiefunctie van de standaarddeviatie in grote steekproeven
26	1. 7	De frequentiefunctie van de steekproefgemiddelden

26	1. 7.1	De frequentiefunctie van de steekproefgemiddelden in grote steekproeven
27	1. 7.2	De frequentiefunctie van de steekproefgemiddelden in kleine steekproeven
27	1. 7.3	Significantie van het verschil tussen twee steekproefgemiddelden (STUDENT's <i>t</i> -toets)
28	1. 8	De G^2 -toets
29	1. 9	De <i>z</i> -toets van FISHER
30	1.10	De variantieanalyse
30	1.10.1	De enkelvoudige variantieanalyse
32	1.10.2	De tweevoudige variantieanalyse
33	1.11	Correlatie en regressie
33	1.11.0	Meerdimensionale universa
33	1.11.1	Enkelvoudige correlatie
35	1.11.1.1	Significantie van het verschil tussen twee correlatiecoëfficiënten
35	1.11.2	Enkelvoudige regressie
36	1.11.3	Meervoudige regressie
37	1.11.4	Meervoudige correlatie
39	1.11.4.1	De frequentiefunctie van de multiële correlatiecoëfficiënt R , wanneer de universumwaarde $P = 0$ is
40	1.12	Persistentie
42	1.12.1	Autocorrelatiecoëfficiënt van de k^{de} orde
42	1.12.2	Het effectieve aantal onafhankelijke waarden $\gamma(g)$
43	1.12.3	Het aequivalente herhalingsgetal $\omega(g)$
46	2	TOEPASSINGEN VAN DE STATISTISCHE METHODEN
46	2. 0	Inleiding
46	2. 1a	Universum en steekproef (population and sample)
51	2. 1b	Idem
54	2. 2.1	Het begrip overschrijdingskans
55	2. 2.2	De begrippen significantie en betrouwbaarheidsdrempel
58	2. 2.3	Selectieeffect
58	2. 3	Enkele veel voorkomende verdelingen
58	2. 3.1	De verdeling van GAUSS
62	2. 3.2	De verdeling van BERNOULLI
65	2. 3.3	De verdeling van POISSON
66	2. 4	Centrum en gemiddelde (universum- en steekproefgemiddelde)
69	2. 5	Het aantal vrijheidsgraden
71	2. 6	Standaarddeviatie en variantie

72	2. 6.1	De frequentiefunctie van de standaarddeviatie van de enkelvoudige waarneming
75	2. 6.2	De frequentiefunctie van de standaarddeviatie in grote steekproeven
76	2. 7	De frequentiefunctie van de steekproefgemiddelden
76	2. 7.1	De frequentiefunctie van de steekproefgemiddelden in grote steekproeven
77	2. 7.2	De frequentiefunctie van de steekproefgemiddelden in kleine steekproeven
80	2. 7.3	Significantie van het verschil tussen twee steekproefgemiddelden (STUDENT's <i>t</i> -toets)
81	2. 8	De G^2 -toets
84	2. 9	De <i>z</i> -toets van FISHER
86	2.10	De variantieanalyse
86	2.10.1	De enkelvoudige variantieanalyse
90	2.10.2	De tweevoudige variantieanalyse
94	2.11	Correlatie en regressie
94	2.11.1	Enkelvoudige correlatie
97	2.11.1.1	Significantie van het verschil tussen twee correlatiecoëfficiënten
98	2.11.2	Enkelvoudige regressie
99	2.11.3	Meervoudige regressie
103	2.11.4	Meervoudige correlatie
104	2.11.4.1	De frequentiefunctie van de multiële correlatiecoëfficiënt <i>R</i> , wanneer de universumwaarde $P = 0$ is
105	2.12	Persistentie
105	2.12.1	Autocorrelatiecoëfficiënt van de k^{de} orde
106	2.12.2	Het effectieve aantal onafhankelijke waarden $\gamma(g)$
107	2.12.3	Het aequivalente herhalingsgetal $\omega(g)$
108	3	ENIGE FOUTENBRONNEN BIJ HET STATISTISCH ONDERZOEK
108	3. 0	Inleiding
109	3. 0.1	Het steekproefeffect
110	3. 0.2	Het selectieeffect
110	3. 0.3	Het persistentieeffect
111	3. 0.4	Het pseudocorrelatieeffect
111	3. 0.5	Het krommingseffect
112	3. 1	Het steekproefeffect
127	3. 2	Het selectieeffect

133	3. 3	Het persistentieeffect
139	3. 4	Het pseudocorrelatieeffect
146	3. 5	Het krommingseffect
153	4	SLOTBESCHOUWINGEN OVER GEOFYSISCH- STATISTISCHE RESULTATEN
153	4. 1	De σ_r -fout
156	4. 2	Een bijzonder ingewikkeld geval van pseudocorrelatie
163	4. 3	De grafieken-methode in het significantie-tijdperk
164	4. 3.1	De invloed van de maan op de temperatuur
167	4. 3.2	De invloed van de maan op de neerslag
171	4. 4	Maandelijkse luchtdrukveranderingen, in het bijzonder die, voorafgaande aan strenge winters
175	4. 5	Proeve van regionale neerslagverwachtingen op lange termijn
185		LITERATUURLIJST
197		SUMMARY

HOOFDSTUK 0

INLEIDING

0.0 OPMERKINGEN VAN ALGEMENE AARD

0.0.1 ONTSTAAN

Deze verhandeling moet gezien worden als resultaat van 5 jaren speurwerk, verricht in de Afdeling Klimatologie van het K.N.M.I.

Het stemt mij tot grote dankbaarheid, dat mij in ruime mate de vrije hand is gelaten in het bepalen van de richting van mijn onderzoekingen. Deze dank geldt niet alleen voor mijn hooggeleerde promotor, Prof. Dr S. W. VISSER, maar ook voor de successieve directeurs van de Afdeling Klimatologie van het K.N.M.I., Dr C. BRAAK voor 1 Jan. 1948 en Dr L. J. L. DEJ na die datum. Prof. VISSER heeft op dit werk zijn stempel gedrukt door zijn waardevolle suggesties en veelvuldige kritische opmerkingen. De heer A. DENKEMA – tijdelijk geholpen door andere rekenaars van de Onderafdeling Landbouwmeteorologie – heeft door zijn grote nauwkeurigheid deze verhandeling een betrouwbare arithmetische ruggegraat geschonken.

0.0.2 GESTELDE DOEL

De overweldigende massa geofysische waarnemingen hebben vele geofysici geïnspireerd tot het zoeken naar causale samenhangen op dit uitgestrekte gebied. Niet alleen de beroepsgeofysici, die hun gehele tijd aan dit fascinerende onderwerp konden wijden, ook tal van verdienstelijke amateurs hebben belangrijke bijdragen tot het ontraadselen van de geheimen van dit gedeelte van de natuur geleverd.

Nu spreekt het wel vanzelf, dat niet alle conclusies volledig betrouwbaar zijn en de schrijver heeft zich dan ook tot taak gesteld, methoden te vinden, die het kaf van het koren kunnen scheiden.

Het hiermede gestelde doel is tweeledig. In de eerste plaats wordt daardoor bereikt, dat het koren – of om een later te definiëren term reeds nu in te voeren, nl. het statistisch betrouwbare materiaal – lijnrecht komt te staan tegenover het kaf of wel: toevallige overeenkomsten. Maar niet alleen daartegenover plaatst zich het statistisch betrouwbare materiaal, neen, tegelijkertijd plaatst het zich mijlenhoog boven die resultaten, waaraan het machtige toeval alleen maar een schijn van werkelijkheid gegeven heeft.

Dat er inderdaad kaf tussen het koren schuilt, blijkt o.a. uit het ontstellend grote aantal perioden in de meteorologische grootheden. N. SHAW (1928, *Manual of Meteorology*, II, 320-4) telde toen reeds 133 door 47 auteurs ontdekte perioden, waarvan de lengten tussen 1 en 260 jaren liggen. De periodenjagers hebben zo'n omvangrijke buit binnengebracht, dat zelfs de meest welwillende criticus achterdochtig moet worden. De buitensporige opbrengst van de periodenjacht mag ongetwijfeld gezien worden als een aanwijzing, dat niet de natuur, maar het toeval een groot deel van de perioden geleverd heeft.

In de tweede plaats zal het bestuderen van dergelijke zeef-methoden de toekomstige productie van het kaf (statistisch onbetrouwbare gevolgtrekkingen) zeker helpen verminderen. Het verhinderen daarvan zal wel een utopie blijken te zijn.

0.0.3 BEREIKTE DOEL

Waarschijnlijkheidsrekening en statistica leverden de methoden, die aan geofysische behoeften aangepast dienden te worden.

Het gestelde doel is nog lang niet bereikt. Tot nu toe veroorloven slechts weinige wetenschappelijke publicaties na hun bestudering een onmiddellijk antwoord op de onafwendbare vraag: „wet of toeval?” Enkele voorbeelden zijn aangegeven, voornamelijk in de hoofdstukken 3 en 4.

Er moet met nadruk op gewezen worden, dat de term „statistisch onbetrouwbaar” nog niet betekent, dat ook de theoretische beschouwingen, waarvan de onderzoeker uitgegaan is, onbetrouwbaar zijn. *De statistica heeft een dienende taak te vervullen*: zij kan de geofysicus wijzen op statistische bezwaren. Deze kunnen in sommige gevallen zo zwaar wegen, dat de geofysicus moet besluiten zijn eigen theorie te verwerpen. Steeds zullen zij hem dwingen zijn conclusies zodanig te funderen, dat zij statistische bezwaren kunnen doorstaan.

Bij de beantwoording van de vraag: wet of toeval? heeft tenslotte de geofysicus het laatste woord. Komt de geofysicus op grond van zijn beschouwingen tot zekere gevolgtrekkingen, die hij in zijn waarnemingsmateriaal bevestigd ziet, dan kan hij desgewenst nog een statisticus raadplegen. Luidt het antwoord, dat er geen sprake van toeval kan zijn, dan is dit een steun voor de geofysicus. Zegt de statisticus echter, dat de waarnemingen wel door het toeval veroorzaakt kunnen zijn, dan moet de geofysicus deze statistische bezwaren afwegen tegen de hechtheid van zijn theoretische beschouwingen.

Er zijn nog veel theorieën en hypothesen, waarvan de betrouwbaarheid, dus de overeenstemming met het experiment, niet statistisch maar alleen subjectief beoordeeld kan worden. Hun aantal zal ongetwijfeld gestadig afnemen; de ontwikkelingsgang van de statistica is daar een goede borg voor.

Naar volledigheid is niet gestreefd. De nieuweling zal na het doorwerken van dit geschrift de eerste schreden op het statistische pad wel durven afleggen. Voor verre reizen, diep het statistische gebied in, moet men Baedekers als KENDALL (1947) en FISHER (1948) nauwkeurig raadplegen.

Wel is getracht, op een zo duidelijk mogelijke wijze tijdig te waarschuwen bij het bereiken van gevaarlijke en doodlopende zijwegen van de statistica. Het bezit van zo vele „pitfalls” is een eigenschap van de statistica, die al menig slachtoffer geëist heeft en ook niet zelden tot een diepe, welhaast onoverbrugbare controverse geleid heeft. Men denke aan de z.g. singulariteiten in de klimatologie!

0.0.4 KORTE INHOUD DER HOOFDSTUKKEN

Na enkele algemene opmerkingen in het inleidende hoofdstuk 0 volgt hoofdstuk 1 met een uitvoerige behandeling van statistische methoden.

In hoofdstuk 2 worden deze methoden op geofysische grootheden toegepast; in de met een a aangegeven paragraaf op de juiste wijze; in de b-paragraaf op een in de literatuur voorgekomen, onjuiste manier. Dit hoofdstuk berust dus op wat KENDALL noemt: „the invaluable principle that example is better than precept”.

Een tussenopmerking betreffende het gebruik van de woorden geofysica en geofysisch: hoewel bijna alle voorbeelden genomen zijn uit het gebied van de klimatologie, zullen de in hoofdstuk 1 aangegeven statistische methoden in het algemeen ook toepasbaar zijn in de overige gebieden van de geofysica.

De algemene fouten op statistisch gebied gemaakt in geofysische onderzoeken worden in hoofdstuk 3 in vijf hoofdgroepen ingedeeld.

Het laatste hoofdstuk critiseert verschillende geofysisch-statistische resultaten. Een verwachtingsmethode voor regionale neerslaghoeveelheden in de maand Augustus bleek weinig succesvol.

0.1 KLIMATOLOGIE EN STATISTICA

De twee takken van wetenschap, die het meest besproken worden in deze verhandeling, zijn de klimatologie en de statistica. Zo gemakkelijk het is een definitie te vinden voor de eerste (de wetenschap, die zich bezighoudt met het bestuderen van het klimaat), zo moeilijk is het, de statistica op een voor iedereen aanvaardbare wijze te definiëren.

KENDALL (1947, I, 1) merkt terecht op, dat tot de vele onderwerpen, waarover de statistici het niet eens zijn, ook de definiëring van hun wetenschap behoort.

Het aantal van 115 definities, door WILLCOX in 1935 reeds opgesomd, noemt KENDALL nog verre van volledig.

Terwijl deze overmaat de wetenschap geen enkele praktische moeilijkheid bezorgt, is de enige definitie van klimatologie juist een bron van onenigheid.

In deze definitie is namelijk het begrip klimaat ingevoerd en dit begrip nu lijdt aan dezelfde overmaat aan omschrijvingen als de statistica.

Weliswaar zullen er geen honderd definities van „klimaat” bij elkaar gebracht kunnen worden – een serieuze poging is voor zover bekend nog niet gewaagd –, maar 35 medewerkers van het Amerikaanse Weather Bureau gaven al niet minder dan 16 uiteenlopende meningen te kennen (THORNTHWAITE 1949, 440). Wat onze eigen klimatologen betreft, VISSER (1949) plaatst 6 formuleringen alleen al van Nederlandse schrijvers naast elkaar zonder op volledigheid aanspraak te willen maken.

In zijn werk (op de volgende bladzijde, p. 4) merkt VISSER terecht op, dat er maar één duidelijke, korte, krachtige definitie van het klimaat is, nl. die welke is opgesteld in 1935 op de bijeenkomst van de Internationale Meteorologische Organisatie te Warschau: „Onder het klimaat verstaat men het geheel der gemiddelde atmosferische verschijnselen voor een bepaald gebied, berekend uit een tijdvak van 30 jaren. Als standaardtijdvak is 1901–1930 aangenomen.”

Deze min of meer willekeurige vaststelling van de „normaal-periode” is het zwakke punt in de definitie. Vele landen houden zich er dan ook niet aan, hoewel het besluit in Warschau destijds algemeen aanvaard is.

Bij gedetailleerde onderzoeken naar klimaatsveranderingen gaat het steekproefeffect een grote rol spelen, zelfs al hebben de gemiddelden betrekking op 30 à 40 jaren.

Zo komt het, dat de statistica te hulp geroepen wordt bij de omschrijving van het begrip klimaat. Een eerste resultaat is afkomstig van Miss CARRUTHERS (1945), die als optimale lengte 35 jaren vond.

0.2 BENAMINGEN VOOR HET BESTUDEERDE GEBIED

Het is niet zo moeilijk voor het in dit werk behandelde wetenschapsgedeelte een aparte naam te bedenken. De vraag is echter, of deze naamgeving verantwoord genoemd mag worden en of zij niet in plaats van vereenvoudiging verwarring zal brengen.

Want naast de in 0.4 en 0.5 aan te stippen klimatologie-categorieën staat een reeks van andere, zoals dynamische, sociale, fysische, synthetische, fysiologische, statische, technische, regionale, toegepaste, algemene, theoretische, micro- en bioklimatologie.

VISTELIUS (zie recensie van TOMKEIEFF 1947) introduceerde de term analytische geologie. Dit begrip heeft betrekking op de toepassing van statistische methoden op numerieke geologische gegevens en de daaruit voortvloeiende interpretatie van de resultaten in het licht van significante geologische feiten. Als deze term zich weet te handhaven, heeft het zin het analoge gebied in de klima-

tologie de naam analytische klimatologie te geven. Helaas heeft CONRAD (1942) deze naam reeds gebruikt in een geheel andere betekenis.

Tenslotte verdient de suggestie van PONCELET (1946) onze aandacht. Deze wil de numerieke en grafische beschrijving van de atmosferische grootheden „klimatografie” noemen. Het begrip klimatologie blijft dan beperkt tot de wetenschap, die een logisch verband weet te leggen tussen de eigenschappen van het klimaat, zijn schommelingen en zijn invloeden op de mens. Wij zullen ons echter onthouden van iedere aanbeveling.

0.3 KLIMATOLOGISCHE PUBLICATIES IN NEDERLAND

Vrijwel alle Nederlandse publicaties, die betrekking hebben op het gebied van de klimatologie, zijn verschenen in de serie „Mededelingen en Verhandelingen” van het Koninklijk Nederlands Meteorologisch Instituut. Van de 67 Mededelingen en Verhandelingen, uitgekomen in de jaren 1905-1949, hielden 37 zich bezig met onderwerpen uit de klimatologie van Nederland en Overzeese Gebiedsdelen.

0.4 KLASSIEKE KLIMATOLOGIE

Al deze Mededelingen en Verhandelingen hebben één ding gemeen: zij beperken zich bijna uitsluitend tot het bepalen van rekenkundige gemiddelden van de talrijke gegevens, waarover de klimatologische dienst nu eenmaal de beschikking heeft.

Op welke wijze zou deze zgn. klassieke klimatologie uitgebouwd kunnen worden tot een klimaatkunde, die meer bevrediging schenkt?

De uitbreidingsplannen kunnen in 2 categorieën verdeeld worden. De ontwerpers van die plannen komen namelijk voort uit twee geheel verschillende wetenschappelijke groeperingen, een statistische en een fysische.

De uitbreiding in statistische richting wordt natuurlijk door statistici gepropageerd.

0.5 LUCHTSOORTEN-KLIMATOLOGIE

De fysische uitbreiding wordt gestimuleerd door de synoptische meteorologen. Met lede ogen moeten zij aanzien, hoe de klassieke klimatologie de vele theoretische en praktische resultaten van de Noorse school ongebruikt laat. De frontentheorie heeft menig duister feit in de meteorologie opgehelderd. In zijn enthousiasme verwacht nu de meteoroloog, dat de Noorse ontdekkingen ook bevruchtend zullen kunnen werken op de andere tak van wetenschap, die zich met de aardse atmosfeer bezig houdt: de klimatologie.

Onlangs hebben ook Nederlandse exponenten van deze fysische richting zich

laten horen: BLEEKER en BIJVOET (1949) begonnen hun artikel met uitvoerig de aandacht te vestigen op de gebreken van de klassieke klimatologie. Een reeks van voor de hand liggende bezwaren werd opgesomd. Vervolgens willen zij de klassieke klimatologie gedeeltelijk vervangen zien door de luchtsoortenklimatologie van LINKE-DINIES-BERGERON-WILLETT-SCHERHAG. Deze fysische ontwikkeling van de klimatologie heeft talrijke voordelen. De belangrijkste zijn: een nader begrip van enkele klimatologische moeilijkheden en een beter inzicht in klimatologische verschillen.

Het blijkt dus, dat de voordelen voornamelijk van theoretische aard zijn. Veel praktisch nut zal deze ontwikkeling voorlopig niet kunnen geven. Oorzaak daarvan is het feit, dat het waarnemingsmateriaal een te geringe omvang heeft om een behoorlijke differentiëring mogelijk te maken. Het gevolg is, dat de onzekerheid van de door de luchtsoortenklimatologie bereikte resultaten groter is dan toegelaten mag worden.

De klimatoloog moet dus een andere wijze van bewerking van zijn gegevens kiezen, één, die vooral ingesteld is op rechtstreekse beantwoording van de vragen, die mens en maatschappij hem stellen.

0.6 STATISTISCHE KLIMATOLOGIE

De klimatologie zal daarom meer baat hebben bij een uitbreiding langs statistische banen.

In 1897 verscheen een verhandeling (PEARSON and LEE) op klimatologisch gebied, waarin de – toen nog in haar kinderschoenen staande – statistica voor de eerste maal een grote rol speelde. In de laatste halve eeuw maakte men in het buitenland – zij het lang niet voor de volle 100% – gebruik van de resultaten van de nieuwe, zich in alle richtingen ontwikkelende statistische theorie. In Nederland bleef dit gebruik beperkt tot de correlatierekening: uit de serie Mededelingen en Verhandelingen waren dit de nummers 50, 51 (in hoge mate) en 55.

Het aantal klimatologische gegevens blijft zich uit de aard der zaak steeds uitbreiden. De statistica nu is de wetenschap, die zich o.a. ten doel stelt, uit een massa gegevens de meest essentiële eigenschappen naar voren te brengen, m.a.w. uit het beschikbare waarnemingsmateriaal die conclusies te trekken, die verantwoord genoemd mogen worden. Wat men onder „verantwoorde conclusies” heeft te verstaan, zal blijken uit 1.1 en 1.2.

Daar in Nederland geen statistische methoden in de klimatologie zijn toegepast en evenmin de mogelijkheid is aangegeven tot die toepassing, is in hoofdstuk 1 een verzameling van niet al te ingewikkelde methoden bijeengebracht. De daarop volgende hoofdstukken behelzen voorbeelden van correcte toepassingen en besprekingen van minder correcte toepassingen.

HOOFDSTUK 1

STATISTISCHE METHODEN

The number of observations in the meteorological register makes the necessity of statistical methods evident. CONRAD 1944

1.0 SAMENVATTING

In de paragrafen 1-6 van dit hoofdstuk wordt getracht een omschrijving te geven van de typisch statistische begrippen als steekproef, universum, significantie, vrijheidsgraden e.d.

De tweede helft van dit hoofdstuk laat zien, hoever steekproefgrootheden als gemiddelden en varianties van de universumwaarden kunnen afwijken. (1.6, 1.7 en 1.9).

De G^2 -toets (1.8) en de variantieanalyse (1.10) zijn uitstekende hulpmiddelen om met zekerheid vast te stellen, of bij bepaalde coincidenties het toeval al dan niet een rol gespeeld kan hebben.

Tien jaar geleden zou aan de correlatie en regressie meer aandacht besteed zijn dan de enkele paragraaf 1.11 nu. Misschien maakt de overschatting van destijds plaats voor een even ongewenste onderschatting in de toekomst.

1.12 behandelt de statistische zijde van de persistentie. De vraag naar het aantal onafhankelijke waarnemingen, dat met een reeks afhankelijke waarnemingen gelijk te stellen is, werd het eerst door Prof. BARTELS beantwoord. Zijn methode wordt in deze slotparagraaf gevolgd.

1.1 UNIVERSUM EN STEEKPROEF (population and sample)

In science we are always faced with the problem of arguing from the particular to the general, or in statistical language, from the sample to the population. MAHALANOBIS 1938

Deze voor de statistica onmisbare begrippen kunnen gezamenlijk behandeld worden, daar ze in nauw verband tot elkaar staan.

Zowel universum als steekproef kunnen opgevat worden als verzamelingen van gelijksoortige elementen. Een steekproef verkrijgt men door een aantal elementen uit een universum af te zonderen. Een steekproef is dus altijd een deel van een universum. Uit een universum kan steeds minstens één steekproef genomen worden. Elk universum kan men splitsen in een aantal steekproeven. (Deze worden dan ook wel deel-universa genoemd).

De grootte van een steekproef is onafhankelijk van de grootte van zijn universum. De steekproef kan alle elementen van het universum omvatten; hij kan ook uit één enkel element bestaan.

Elk universum kan op zijn beurt weer opgevat worden als steekproef van een ander universum. Ter toelichting geven we nu enkele voorbeelden uit de klimatologie: als element kiezen we de temperatuur, die om 8.40 is afgelezen op een bepaalde thermometer in De Bilt. Een steekproef vormen dan de temperaturen van de 365 ochtend-waarnemingen van het jaar 1901. Als universum kan men nemen alle ochtend-waarnemingen uit het tijdvak 1901-30. Dit laatste universum kan men weer opvatten als een steekproef uit het universum der ochtend-waarnemingen van een veel langer tijdvak, zoals bv. dat van 1735-1949.

Men kan onderscheid maken tussen eindige en oneindige universa. Een universum met een eindig aantal elementen wordt bv. gevormd door de herleide reeks waarnemingen van de neerslaghoeveelheden te Zwanenburg-Hoofddorp (1735-1949). Tot nu toe zijn voorbeelden behandeld, waarbij de tijd een variabele factor was, terwijl de ruimtecoördinaten van het meetpunt constant waren. Het omgekeerde is ook mogelijk. De verzameling neerslaghoeveelheden, afgetapt op de meer dan 300 Nederlandse regenstations te 8 uur op 1 Sept. 1949 is een voorbeeld van een eindig universum, waarbij de ruimtecoördinaten veranderlijk zijn en de tijd constant is.

De geofysica kent ook universa met een oneindig aantal elementen. Voorbeelden zijn: de temperaturen van alle punten van de Atlantische Oceaan op een bepaald ogenblik en alle gemiddelde jaarlijkse luchttemperaturen te De Bilt, zowel uit het verleden als uit de toekomst. Ook het diagram van het temperatuurverloop van één enkele dag levert een universum met oneindig vele temperaturen.

Het begrip steekproef is in de statistica daarom van zoveel betekenis, omdat men meestal niet op bevredigende wijze met universa kan werken. Dit bezwaar zou onbelangrijk zijn, indien men niets van en over de universa behoefde te weten. Maar het omgekeerde is juist het geval: onze belangstelling gaat altijd naar het universum uit.

In het algemeen kan men wel zeggen, dat hetgene, dat we willen kennen, het universum is. Er is echter geen enkele mogelijkheid dit universum exact te leren kennen.

Wel kunnen we alle eigenschappen te weten komen van de steekproef. De stap steekproef-universum is een zeer moeilijke; om die mogelijk te maken is de gehele statistica opgebouwd. Zou elk universum zich tot elke bewerking lenen, dan waren er geen steekproeven nodig en zou de gehele statistica overbodig zijn. Onder deze ideale omstandigheden zouden de lichamen, die zich met het onder-

zoek van de publieke opinie bezighouden, „slechts” absoluut betrouwbare resultaten kunnen publiceren.

Het kost weinig moeite de oorzaken van de „onhandelbaarheid” der universa op te sporen:

1. het universum kan zodanig zijn, dat we niet alle elementen kennen. Dit geval doet zich voor, wanneer de waarnemingen niet allemaal verricht kunnen worden. Dit is duidelijk, wanneer het een oneindig universum betreft.
2. men kan zich een geval indenken, waarbij alle elementen van het universum bekend zijn, althans kunnen zijn. Dit universum moet dus eindig zijn. Het aantal elementen kan nu zo groot zijn, dat een bewerking van deze gegevens dusdanige eisen aan mensen en materiaal stelt, dat een volledige bewerking niet gemakkelijk realiseerbaar zou zijn.

De enige oplossing voor de moeilijkheden in deze beide gevallen is: genoeg te nemen met een kleiner aantal elementen en met deze de diverse bewerkingen uit te voeren. Op deze wijze wordt men gedwongen steekproeven te nemen.

De wijze van steekproeven nemen moet aan twee voorwaarden voldoen. De eerste betreft de manier, waarop de elementen voortgebracht worden. Hierbij komen de meetfouten ter sprake. Deze mogen een bepaalde grens, die van geval tot geval verschillend is, niet overschrijden. Onder de verzamelnaam „meetfouten” verstaan we niet alleen de afleesfouten, maar ook alle andere fouten, die optreden, wanneer we het resultaat van de waarneming of meting beschouwen als element van het te onderzoeken universum. De afleesfouten kunnen we ook subjectieve fouten noemen, de overige de objectieve of instrumentele. Altijd zal het door mensenhanden gemaakte instrument onvolmaakt zijn, of het nu een dobbelsteen is of een thermometer.

De tweede voorwaarde behandelt de vraag, welke elementen in de steekproef opgenomen moeten worden.

Hoe aan deze eisen al of niet voldaan kan zijn, zullen we eerst demonstreren aan de hand van een eenvoudig voorbeeld: het universum der ideale kruis- of muntworpen.

Wanneer we enkele eigenschappen van dit universum experimenteel willen onderzoeken, moeten we met een muntstuk gaan werpen. De resultaten vormen een steekproef uit het universum der worpen met dit onvolmaakte muntstuk. We willen deze steekproef echter graag beschouwen als een steekproef uit het ideale universum. Dit mag echter alleen dan, wanneer het hier gebruikte muntstuk een slechts geringe afwijking van het ideale („unbiased”) muntstuk vertoont. De grootte van de maximaal toelaatbare afwijking hangt af van de te be-

schouwen eigenschappen van steekproef en universum en van het aantal elementen van de steekproef.

De tweede voorwaarde betreft de samenstelling van de steekproef. Deze hangt geheel af van het doel, waartoe men de steekproef wil nemen. Ligt het in de bedoeling speciale elementen te selecteren, dan kan dat natuurlijk gebeuren. Maar men moet dan niet verwachten, dat zo'n bijzondere steekproef algemene inlichtingen kan geven over het universum.¹⁾

In bijna alle gevallen zullen we te maken hebben met toevallige steekproeven (random samples). Elke steekproef zal in het vervolg als toevallig beschouwd worden, tenzij uitdrukkelijk anders vermeld.

Wanneer mogen we een steekproef toevallig noemen? Dat mag, wanneer de elementen van een steekproef op volkomen willekeurige wijze genomen zijn, dus wanneer elk element een even grote kans heeft getrokken te worden.

Uit het kruis- of munt-universum krijgt men een niet-toevallige steekproef door een aantal worpen op niet-willekeurige wijze te annuleren, bv. alle of een gedeelte van alle kruis-resultaten. Nu kan men niet meer zeggen, dat de elementen van de steekproef volkomen willekeurig genomen zijn. De steekproef is niet meer representatief voor het universum.

Men zou kunnen denken, dat elke steekproef toevallig is, indien men maar geen bijzondere dingen doet en op de gewone, normale wijze de steekproef neemt. In dit eenvoudige geval is dit inderdaad zo. De geofysica geeft echter voorbeelden, waarin slechts een diepgaand onderzoek het niet-toevallige karakter van de steekproef kan ontdekken.

Theoretisch zouden ook nog afwijkingen van de toevalligheid (randomness) kunnen ontstaan door steeds te werpen met dezelfde beginvoorwaarden, zelfde beginstand, zelfde aanvangssnelheid. De praktijk leert echter, dat de beginvoorwaarden niet voldoende reproduceerbaar zijn. Zelfs het meest methodische (kruis of munt) werpen geeft nooit anders dan toevallige resultaten. Het imiteren van het toeval gaat met grote moeilijkheden gepaard; een voorbeeld daarvan vindt men bij VISSER en VAN DER BIJL (1952).

¹⁾ Vergelijk deze tweede voorwaarde met Bok (1946, 111):

„De materiaalselectie moet ten opzichte van de te onderzoeken eigenschap zuiver toevallig tot stand komen.”

„De cardinale eis, die aan een statistisch onderzoek dient te worden gesteld, is dus, dat er geen correlatie mag bestaan tussen de te onderzoeken eigenschappen en de eigenschap, die heeft beslist over de begrenzing van het materiaal, waaraan de gegeven meet- of telgetallen ontleend zijn.”

En op p. 119: „..... de cardinale eis, dat de onderzochte steekproef uitsluitend begrensd mag zijn door factoren, die onafhankelijk variëren van de te onderzoeken factoren.”

In het eenvoudige kruis- of munt-geval vormt de onvolkomenheid van het materiaal dus de enige foutenbron.

Zoals reeds gezegd: in de geofysica wordt het ingewikkelder. Het nemen van steekproeven kan niet op volkomen willekeurige wijze geschieden. Steekproeven kunnen slechts uit bestaande waarnemingen gedaan worden, geofysische steekproeven dus uit het materiaal van de laatste n jaar. Hierbij is n slechts zelden groter dan 200.

Aan de geofysische onderzoeksmogelijkheden wordt door deze historische tekortkomingen een grote beperking opgelegd. Hierin en in de ingewikkelde onderlinge verbondenheid der geofysische grootheden vindt men de twee hoofdoorzaken van de betrekkelijk onbevredigende ontwikkelingsgang van de geofysica.

De geofysica is niet de enige tak van wetenschap, die deze feiten als verontschuldiging mag laten gelden: hetzelfde geldt voor alle wetenschappen, waarin experimenteren door de mens zo goed als geheel uitgesloten is.

Naast het relatief geringe aantal waarnemingen bestaan nog andere factoren, die de kwaliteit van de steekproeven nadelig kunnen beïnvloeden. Met de behandeling hiervan komen we aan de geofysische parallel van het hiervoor genoemde eenvoudige model (kruis- of munt-universum).

De eerste foutenbron vinden we weer in de categorie meetfouten (in de meest uitgebreide zin des woords): men noteert het resultaat van de aflezing van een instrument, men interpreteert het als element van het universum, dat we wensen te onderzoeken.

De meetfouten kunnen we splitsen in: 1) afleesfouten en 2) fouten in de opstelling van het instrument. Deze laatste noemt men ook wel fouten, veroorzaakt door de invloed van de omgeving, omdat de oorzaak van een verkeerde opstelling der instrumenten altijd een vreemde invloed moet zijn. Denk aan de grote thermometerhut in De Bilt! Deze in 1897 geconstrueerde hut was aan de onderkant open. In 1942 werd op 4 m afstand van de grote hut een stensonhut geplaatst. Deze laat van geen enkele zijde straling naar binnen. Reeds spoedig opgemerkte verschillen gaven aanleiding tot het trekken van de conclusie, dat straling, afkomstig van het aardoppervlak, de nauwkeurigheid van de temperatuurmeting in de grote hut nadelig beïnvloedde. Zo wees bv. de thermometer in de grote hut op 22 April 1942 om 14 uur $3,0^{\circ}\text{C}$ hoger aan dan de thermometer in de stensonhut. Oorzaak was de sterke verhitte van het aangrenzende grasveld.

Betreft het niet-instrumentele waarnemingen, dan heeft men alleen met schattingsfouten te doen. Maar deze fouten zullen uiteraard groter zijn dan de beide in de vorige alinea genoemde soorten, die op instrumentele waarnemingen betrekking hebben.

De tweede voorwaarde, betreffende de samenstelling van de steekproef, komt in de geofysica meestal hierop neer, dat universum en steekproef zoveel mogelijk gelijkvormig moeten zijn. De steekproef moet in de meeste gevallen een afschaduwing, een min of meer conforme afbeelding van het universum zijn. Daaruit volgt, dat men bij het bepalen van een jaartemperatuur niet een steekproef moet nemen, waarin de temperaturen van de zomermaanden overheersen. Bij het bepalen van een gemiddelde wintertemperatuur mag de steekproef niet samengesteld zijn uit de temperaturen van uitsluitend strenge winters.

Men moge dit triviale fouten vinden, BARTELS (1943) geeft gevallen aan, waarin het opsporen van gemaakte onjuistheden veel meer hersenenenergie kost. Zo bv. de gemiddelden van windsterkten en windrichtingen, die berekend zijn uit loodsballonwaarnemingen. De resultaten voor de hogere luchtlagen zijn hierbij in sterke mate gecorreleerd met mooi-weer-situaties.

In een tweede voorbeeld behandelt BARTELS de gemiddelde barometerstand in afhankelijkheid van de maanstand. Daar grote drukschommelingen de maansinvloed al te zeer verdoezelen, worden dagen met zulke grote drukveranderingen uitgeschakeld. Het gemiddelde over de rustige dagen vertoont dan een gladde convexe kromme, waarvan de hoogste waarde bij het bovenste culminatiepunt valt en ongeveer 0,2 mbar meer bedraagt dan de laagste waarde, die tijdens het passeren van het onderste culminatiepunt bereikt wordt.

Hoe moet deze kromme geïnterpreteerd worden?

Het ligt voor de hand als volgt te redeneren: is er geen verband tussen maanstand en luchtdruk, dan verwacht men op elk „maanuur” dezelfde gemiddelde luchtdrukwaarde, afgezien van toevallige schommelingen. Deze toevalsafwijkingen nemen af in grootte, wanneer het aantal waarnemingen toeneemt (de foutenvoortplantingswet of $\sqrt{n}$ -wet).

Het maan-luchtdruk-onderzoek gaf inderdaad toevalsafwijkingen te zien, die kleiner werden bij een toenemend aantal waarnemingen. Tenslotte ontstond echter geen horizontale rechte (wat onafhankelijkheid zou impliceren), maar een fraaie, symmetrische convexe kromme (als in fig. 17 f en f' op p. 149).

En toch is het onjuist, dit als een bewijs te zien van de invloed van de maan op de barometerstand!

De bijzondere vorm is alleen ontstaan door de speciale statistische behandeling. Dagen, waarop de barometer zich rustig gedraagt, zijn namelijk in het algemeen gekenmerkt door de aanwezigheid van een hoge luchtdrukwaarde. Een voorbijtrekkend hogedrukgebied geeft de luchtdrukcurve een convexe kromming. Bij het middelen blijft deze kromming bestaan. (Zie ook 3.5, waar dit krommingseffect uitvoeriger behandeld wordt).

BARTELS beweert, dat de publicaties, die dit maaneffect als reëel – beter gezegd: dit effect als maaneffect – beschouwen, een gezamenlijke omvang hebben van verscheidene honderden bladzijden!

1.1.1 VOORWAARDEN, WAARAAN EEN STEEKPROEF BEHOORT TE VOLDOEN

Door welke maatregelen valt te bereiken, dat de twee voorwaarden, waaraan een steekproef moet voldoen, zo goed mogelijk vervuld worden?

De eerste foutenbron, de meetfouten, kunnen we reduceren door de meetnauwkeurigheid op te voeren, dus door ongewenste invloeden zoveel mogelijk uit te schakelen.

De tweede voorwaarde eist allereerst het gebruik van gezond verstand.¹⁾ Bij ingewikkelde situaties is zelfs meer dan dat nodig. Zie het hierboven behandelde voorbeeld van de pseudo-invloed van de maanstand op de hoogte van de kwikkolom.

Is men er niet zeker van, dat de steekproef willekeurig is, dan doet men het beste gebruik te maken van tabellen van toevalscijfers (tables of random sampling numbers, vervaardigd door FISHER and YATES 1949, TIPPETT 1927, KENDALL and BABINGTON SMITH 1939, BARTELS 1948a, HAMAKER 1948a).

Zo'n tabel is een verzameling van getallen, waartussen geen enkele relatie bestaat. Elk getal is volkomen onafhankelijk van voorafgaande en volgende getallen. Deze tabellen met toevalscijfers dienen om het trekken van genummerde briefjes uit hoge hoeden, het werpen met dobbelstenen e.d. te vervangen. Het is nl. gebleken, dat dergelijke trekkingen dikwijls niet aan de eisen van volkomen toevalligheid voldoen. Bepaalde waarden zullen teveel, andere weer te weinig voorkomen en deze afwijkingen zijn zo groot, dat het in hoge mate onwaarschijnlijk is, dat alleen het toeval de oorzaak daarvan is geweest.

Uit de geofysische literatuur zijn mij slechts 4 gevallen bekend, waarin van de „ouderwetse” manier gebruik werd gemaakt. KLEINSCHMIDT (1932) gebruikte dobbelstenen, VISSER (1946, 15-7) trok 840 maal uit een zak met 99 nummers en WEISS (1944) schreef op 810 stukjes papier „rain” en op 810 andere „fair” en trachtte door trekking uit een doos een toevallige volgorde van mooie en regendagen te krijgen. ELKINS (1952) trok 361 maal uit een doos met 111 ballen.

Nu doet zich het opmerkelijke feit voor, dat deze vier onderzoekers resultaten bereikten, die ernstige twijfel deden rijzen over het zuiver toevallige karakter van hun methoden.

Dit is dus een reden te meer om steeds gebruik te maken van bestaande tabellen met toevalscijfers. Is het universum zodanig, dat elementen genummerd kunnen worden, dan kan men met behulp van een tabel van toevalscijfers die elementen uitkiezen, die tezamen een toevallige steekproef zullen vormen.

¹⁾ Men leze echter ook de discussie bij de voordracht van Prof. KOPPE (VAN DER BIJL 1951a).

Hebben we tot nu toe fouten behandeld, die min of meer geëlimineerd konden worden, er is een soort van fouten, die aan elke poging tot eliminatie hardnekkig weerstand biedt. Alleen met statistische methoden kan aan deze fouten paal en perk gesteld worden. Deze fouten vinden hun oorzaak in het feit, dat steekproeven altijd minder inlichtingen verschaffen dan universa.

Omdat dit zo buitengewoon belangrijk is, zullen we het voorgaande op een enigszins andere wijze kort weergeven:

Men heeft de beschikking over een aantal waarnemingen. Deze geven aanleiding tot het trekken van een voorlopige conclusie.

Nu wil men generaliseren: de conclusie wenst men geldig te verklaren voor een grotere groep waarnemingen. (De eerste verzameling elementen wordt steekproef genoemd, de tweede universum). Generaliseren is dikwijls gevaarlijk en ook hier moet men oppassen. Immers, de conclusie kan betrekking hebben op een eigenschap van de steekproef, die geheel te danken is aan de toevallige samenstelling van de steekproef, maar die volstrekt geen eigenschap van het universum behoeft te zijn.

Een overdreven voorbeeld: Juli 1919 was in De Bilt kouder dan October 1831. Hieruit volgt nu geenszins, dat de conclusie, die juist is voor deze bijzondere steekproef, ook geldigheid heeft voor het universum, bestaande uit de maandtemperaturen van Juli en October. Het is daarom nog lang niet zeker, dat in De Bilt de maand Juli in het algemeen kouder is dan de maand October.

Het is nu de mathematische statistiek, die in een zich steeds uitbreidend aantal gevallen berekenen kan, wanneer de door de steekproef gegeven aanwijzing te zwak is om de conclusie de gewenste werkingssfeer te geven.

Maar al te vaak – vooral in de geofysica – handelt men, alsof de gegeven aanwijzing steeds een reële oorzaak heeft.

Deze fout duikt ook in een andere vorm op: men berekent een aantal parameters van een steekproef en stelt de onbekende waarden van de overeenkomstige parameter van het universum precies gelijk aan de steekproefuitkomsten. De gemaakte onjuistheden kunnen we korthedshalve aangeven door te zeggen, 1) dat men doet, alsof er geen steekproeffouten bestaan of 2) dat men het steekproefeffect verwaarloost.

Dit en het selectieeffect zijn de oorzaken geweest van verschillende geofysische onjuistheden (Hfdst. 3). Ook het geringe succes van weersverwachtingen op lange termijn kan hierdoor verklaard worden.

Hiermee zijn we aan het eind gekomen van de bespreking van universum en steekproef en is de basis gelegd voor de behandeling en de omschrijving van het begrip overschrijdingskans.

1.2 VERSCHILLEN TUSSEN STEEKPROEF- EN UNIVERSUMWAARDEN

We gaan ons nu uitsluitend bezighouden met universa, waarvan de rechtstreekse behandeling te veel van de ons ten dienste staande middelen zou vergen. We moeten dus onze toevlucht nemen tot een hulpmiddel. Reeds toonden we aan, dat dit hulpmiddel bestaat uit het nemen van steekproeven.

Dit hulpmiddel zou ons niets baten, wanneer onze kennis en vermoedens aangaande het universum zich zouden beperken tot de waarnemingen van de elementen van de verrichte steekproef. Maar gelukkig komt dit nooit voor. Steeds is er wel een min of meer vaag vermoeden over het universum aanwezig. Deze vermoedens worden geconcretiseerd tot basisthesen (Nass 1948).

Het tegenovergestelde komt wel eens voor. Een tot in alle finesses uitgewerkte theorie weet dan precies te vertellen, uit welke elementen het universum opgebouwd zal zijn. De steekproef heeft nu een controlerende taak. Heeft de controle een positief resultaat, dan behoeft dit nog niet te betekenen, dat de theorie voor 100% juist is. Een eventueel bestaand verschil tussen theorie en werkelijkheid kan zo gering zijn, dat de mazen van het door de steekproef gespannen net te wijd waren om dat verschil aan het daglicht te brengen. De conclusie mag dan luiden, dat de theorie niet in strijd is met het door de steekproef verschaft materiaal.

Leidt de controle tot een negatief resultaat, dan is statistisch zeker,¹⁾ dat de theorie niet strookt met de werkelijkheid en dus verworpen moet worden – althans gedeeltelijk.

Op welke wijze vindt die controle plaats?

We gaan als volgt te werk: Uit het hypothetische, eventueel theoretisch vastgelegde universum nemen we alle mogelijke steekproeven met hetzelfde aantal elementen als de door de waarnemingen verschaft steekproef. Deze laatste *moet* tot de verzameling steekproeven behoren, want anders zou het hypothetische universum een element missen, dat in de werkelijke steekproef wel aanwezig is.²⁾ Hoe moeten we nu beoordelen, of de steekproef afkomstig kan zijn uit dat hypothetische universum? Het ligt voor de hand te vragen naar de waarschijnlijkheid, waarmee een steekproef van de waargenomen samenstelling voorkomt. Bestaat het universum geheel uit onderling ongelijke elementen, dan heeft elke steekproef van n elementen dezelfde kans om getrokken te worden. Telt het universum N elementen, dan is die kans N^{-n} . Heeft het universum gelijke elementen, dan wordt de kansberekening ingewikkeld of zelfs prohibitief.

¹⁾ Statistische zekerheid = zekerheid, afgezien van een kleine waarschijnlijkheid, gelijk aan de overschrijdingskans. (1.2.1).

²⁾ Hier ontmoeten we dus de eerste voorwaarde, waaraan het hypothetische universum moet voldoen. (De tweede voorwaarde bevindt zich op p. 16, regel 10).

Door het vergrote aantal mogelijkheden wordt de absolute waarde van de kans gering en men heeft geen vergelijkingsmateriaal om te constateren, of die waarde ook relatief gering is. Daartoe zou men immers de waarschijnlijkheden van het voorkomen van alle steekproeven moeten berekenen.

Omdat we langs deze weg niet verder kunnen komen, zijn we verplicht een andere oplossing te zoeken. We gaan nu niet de waarschijnlijkheid van de samenstelling van de steekproef bepalen, maar de waarschijnlijkheid van een of andere parameter. Natuurlijk komen de eenvoudigste: gemiddelde (1.4), standaarddeviatie (1.6) daarvoor het eerst in aanmerking. Deze zijn voor de steekproef gemakkelijk te berekenen. Dit moet ook gelden voor het universum,¹⁾ anders kunnen deze parameters van steekproef en universum niet met elkaar vergeleken worden.

Niet altijd luidt de vraag: kan deze steekproef uit een zeker hypothetisch universum getrokken zijn? Vaak is onze aandacht uitsluitend gericht op een bepaalde parameter, zo bv. als een experimentele meetnauwkeurigheid onderzocht moet worden. In zo'n geval is de standaarddeviatie de aangewezen parameter. In andere gevallen behoeft alleen het gemiddelde vergeleken te worden met een theoretische waarde. Zo krijgen alle vraagstukken eenzelfde methode van behandeling: het toetsen van parameters.

Uit 1.1 weten we reeds, dat steekproef- en universumwaarde in het algemeen niet zullen samenvallen. We gaan nu onderzoeken, of het steekproefeffect deze afwijking kan verklaren, nl. als toevalseffect. Zo niet, dan moet er een andere oorzaak werkzaam geweest zijn. Is aan de twee voorwaarden voor het steekproefnemen voldaan, dan kan de oorzaak alleen liggen in het aannemen van een onjuist hypothetisch universum. Minstens één van de hypothesen, die tot de opstelling van het universum hebben geleid, moet dan verworpen worden.

Onze aandacht gaat zich nu richten op de grens tussen de twee alternatieven. Aan de éne kant van de grens *kunnen* alle afwijkingen aan het steekproefeffect toegeschreven worden, m.a.w. een statisticus is niet in staat te beslissen, of de gevonden afwijkingen op het steekproefeffect berusten of op een geheel ander, een „reëel” effect; aan de andere kant van die grens is het onmogelijk, dat de afwijkingen voor de volle 100% door het steekproefeffect veroorzaakt zijn. De verschillen tussen steekproef- en universumwaarde moeten dan op reële, niet-toevallige gronden berusten. (Men zegt ook dikwijls: op fysische gronden, wat m.i. minder juist is. Zie VAN DER BIJL 1950).

Deze grens kunnen we op twee manieren vinden: experimenteel en theoretisch. De experimentele weg leidt – via het nemen van „tallose” steekproeven uit het

¹⁾ Zie noot 2 op p. 15.

hypothetische universum en het berekenen van de waarde van de desbetreffende parameter voor elke steekproef – tot het opmaken van een frequentieverdeling¹⁾ van deze waarden. Soms lukt het deze verdeling te identificeren met een bekende mathematische functie (zoals STUDENT in 1908 deed: het klassieke voorbeeld). Geheel veilig is deze weg niet, want de nauwkeurigheid is afhankelijk van het aantal steekproeven, dat genomen is om de experimentele frequentieverdeling te bepalen.

Wel zeker is men van het eindresultaat, wanneer men van de tweede manier gebruik kan maken. In dat geval is het mogelijk de verdeling van de parameter in steekproeven geheel langs theoretische weg af te leiden. Men kan dan onmiddellijk een indruk krijgen, op welke wijze de steekproefwaarde afwijkt van de universumwaarde. Het aangeven van de absolute waarde van de afwijking geeft weinig houvast bij gebrek aan vergelijkingsmateriaal; een maat voor de relatieve waarde van de afwijking krijgt men pas door het invoeren van het begrip overschrijdingskans.

Doch alvorens dit in 1.2.1 te definiëren moeten we eerst nog die universa behandelen, waarvoor geen volledige theorieën aanwezig zijn. Op grond van diverse overwegingen, zoals kwalitatieve theoretische beredeneringen en losse aanwijzingen, door een steekproef gegeven en met behulp van een binnen redelijke perken gehouden fantasie zal het voorstellingsvermogen van de onderzoeker het universum althans gedeeltelijk kunnen vastleggen.

Heel vaak wordt als basisthese (NASS 1948) genomen de bewering, dat de elementen in het universum volgens GAUSS verdeeld zijn (1.3.1). Twee eigenschappen van het universum blijven dan nog onbepaald: gemiddelde (1.4) en standaarddeviatie (1.6). Hebben we voor de grootte van beide parameters geen enkele suggestie, dan leert de statistische theorie, op welke wijze de „beste schatting” voor deze parameters uit de steekproefgegevens berekend kan worden.

Het begrip „beste schatting” zal in deze verhandeling niet gedefiniëerd worden; de woorden geven reeds een vage indruk, van wat zij pogen weer te geven.

1.2.1 HET BEGRIIP OVERSCHRIJDINGSKANS

Wanneer nu de oorspronkelijke basisthesen uitgebreid worden tot een volledige theorie, kunnen we de parameters van die nieuwe theorie onmiddellijk toetsen aan de reeds bewerkte steekproefuitkomsten en wel met behulp van de overschrijdingskans. Om dit begrip te verduidelijken moeten we de frequentiefuncties van de steekproefparameters in twee groepen verdelen. De eerste groep omvat die functies, waarbij de universumwaarde in de nabijheid van het centrum

¹⁾ Een frequentieverdeling, ook wel waarschijnlijkheidsverdeling, frequentiefunctie of kansdichtheid geheten, is een functie $y = f(x)$, waarbij y de waarschijnlijkheid van het optreden van de gebeurtenis x is.

van de functie ligt. De tweede groep wordt gevormd door de functies, die

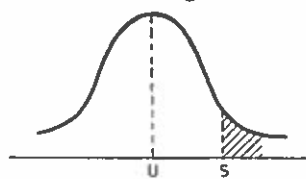


fig. 1

begrensd worden door de universumwaarde. Zij $f(x)$ de frequentiefunctie, U en S resp. de universum- en steekproefwaarde van de parameter, dan wordt in de eerste groep de overschrijdingskans P gegeven door de integraal $\int_S^{\infty} f(x) dx$. Dit

is gelijk aan de verhouding van gearceerd en totaal oppervlak (fig. 1).

We zien, dat P de kans is op het voorkomen van die afwijkingen, die een kleinere of even grote waarschijnlijkheid en bovendien hetzelfde teken bezitten als de uit deze steekproef berekende afwijking. De laatste eis – hetzelfde teken – maakt het duidelijk, waarom we dit de enkelzijdige overschrijdingskans noemen.

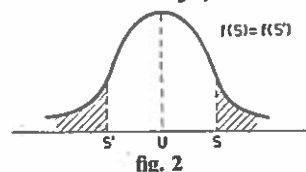


fig. 2

In de eerste groep kunnen zich ook negatieve afwijkingen voordoen. Nemen we deze eveneens in onze beschouwing op, m.a.w. beschouwen we de afwijkingen in absolute zin, dan moeten we spreken van een dubbelzijdige overschrijdingskans.

Deze wordt berekend uit $\int_{-\infty}^{S'} f(x) dx + \int_S^{\infty} f(x) dx$ en is dus gelijk aan de som der gearceerde oppervlakken gedeeld door het totale oppervlak (fig. 2). In de tweede groep hebben we alleen te maken met enkelzijdige overschrijdingskansen en daarom



fig. 3

kan het adjectief hier verdwijnen. $P = \int_S^{\infty} f(x) dx =$ gearceerd gedeeld door totaal oppervlak (fig. 3).

Ligt U aan het rechteruiteinde van de waarschijnlijkheidsverdeling, dan kan P op analoge wijze gedefiniëerd worden. Hetzelfde geldt ook voor de eerste groep, wanneer $S < U$.

1.2.2 DE BEGRIPPEN SIGNIFICANTIE EN BETROUWBAARHEIDSDREMPEL

Rest nu nog het onderzoek naar de ligging van de betrouwbaarheidsdrempel. D.w.z. de vraag wordt gesteld: hoe klein moet de overschrijdingskans minstens zijn om te kunnen spreken van een reëel of significant verschil tussen steekproef- en universumwaarde? Vrij algemeen wordt de grens bij 5% gelegd. Is de overschrijdingskans $P < 5\%$, dan moet aangenomen worden, dat de gebruikte theorie in strijd is met de werkelijkheid. Immers: nemen we meer dan 20 onderling onafhankelijke en even grote steekproeven uit het hypothetische universum, dan zou in dit geval ($P < 5\%$) slechts één steekproef een grotere afwijking vertonen dan „onze” steekproef (nl. de in werkelijkheid uitgevoerde steekproef). Men oordeelt het dan als wel heel toevallig, dat de eerste de beste steekproef, die

genomen werd, juist die ene met zulke grote, weinig waarschijnlijke afwijkingen is. Is de overschrijdingskans $P > 5\%$, dan bevindt zich onder minder dan 20 uit eenzelfde universum genomen steekproeven een steekproef met even grote of nog grotere afwijkingen. Nu oordeelt men, dat er voldoende waarschijnlijkheid is om aan te nemen, dat het gemeten verschil geheel aan het steekproefeffect te danken is, m.a.w. het verschil behoeft niet reëel of niet significant te zijn.¹⁾

Hiermee is aangetoond, dat een $P < 5\%$ tot de conclusie moet leiden, dat er slechts een geringe waarschijnlijkheid bestaat, dat de steekproef uit het hypothetische universum afkomstig is, m.a.w. dat er een grote waarschijnlijkheid is, dat de hypothese onjuist is.

Nu zal blijken, dat ook hoge overschrijdingskansen bezwaren met zich meebrengen. Immers, de bewering, dat het onwaarschijnlijk is, dat de eerste de beste steekproef heel geringe afwijkingen vertoont, is even plausibel te maken als de zoëven geuite bewering, dat het onwaarschijnlijk is, dat juist de door ons gekozen steekproef zulke grote afwijkingen bezit. De afwijkingen, die te wijten zijn aan het steekproefeffect, schommelen om een zekere gemiddelde waarde. Hoe groter *en* hoe kleiner de afwijkingen, des te geringer is de waarschijnlijkheid, waarmee die afwijkingen zullen optreden. Wanneer nu $P > 95\%$ is, liggen steekproef- en universumwaarde zo dicht bij elkaar, dat het lijkt, alsof er geen steekproefeffect werkzaam geweest is. Aangezien dit effect nu eenmaal altijd aanwezig is, moeten we de hypothese, met behulp waarvan het universum bekend was geworden, afkeuren. De overeenstemming tussen theorie en werkelijkheid is dan al te nauwkeurig geweest.

1.2.3 SELECTIEEFFECT

De lezer zal zo langzamerhand wel vertrouwd geraakt zijn met de betekenis van de drempelwaarde van 5% . Toch moet er nog enige aandacht aan besteed worden. Want feitelijk gelden de tot nu toe gemaakte opmerkingen alleen voor het geval, waarin één en slechts één onderzoeker één en slechts één steekproef te onderzoeken heeft.

De situatie wordt anders, wanneer verscheidene – laten we zeggen a – personen elk met het onderzoek van verscheidene – b – steekproeven belast zijn. Laat bekend zijn, dat alle steekproeven uit hetzelfde universum afkomstig zijn. Voor $a \cdot b > 20$ moet – gemiddeld genomen – minstens één $P < 5\%$ zijn. Gaat een statisticus op formele wijze volgens het aangegeven recept te werk, dan zou deze statisticus – wanneer hij niet wist, dat de steekproef met een $P < 5\%$ tot het uni-

¹⁾ Bij ingewikkelde statistische onderzoeken maken we verschil tussen de begrippen 'reëel' en 'significant'. (Zie 2.2.2).

versum behoorde – moeten ontkennen, dat deze steekproef uit het universum genomen was. Dit verschijnsel treedt op, zodra men moet beslissen over de vraag, of *enkele* steekproeven al of niet deel uit maken van een bepaald universum. Het maakt het statistisch onderzoek nog gecompliceerder dan het al is.

Het is jammer, dat toegegeven moet worden, dat men in de geofysica de genoemde formele weg wel eens volgt. Men vergeet dan, dat vergroting van het aantal steekproeven de kans op een steekproef met grote afwijkingen doet toenemen. Kiest men uit de parameterwaarden van een aantal steekproeven de grootste en heeft deze een $P < 5\%$, dan mag niet onmiddellijk beweerd worden, dat deze significant afwijkt van de universumwaarde. Men moet rekening houden met het selectieeffect, zoals we dit kortweg kunnen noemen. In 3.2 gaan we dieper op dit onderwerp in. Dan is er ook gelegenheid te wijzen op de vele manieren, waarop het selectieeffect geëlimineerd kan worden.

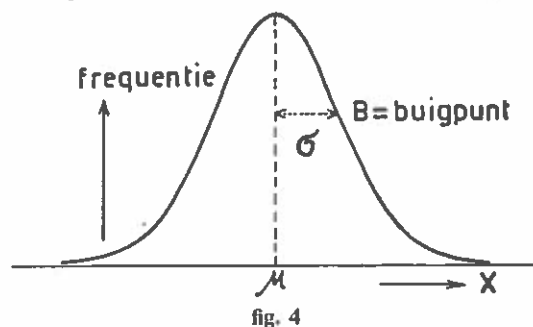
1.3 ENKELE VEEL VOORKOMENDE VERDELINGEN

1.3.1 DE VERDELING VAN GAUSS (In andere werken ook wel genoemd de normale verdeling. Zie noot op p. 23)

De elementen van een universum heten gaussisch verdeeld, wanneer de frequentiefunctie aangegeven kan worden door de formule

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

De onafhankelijk veranderlijke grootte x is de uitkomst van de meting, vericht aan elk element van universum of steekproef. Deze symmetrische klok-vormige kromme heeft een maximum voor $x = \mu$.



μ is het rekenkundig gemiddelde (1.4) van alle elementen en wordt ook wel „centrum” genoemd. σ is de standaarddeviatie (1.6). In de klimatologie treedt de gaussverdeling veelvuldig op, althans bij benadering (KÖPPEN-GEIGER 1936). Deze is vrij goed bij de temperatuur (zie bv. BILANCINI 1948)

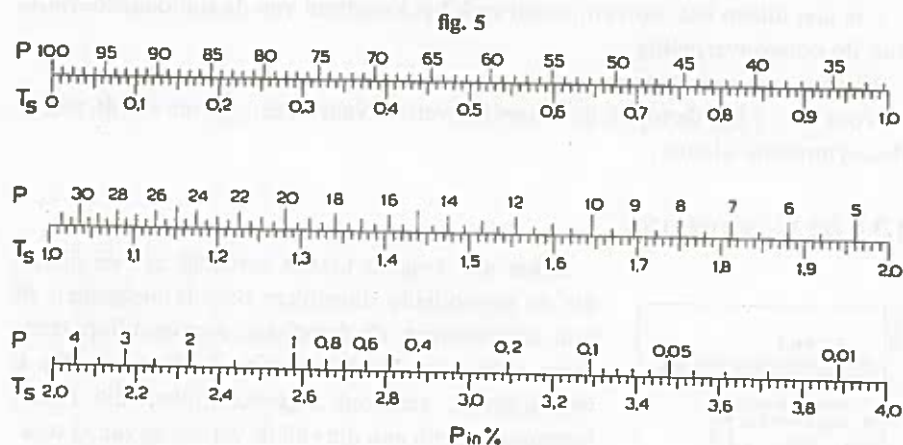
en bij de luchtdruk (PEARSON and LEE 1897, hoewel zij het zelf nog „a marked exception to this rule” noemen). Minder goed kunnen de neerslaghoeveelheden aan een gaussverdeling aangepast worden. Dit kan verholpen worden door de

frequentie van de regensommen uit te zetten tegen de logaritmie van die regenhoeveelheden en niet tegen de hoeveelheden zelf (LEVERT 1948).

Geringe afwijkingen van de verdeling van GAUSS zullen geen nadelige invloed op onze onderzoeken hebben (COCHRAN 1947). In het vervolg mogen we dan ook aannemen, dat de meeste der door ons bestudeerde elementen volgens GAUSS verdeeld zijn (VISSER 1946, 12). Behalve aan het veelvuldige optreden in de natuur heeft de verdeling van GAUSS de eerste plaats onder de frequentiefuncties in de statistica ook te danken aan zijn eenvoudige meetkundige eigenschappen. De grootte van de overschrijdskansen hangt nl. alleen af van de verhouding van de afstand tot het centrum en de standaarddeviatie. Deze verhouding noemen we de relatieve afwijking T . De dubbelzijdige overschrijdskans

$$\text{is } P = \frac{2}{\sigma \sqrt{2\pi}} \int_{\frac{x_s - \mu}{\sigma}}^{\infty} e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \sqrt{\frac{2}{\pi}} \int_{T_s}^{\infty} e^{-\frac{T^2}{2}} dT \quad \text{met } T_s = \frac{x_s - \mu}{\sigma}.$$

Het verband tussen T_s en P wordt gedemonstreerd door het volgende diagram (fig. 5).



1.3.2 DE VERDELING VAN BERNOULLI (Ook genoemd binomiale verdeling)

Deze verdeling heeft tot frequentiefunctie $f(x) = \binom{n}{x} p^x q^{n-x}$, waarbij $q = 1-p$.

Wij beschouwen een universum met N elementen, pN daarvan bezitten een zekere eigenschap (A), qN niet. Trekken we een steekproef van n elementen, dan kan het aantal elementen, dat die eigenschap bezit, uiteenlopen van 0 tot n .

De kans, dat de steekproef met grootte n precies x elementen bevat met de genoemde eigenschap A , wordt weergegeven door de bovenstaande frequentiefunctie.

Het centrum van de verdeling van BERNOULLI is np , de standaarddeviatie $\sqrt{npq}$.

Dit betekent, dat de aantallen elementen met eigenschap A , die zich in steekproeven met lengte n bevinden, zullen schommelen om de waarde np . De standaarddeviatie is een maat voor de grootte van de schommeling.

De verdeling van BERNOULLI komt niet veel voor in de klimatologie. We zullen in 2.3.2 als voorbeeld de onweersdagen in Januari behandelen.

Deze verdeling kan vaak met voordeel benaderd en vervangen worden door de gauss- en in andere gevallen door de poissonverdeling. Dit verklaart dan ook gedeeltelijk het betrekkelijk geringe optreden van de bernoulliverdeling.

Wordt n steeds groter, dan is de gaussverdeling de limiet. Wanneer p en q in de buurt van $\frac{1}{2}$ liggen, is die benadering bijzonder goed. Dit komt, omdat de symmetrie (een noodzakelijke eis) in die gevallen steeds aanwezig is.

Wordt p steeds kleiner met $np = \text{constant}$, dan is de poissonverdeling de limiet.

1.3.3 DE VERDELING VAN POISSON

De frequentiefunctie van deze verdeling is $f(x) = e^{-\mu} \frac{\mu^x}{x!}$, waarbij x alleen de gehele waarden $0, 1, 2, \dots$ kan aannemen.

μ is niet alleen het centrum, maar ook het kwadraat van de standaarddeviatie van de poissonverdeling.

Voor $\mu < 1$ is $f(x)$ maximaal bij $x = 0$.

Voor $\mu > 1$ ligt de top in de omgeving van μ . Naarmate μ groter wordt, wordt de asymmetrie kleiner.

1.3.4 DE U-VERDELING

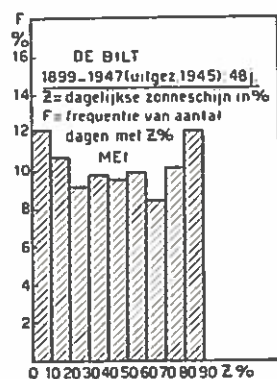


fig. 6

Zeker niet volgens GAUSS verdeeld zijn de cijfers, die de gemiddelde dagelijkse bewolkingsgraden of hun complement, de dagelijkse zonneshijnpercentages aangeven. De U-vormige kromme, in fig. 6 weergegeven, vertoont eigenschappen, die totaal tegengesteld zijn aan die van de verdeling van GAUSS. Immers, de U-kromme heeft 2 maxima en 1 minimum, de gaussverdeling 1 maximum en 2 minima. De eerste heeft een minimum in het centrum, de laatste een maximum.

De bewolkingscijfers leveren in vele statistische leerboeken het enige voorbeeld van U-krommen. Wij geven hier een voorbeeld van een frequentieverdeling van dagelijkse zonneshijnpercentages.

1.4 CENTRUM EN GEMIDDELDE (universum- en steekproefgemiddelde)

In het volgens GAUSS verdeelde universum is de waarschijnlijkste waarde gelijk aan het centrum. De beste schatting voor die waarde kan men uit het steekproefmateriaal bepalen door het rekenkundig gemiddelde van de steekproefelementen te berekenen. Van deze stelling wordt – zij het onbewust – reeds meer dan twee eeuwen lang gebruik gemaakt in de klassieke klimatologie. Het rekenkundig gemiddelde heeft daarin zo'n overheersende functie gekregen, dat men het dikwijls met de naam „normaal” betitelt.¹⁾ Helaas wordt al te vaak uit het oog verloren, dat ook het gemiddelde aan steekproeffouten onderhevig is. Een klimatologisch voorbeeld: de gemiddelde jaartemperatuur te De Bilt van 1901–1920 zal niet wiskundig gelijk zijn aan de gemiddelde jaartemperatuur te De Bilt van 1921–1940. Hebben de het weer in De Bilt bepalende factoren geen wijzigingen ondergaan – m.a.w. is er in De Bilt in het tijdvak 1901–1940 geen klimaatverandering opgetreden – dan moet het verschil der beide gemiddelde temperaturen aan zekere statistische wetten gehoorzamen. Dit wordt uitvoeriger besproken in 1.7.3.

1.5 HET AANTAL VRIJHEIDSGRADEN

Van sommige statistische parameters zullen we zeggen, dat ze vrijheidsgraden bezitten. Dit begrip is moeilijk exact te definiëren, maar een min of meer vage omschrijving laat zich wel geven. Bovendien zal bij elke parameter aangegeven worden, hoe groot ν , het aantal vrijheidsgraden, is.

Steeds geldt $\nu = N - a - b$. Hierin is:

N = het aantal elementen, waaruit de statistische parameter berekend is. Deze elementen kunnen zijn: waarnemingen, herleide waarnemingen, frequenties e.d.

a = het aantal onafhankelijke, lineaire voorwaarden, waaraan de elementen moeten voldoen.

b = het aantal parameters in de frequentiefunctie, dat uit het steekproefmateriaal geschat moet worden.

Voor $b = 0$ is ν gelijk aan het aantal elementen, waaruit de statistische parameter berekend is, verminderd met het aantal onafhankelijke, lineaire voorwaarden, waaraan de waarnemingen moeten voldoen. Zijn er N elementen geweest en $(N - \nu)$ voorwaarden, dan kan men ν elementen door geheel willekeurige vervangen. De overige $(N - \nu)$ elementen zijn door de $(N - \nu)$ beperkingen geheel

¹⁾ Dit gebruik van het woord 'normaal' is de reden geweest, waarom we de in 1.3.1 behandelde verdeling niet de vrijwel algemeen aanvaarde benaming (de normale verdeling) gegeven hebben.

vastgelegd. Voorbeeld: De afwijkingen van N waarnemingen t.o.v. hun rekenkundig gemiddelde hebben $(N-1)$ vrijheidsgraden. Nu kunnen $(N-1)$ afwijkingen willekeurig gekozen worden, maar de laatste afwijking is bepaald door de voorwaarde, dat de algebraïsche som van de afwijkingen 0 moet zijn.

Het geval: $b \neq 0$ komt voornamelijk voor bij G^2 -toetsen (1.8). Daar zullen we waargenomen frequenties vergelijken met theoretische frequenties. N is nu niet meer gelijk aan het totale aantal waarnemingen, maar gelijk aan het aantal cellen, waarover alle waarnemingen verdeeld zijn.

Is de frequentiefunctie geheel door de theorie bepaald, dan is $b = 0$.

Zijn van de frequentiefunctie nog enkele parameters onbekend (zij bv. alleen bekend, dat de verdeling gaussisch is), dan kunnen we toch nog de G^2 -toets toepassen door voor de onbekende parameters waarden te substitueren, die uit het steekproefmateriaal geschat worden. Voor b moet dan het aantal geschatte parameters genomen worden. Dus bij de gaussverdeling is $b = 2$, bij de poissonverdeling $b = 1$.

Dat het verminderen van het aantal vrijheidsgraden met het aantal geschatte parameters zin heeft, blijkt uit het volgende: Neem een frequentieverdeling aan met N hokjes of intervallen. Door de N punten van de grafische voorstelling van deze frequentieverdeling leggen we met behulp van de theorie der kleinste kwadraten een kromme van de $(N-2)^{\text{de}}$ graad. Het steekproefmateriaal moet dus schattingen van $(N-2)$ parameters leveren ($b = N-2$). Het totale aantal waarnemingen moet bij de waargenomen en theoretische verdeling gelijk zijn. Dus $a = 1$ en $v = N - a - b = N - 1 - (N-2) = 1$.

De aanpassing van de $(N-2)^{\text{de}}$ graadskromme aan de gevonden frequentieverdeling zal vrij goed zijn. De G^2 -som (zie 1.8) is dan gering. Het aantal vrijheidsgraden is 1. De overschrijdingskans heeft in dergelijke gevallen een vrij redelijke waarde.

Anders wordt het, wanneer we b niet van het aantal vrijheidsgraden hadden afgetrokken. Dan zou $v = N-1$ zijn. Vooral bij grote N zou een kleine G^2 -som een P geven, die heel dicht bij de 100% gelegen is. We mochten dan de kleine verschillen tussen waargenomen en theoretische frequenties niet als toevallig beschouwen (zie slot 1.2.2).

1.6 STANDDAARDDEVIATIE EN VARIANTIE

Leert het rekenkundig gemiddelde of wel de „normaal” ons iets over het gemiddelde niveau, waaromheen de waarnemingen schommelen, de volgende statistische parameter dient ons in te lichten over de grootte van die schommelingen, dus over de grootte van de spreiding. De parameters, die daarvoor in aanmerking komen, zullen we in volgorde van eenvoudigheid behandelen.

De eerste is het verschil tussen minimale en maximale waarde, de zgn. spreidingsbreedte. De steekproefwaarde voor deze grootheid geeft weinig inlichtingen over de spreiding in het universum, vooral wanneer de steekproef meer dan tien elementen telt. Meer geschikt is de gemiddelde absolute afwijking ten opzichte van het rekenkundig gemiddelde. Een nadeel van deze grootheid is, dat weinig bekend is over zijn statistische eigenschappen.

Reeds in 1897 schreven PEARSON and LEE (p. 443):

„The only method hitherto used by meteorologists to express briefly the variability of the atmospheric pressure is, so far as we are aware, the statement of the maximum and minimum heights reached during any given period. It gives no real impression whatever of the manner in which the bulk of the variation is distributed, yet, for most climatological purposes this is precisely what we require . . . One satisfactory and useful measure . . . is σ , the standard deviation.”

Inderdaad verenigt de standaarddeviatie de meeste voordelen in zich.

Daarom is deze *de* aangewezen maat voor de spreiding. De standaarddeviatie in het universum is gelijk aan de wortel van het rekenkundig gemiddelde der kwadratische afwijkingen. We duiden deze grootheid met de letter σ aan. Beperkt onze kennis zich tot de gegevens van een steekproef, dan leert de statistica, dat hieruit *de beste schatting voor de standaarddeviatie van de enkelvoudige waarneming* berekend kan worden met behulp van de formule

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2}$$

x_i ($i = 1, 2, \dots, n$) is het resultaat van een meting, uitgevoerd aan een steekproefelement. n is het aantal elementen in de steekproef. m is het gemiddelde van de steekproefmetingen, dus $m = \frac{1}{n} \sum_{i=1}^n x_i$. Het aantal vrijheidsgraden is kenmerkend $(n-1)$, omdat de afwijking van het laatste steekproefelement door de eerste $(n-1)$ volkomen vastgelegd is. Deze beste schatting s zal voortaan de naam steekproefstandaarddeviatie dragen en heeft dus altijd betrekking op steekproefmateriaal. De overeenkomstige parameter in het universum zullen we standaarddeviatie blijven noemen, slechts voorzien van het voorvoegsel universum, wanneer misverstanden te vrezen zijn. s^2 heeft een afzonderlijke naam gekregen en wel: de steekproefvariantie. Het universumaequivalent is σ^2 , dat de naam draagt van universumvariantie. De naam variantie heeft zich ingeburgerd door de enorme ontwikkeling van de variantieanalyse (1.10).

De volgens de aangegeven wijze berekende steekproefstandaarddeviatie zal in het algemeen niet samenvallen met zijn universumwaarde, omdat het steekproef-effect ook hier zijn invloed doet gelden.

1.6.1 DE FREQUENTIEFUNCTIE VAN DE STANDAARDDEVIATIE VAN DE ENKELVOUDIGE WAARNEMING

Deze is: $f(s) = 2^{-\frac{n-3}{2}} \left\{ \Gamma\left(\frac{n-1}{2}\right) \right\}^{-1} (n-1)^{-\frac{n-1}{2}} \left(\frac{s}{\sigma}\right)^{n-2} \frac{1}{\sigma} e^{-\frac{1}{2}(n-1)\left(\frac{s}{\sigma}\right)^2}$

Met behulp hiervan kunnen de overschrijdingskansen als volgt bepaald worden.

Vervang $(n-1)\left(\frac{s}{\sigma}\right)^2$ door G^2 . Dan gaat $f(s)$ over in

$$f(G^2) = 2^{-\frac{n-1}{2}} \left\{ \Gamma\left(\frac{n-1}{2}\right) \right\}^{-1} e^{-\frac{1}{2}G^2} (G^2)^{\frac{n-3}{2}}$$

Het aantal vrijheidsgraden is $v = n-1$. Voor zekere n , s en σ kan G^2 berekend worden. Een G^2 -tabel levert ons de gevraagde overschrijdingskansen (zie 1.8).

1.6.2 DE FREQUENTIEFUNCTIE VAN DE STANDAARDDEVIATIE IN GROTE STEEKPROEVEN

Voor $n \rightarrow \infty$ nadert de verdeling van s tot de gaussverdeling. Deze heeft dan een standaarddeviatie $= \frac{\sigma}{\sqrt{2n-2}}$ en het centrum in $x = \sigma$. Het bepalen van overschrijdingskansen is in dit geval dus veel eenvoudiger.

1.7 DE FREQUENTIEFUNCTIE VAN DE STEEKPROEFGEMIDDELDEN

Van een gaussisch verdeeld universum met centrum μ en standaarddeviatie σ gehoorzamen de steekproefgemiddelden eveneens een gaussverdeling. Hun centrum ligt in μ , maar hun standaarddeviatie, dus de standaarddeviatie van het gemiddelde is $\frac{\sigma}{\sqrt{n}}$. Moet de overschrijdingskans van een zeker steekproefgemiddelde m bepaald worden, dan zal in het algemeen eerst een schatting van μ en σ verkregen moeten worden.

1.7.1 DE FREQUENTIEFUNCTIE VAN DE STEEKPROEFGEMIDDELDEN IN GROTE STEEKPROEVEN

In grote steekproeven kan σ met een voor ons doel voldoende nauwkeurigheid aan de steekproefstandaarddeviatie gelijkgesteld worden. Uit het quotiënt van $m-\mu$ en $\sigma/\sqrt{n}$ wordt de dubbelzijdige overschrijdingskansen P met behulp van het diagram in fig. 5 van p. 21 bepaald.

1.7.2 DE FREQUENTIEFUNCTIE VAN DE STEEKPROEFGEMIDDELDEN IN KLEINE STEEKPROEVEN

In kleine steekproeven kan het verschil tussen steekproef- en universumstandaarddeviatie door het steekproefeffect zo grote afmetingen aannemen, dat P een te grote fout vertoont, die op geen enkele wijze gecorrigeerd kan worden.

STUDENT loste deze moeilijkheden op. Zijn methode is aanvankelijk precies dezelfde als die van 1.7.1. Het quotiënt $\frac{m-\mu}{s/\sqrt{n}}$ wordt bepaald, maar nu niet gelijk gesteld aan $\frac{m-\mu}{\sigma/\sqrt{n}}$. STUDENT's quotiënt wordt aangeduid met de letter t . Het mag als STUDENT's grootste verdienste gerekend worden, dat hij erin slaagde de frequentiefunctie van t te vinden.

$$\text{Deze is: } f(t) = \frac{1}{(n-1)^{1/2} B\left(\frac{n-1}{2}, \frac{1}{2}\right) \left(1 + \frac{t^2}{n-1}\right)^{n/2}}$$

Het aantal vrijheidsgraden is ook hier één minder dan het aantal elementen in de steekproef, dus $\nu = n-1$. De aansluiting met grote steekproeven blijkt uit de limietovergang van $n \rightarrow \infty$. Zoals men gemakkelijk kan inzien, gaat STUDENT's

t -verdeling dan over in $ce^{-\frac{t^2}{2}}$ (met $c = \text{constante}$). Hierin mag t wel gelijk aan $\frac{m-\mu}{\sigma/\sqrt{n}}$ gesteld worden. t is dan de afstand tot het centrum, uitgedrukt in de standaarddeviatie van het gemiddelde en kan dus vergeleken worden met de relatieve afwijking T uit het slot van 1.3.1.

De overschrijdingskansen kunnen door integratie van de frequentiefunctie van t berekend worden. Voor bepaalde waarden van t en ν is dit reeds geschied en de resultaten zijn in tabellen en grafieken vastgelegd.

Het onderzoek naar een significant verschil tussen steekproef- en universum-gemiddelde is niet het enige gebied, waarop de t -toets van STUDENT toegepast kan worden.

1.7.3 SIGNIFICANTIE VAN HET VERSCHIL TUSSEN TWEE STEEKPROEFGEMIDDELDEN (STUDENT's t -toets)

Ook een eventueel verschil tussen twee steekproefgemiddelden kan met deze t -verdeling vastgesteld worden. Laat x_i , n_1 , m_1 en s_1 , resp. een element, het aantal elementen, het gemiddelde en de standaarddeviatie van de eerste en y_j , n_2 , m_2 en s_2 de overeenkomstige parameters van de tweede steekproef voorstellen, dan is

$$(n_1 - 1) s_1^2 = \sum_{i=1}^{n_1} (x_i - m_1)^2 \quad \text{en} \quad (n_2 - 1) s_2^2 = \sum_{j=1}^{n_2} (y_j - m_2)^2 \quad (\text{zie 1.6})$$

We gaan uit van de nulhypothese, dat beide steekproeven afkomstig zijn uit één en hetzelfde gaussisch verdeelde universum met centrum μ en standaarddeviatie σ (μ en σ beide onbekend). Het stellen van deze nulhypothese betekent onder meer, dat we aannemen, dat de steekproefgemiddelden m_1 en m_2 beide beste schattingen zijn voor het centrum μ van het universum en dus niet significant verschillen. Blijkt de nulhypothese in strijd met de werkelijkheid, dan is het onmogelijk, dat de twee steekproeven uit hetzelfde universum afkomstig zijn.

Het is goed, vooraf de juistheid van het tweede deel van de nulhypothese te onderzoeken. De z -toets van FISHER (1.9) kan beslissen over de vraag, of s_1 en s_2 significant verschillend zijn. De beste schatting s voor σ wordt berekend uit

$$s^2 = \left[(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2 \right] / (n_1 + n_2 - 2) = \left[\sum_{i=1}^{n_1} (x_i - m_1)^2 + \sum_{j=1}^{n_2} (y_j - m_2)^2 \right] / (n_1 + n_2 - 2)$$

Dus stelt s^2 het gewogen gemiddelde van s_1^2 en s_2^2 voor. Nu moet t gelijk gesteld worden aan $\frac{m_1 - m_2}{s_{m_1 - m_2}}$. Daar $s_{m_1 - m_2}^2 = s_{m_1}^2 + s_{m_2}^2 = \frac{s^2}{n_1} + \frac{s^2}{n_2} = s^2 \left(\frac{n_1 + n_2}{n_1 n_2} \right)$,

$$\text{geldt } t = \frac{m_1 - m_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}.$$

De overschrijdingskans kan ook hier uit een tabel of grafiek afgelezen worden, waarbij men $v = n_1 + n_2 - 2$ moet nemen. Is $P < 5\%$, dan moet de nulhypothese verworpen worden en is langs statistische weg bewezen, dat er een significant verschil bestaat tussen deze twee steekproefgemiddelden.

1.8 DE G^2 -TOETS

De G^2 -toets (in de Angelsaksische literatuur: χ^2) werd in 1.6.1 reeds gebruikt, maar hij dient toch voornamelijk om waargenomen frequenties te vergelijken met frequenties, die de theorie verwacht. Het toepassen van deze toets eist, dat alle theoretische frequenties $> b$ zijn. BROWNEE (1947) stelt $b = 5$, AITKEN (1947) $b = 10$, KENDALL (1947, 292) $b = 20$. Wij houden ons, enkele uitzonderingen daargelaten, bij het meetkundige gemiddelde van al deze eisen, dus $b = 10$. Dit aantal kan men bereiken door zoveel intervallen bij elkaar te nemen, tot de daarbij behorende totale frequentie volgens de theorie > 10 is. Is x_i het aantal waargenomen en X_i het door de theorie vereiste aantal elementen in het i^{de} interval, dan wordt de G^2 -som berekend met behulp van de formule $G^2 = \sum_{i=1}^n \frac{(x_i - X_i)^2}{X_i}$.

n is het aantal intervallen of hokjes, waarover de resultaten van de waarnemingen verdeeld worden.

Volgt elke x_i de gaussverdeling, dan wordt de frequentiefunctie van G^2 gegeven door

$$2^{-\frac{\nu}{2}} \left\{ \Gamma\left(\frac{\nu}{2}\right) \right\}^{-1} G^{\nu-2} e^{-\frac{1}{2} G^2}$$

ν , het aantal vrijheidsgraden, is gelijk aan het aantal frequenties, dat op willekeurige wijze gekozen kan worden.

De overschrijdingskans van een bepaalde G_s^2 , berekend uit de frequentieverdeling van een steekproef, kan door integreren verkregen worden.

$$P_{G_s^2} = \int_{G_s^2}^{\infty} 2^{-\frac{\nu}{2}} \left\{ \Gamma\left(\frac{\nu}{2}\right) \right\}^{-1} G^{\nu-2} e^{-\frac{1}{2} G^2} dG^2. \text{ Voor bepaalde waarden van}$$

ν en G_s^2 zijn de overschrijdingskansen in tabel en grafiek vastgelegd.

1.9 DE z-TOETS VAN FISHER

De z -toets van FISHER onderzoekt de mogelijkheid, of twee steekproefstandaarddeviaties s_1 en s_2 als beste schattingen beschouwd mogen worden van één en dezelfde universumstandaarddeviatie σ . Deze toets kan dus de vraag beantwoorden, of s_1 en s_2 significant verschillen en is dus in zekere zin een parallel van de toets, die beschreven is in 1.7.3. Daar werd nl. de hypothese opgesteld, dat twee steekproefgemiddelden de beste schattingen zijn van één en hetzelfde universumgemiddelde.

FISHER stelt $z = \frac{1}{2} \ln s_1/s_2 = \frac{1}{2} \ln F$. Met s_1 wordt altijd de grootste standaarddeviatie bedoeld, zodat z steeds positief is. De frequentiefunctie van z is

$$f(z) = \frac{2^{\frac{1}{2} \nu_1} \nu_1^{\frac{1}{2}} \nu_2^{\frac{1}{2}}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} e^{\nu_1 z} \left| (\nu_1 e^{2z} + \nu_2)^{\frac{1}{2}} (\nu_1 + \nu_2) \right| \quad \begin{array}{l} \text{met } \nu_1 = n_1 - 1 \\ \text{en } \nu_2 = n_2 - 1. \end{array}$$

Door integraties kunnen de overschrijdingskansen berekend worden. Omdat er drie veranderlijken (ν_1 , ν_2 en z) zijn, is één tabel of grafiek niet voldoende. Het bleek efficiënter te zijn, z op te vatten als functie van ν_1 , ν_2 en P . Door P achtereenvolgens gelijk te stellen aan 20, 10, 5, 1 en 0,1 % kon z en daarmee ook $F = s_1/s_2$ voor variërende waarden van ν_1 en ν_2 berekend worden.

Zoals steeds dwingt een overschrijdingskans kleiner dan een geschikt gekozen drempelwaarde ons tot het trekken van de conclusie, dat de gebezigde theorie in strijd is met de werkelijkheid en dus herzien moet worden. In dit geval moet dan aangenomen worden, dat het verschil tussen s_1 en s_2 niet uitsluitend aan het steekproefeffect mag worden toegeschreven. Het kan ook zo uitgedrukt worden: s_1 is significant groter dan s_2 .

1.10 DE VARIANTIEANALYSE

Een meting wordt bepaald door talloze oorzaken. Zou elke oorzaak een constante invloed oefenen, dan zou het resultaat van elke meting gelijk zijn. Dit doet zich nooit voor. Het ligt ook wel voor de hand om aan te nemen, dat sommige (misschien wel alle) oorzaken voortdurend in grootte variëren. Men noemt deze oorzaken dikwijls onafhankelijke veranderlijken en het meetresultaat de afhankelijk veranderlijke.

Het doel van vele onderzoeken is het bepalen van de grootte van de invloed, die elke oorzaak apart oefent. Is dit doel bereikt, m.a.w. is men er in geslaagd, het verband tussen de afhankelijk en elke onafhankelijk variabele op te sporen, dan kan men deze in wetten vastgelegde ontdekkingen gebruiken om toekomstige veranderingen van de afhankelijk variabele te voorspellen uit aan te brengen veranderingen in de onafhankelijke variabelen.

In bevoorrechte gevallen (o.a. in de fysica) kan men één bepaalde oorzaak sterk laten variëren en tegelijkertijd zodanige voorzorgsmaatregelen treffen, dat de overige oorzaken zo goed mogelijk gelijk blijven – althans een veel geringere, geheel te verwaarlozen invloed oefenen dan de ene sterk variërende. De veranderingen van zowel de afhankelijk als de onafhankelijk variabele kunnen geregistreerd worden. Het uitwerken van de metingen levert het nauwe verband tussen deze twee veranderlijken. Zo leert men de invloed van één oorzaak kennen. Desgewenst kan nu ook de invloed van andere onafhankelijk variabelen opgespoord worden. In de minder bevoorrechte gevallen is er geen sprake van, dat men een bepaalde onafhankelijk variabele naar eigen willekeur kan regelen, terwijl de overige zo goed mogelijk constant blijven. De atmosfeer, die zich tot dusverre vrijwel geheel aan de gevolgen van menselijke experimenteerzucht heeft kunnen onttrekken, bezorgt daarmee de geofysica een plaats in de minder bevoorrechte klasse. De statistica komt nu te hulp snellen met een vrij nieuw aanvalswapen, de variantieanalyse. Met behulp hiervan kan het toeval over een gedeelte van het gebied, waarover het heerst, worden teruggedrongen en kan dat gebiedsdeel aan een definitieve bezetting door de wetenschap onderworpen worden. Het begrip toeval moet dan opgevat worden, zoals VOLTAIRE het deed: „Il n'est point de hasard. Nous avons nommé ainsi l'effet que nous voyons d'une cause que nous ne voyons pas”. (Lettres de Memmius).

1.10.1 DE ENKELVOUDIGE VARIANTIEANALYSE

Deze variantieanalyse kan zijn diensten verlenen bij het onderzoek van k steekproefgemiddelden. Als nulhypothese wordt gesteld: de k steekproeven zijn alle afkomstig uit één universum. Geeft de variantieanalyse een kleine over-

schrijdingskans, dan moet de nulhypothese verworpen worden en is aangetoond, dat de gemiddelden onderling significant verschillen. Is $k = 2$, dan is het beter STUDENT's t -toets (1.7.3) rechtstreeks toe te passen. Voor $k > 2$ zou deze toets ook gebruikt kunnen worden, maar daaraan zijn enkele bezwaren verbonden. We moeten immers de k gemiddelden paarsgewijze vergelijken. Dit leidt tot het berekenen van $\frac{1}{2}k(k-1)$ waarden van t . Niet alleen vergroot dit het rekenwerk, het brengt ook principiële moeilijkheden met zich mee bij het bepalen van de drempelwaarde. Immers, we krijgen $\frac{1}{2}k(k-1)$ overschrijdingskansen. De 5%-grens moet kleiner worden, maar het is niet gemakkelijk uit te maken, welke grenswaarde dan aangenomen moet worden. (In 3.2 gaan we er nader op in). De variantieanalyse geeft als eindresultaat van de bewerking slechts één overschrijdingskans, zodat we onmiddellijk kunnen beslissen over de vraag, of de nulhypothese al dan niet verworpen moet worden.

De steekproefelementen worden in kolommen opgesteld. Zo ontstaat het volgende schema

$$\begin{array}{ccccccc} & & x_{11} & x_{12} & \dots & x_{1k} \\ \text{De } j^{\text{de}} \text{ kolom heeft } r_j \text{ elementen} & & x_{21} & x_{22} & \dots & x_{2k} \\ & & x_{31} & x_{32} & \dots & x_{3k} \end{array}$$

$$\sum_{j=1}^k r_j = N$$

$$x_{.j} = \frac{1}{r_j} \sum_{i=1}^{r_j} x_{ij}$$

$$x_{r_1 1} \quad x_{r_2 2} \quad \dots \quad x_{r_k k}$$

$$\text{Kolomgemiddelden: } x_{.1} \quad x_{.2} \quad \dots \quad x_{.k}$$

$$\text{Algemeen gemiddelde} = x_{..} = \frac{1}{N} \sum_{j=1}^k \sum_{i=1}^{r_j} x_{ij}$$

$$\text{Nu geldt: } \sum_{j=1}^k \sum_{i=1}^{r_j} (x_{ij} - x_{..})^2 = \sum_{j=1}^k \sum_{i=1}^{r_j} (x_{ij} - x_{.j})^2 + \sum_{j=1}^k r_j (x_{.j} - x_{..})^2$$

De aantallen vrijheidsgraden, die bij deze kwadratsommen behoren, zijn resp. $n-1$, $n-k$ en $k-1$. Daaruit blijkt, dat zowel kwadratsommen als aantallen vrijheidsgraden additief zijn. Elke kwadratsom gedeeld door het bijbehorende aantal vrijheidsgraden is een schatting van σ^2 en wordt daarom met de naam variantie aangeduid. Dit geldt alleen, indien de nulhypothese waar is. De totale kwadratsom kan dus gesplitst worden in een som, die uitsluitend op meetfouten betrekking heeft – elke afwijking is ten opzichte van zijn eigen steekproefgemiddelde bepaald – en een som, opgebouwd uit de gewogen kwadraten van de afwijkingen van de gemiddelden t.o.v. het algemeen gemiddelde. Zijn de steekproefgemiddelden niet reëel verschillend, dan moet de laatstgenoemde som eveneens aan meetfouten – of nog eenvoudiger: aan het toeval – worden toegeschreven. De variantie binnen de kolommen en de variantie tussen de kolommen

worden gedefinieerd door $s_{bik}^2 = \frac{1}{n-k} \sum \sum (x_{ij} - x_{.j})^2$ en $s_{tuk}^2 = \frac{1}{k-1} \sum r_j (x_{.j} - x_{..})^2$.

s_{bik} is steeds een schatting voor de toevalsspreiding, s_{tuk} alleen dan, wanneer de gemiddelden niet significant verschillen. Is s_{tuk} significant groter dan s_{bik} , dan moet de nulhypothese verworpen worden. De steekproefgemiddelden zijn dan reëel verschillend. Met behulp van FISHER's z -toets kan het verschil tussen s_{tuk} en s_{bik} op statistische wijze worden aangetoond. Passen we deze toets toe, dan moet voldaan zijn aan de voorwaarde, dat de elementen in de kolommen gaussisch verdeeld zijn met eenzelfde standaarddeviatie. Deze variantieanalyse wordt enkelvoudig genoemd, omdat hiermee alleen verschillen tussen de kolommen (het zogenaamde kolomeffect) aan het daglicht gebracht kunnen worden. De invloed van slechts één onafhankelijk veranderlijke kan – al of niet – worden aangetoond.

1.10.2 DE TWEEVOUDIGE VARIANTIEANALYSE

Deze wordt toegepast, wanneer een antwoord gegeven moet worden op de vraag, of ook de rijen significant verschillen. Hetzelfde schema kan gebruikt worden, alleen moet $r_j = r$ zijn. Elke kolom telt dus evenveel elementen. Nu geldt:

$$\sum_{i=1}^r \sum_{j=1}^k (x_{ij} - x_{..})^2 = k \sum (x_{i.} - x_{..})^2 + r \sum (x_{.j} - x_{..})^2 + \sum \sum (x_{ij} - x_{i.} - x_{.j} + x_{..})^2$$

De aantallen vrijheidsgraden zijn resp. $rk-1$, $r-1$, $k-1$ en $(r-1)(k-1)$. Ook hier dus additiviteit. De nulhypothese bestaat hier uit twee delen. Het eerste deel onderstelt, dat de rijen niet onderling significant verschillen. Het tweede deel veronderstelt hetzelfde van de kolommen. Op een andere wijze uitgedrukt: men neemt aan, dat zowel rij- als kolomeffect ontbreken. De overeenkomstige tellers en noemers vormen 4 quotiënten, die elk een schatting s^2 geven voor de universele variantie van de enkelvoudige waarneming – natuurlijk onder de voorwaarde, dat de nulhypothese juist is. s_{residu} is steeds een schatting voor de toevalsspreiding – ook al bestaat er een rij- of kolomeffect – omdat $x_{ij} - x_{i.} - x_{.j} + x_{..} = (x_{ij} - x_{i.}) - (x_{i.} - x_{..}) - (x_{.j} - x_{..})$ de afwijking van een willekeurig element ten opzichte van het algemeen gemiddelde voorstelt, ontdaan van rij- en kolominvloed. s_{residu} neemt de plaats in van s_{bik} uit de enkelvoudige variantieanalyse. s_{tur} (tur = tussen rijen) geeft slechts dan een schatting voor de toevalsspreiding, indien de rijgemiddelden niet significant verschillen. De z -toets van FISHER is het middel om de juistheid van het eerste deel van de nulhypothese te onderzoeken. Is s_{tur} significant groter dan s_{residu} , dan moet de aanwezigheid van een rijeffect aangenomen worden. Dezelfde methode, toegepast op s_{tuk} , geeft ons statistische zekerheid (zie 1) p. 15) over het al of niet bestaan van een kolomeffect.

Als voorbeeld van een tweevoudige variantieanalyse nemen we de dagelijkse

zonneshijnpercentages in afhankelijkheid van de geografische ligging van de stations en de datum. We stellen 4 kolommen naast elkaar, die de percentages weergeven van de zonneshijnstations A, B, C en D. Elke rij geeft de percentages van de 4 stations op één bepaalde dag. Omdat de zonneshijnpercentages van dag tot dag variëren, is er zeker een datumeffect. Dit is algemeen bekend, zodat het niet eens strikt noodzakelijk zou zijn, het datumeffect statistisch aan te tonen. Wel twijfelen sommige klimatologen aan de invloed van de geografische ligging van de stations (stationseffect). Zou de variantieanalyse een significant verschil kunnen ontdekken tussen de gemiddelde zonneshijnpercentages op de vier stations? Zo ja, dan zou dit in overeenstemming zijn met wat gesuggereerd wordt het stadsklimaat te zijn; vooral als het station, gelegen in de grootste stad, de minste zonneshijn zou ontvangen. Zo neen, dan mag men toch niet beweren, dat er in het geheel geen apart stadsklimaat zou zijn. Het is best mogelijk, dat verschillende windrichtingen verschillende zonneshijnpercentages zouden kunnen teweegbrengen. Een drievoudige variantieanalyse zal ons in staat stellen, een eventueel windeffect te ontdekken.

Niet vergeten mag worden, dat het toepassen van de z -toets van FISHER eist, dat de elementen een gaussverdeling volgen. Daar de zonneshijnpercentages zeker niet aan deze voorwaarde voldoen (zie 1.3.4), verdient het aanbeveling, een verdelingsvrije variantieanalyse toe te passen.

Verdelingsvrije methoden zijn methoden, die geen enkele veronderstelling over de frequentieverdeling van de universelementen maken. Hoewel hun toepassingsgebied dus veel groter is dan dat van de in deze verhandeling besproken methoden, is er toch geen plaats voor ingeruimd.

Redenen daarvoor zijn o.a. het geringe voorkomen van niet-gaussische verdelingen en het ontbreken van ervaring met verdelingsvrije methoden op geofysisch gebied. Voorts zijn – zoals reeds in 1.3.1 is vermeld – de klassieke methoden vrij ongevoelig voor verschillen tussen onderstelde en ware verdelingen.

1.11 CORRELATIE EN REGRESSIE

1.11.0 MEERDIMENSIONALE UNIVERSA

Tot nog toe behandelden we universa met elementen, die elk de meting van slechts één eigenschap toelieten. We gaan nu ook kennis maken met verzamelingen van zulke elementen, die elk twee of meer meetbare eigenschappen bezitten. Dergelijke universa noemen we 2- of m -dimensionale universa.

1.11.1. ENKELVOUDIGE CORRELATIE

Zij gegeven een 2-dimensionaal universum met N elementen. De resultaten van

de metingen van de twee eigenschappen zijn resp. x_i en y_i . Hun centra zijn μ_1 en μ_2 . Daar bij elke x_i één en slechts één y_i behoort, kunnen we de correlatiecoëfficiënt ρ definiëren door

$$\rho = \frac{\sum_{i=1}^N (x_i - \mu_1) (y_i - \mu_2)}{\left[\sum_{i=1}^N (x_i - \mu_1)^2 \sum_{i=1}^N (y_i - \mu_2)^2 \right]^{1/2}}$$

Is er tussen x_i en y_i een lineair verband ($y_i = bx_i + c$), dan is $\rho = +1$ voor $b > 0$ en $\rho = -1$ voor $b < 0$. Geldt de vergelijking $y_i = bx_i + c$ slechts met een min of meer goede benadering, dan neemt ρ tussenliggende waarden in. Hoe beter het lineaire verband benaderd wordt, des te verder ligt ρ van 0 af. Is $\rho = 0$, dan mag alleen geconcludeerd worden tot *lineaire* onafhankelijkheid, *niet tot algemene* onafhankelijkheid. Er kan dan nog wel een andere, bv. kwadratische afhankelijkheid bestaan.

Is het universum onbekend, dan moeten we ons weer met een steekproef behelpen. Uit het steekproefmateriaal kan de beste schatting r voor de parameter ρ berekend worden met behulp van de formule

$$r = \frac{\sum_{i=1}^n (x_i - m_1) (y_i - m_2)}{\left[\sum_{i=1}^n (x_i - m_1)^2 \sum_{i=1}^n (y_i - m_2)^2 \right]^{1/2}}$$

waarin m_1 en m_2 de steekproefgemiddelden zijn en n het aantal elementen in de steekproef. Ook hier is het de bedoeling uit de grootte van de steekproefparameter inlichtingen te winnen over de parameter van het universum. Het was FISHER (1915), die er in slaagde – uitgaande van een gaussisch verdeeld tweedimensionaal universum – de frequentiefunctie van r voor elke ρ te vinden. Die voor $\rho = 0$ is voor ons de belangrijkste. Reeds in 1908 maakte STUDENT langs experimentele weg plausibel, dat deze verdeling gelijk is aan

$$f(r, 0) = \frac{(1-r^2)^{\frac{n-4}{2}}}{B\left(\frac{n-2}{2}, \frac{1}{2}\right)}$$

Het aantal vrijheidsgraden is $\nu = n-2$. Levert de steekproef een $|r| = r_s$ en nemen we als nulhypothese $\rho = 0$, dan is de dubbelzijdige overschrijdingskans

$$P = \int_{r_s}^1 f(r, 0) dr + \int_{-1}^{-r_s} f(r, 0) dr = 2 \int_{r_s}^1 \text{const.} (1-r^2)^{\frac{n-4}{2}} dr.$$

In tabel VI van FISHER and YATES (1949) zijn voor $P = 0,1, 0,05, 0,02, 0,01$ en $0,001$ en $\nu = 1(1)20(5)50(10)100$ de bijbehorende waarden van r_s aangegeven.

Na berekening van de correlatiecoëfficiënt van een steekproef en bepaling van het aantal vrijheidsgraden geeft raadpleging van tabel VI onmiddellijk het interval, waarin de overschrijdingskans P ligt. Is r_s zo groot, dat $P < 5\%$, dan moet de nulhypothese $\rho = 0$ verworpen worden. De steekproef heeft ons dan geleerd, dat x_i en y_i niet onafhankelijk van elkaar zijn.

F. N. DAVID (1938) heeft tabellen en diagrammen opgesteld, waarmee de nulhypothese $\rho = 0,1(0,1)0,9$ getoetst kunnen worden.

Heeft men niet de beschikking over het exacte werk van Miss DAVID, dan moet men gebruik maken van een benaderingsmethode. De beste methode vond FISHER (1921). Hij transformeert eerst r in z volgens $2z = \ln(1+r) - \ln(1-r)$ en dan ρ in ζ volgens $2\zeta = \ln(1+\rho) - \ln(1-\rho)$. Tabel VII van FISHER and YATES (1949) geeft deze omzettingen onmiddellijk. $z-\zeta$ is bij zeer goede benadering volgens GAUSS verdeeld met de standaarddeviatie $\frac{1}{\sqrt{v-1}}$ en centrum $\frac{\rho}{2v+2}$.

1.11.1.1 SIGNIFICANTIE VAN HET VERSCHIL TUSSEN TWEE CORRELATIE-COEFFICIENTEN

Deze transformatie van FISHER is het enige middel om te onderzoeken, of twee correlatiecoëfficiënten significant van elkaar afwijken. Berust r_1 op $n_1 = (v_1 + 2)$ paren waarnemingen en r_2 op $n_2 = (v_2 + 2)$ paren waarnemingen en zijn beide steekproeven afkomstig van een universum, waarin de correlatiecoëfficiënt een waarde $= \rho$ heeft, dan is $z_1 - z_2$ gaussisch verdeeld met standaarddeviatie $\sqrt{\frac{1}{v_1-1} + \frac{1}{v_2-1}}$ en centrum 0.

z_1 en z_2 zijn beide weer bepaald met behulp van $2z = \ln(1+r) - \ln(1-r)$.

1.11.2 ENKELVOUDIGE REGRESSIE

Geeft de enkelvoudige correlatierekening antwoord op de vraag, in hoeverre van een lineair verband tussen 2 variabelen gesproken mag worden, de regressie-analyse leert ons, hoe de beste benadering van dit lineaire verband berekend kan worden. De elementen (x_i, y_i) van het tweedimensionale universum gaan we in gedachten vervangen door de elementen (x_i, Y_i) , waarbij $Y_i - \mu_2 = \beta(x_i - \mu_1) + \alpha$. α en β zijn constanten, die zo gekozen zijn, dat de nieuwe elementen zo veel mogelijk op de oude gelijken. x_i blijft dezelfde. We behoeven dus alleen maar het verschil tussen y_i en Y_i zo gering mogelijk te maken. Door de eis, dat $\sum_{i=1}^N (y_i - Y_i)^2$ minimaal moet zijn, zijn α en β éénduidig bepaald en worden de kwalitatieve eisen: „zo veel mogelijk” en „zo gering mogelijk” gepreciseerd. Deze methode wordt de methode der kleinste kwadraten genoemd. Het resultaat is, dat

$$\beta = \frac{\sum_{i=1}^N (x_i - \mu_1) (y_i - \mu_2)}{\sum_{i=1}^N (x_i - \mu_1)^2} \text{ en } \alpha = 0.$$

β draagt de naam van regressiecoëfficiënt, waaraan het adjectief „totaal” voorafgaat, indien we onderscheid willen maken met andere regressiecoëfficiënten, nl. de „partiële”, die in 1.11.3 gedefiniëerd worden.

Het is gebleken, dat het het beste is, het steekproefmateriaal ook met de methode van de kleinste kwadraten te behandelen. De beste schatting b voor β kunnen we berekenen uit de vergelijking

$$b = \frac{\sum_{i=1}^n (x_i - m_1) (y_i - m_2)}{\sum_{i=1}^n (x_i - m_1)^2} \quad \begin{aligned} a &= \alpha = 0 \\ (n-1) s_x^2 &= \sum_{i=1}^n (x_i - m_1)^2 \\ (n-1) s_y^2 &= \sum_{i=1}^n (y_i - m_2)^2 \end{aligned}$$

Hieruit blijkt, dat $b = r \frac{s_y}{s_x}$.

Is β langs theoretische weg gevonden, dan kan β aan de steekproefwaarde b getoetst worden. Daartoe moet berekend worden

$$\begin{aligned} \sum_{i=1}^n (y_i - Y_i)^2 &= \sum_{i=1}^n \{y_i - m_2 - b(x_i - m_1)\}^2 = \\ &= \sum_{i=1}^n (y_i - m_2)^2 + b^2 \sum_{i=1}^n (x_i - m_1)^2 - 2b \sum_{i=1}^n (y_i - m_2)(x_i - m_1) \\ &= (n-1) s_y^2 + b^2 (n-1) s_x^2 - 2b^2 (n-1) s_x^2 = \\ &= (n-1) s_y^2 - b^2 (n-1) s_x^2 = (n-1) u^2 \end{aligned}$$

Stelt men $t = \frac{1}{u} (b - \beta) s_x \sqrt{n-2}$, dan kan aangetoond worden, dat deze t dezelfde waarschijnlijkheidsverdeling heeft als de t van STUDENT. Het aantal vrijheidsgraden is hier $v = n - 2$.

Voor grote n is b dus gaussisch verdeeld met standaarddeviatie

$$\sigma_b = \frac{\sigma_y}{\sigma_x} \sqrt{\frac{1-\rho^2}{n}}$$

Tot nu toe hebben we getracht y_i uit te drukken in x_i , maar uit overwegingen van symmetrie blijkt, dat het omgekeerde ook kan plaats vinden. De eis, dat $\sum_{i=1}^n (x_i - X_i)^2$ minimaal zij (met $x_i = \gamma y_i + \delta$) voert tot analoge formules.

1.11.3 MEERVOUDIGE REGRESSIE

Bij een tweedimensionaal universum kan de éne veranderlijke uitgedrukt worden in de andere. Bij een $(m+1)$ -dimensionaal universum kan elk der

$(m + 1)$ veranderlijken uitgedrukt worden in de m overige. Het algemene element van het $(m + 1)$ -dimensionale universum zij weergegeven door

$$(x_{0i}, x_{1i}, x_{2i}, \dots, x_{mi})$$

Dit gaan we weer vervangen door

$$(X_{0i}, x_{1i}, x_{2i}, \dots, x_{mi}) \text{ met } X_{0i} = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} = \sum_{j=1}^m \beta_j x_{ji}$$

De eis, dat $\sum_{i=1}^N (x_{0i} - X_{0i})^2$ minimaal moet zijn, legt β_j weer éénduidig vast. De

coëfficiënten β_j noemen we bij de meervoudige regressie partiële regressiecoëfficiënten. Is het universum geheel bekend, dan kan β_j uit de m normaalvergelijkingen berekend worden:

$$\sum_{i=1}^N (x_{0i} x_{ji}) = \beta_1 \sum_{i=1}^N (x_{1i} x_{ji}) + \beta_2 \sum_{i=1}^N (x_{2i} x_{ji}) + \dots + \beta_m \sum_{i=1}^N (x_{mi} x_{ji}), \quad j = 1, 2, \dots, m$$

Is alleen een steekproef met n elementen beschikbaar, dan kunnen schattingen b_j voor de regressiecoëfficiënten β_j berekend worden uit de m normaalvergelijkingen.

$$\frac{\partial}{\partial b_j} \left[\sum_{i=1}^n (x_{0i} - X_{0i})^2 \right] = \frac{\partial}{\partial b_j} \left[\sum_{i=1}^n (x_{0i} - b_1 x_{1i} - b_2 x_{2i} - \dots - b_m x_{mi})^2 \right] = 0. \text{ Dit geeft:}$$

$$\sum_{i=1}^n x_{0i} x_{ji} = b_1 \sum_{i=1}^n x_{1i} x_{ji} + b_2 \sum_{i=1}^n x_{2i} x_{ji} + \dots + b_m \sum_{i=1}^n x_{mi} x_{ji}, \quad j = 1, 2, \dots, m$$

Deze m normaalvergelijkingen bevatten de m onbekenden b_j en laten in het algemeen één oplossing toe.

Voor het significantieonderzoek hebben we nodig de determinant D met de elementen $e_{kl} = \sum_{i=1}^n x_{ki} x_{li}$ ($k = 1 \dots m; l = 1 \dots m$). D_{jj} is de onderdeterminant, toegevoegd aan e_{jj} . Voorts is nodig

$$\sum_{i=1}^n (x_{0i} - X_{0i})^2 = \sum_{i=1}^n x_{0i}^2 - b_1 \sum_{i=1}^n x_{0i} x_{1i} - b_2 \sum_{i=1}^n x_{0i} x_{2i} - \dots - b_m \sum_{i=1}^n x_{0i} x_{mi}$$

$$\text{De grootheden } t_j = \frac{(b_j - \beta_j) \sqrt{n-m-1}}{\sqrt{\sum_{i=1}^n (x_{0i} - X_{0i})^2}} \frac{\sqrt{D}}{\sqrt{D_{jj}}} \text{ zijn volgens de } t\text{-verdeling van}$$

STUDENT verdeeld met $v = n-m-1$.

Voor $m = 1$ blijkt, dat alle formules van deze onderparagraaf overgaan in die van 1.11.2.

1.11.4 MEERVOUDIGE CORRELATIE

De zojuist bereikte resultaten over de meervoudige regressie stellen ons in staat de meervoudige correlatierekening op te bouwen.

Het $(m + 1)$ -dimensionale universum brengen we terug tot een m -dimensio-

naal universum door van elk element de meting van één eigenschap, de x_{1i} , buiten beschouwing te laten. In dit m -dimensionale universum kunnen we de regressievergelijking van x_{0i} ten opzichte van $x_{2i}, \dots, x_{mi}$ berekenen: $X_{0i} = \beta_{02}x_{2i} + \dots + \beta_{0m}x_{mi}$. Een ander m -dimensionaal universum krijgen we door de metingen x_{0i} buiten beschouwing te laten. In dit universum bepalen we de regressievergelijking van x_{1i} t.o.v. $x_{2i}, \dots, x_{mi}$, nl. $X_{1i} = \beta_{12}x_{2i} + \dots + \beta_{1m}x_{mi}$. Trekken we de rechter leden van deze zo juist uitgerekende regressievergelijkingen af van x_{0i} resp. x_{1i} , dan houden we over $x_{0i} - X_{0i} = x_{0.23\dots m}$ en $x_{1i} - X_{1i} = x_{1.23\dots m}$. Nu zijn x_0 en x_1 elk bevrijd van de invloed van $x_2, x_3, \dots$ en x_m , nl. door de bij x_0 resp. x_1 behorende lineaire regressievergelijkingen.

Onder de partiële correlatiecoëfficiënt $\rho_{01.23\dots m}$ tussen x_0 en x_1 bij eliminatie van $x_2, \dots, x_m$ verstaan we nu de correlatiecoëfficiënt tussen $x_{0.23\dots m}$ en $x_{1.23\dots m}$, dus

$$\rho_{01.23\dots m} = \frac{\sum_{i=1}^N x_{0.23\dots m} x_{1.23\dots m}}{\left\{ \sum_{i=1}^N (x_{0.23\dots m})^2 \right\}^{1/2} \left\{ \sum_{i=1}^N (x_{1.23\dots m})^2 \right\}^{1/2}}$$

0 en 1 worden de primaire indices genoemd, 2, 3, ... m de secundaire. Daar er $(m-1)$ secundaire indices zijn, wordt de partiële correlatiecoëfficiënt van de $(m-1)^{\text{de}}$ orde genoemd.

De multiële correlatiecoëfficiënt $P_{0.123\dots m}$ (spreek uit: rho) kunnen we definiëren met behulp van de vergelijking

$$1 - P_{0.12\dots m}^2 = (1 - \rho_{01}^2) (1 - \rho_{02.1}^2) (1 - \rho_{03.12}^2) \dots (1 - \rho_{0m.12\dots(m-1)}^2)$$

Partiële correlatiecoëfficiënten van de p^{de} orde kunnen we altijd uitdrukken in enige correlatiecoëfficiënten van de $(p-1)^{\text{de}}$ orde. De formule luidt:

$$\rho_{01.23\dots p(p+1)} = \frac{\rho_{01.23\dots p} - \rho_{0(p+1).23\dots p} \rho_{1(p+1).23\dots p}}{(1 - \rho_{0(p+1).23\dots p}^2)^{1/2} (1 - \rho_{1(p+1).23\dots p}^2)^{1/2}}$$

Zo kunnen we, uitgaande van de correlatiecoëfficiënten van de 0^{de} orde (de zgn. totale correlatiecoëfficiënten), stap voor stap alle benodigde partiële correlatiecoëfficiënten berekenen. Wanneer geen grotere nauwkeurigheid dan 0,01 in de coëfficiënten vereist is, vervangt gebruikmaking van Tafel 12 van KOLLER (1943) de ingewikkelde berekeningen door een eenvoudige aflezing. Is een grotere nauwkeurigheid nodig, dan construeer men KOLLER's diagram op een andere schaal.

Pas als de partiële correlatiecoëfficiënt van de hoogste orde, nl. $\rho_{0m.12\dots(m-1)}$ bekend is, kan $P_{0.12\dots m}$ bepaald worden met behulp van de vergelijking, die hem definiëert.

$P_{0.123\dots m}$ is inderdaad een correlatiecoëfficiënt, nl. tussen x_0 en X_0 , dus tussen de waargenomen en de uit de regressievergelijking berekende waarden.

De volgende betrekkingen gelden:

$$P_{0.123\dots m} = \frac{\sum_{i=1}^N x_{0i} X_{0i}}{\left(\sum_{i=1}^N x_{0i}^2\right)^{1/2} \left(\sum_{i=1}^N X_{0i}^2\right)^{1/2}}$$

$$P_{0.1\dots m}^2 = \frac{\sum_{i=1}^N x_{0i} X_{0i}}{\sum_{i=1}^N x_{0i}^2} = 1 - \frac{\sum_{i=1}^N (x_{0i} - X_{0i})^2}{\sum_{i=1}^N x_{0i}^2} = \frac{\sum_{i=1}^N X_{0i}^2}{\sum_{i=1}^N x_{0i}^2}$$

Vooraf de laatste identiteit geeft een buitengewoon eenvoudig controlemiddel op de grootte van P . Bovendien geeft zij twee kostbare inlichtingen. In de eerste plaats blijkt, dat steeds geldt:

$$\sum_{i=1}^N X_{0i}^2 < \sum_{i=1}^N x_{0i}^2. \quad \text{Voor bijna alle verdelingen volgt hieruit:}$$

$$\sum_{i=1}^N |X_{0i}| < \sum_{i=1}^N |x_{0i}|. \quad \text{De voorspelde waarden liggen dus gemiddeld}$$

dichter bij de normaal dan de waargenomen waarden. Geeft de regressievergelijking een extreme waarde, dan zou men dus eigenlijk dat resultaat een zekere toeslag mogen geven.

In de tweede plaats kunnen we opmerken, dat voor een hoge P de extreme waarden zeer goed voorspeld zullen worden. Immers, alleen in dat geval houdt $\sum X_{0i}^2$ gelijke tred met $\sum x_{0i}^2$.

De slechts weinig afwijkende, de zgn. „normale” gevallen zullen niet zo goed voorspeld kunnen worden. Deze hoedanigheden zijn in goede overeenstemming met de eisen, die de maatschappij aan de weersverwachtingen op lange termijn stelt. De synoptische prognosen zullen daarentegen grotere successen boeken juist bij de normale gevallen, maar machteloos blijven ten opzichte van de extreme ontwikkelingen in de atmosfeer.

Blijft het universum gedeeltelijk onbekend, dan moeten we ons weer behelpen met steekproeven. De vermelde formules blijven geldig, alleen dienen de Griekse letters door Romeinse vervangen te worden.

Moeten we terugvallen op het steekproevenonderzoek, dan gaat daarmee onherroepelijk gepaard het instellen van een onderzoek naar de significantie van de verkregen steekproefwaarden.

1.11.4.1 DE FREQUENTIEFUNCTIE VAN DE MULTIPLE CORRELATIE-COEFFICIENT R , WANNEER DE UNIVERSUMWAARDE $P=0$ IS

FISHER vond in 1928 deze functie. Zij geldt strikt genomen alleen, wanneer alle betrokken verdelingen gaussisch zijn.

De formule luidt:

$$f(R, 0) = \frac{R^{m-1} (1-R^2)^{\frac{n-m-3}{2}}}{B\left(\frac{n-m-1}{2}, \frac{m}{2}\right)}$$

n is het aantal elementen van de steekproef. $(m+1)$ is het aantal eigenschappen, dat van de elementen gemeten is.

Voor $m=1$ gaat $f(R, 0)$ over in de laatste formule van 1.11.1, die de frequentiefunctie $f(r, 0)$ van de totale correlatiecoëfficiënt uitdrukt in r en n .

De integraal $\int_{R_S}^1 f(R, 0) dR$ is de kans op het verkrijgen van een nog hogere steekproefwaarde dan R_S , wanneer $P=0$.

Voor $m=3, 4, 5$ en 6 en $n=39$ zijn deze overschrijdingskansen uitgerekend en in grafiek gebracht en wel speciaal voor waarden van $R_S > 0,50$ (fig. 7).

1.12 PERSISTENTIE

Zij gegeven een reeks waarnemingen $w_1, w_2, \dots, w_i, \dots, w_n$. Dit kunnen bv. zijn de temperaturen van de 1200 achtereenvolgende maanden in de 19^{de} eeuw. Door de jaarlijkse gang zijn deze niet dadelijk onderling vergelijkbaar. We kunnen de jaarlijkse gang verwijderen door elke maandtemperatuur te verminderen met de „normale” waarde voor de temperatuur in de desbetreffende maand. De nieuwe reeks $x_1, x_2, \dots, x_i, \dots, x_n$ heeft dan de eigenschap $E(x_i) = 0$.¹⁾

Door steeds de afwijkingen te beschouwen in plaats van de waarnemingen zelf, kunnen we elke reeks transformeren in een x -reeks.

Verder stellen we aan een x -reeks de eis, dat $E(x_i^2) = \sigma^2$. Iedere x_i moet dus dezelfde universumstandaarddeviatie hebben. Het is duidelijk, dat slechts weinige reeksen aan deze eis kunnen voldoen. Het in het begin van deze paragraaf genoemde voorbeeld voldoet zeker niet aan de gestelde voorwaarde. Immers de Januari-temperatuur vertoont grotere schommelingen dan de Juli-temperatuur. In sommige onderzoeken zal het geoorloofd zijn, de standaarddeviaties van de maandtemperaturen aan elkaar gelijk te stellen en indien het niet geoorloofd is, kan men – door de afwijkingen te delen door de respectieve standaarddeviaties – bereiken, dat $E(x_i^2) = 1$, dus onafhankelijk van i is.

Wanneer de x -reeks aan een derde voorwaarde voldoet, nl. $E(x_i x_j) = 0$ voor $i \neq j$, dan zijn de waarnemingen onderling ongecorreleerd, d.w.z. zij vertonen geen lineaire afhankelijkheid.

Voor deze x -reeksen geldt de foutenvoortplantingswet: $\sigma(g) = \frac{\sigma}{\sqrt{g}}$.

¹⁾ $E(x_i^k y_j^l) = \iint x_i^k y_j^l f(x_i, y_j) dx_i dy_j$

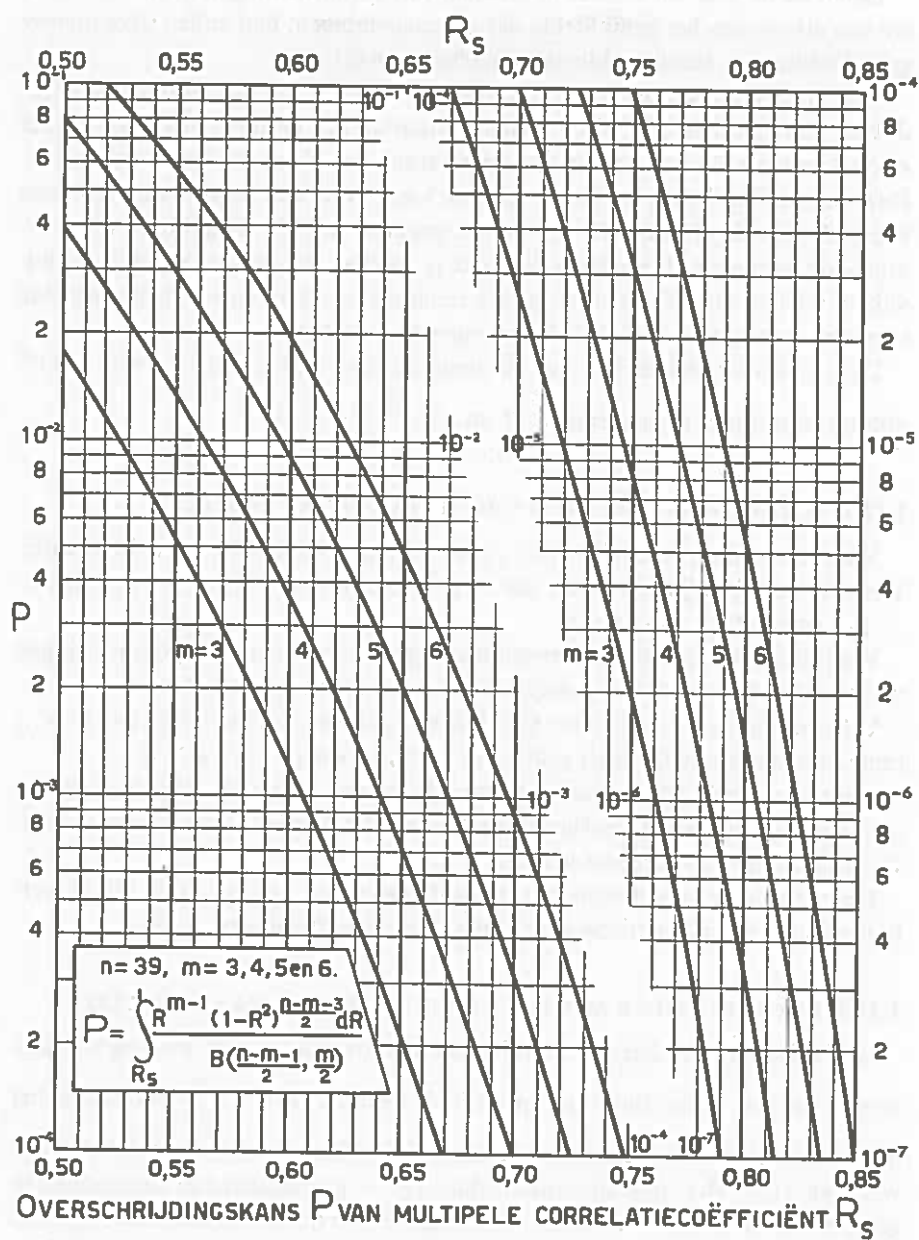


fig. 7

Delen we de gegeven x -reeks in groepen van g waarnemingen in en berekenen we van elke groep het gemiddelde der g waarnemingen, dan zullen deze nieuwe gemiddelden een standaarddeviatie hebben $= \sigma(g)$.

Zoals reeds aangetoond werd kan ons voorbeeld zo genormaliseerd worden, dat het aan de eerste twee eisen voldoet. Maar aan de derde voorwaarde zal het nooit kunnen voldoen, omdat de temperatuur van de atmosfeer persistent is. Deze eigenschap houdt in, dat de grootte van een temperatuurafwijking in een bepaald tijdvak afhankelijk is van de grootte van de afwijking in de voorgaande perioden. De afhankelijkheid is sterker, naarmate de gekozen tijdvakken kleiner zijn. Zo is er natuurlijk een hoge correlatie tussen de gemiddelde temperaturen tijdens twee achtereenvolgende minuten.

De correlatiecoëfficiënt tussen de temperaturen tijdens de i^{de} en $(i+k)^{\text{de}}$ minuut neemt met toenemende k af, m.a.w. $\frac{\partial \rho_{i,k}}{\partial k} < 0$.

1.12.1 AUTOCORRELATIECOEFFICIENT VAN DE k^{de} ORDE

We noemen deze coëfficiënten $\rho_{i,k}$ autocorrelatiecoëfficiënten van de k^{de} orde. Het valt gemakkelijk in te zien, dat $\rho_{i,k}(\text{minuut}) > \rho_{i,k}(\text{uur}) > \rho_{i,k}(\text{dag}) > \rho_{i,k}(\text{maand}) > \rho_{i,k}(\text{jaar})$.

Voor de meeste geofysische verschijnselen geldt $\rho_{i,k}(\text{jaar}) = 0$. Uitzonderingen vormen o.a. de met de zonnevlekken samenhangende grootheden.

Uit een onderzoek, dat later gepubliceerd zal worden, leren we, dat we $\rho_{i,k}$ (maandtemperatuur De Bilt) gelijk aan 0 mogen stellen voor $k > 2$.

Alleen wanneer alle autocorrelatiecoëfficiënten $= 0$ zijn, mogen de waarnemingen als onderling onafhankelijk beschouwd worden. Dan kan de foutenvoortplantingswet toegepast worden.

De geofysische grootheden zijn meestal continue functies van de tijd en vertonen daarom persistentieneiging (autocorrelatiecoëfficiënten > 0).

1.12.2 HET EFFECTIEVE AANTAL ONAFHANKELIJKE WAARDEN $\gamma(g)$

Ondanks deze persistentie kan de foutenvoortplantingswet nu nog wel toegepast worden, maar dan moeten we in de formule $\sigma(g) = \frac{\sigma}{\sqrt{g}}$ voor g niet het aantal waarnemingen substitueren, maar het effectieve aantal onafhankelijke waarden $\gamma(g)$. Het spreekt vanzelf, dat $\gamma(g) < g$ (afgezien van uitzonderlijke gevallen als de c -reeks van 1.12.3). Substitueren we dus g in plaats van $\gamma(g)$, dan krijgen we een te kleine $\sigma(g)$. We geven het gemiddelde daarmee een te geringe spreiding, dus een te grote nauwkeurigheid.

De gegeneraliseerde foutenvoortplantingswet luidt: $\sigma(g) = \frac{\sigma}{\sqrt{\gamma(g)}}$. Hieruit volgt: $\gamma(g) = \frac{\sigma^2}{\sigma^2(g)}$. Kennen we σ en $\sigma(g)$, dan is $\gamma(g)$ bepaald.

Meestal kennen we de universumwaarden σ en $\sigma(g)$ niet en moeten we ons behelpen met de schattingen s en $s(g)$. Uit deze steekproefwaarden berekenen we dan de schatting $c(g)$ voor het effectieve aantal onafhankelijke waarden in groepen van g elementen.

1.12.3 HET AEQUIVALENTE HERHALINGSGETAL $\omega(g)$

Dit aequivalente herhalingsgetal heeft BARTELS (1943) als volgt gedefiniëerd:

$$\omega(g) = \frac{g}{\gamma(g)} = g \frac{\sigma^2(g)}{\sigma^2}$$

Bij volkomen onafhankelijkheid is $\gamma(g) = g$ en dus $\omega(g) = 1$. Dit blijkt ook uit de formule

$$\omega(g) = 1 + 2 \left(\frac{g-1}{g} p_1 + \frac{g-2}{g} p_2 + \dots + \frac{1}{g} p_{g-1} \right)$$

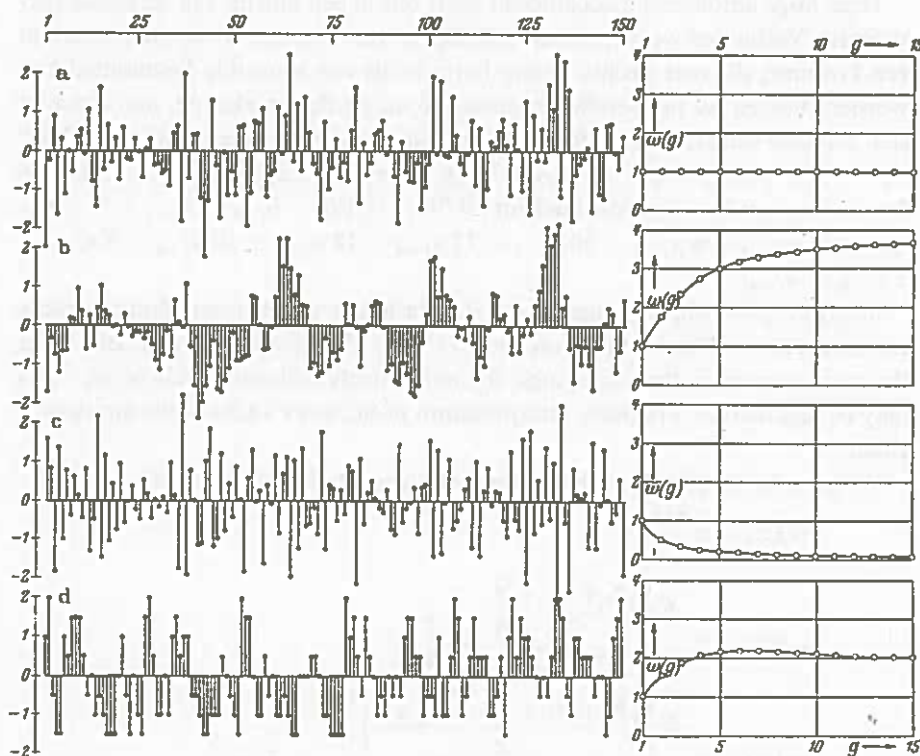


fig. 8

BARTELS (1943) gaf in zijn fig. 5 (= onze fig. 8) vier voorbeelden van waarnemingsreeksen met verschillende persistentieneiging. De grootte van de persistentieneiging werd aangegeven door $\omega(g)$.

Reeks *a* wordt gevormd door toevalsgetallen, dus een *x*-reeks. $E(x_i) = 0$, $E(x_i^2) = 1$ en $E(x_i x_j) = 0$ voor $i \neq j$. Het effectieve aantal onafhankelijke waarden is steeds gelijk aan het totale aantal waarden en daarom is het aequivalente herhalingsgetal $\omega(g) = 1$.

Reeks *b* vertoont persistentie, d.w.z. na hoge waarden komen meer hoge dan lage waarden voor: dit in tegenstelling tot rij *a*, waarin elk getal onafhankelijk van voorafgaande waarden is. De persistentieneiging vindt zijn verklaring in de wijze van ontstaan van deze reeks: $b_i = 1/2 (x_i + x_{i+1} + x_{i+2} + x_{i+3})$. Elke term in de *b*-reeks heeft 3 van de 4 componenten gemeen met zijn voorganger.

De autocorrelatiecoëfficiënt van de eerste orde is $\rho_{1,1} = \frac{E(b_i b_{i+1})}{\sqrt{E(b_i^2) E(b_{i+1}^2)}} = \frac{3}{4}$ (de volledige afleiding, ook van de volgende formules, is te vinden bij v.d. BUL 1951).

Deze hoge autocorrelatiecoëfficiënt geeft ons al een indruk van de sterke persistentie. Verbinden we de punten in de figuur van rij *b* met elkaar, dan ontstaat een kromme, die met slechts weinig fantasie als een sinusoïde beschouwd kan worden. Passen we het vereffeningsprocédé nu op de *b*-reeks toe, dan ontstaat een '*b*-reeks, waarvan de grafische voorstelling nog meer op een sinusoïde lijkt. ' $b_i = 1/2 (b_i + b_{i+1} + b_{i+2} + b_{i+3}) = 1/4 (x_i + 2x_{i+1} + 3x_{i+2} + 4x_{i+3} + 3x_{i+4} + 2x_{i+5} + x_{i+6})$. De volgende stap leidt tot ' $b_i = 1/2 ('b_i + 'b_{i+1} + 'b_{i+2} + 'b_{i+3}) = 1/8 (x_i + 3x_{i+1} + 6x_{i+2} + 10x_{i+3} + 12x_{i+4} + 12x_{i+5} + 10x_{i+6} + 6x_{i+7} + 3x_{i+8} + x_{i+9})$.

SLUTSKY (1937) heeft aangetoond, dat de limiet van het vereffeningsproces een sinusoïde is. Deze stelling vat DODD (1941, 25) als volgt kort samen: "But the more numerous the averagings, the more firmly will some cycle be set - this may be regarded as a popular interpretation of SLUTSKY's Law of the Sinusoidal Limit."

We gaan het aequivalente herhalingsgetal voor de *b*-reeks bepalen.

$$\begin{aligned}\omega(g) &= g \frac{\sigma^2(g)}{\sigma^2} \\ \omega(2) &= 2 \frac{E\left\{\left(\frac{b_i + b_{i+1}}{2}\right)^2\right\}}{E(b_i^2)} = \frac{7}{4} \\ \omega(g) &= g \frac{E\left\{\left(\frac{b_i + b_{i+1} + \dots + b_{i+g-1}}{g}\right)^2\right\}}{E(b_i^2)} = 4 - \frac{5}{g}\end{aligned}$$

Dit is de formule voor de $\omega(g)$ -kromme, die correspondeert met de b -reeks.

Werken we niet met de voortschrijdende gemiddelden over 4 elementen, maar over k elementen $-b_i^* = \frac{1}{k}(x_i + x_{i+1} + \dots + x_{i+k-1})$ - dan komen we op de volgende wijze tot de algemene formule:

$$\omega(g) = g \frac{\sigma^2(g)}{\sigma^2} = g \frac{E \left\{ \frac{1}{g^2} (b_i^* + \dots + b_{i+g-1}^*)^2 \right\}}{E \{ (b_i^*)^2 \}} = k - \frac{k^2 - 1}{3g}$$

(alleen geldig voor $g > k-2$).

In reeks c is anti-persistentie aanwezig. Na positieve waarden is er voorkeur voor negatieve waarden en andersom.

De algemene term is $c_i = \frac{a_{i+1} - a_i}{\sqrt{2}}$. Nu is

$$\omega(g) = g \frac{E \left\{ \frac{1}{g^2} (c_i + c_{i+1} + \dots + c_{i+g-1})^2 \right\}}{E \{ (c_i)^2 \}} = \frac{1}{g} \frac{E \left\{ \left(\frac{a_{i+g-1} - a_i}{\sqrt{2}} \right)^2 \right\}}{E \left\{ \left(\frac{a_{i+1} - a_i}{\sqrt{2}} \right)^2 \right\}} = \frac{1}{g}$$

en $\gamma(g) = \frac{g}{\omega(g)} = g^2$. Slechts in gevallen als deze is het effectieve aantal onafhankelijke waarden in groepen van g elementen groter dan g .

Zeer interessant is, dat de leek rij c voor de toevalsreeks houdt en aan rij a persistentie toekent. Onvoldoende kennis van wat het machtige toeval tot stand kan brengen, heeft er vaak toe geleid, dat men geofysische waarnemingsreeksen, die zich als reeks a gedroegen, niet als toevalsreeksen herkende.

Tenslotte toont de d -reeks karaktergetallen van de dagelijkse aardmagnetische storingen voor de periode van 1 Mei tot 27 September 1932. De karaktergetallen zijn zodanig genormeerd, dat $E(d_i) = 0$ en $E(d_i^2) = 1$. Het blijkt, dat $\omega(2) = 1,55$, $\omega(3) = 1,88$ en $\omega(4) = 2,06$ is. Dit wijst op persistentie, maar toch van een iets ander karakter dan bij de b -reeks, daar $\omega(g)$ later weer afneemt.

Berekening van $\omega(g)$ leert onmiddellijk, welke rij als toevalsrij beschouwd mag worden. Het verdient dus aanbeveling naast elke grafische voorstelling van waarnemingsreeksen de grafiek van $\omega(g)$ weer te geven, zoals dit gedaan is in fig. 8.

HOOFDSTUK 2

TOEPASSINGEN VAN DE STATISTISCHE METHODEN

However, the application, by rote, of procedures extracted from textbooks may place pitfalls in the path of the forecaster, who is expected to treat his data and select his procedures with discernment.

GARSTKA 1950

2.0 INLEIDING

In het voorafgaande hoofdstuk werden de methoden aangegeven, volgens welke de onafzienbare hoeveelheid geofysisch waarnemingsmateriaal gereduceerd kan worden tot een gering aantal voor ons van belang zijnde parameters.

Gewaarschuwd door zijn motto wil dit hoofdstuk deze methoden verduidelijken aan de hand van geschikte voorbeelden. Ook gemaakte fouten in de geofysische literatuur zullen behandeld worden. Het weten, hoe het niet moet, is vaak van evenveel belang als het kennen van de correcte toepassingen.

De ondernummering van dit hoofdstuk loopt parallel met die van het eerste hoofdstuk. Zo vindt men de praktische voorbeelden van bv. 1.4 in 2.4. Wil omgekeerd iemand meer weten van de methoden, toegepast in 2.6, dan behoeft hij slechts 1.6 op te slaan.

De eerste paragraaf van dit hoofdstuk is in twee gedeelten gesplitst; het gedeelte met een a voorzien bevat de juiste toepassingen, dat met een b verkeerde toepassingen.

2.1a UNIVERSUM EN STEEKPROEF (population and sample)

Het universum, dat de resultaten van alle mogelijke geofysische waarnemingen omvat, kortweg gezegd: het geofysische universum, zou moeten bestaan uit het materiaal, dat een feilloos werkend observator met een perfect en volledig instrumentarium vanaf de schepping der aarde tot aan haar vernietiging bijéén zou kunnen brengen. We zullen trachten duidelijk te maken, dat zelfs het in gedachten construeren van dit geofysische universum weinig zin heeft, omdat dit universum inhomogeen is (oorzaak o.a.: het voorkomen van ijstijden).

Een universum noemen we homogeen, als de kansen op het optreden van bepaalde gebeurtenissen onafhankelijk zijn van het beschouwde tijdvak (GRANT 1940, HAMMING 1952). Is de gebeurtenis onderhevig aan een periodiciteit, dan moeten die tijdvakken minstens een veelvoud van de periode omvatten.

Zo mag ons model, het universum der kruis- en muntworpen, homogeen genoemd worden, omdat de kans op een kruisworp steeds $\frac{1}{2}$ is.

De kans op een gemiddelde zomertemperatuur van 16°C of meer is in De Bilt in deze eeuw $= 1/2$. In de twee daaraanvoorgaande eeuwen was deze kans ook $= 1/2$. We mogen het universum, bestaande uit de zomertemperaturen van de 18^{de}, 19^{de} en 20^{ste} eeuw dus als homogeen beschouwen.

Dit mag nu in geen geval met het geofysische universum, want in de ijstijdperioden zal de kans op zomers met gemiddelde temperaturen van 16°C of meer wel zo goed als gelijk aan 0 geweest zijn (SCHWARZBACH 1950, 146).

Wil men van het geofysische universum een representatieve steekproef hebben, dan zou men bv. uit elke eeuw de zomertemperatuur van een willekeurig jaar moeten kiezen. Daar de oudste gegevens geheel onbetrouwbaar of zelfs onbekend zijn, is het duidelijk, dat er geen representatieve steekproef uit het grote universum getrokken kan worden.

Zou men als elementen van de steekproef de temperaturen van de zomers 1735–1949 nemen, dan is deze steekproef niet representatief voor het universum, omdat dit niet homogeen is.

Heeft men redelijke zekerheid, dat er sinds het jaar 1000 generlei verandering is opgetreden in de factoren, die het klimaat beïnvloeden, dan mag men het universum van de zomertemperaturen uit de jaren 1000–2000 als homogeen beschouwen. De steekproef, bestaande uit de zomertemperaturen van de jaren 1735–1949, mag nu wel als representatief beschouwd worden voor het universum 1000–2000.

Evenzo kan de steekproef 1735–1949 als universum optreden. Als steekproef uit dit universum (1735–1949) kan nu bv. gekozen worden een steekproef met als elementen de zomertemperaturen van die jaren, die deelbaar zijn door 50. Hiertoe behoren dus 1750, 1800, 1850 en 1900. Deze zomers hadden temperaturen van resp. $16,5$, $15,1$, $15,9$ en $17,0^{\circ}\text{C}$. Hun gemiddelde is $64,5 : 4 = 16,12^{\circ}\text{C}$. Het universumgemiddelde (1735–1949) is $16,16^{\circ}\text{C}$. Hieruit blijkt dus, dat een zo kleine steekproef van vier elementen al heel wat vertellen kan van het universum.

De afwijkingen zijn echter niet altijd zo gering. Neemt men bv. het jaar 1950 erbij (zomertemperatuur $= 17,2^{\circ}\text{C}$), dan wordt het nieuwe steekproefgemiddelde $16,34^{\circ}\text{C}$. Het gemiddelde van het nieuwe universum (1735–1950) verandert echter heel weinig, want het wordt $16,17^{\circ}\text{C}$. In 1.7 is reeds gesproken over de grootte van de afwijkingen, die men tussen universum- en steekproefgemiddelden mag verwachten; in 2.7 worden voorbeelden gegeven.

De bovenstaande temperatuurgegevens zijn ontleend aan Tabel A van LABRIJN (1945). Zijn Tabel C kan ons de gegevens van de zomerneerslag in Zwanenburg-Hoofddorp leveren. In de zomers van 1750, 1800, 1850 en 1900 vielen respectievelijk 193, 171, 232 en 218 mm. Het steekproefgemiddelde is dus 204 mm. Het gemiddelde over het universumtijdvak 1735–1944 is 219 mm.

Men ziet, dat men toch nog wel min of meer bruikbare resultaten kan verkrijgen, als men het 210 jaren omvattende tijdvak tot op minder dan $1/50$ van zijn omvang reduceert, maar men moet zich echter alleen in noodgevallen met dergelijke kleine steekproeven tevreden stellen. Het spreekt vanzelf, dat hoe groter de steekproef, hoe groter de nauwkeurigheid van het eindresultaat.

Er kunnen bezwaren tegen deze wijze van steekproefnemen naar voren gebracht worden.

Immers, waaruit blijkt de toevalligheid van de steekproef? Bestaat er tussen de elementen van het universum geen verband, dan zal elke steekproef uit een dergelijk universum toevallig genoemd mogen worden, ook al is de samenstelling van de steekproef op een systematische manier geschied zoals door de eerste a elementen te nemen, of alleen de even elementen of — zoals hierboven — de $(50n)^{\text{de}}$ elementen.

Vertoont de rij van jaarlijkse temperaturen een periode van bv. 2 jaren, dan zal een steekproef, die alleen de waarnemingen van de even jaartallen omvat, een systematisch hogere (of lagere) uitkomst geven dan de steekproef van de overige, de oneven jaren.

Zouden de oude meteorologische reeksen van Nederland een periode van 50 jaren bezitten en zou het maximum steeds bij de eeuwwisseling en in het midden van de eeuw vallen, dan zou de zojuist gebruikte steekproef geheel gevormd worden door de elementen met de grootste afwijkingen. Het verschil tussen het gemiddelde van zo'n steekproef en het universumgemiddelde kan niet meer toevallig genoemd worden; het is een systematisch verschil geworden.

Daar er in de klimatologische wereld nog steeds geen overeenstemming bereikt is over de kwestie van het al of niet aanwezig zijn van perioden in de weersverschijnselen, kan men elke opmerking, dat de steekproef niet representatief zou zijn, omzeilen door gebruik te maken van toevallige steekproeven. Hiertoe moet men zijn toevlucht nemen tot tabellen met toevalscijfers.

Als voorbeeld kiezen we de tabellen van FISHER and YATES (1949). We gebruiken het linker bovengedeelte van de eerste bladzijde (p. 104) van hun toevalstabellen.

Dit gedeelte wordt hier gereproduceerd, waarbij in afwijking van het origineel sommige cijfers vet gedrukt zijn. Deze worden gebruikt voor het samenstellen van enkele steekproeven.

Wij combineren de elementen van deze tabel tot getallen van drie cijfers. Zo vormen de eerste drie kolommen 15 onder elkaar staande getallen 034, 977, 167, 125, enz.

Als we het jaar 1735 laten samenvallen met bv. 000, dan stelt 034 het jaar 1769 voor; 977 een jaar, dat nog komen moet; 167 het jaar 1902 en 125 het jaar 1860.

03	47	43	73	86	36	96	47	36	61	46	98	63	71	62
97	74	24	67	62	42	81	14	57	20	42	53	32	37	32
16	76	62	27	66	56	50	26	71	07	32	90	79	78	53
12	56	85	99	26	96	96	68	27	31	05	03	72	93	15
55	59	56	35	64	38	54	82	46	22	31	62	43	09	90
16	22	77	94	39	49	54	43	54	82	17	37	93	23	78
84	42	17	53	31	57	24	55	06	88	77	04	74	47	67
63	01	63	78	59	16	95	55	67	19	98	10	50	71	75
33	21	12	34	29	78	64	56	07	82	52	42	07	44	38
57	60	86	32	44	09	47	27	96	54	49	17	46	09	62
18	18	07	92	46	44	17	16	58	09	79	83	86	19	62
26	62	38	97	75	84	16	07	44	99	83	11	46	32	24
23	42	40	64	74	82	97	77	77	81	07	45	32	14	08
52	36	28	19	95	50	92	26	11	97	00	56	76	31	38
37	85	94	35	12	83	39	50	08	30	42	34	07	96	88

Het eerstvolgende bruikbare getal is 162, wat correspondeert met het jaar 1897. Op deze wijze is een geheel toevallige steekproef verkregen, bestaande uit de jaren 1769, 1902, 1860 en 1897. De zomertemperaturen van deze jaren zijn resp. 15,6, 15,4, 14,6 en 17,3°C. Hun gemiddelde, het steekproefgemiddelde is 15,7°C.

Een tweede steekproef krijgt als eerste element het jaar 1735 + 181. Daarna bevinden zich er voor ons doel geen bruikbare getallen meer onder de 15 getallen, gevormd door de eerste drie kolommen. De vierde, vijfde en zesde kolom geven weer nieuwe getallen, waaruit we 163, 112 en 086 halen ter completering van de tweede steekproef. Deze bestaat dus uit de jaren 1916, 1898, 1847 en 1821 met zomertemperaturen van resp. 14,6, 16,0, 16,3 en 14,7°C. Hun gemiddelde is 15,4°C.

De volgende drietallen kolommen leveren ons de elementen van de derde steekproef: 199, 157, 171 en 160, wat overeenkomt met de jaren 1934, 1892, 1906 en 1895. De temperaturen van de zomers in deze jaren waren resp. 16,6, 15,6, 16,1 en 16,4°C. Het derde steekproefgemiddelde is 16,2°C.

We kunnen deze drie steekproefschema's ook nog gebruiken bij de bepaling van de gemiddelde hoeveelheid neerslag in de zomer.

Steekproef 1			Steekproef 2			Steekproef 3		
Toevals- getal	jaartal	som zomer- neerslag in mm	Toevals- getal	jaartal	som zomer- neerslag in mm	Toevals- getal	jaartal	som zomer- neerslag in mm
034	1769	291	181	1916	236	199	1934	118
167	1902	229	163	1898	213	157	1892	233
125	1860	217	112	1847	199	171	1906	199
162	1897	284	086	1821	131	160	1895	256
Gemiddeld		255	Gemiddeld		195	Gemiddeld		202

Overeenkomstig de theorie in 1.7 zien we al dadelijk, dat deze steekproef-gemiddelden dichter om het universumgemiddelde van 219 mm liggen dan de jaarlijkse zomerhoeveelheden zelf. De theorie in 1.7 zegt immers, dat de standaarddeviatie van de steekproefgemiddelden $\sqrt{n}$ maal zo klein is als de standaarddeviatie in het oorspronkelijke, volgens GAUSS verdeelde universum (n is het aantal elementen in de steekproef gebruikt, dus hier 4). Weliswaar zijn hier-tegen enkele bezwaren aan te voeren, – zo gaat het hier bv. om de steekproef-niet om de universumstandaarddeviatie; ons universum zal niet gaussisch verdeeld zijn – maar deze bezwaren leggen weinig gewicht in de schaal.

De tabellen met toevalscijfers kunnen ons ook op andere wijze aan representatieve steekproeven helpen, nl. door het laagste jaartal (hier 1735) niet te laten corresponderen met het laagste getal (= 000), maar door een andere geschikte keuze te doen.

Het minste rekenwerk moet verricht worden, wanneer men het jaartal X met het toevalsgetal ($X-1000$) laat samenvallen. In dit geval behoeft men dus alleen te letten op de getallen tussen de grenzen 735 en 944.

Het hierboven weergegeven gedeelte van de tabellen met toevalsgetallen van FISHER and YATES geeft voor de eerste (1a) steekproef van vier elementen de volgende jaren: 1844, 1743, 1807 en 1738. De tweede en derde steekproef (2a en 3a) worden met behulp van de daaraan aansluitende toevalsgetallen bepaald. Zo ontstaat dit schema.

Steekproef 1a			
Toevalsgetal	jaartal	gem. zomertemp. in °C	som zomerneerslag in mm
844	1844	15,0	352
743	1743	16,2	152
807	1807	17,2	122
738	1738	15,9	316
	Gemiddeld	16,1	236

Steekproef 2a			
Toevalsgetal	jaartal	gem. zomertemp. in °C	som zomerneerslag in mm
943	1943	16,1	225
785	1785	16,0	230
924	1924	15,2	274
916	1916	14,6	236
	Gemiddeld	15,5	241

Toevalsgetal	jaartal	Steekproef 3a	
		gem. zomertemp. in °C	som zomerneerslag in mm
811	1811	17,0	257
922	1922	15,1	201
736	1736	17,0	163
827	1827	16,0	152
	Gemiddeld	16,3	194

2.1b UNIVERSUM EN STEEKPROEF (population and sample)

Waren de verschillen tussen steekproef- en universumgemiddelden in de zo juist genoemde voorbeelden klein, dan was dat te danken 1) aan de relatief geringe variabiliteit van de klimatologische grootheden in de zomer en 2) aan de onderlinge onafhankelijkheid van die zomers.

Deze twee voorwaarden zijn niet vervuld in een voorbeeld als dat van de dagelijkse neerslaghoeveelheden in Augustus 1950 te De Bilt. Deze 31 hoeveelheden vormen het universum. Als steekproef zij steeds een serie van drie opeenvolgende dagen gekozen.

De eerste steekproef: 1—3 Aug. met resp. 1,0, 3,2 en 12,0 mm.

Gemiddeld: 5,4 mm per dag.

De tweede steekproef: 4—6 Aug. met resp. 27,3, 3,6 en 0,0 mm.

Gemiddeld: 10,3 mm per dag.

De derde steekproef: 7—9 Aug. met elk 0,0 mm.

Gemiddeld: 0,0 mm per dag.

Het universumgemiddelde is 3,6 mm per dag (de totale gemeten regenhoeveelheid in Augustus 1950 bedroeg 113 mm). Men ziet, dat de drie steekproefgemiddelden er vrij belangrijk naast liggen, terwijl hier toch in elke steekproef ca 10% van het materiaal gebruikt werd en in de voorbeelden van 2.1a slechts 2%.

Een ander onjuist gebruik van de steekproefmethode werd in 1.1 vermeld. Het betrof het volgende: Met behulp van dagelijkse loodsbloodwaarnemingen kan men richting en sterkte van de wind in de hogere luchtlagen berekenen. Men handelt nu verkeerd door de uit deze waarnemingen verkregen gemiddelden gelijk te stellen aan de gemiddelden van windrichting en -sterkte in die niveaus over dat gehele jaar.

Dit zou wel juist zijn, wanneer voor elke hoogte de 365 dagelijkse waarnemingen aanwezig waren.

Bevindt de waarnemingsplaats zich in een land met een klimaat, waarin wolken voorkomen, dan kunnen fouten ontstaan. Onder dergelijke klimatologische omstandigheden zal het aantal waarnemingen in de hogere luchtlagen zeker kleiner zijn dan het aantal waarnemingen in de onderste luchtlagen.

Zou nu het al of niet voorkomen van wolken onafhankelijk zijn van de windrichting en -sterkte, dan zouden de resultaten verkregen voor de hoge niveaus alleen maar iets onnauwkeuriger zijn dan die voor de lagere niveaus. Maar daar er zeker wel afhankelijkheid bestaat tussen de wind en het al of niet aanwezig zijn van wolken, zullen er ook reële verschillen optreden tussen de windgemiddelden in de bovenste luchtlagen, berekend uit de loodsballonwaarnemingen en de ware jaarlijkse gemiddelden van deze aerologische grootheden, zoals ze uit radarmetingen afgeleid zouden kunnen worden.

Inderdaad blijkt uit 1.1, dat deze wijze van steekproefnemen onjuist is. De drie formuleringen van BOK (noot p. 10) geven elk duidelijk te kennen, waaraan we de gemaakte fout te danken hebben.

A. WEGENER (1910) was de eerste, die pogingen in het werk stelde, de invloed van deze soort fouten te verkleinen door gebruik te maken van de „verschilmethode”. BLEEKER (1936) bracht niet alleen het twijfelachtige karakter van WEGENER's methode aan het licht, maar gaf bovendien een weg aan, die meer bevrediging schonk. Men mag wel aannemen, dat BLEEKER's „verbeterde verschilmethode” tevens de beste methode is om de niet-representatieve steekproef (het aerologisch waarnemingsmateriaal) om te zetten in een representatieve steekproef.

CONRAD (1944, tweede editie in 1950) geeft in zijn uitgebreid boek „Methods in Climatology” een geval aan, waarin wij kunnen aantonen, hoe gevaarlijk het is de grens tussen steekproef en universum te verdoezelen of zelfs geheel weg te schrappen. Het betreft table 7 (tweede editie), die hier in zijn geheel volgt.

Table 7

Year (1)	Date (2)	Vancouver Day of the year (3)	d (4)	d^2 (5)
1901	Apr. 24	114	+ 6	36
2	12	102	— 6	36
3	15	105	— 3	9
4*	23	114*	+ 6	36
5	Mar. 30	89	—19	361
6	Apr. 12	102	— 6	36
7	29	119	+11	121
8*	28	119*	+11	121
9	21	111	+ 3	9
10	14	104	— 4	16
		1079	+37	781
			—38	
	$m = \text{April 18}$	108	$ \bar{d} = 7,5$	$9,32 = s$

	Date (6)	Bismarck Day of the year (7)	d (8)	d^2 (9)
1901	June 7	158	+24	576
2	April 29	119	-15	225
3	May 5	125	-9	81
4*	14	135*	+1	1
5	12	132	-2	4
6	27	147	+13	169
7	14	134	0	0
8*	2	123*	-11	121
9	13	133	-1	1
10	17	137	+3	9
		1343	+41	1187
			-38	
	$m = \text{May 14}$	134	$\bar{d} = 7,9$	$11,49 = s$

Dates of the last killing frost in Vancouver, Wash., and in Bismarck, N.D., and deviations (d) from the mean date (m). * = leap year

$$s = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1}}$$

$$(\sum d_i^2)_{Van} = 781 \quad (\sum d_i^2)_{Bis} = 1187$$

$$s_{Van} = \sqrt{\frac{781}{9}} = 9,32 \quad s_{Bis} = \sqrt{\frac{1186}{9}} = 11,49$$

„Thus, the scattering of the date of the last killing frost is about 23 percent greater inland than in the coastal region”.

Deze algemene conclusie trekt CONRAD uit het materiaal van zijn steekproef.

Op dezelfde manier kunnen we nu precies het tegengestelde van CONRAD's bewering „bewijzen”. Dit leiden we uit een andere steekproef af. Deze telt 8 elementen, nl. de gegevens van de jaren 1903—1910.

Voor deze steekproef geldt:

$$s_{Van} = \sqrt{\frac{709}{7}} = 10,1 \quad s_{Bis} = \sqrt{\frac{382}{7}} = 7,4$$

Wanneer wij op onze beurt de grens tussen steekproef en universum mogen uitwissen, dan geeft deze steekproef ons evenveel recht te beweren, dat „the scattering of the date of the last killing frost is about 36 percent smaller inland than in the coastal region”.

Het „Department of Agriculture” in de Verenigde Staten van Amerika geeft

sinds 1936 jaarboeken uit, die speciale gebieden behandelen. In 1941 verscheen „Climate and man”. Daarin bevindt zich een overweldigende hoeveelheid gegevens, o.a. de data van de laatste „killing frosts” in Vancouver, Wash. (p. 1175) en Bismarck, N. D. (p. 1048–9) voor het tijdvak 1899–1938.

Bewerken we deze gegevens op dezelfde wijze als CONRAD in zijn Table 7, dan krijgen we de volgende resultaten.

Tijdvak	Vancouver		Bismarck	
1899–1938	$m_{40} = 89$	$s_{40} = 22,5$	$m_{40} = 130$	$s_{40} = 9,7$
1911–1938	$m_{28} = 83$	$s_{28} = 21,2$	$m_{28} = 128$	$s_{28} = 9,0$
1901–1910	$m_{10} = 108$	$s_{10} = 9,3$	$m_{10} = 134$	$s_{10} = 11,5$
1903–1910	$m_8 = 108$	$s_8 = 10,1$	$m_8 = 133$	$s_8 = 7,4$

CONRAD trok zijn conclusie uit de resultaten van de derde regel. Dat deze conclusie voorbarig was, toonden we aan door de corresponderende parameters te berekenen voor het tijdvak 1903–10. Dat de conclusie waarschijnlijk onjuist is, blijkt nu uit de eerste en tweede regel.

Dat de spreiding in Vancouver significant groter is dan in Bismarck zal aan het slot van 2.9 door de z -toets uitgemaakt worden. Daarmee is CONRAD's conclusie veroordeeld.

Een ander interessant punt is het grote verschil tussen m_{10} en m_{28} in Vancouver. De gemiddelde datum voor de laatste „killing frosts” kwam in de periode 1901–1910 bijna een maand later dan de gemiddelde datum in het interval 1911–1938. De oorzaak van het verschil kan natuurlijk gezocht worden in een andere wijze van waarnemen, bv. verplaatsing van het proefterrein. Men mag echter niet vergeten, de vraag te stellen: Kan het steekproefeffect het verschil verklaren?

De in 2.7.3 berekende overschrijdingskans $= 1\%$ geeft ons statistische zekerheid, dat het verschil tussen m_{10} en m_{28} niet door het toeval veroorzaakt is. Nu is het vermoeden gerechtvaardigd, dat de waarnemingsmethode in de loop der jaren gewijzigd is.

2.2.1 HET BEGRIIP OVERSCHRIJDINGSKANS

Neem aan, dat een meteorologische hypothese het bepalen van de kans p op het voorkomen van een natte eeuw mogelijk maakt. Zij deze kans $= 0,4$. Als we nu het meteorologische gedrag van de laatste 10 eeuwen weten – bv. 8 van de 10 eeuwen waren nat geweest – dan kunnen we de vraag stellen, of deze enkelvoudige experimentele bepaling p_1 van $p(p_1 = 0,8)$ beschouwd mag worden als toevallig afwijkend van de waarde van $p (= 0,4)$, die de theorie voorschrijft.

De frequentiefunctie van het aantal natte eeuwen in een millennium wordt

weergegeven door $f(k) = \binom{n}{k} p^k q^{n-k}$ (de binomiale verdeling of verdeling van BERNOULLI. Zie 1.3.2 en 2.3.2).

In ons geval is $n=10$, $p=0,4$ en $q=0,6$.

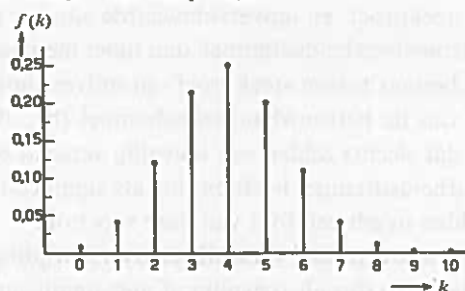


fig. 9

Deze frequentiefunctie is in fig. 9 grafisch voorgesteld. De kansen op elke waarde van k (0—10) zijn in onderstaande tabel aangegeven.

k	$f(k)$
0	0,006
1	0,040
2	0,121
3	0,215
4	0,251
5	0,201
6	0,111
7	0,042
8	0,011
9	0,002
10	0,0001

Na deze noodzakelijke, voorbereidende werkzaamheden kunnen we aan de eigenlijke toets beginnen.

Uit de definitie van overschrijdingskans volgt, dat we daartoe de waarschijnlijkheid van het steekproefresultaat moeten bepalen. Deze is $f(8) = 0,011$. Vervolgens moeten alle kleinere waarschijnlijkheden bepaald worden. Deze zijn $f(0) = 0,006$, $f(9) = 0,002$ en $f(10) = 0,0001$.

Hun som ($= 0,019$) is de dubbelzijdige overschrijdingskans, dus de kans op het voorkomen van afwijkingen met even grote of kleinere waarschijnlijkheden.

Deze overschrijdingskans van 1,9% is zeer gering. Het betekent, dat, als we het gedrag van 100 millennia kenden, er gemiddeld nog geen 2 millennia bij zouden zijn met dergelijke afwijkingen ($k = 8$ of 0, 9 of 10). We kennen slechts het verloop van één geval (de laatste 10 eeuwen). Dat dit éne geval zich juist zeer uitzonderlijk gedraagt, kan moeilijk aan het toeval geweten worden. Er is voldoende reden om te beweren, dat het verschil tussen steekproef- en universumwaarde niet door het toeval verklaard kan worden en dat bijgevolg de gebruikte hypothese – in haar geheel bezien – niet juist kan zijn.

2.2.2 DE BEGRIPPEN SIGNIFICANTIE EN BETROUWBAARHEIDSDREMPEL

Het is duidelijk, dat hoe lager de overschrijdingskans is, hoe zekerder onze conclusie zal zijn.

Wil men dus niet het risico lopen, een toevallig verschil tot realiteit te moeten

verklaren, dan stelle men een zo laag mogelijke waarde voor de betrouwbaarheidsdrempel vast. Levert nu de steekproef een overschrijdingskans P groter dan de aangenomen waarde voor de betrouwbaarheidsdrempel, dan mag men het verschil tussen steekproef- en universumwaarde aan het toeval toeschrijven. Ligt P onder de betrouwbaarheidsdrempel, dan moet men vaststellen, dat er een significant verschil bestaat tussen steekproef- en universumwaarde.

De lage waarde van de betrouwbaarheidsdrempel (bv. $10/0$, $2,70/00$ of $10/00$) heeft tot voordeel, dat slechts zelden een toevallig verschil een overschrijdingskans $<$ betrouwbaarheidsdrempel heeft en dus als significant aangemerkt moet worden. Slechts zelden treedt een fout van deze soort op.

Het nadeel is, dat vaak een reëel verschil een overschrijdingskans $>$ betrouwbaarheidsdrempel heeft en dus als toevallig of niet-significant aangemerkt moet worden. Fouten van deze soort treden dus vrij geregeld op bij een lage waarde van de betrouwbaarheidsdrempel.

Bij algemene statistische onderzoeken verdient het daarom aanbeveling het juiste midden op te zoeken. Men neemt aan, dat het bij 0,05 ligt.

Bij bijzondere onderzoeken verlegt men de grens.

Zo bv. bij het opstellen van een verwachting ten behoeve van een speciale gelegenheid. Het mislukken van deze verwachting zou een catastrofe kunnen veroorzaken. Het is duidelijk, dat men dan bij het toetsen van de beschikbare hypothesen een zo laag mogelijke drempelwaarde moet aannemen. Anders loopt men kans, de verwachting te baseren op een aantal hypothesen van twijfelachtige waarde.

Hiertegenover staat het geval van de tabaksplanters bij Djokjakarta. Zij vinden bij een juiste verwachting van het intreden van de regens bij het begin van de westmoesson zulk een baat, dat de nadelige gevolgen van het totaal mislukken van zo'n verwachting op lange termijn ruimschoots gecompenseerd worden door de extra-opbrengsten bij het redelijk slagen van een dergelijke verwachting in twee andere jaren.

In dit geval kan de statisticus gebruik maken van een hoge waarde van de betrouwbaarheidsdrempel (0,20 of zelfs 0,40). Weliswaar vindt hij dan een aantal significante resultaten, die niet reëel zijn, maar dit aantal wordt overschaduwd door het aantal significante resultaten, die wel een reële basis hebben, maar die niet ontdekt zouden zijn, als de grens bij 0,05 lag.

Een lage grens (bv. 0,001) ontnemt aan vele reële samenhangen het significante karakter. De verwachting heeft nu een smalle basis en kan dus maar weinig gedetailleerd zijn.

In het algemeen zal de opsteller van de verwachtingen op lange termijn een lagere drempel moeten vaststellen dan zijn collega van de 36-uur-verwachtingen.

Dit blijkt uit het volgende: neem eens aan, dat beiden een zelfde grens gebruikten. Voor beiden geldt dan, dat er bv. van zeven verwachtingen één mislukt.

De synopticus zal door zijn mislukte prognose firma A. een schade bezorgen van bv. f10.000,—. De volgende zes verwachtingen, die gemiddeld genomen alle zullen slagen, geven deze firma een voordeel van f20.000,—. De geslaagde verwachtingen volgen zo snel op de mislukte, dat de nadelige gevolgen van de foutieve prognose spoedig worden opgevangen.

Geheel anders staat het met de seizoenvoorspeller. Zijn mislukte verwachting zal firma A. een schade bezorgen van f1.000.000,— (bv. bij een verwachting voor de winter = ongeveer 100 dagen). Weliswaar zullen de volgende zes winters dan een winst brengen van f2.000.000,— maar er zullen slechts weinig crediteuren gevonden kunnen worden, die zo lang op het voldoen van hun vorderingen willen wachten.

De zomer van 1947 was in Europa zo uitzonderlijk, dat vele klimatologen zich afvroegen, wat nu wel de kans zou zijn op het voorkomen van zulk een extremum. Om deze vraag te kunnen beantwoorden moet het universum geheel of gedeeltelijk bekend zijn.

Het ideaal zou bereikt zijn, wanneer de theorie ons de frequentiefunctie van het universum kan leveren. We moeten ons echter tevreden stellen met een steekproef van 100 à 200 elementen.

BIDER (1948) vergeleek de temperaturen van de maanden April—September 1947 in Basel met die van de reeks van 1826—1945 (zie tabel 1).

Tabel 1

Temperaturen te Basel											
	April	Mei	Juni	Juli	Aug.	Sep.	Jn-Au	Jl-S	Jn-S	M-S	Ap-S
1)	9,4	13,6	17,0	18,7	17,9	14,5	17,9	17,0	17,0	16,3	15,2
2)	4,3	4,8	3,4	3,2	3,8	3,8	2,0	3,0	2,1	2,1	2,0
3)	3,7	2,5	3,3	3,1	4,1	3,5	3,5	3,5	3,5	3,3	3,4
4)	2,4	1,5	2,4	2,0	2,8	2,4	3,7	3,3	3,9	4,0	4,3
5)	16	140	16	45	5	16	0,2	1	0,1	0,06	0,02

Eerste rij: gemiddelde temperatuur 1826–1945 = T_{gem}

Tweede rij: grootste positieve afwijking van dat gemiddelde, voorgekomen in het tijdvak 1826–1946.

Derde rij: de afwijking van 1947 = $T_{1947} - T_{gem}$

Vierde rij: $\frac{T_{1947} - T_{gem}}{s}$ (s = standaarddeviatie voor 1826–1945)

Vijfde rij: dubbelzijdige overschrijdingskansen in ‰, bepaald met behulp van het diagram in fig. 5 (p. 21).

De individuele maanden hebben overschrijdingskansen tussen 140 (Mei) en 50/100 (Augustus). Daar de reeks 120 jaren telt, is de overschrijdingskans van Augustus niet als uitzonderlijk laag te beschouwen.

Geheel anders is het bij de combinatie van opvolgende maanden. De kroon spant het tijdvak April—September met 0,029/100.

Dit zou betekenen, dat een dergelijke reeks warme maanden slechts éénmaal in 50.000 jaren voorkomt. Dit hoge aantal demonstreert wel duidelijk de grote uitzonderlijkheid van de zomer in 1947.

Liever dan dit als een toevallige afwijking te beschouwen, zien we dit als een aanwijzing, dat de zomertemperaturen niet gaussisch verdeeld zijn. Dat de verschillen tussen de werkelijke verdeling van de zomertemperaturen en de best aangepaste gaussverdeling niet gering zijn, zal op een andere wijze in 2.3.1 blijken.

2.2.3 SELECTIEEFFECT

Het was wenselijk, met de bespreking van het selectieeffect reeds in 1.2.3 te beginnen. De verdere behandeling vindt plaats in 3.2.

2.3 ENKELE VEEL VOORKOMENDE VERDELINGEN

2.3.1 DE VERDELING VAN GAUSS (zie noot op p. 23)

De frequentieverdeling van de zomertemperatuur in De Bilt voor het tijdvak 1735–1944 wordt in 2.4 bepaald. We willen nu al gebruik maken van die verdeling, maar dan in gewijzigde vorm. De intervallen zijn daar 0,1°C groot. De hier te bespreken verdeling heeft intervallen ter grootte van 0,5°C.

Kolom *a* geeft een aanduiding van de grenzen van de intervallen in °C. In interval 16,2 – 16,6 bevinden zich alle zomers, waarvoor een gemiddelde temperatuur

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
— ∞ — 14,1	2	— ∞	— 0,5000	0,0113	2,4	— 0,4
14,2 — 14,6	4	— 2,28	— 0,4887	0,0323	6,8	— 2,8
14,7 — 15,1	17	— 1,71	— 0,4564	0,0835	17,5	— 0,5
15,2 — 15,6	40	— 1,14	— 0,3729	0,1572	33,0	7,0
15,7 — 16,1	54	— 0,57	— 0,2157	0,2157	45,3	8,7
16,2 — 16,6	35	0,57	0,2157	0,2157	45,3	— 10,3
16,7 — 17,1	35	1,14	0,3729	0,1572	33,0	2,0
17,2 — 17,6	13	1,71	0,4564	0,0835	17,5	— 4,5
17,7 — 18,1	3	2,28	0,4887	0,0323	6,8	— 3,8
18,2 — 18,6	6	2,85	0,4978	0,0091	1,9	4,1
18,7 — ∞	1	∞	0,5000	0,0022	0,5	0,5
	210			1,0000	210,0	+ 22,3 — 22,3

van afgerond 16,2, 16,3, 16,4, 16,5 of 16,6°C gemeten is. De juiste grenzen van dit interval zijn dus 16,15 en 16,65.

Kolom *b* bevat de aantallen zomers voor elk van de elf in kolom *a* aangegeven intervallen en kan dus als een frequentieverdeling beschouwd worden.

Wanneer we aannemen, dat de verdeling van de zomertemperaturen in het universum met die van GAUSS overeenkomt, dan kan ons steekproefmateriaal inlichtingen geven over de gaussverdeling in het universum. De beste schattingen van de parameters centrum en universumstandaarddeviatie zijn het gemiddelde m ($= 16,15^\circ\text{C}$) en de steekproefstandaarddeviatie s ($= 0,88^\circ\text{C}$). Zij worden berekend in 2.4 en 2.6.

Met behulp van deze gegevens kunnen we de gevraagde kromme van GAUSS tekenen.

De voor dit doel zeer geschikte Appendix VI van CONRAD and POLLAK (1950) helpt ons bij het bepalen van de aantallen zomers, die in elk interval verwacht mogen worden. Beschikt men niet over dit werk, dan kan onze fig. 5 (p. 21) helpen, zoals gemakkelijk is in te zien.

De getallen in kolom *c* zijn $= \frac{u-m}{s} = c$. De grootte u is de abscis van dat uiteinde van een interval, dat het verst van het gemiddelde verwijderd is.

De getallen d_i , e_i , f_i en g_i in de volgende kolommen hangen met die uit kolom *c* als volgt samen.

$$d_i = \frac{1}{\sqrt{2\pi}} \int_0^{c_i} e^{-\frac{1}{2}T^2} dT$$

$e_i = |d_i - d_{i+1}|$, behalve voor de twee centrale intervallen, nl. voor de intervallen 15,7 - 16,1 en 16,2 - 16,6. Voor deze geldt $e_i = |d_i|$.

De e -waarde voor een interval is steeds gelijk aan het oppervlak van de gausskromme, begrensd door de intervaluiteinden. Valt het gemiddelde niet samen met een intervalgrens, dus in een zeker interval, dan moeten de d -waarden (d_l en d_r) van beide grenzen van dit interval bepaald worden. De e -waarde van dit centrale interval is nu $|d_l| + |d_r| = |d_l - d_r|$, d.w.z. weer het oppervlak van de kromme van GAUSS, begrensd door de intervaluiteinden.

Verder geldt voor alle gevallen: $f_i = Ne_i = 210 e_i$ en $g_i = b_i - f_i$.

In fig. 10 zijn de frequenties van de steekproefgegevens en de best aangepaste gausskromme weergegeven.

Het is moeilijk om zich een subjectief oordeel te vormen over de afwijkingen tussen beide verdelingen. Een objectief oordeel levert de G^2 -toets van 1.8. We

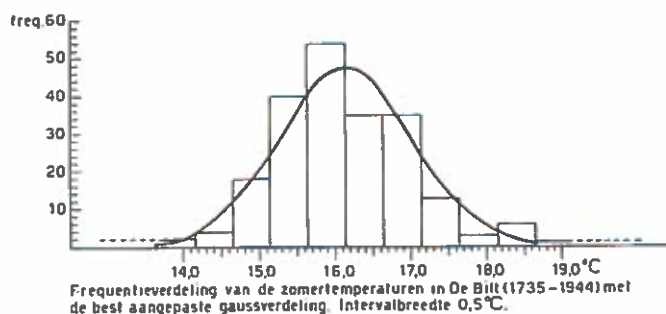


fig. 10

moeten $\sum \frac{g_i^2}{f_i}$ bepalen. Deze som = 7,97. Het aantal vrijheidsgraden $\nu = N - a - b = 8 - 1 - 2 = 5$. $P = 0,16$. Deze overschrijdingskans is niet laag genoeg om de hypothese te verwerpen, die inhoudt, dat de zomertemperaturen te De Bilt in het universum de gaussverdeling volgen.

Een geheel andere methode om de waargenomen verdeling der 210 zomertemperaturen met de aangepaste gaussverdeling te vergelijken is de volgende. Zij gaat uit van een vaste frequentie in de intervallen. Alleen de grenzen van de intervallen moeten nu bepaald worden.

Dit is juist het tegengestelde van de zojuist besproken manier. Daar werden de intervalgrenzen eerst vastgelegd en vervolgens met behulp van tabellen de theoretische frequenties in die intervallen berekend.

Bij de tweede methode is dus de keuze van de frequentie in de intervallen vrij. Laten we als eis stellen, dat elk interval evenveel waarnemingen omvat en wel een aantal, dat dicht bij 10 ligt. Dit aantal van ca 10 houdt verband met de G^2 -toets (1.8). Om het rekenwerk te vergemakkelijken moet dit aantal ook een veelvoud zijn van $\frac{1}{100}$ van het totale aantal, hier dus een veelvoud van 2,1. Het inwilligen van de laatste eis stelt ons in staat gebruik te maken van tabel I van FISHER and YATES (1949). In deze tabel zijn de dubbelzijdige overschrijdingskansen aangegeven van de gaussverdeling met P in veelvouden van 0,01. $P = 0,90$ heeft een absciswaarde = 0,126 of $0,126\sigma$, zo $\sigma \neq 1$. Dit houdt in, dat 90% van het oppervlak van de gausskromme zich bevindt buiten het middenstuk, dat tussen de absciswaarden $-0,126\sigma$ en $+0,126\sigma$ in ligt. Het middenstuk zelf heeft dus een oppervlak van 10%. De intervallen van $-0,126\sigma$ tot 0 en van 0 tot $0,126\sigma$ moeten elk de helft, dus 5% van de totale frequentie herbergen.

Bij $P = 0,80$ is in de tabel I van FISHER and YATES de waarde 0,253 aangegeven. Het middenstuk met de grenzen $0,253\sigma$ heeft een oppervlak van 20%. De intervallen van $-0,253\sigma$ tot $-0,126\sigma$ en van $0,126\sigma$ tot $0,253\sigma$ omvatten ook elk

5% van het aantal waarnemingen, zoals gemakkelijk is in te zien. De grenzen van de overige intervallen kunnen eveneens uit de tabel I (FISHER and YATES) overgenomen worden. Het resultaat wordt:

$-\infty$	tot	$\mu - 1,645 \sigma$	$-\infty$	tot	14,70	6	Voor μ en σ worden het gemiddelde $m = 16,15$ en de standaarddeviatie $s = 0,88$ genomen. De middelste kolom geeft de uiteinden der intervallen, uitgedrukt in °C. De laatste kolom bevat de in die intervallen aanwezig zijnde aantallen zomers afgeleid uit tabel 4 (p. 67). Nauwkeurig kan deze methode in dit geval niet zijn, daar de zomertemperaturen slechts tot in tienden van °C bekend zijn. Zo kwamen de twee zomers met een
$\mu - 1,645 \sigma$	tot	$\mu - 1,282 \sigma$	14,70	tot	15,02	10	
$\mu - 1,282 \sigma$	tot	$\mu - 1,036 \sigma$	15,02	tot	15,24	13	
$\mu - 1,036 \sigma$	tot	$\mu - 0,842 \sigma$	15,24	tot	15,41	17	
$\mu - 0,842 \sigma$	tot	$\mu - 0,674 \sigma$	15,41	tot	15,56	8	
$\mu - 0,674 \sigma$	tot	$\mu - 0,524 \sigma$	15,56	tot	15,69	9	
$\mu - 0,524 \sigma$	tot	$\mu - 0,385 \sigma$	15,69	tot	15,81	16	
$\mu - 0,385 \sigma$	tot	$\mu - 0,253 \sigma$	15,81	tot	15,93	16	
$\mu - 0,253 \sigma$	tot	$\mu - 0,126 \sigma$	15,93	tot	16,04	13	
$\mu - 0,126 \sigma$	tot	μ	16,04	tot	16,15	9	
μ	tot	$\mu + 0,126 \sigma$	16,15	tot	16,26	9	
$\mu + 0,126 \sigma$	tot	$\mu + 0,253 \sigma$	16,26	tot	16,37	6	
$\mu + 0,253 \sigma$	tot	$\mu + 0,385 \sigma$	16,37	tot	16,49	4	
$\mu + 0,385 \sigma$	tot	$\mu + 0,524 \sigma$	16,49	tot	16,61	16	
$\mu + 0,524 \sigma$	tot	$\mu + 0,674 \sigma$	16,61	tot	16,74	5	
$\mu + 0,674 \sigma$	tot	$\mu + 0,842 \sigma$	16,74	tot	16,89	3	
$\mu + 0,842 \sigma$	tot	$\mu + 1,036 \sigma$	16,89	tot	17,06	18	
$\mu + 1,036 \sigma$	tot	$\mu + 1,282 \sigma$	17,06	tot	17,28	16	
$\mu + 1,282 \sigma$	tot	$\mu + 1,645 \sigma$	17,28	tot	17,60	4	
$\mu + 1,645 \sigma$	tot	∞	17,60	tot	∞	12	

gemiddelde temperatuur = 14,7°C in het tweede interval terecht. De 10 zomers van het tweede interval hebben temperaturen, waarvan de begrenzingsgrenzen dus eigenlijk 14,65 en 15,05°C zijn en niet 14,70 en 15,02. De daardoor ontstane verschillen zullen niet groot zijn. Ze verdwijnen geheel, wanneer de temperaturen tot in honderdsten van °C bekend zijn. Een betere toepassing van deze tweede methode wordt in 2.6.1 gegeven.

Een voordeel van de tweede methode is het geringe rekenwerk, dat nodig is voor de bepaling van de G^2 -som (zie 1.8 en 2.8). Immers, de noemer in alle breuken is nu constant (= 10,5). Verder behoeven we de aantallen in de laatste kolommen slechts te kwadrateren en te sommeren (uitkomst = 2664). Aftrekking (zie soortgelijke bewerking in 2.6) van de correctiefactor $210^2/20 = 2205$ en deling door de gemeenschappelijke noemer geeft $459/10,5 = 43,7$. Het aantal vrijheidsgraden is $N - a - b = 20 - 1 - 2 = 17$. Daarbij behoort een overschrijdingskans = 0,0005. Deze is zo gering, dat het reeds aan het slot van 2.2.2 uitgesproken vermoeden, dat de zomertemperaturen in De Bilt niet volgens GAUSS verdeeld zijn, nu afdoende bewezen is.

Het mag geen verbazing wekken, dat de overschrijdingskans van de G^2 -toets bij de eerste methode veel hoger was (nl. 0,25), want daar werden de theore-

tische en waargenomen aantallen van slechts 8 intervallen met elkaar vergeleken. Bij de tweede methode is het gehele materiaal verdeeld over niet minder dan 20 intervallen. De daarbij naar voren getreden afwijkingen tussen verwachte en waargenomen waarden heffen elkaar gedeeltelijk op, wanneer intervallen samengevoegd worden.

2.3.2 DE VERDELING VAN BERNOULLI (Ook genoemd binomiale verdeling)

Dr H. TEN KATE, adjunct-directeur van de Afdeling Klimatologie, was zo vriendelijk om een tabel met de maandelijks aantallen onweersdagen uit het tijdvak 1889–1944 ter beschikking te stellen. Voor elke maand en ook voor het jaar werden de frequentieverdelingen opgemaakt. Het resultaat is in tabel 2 opgesteld.

Wanneer achtereenvolgende dagen, wat betreft het voorkomen van onweer, als onderling onafhankelijk beschouwd mogen worden en de jaarlijkse gang van de onweerskansen binnen de afzonderlijke maanden verwaarloosd kan worden, dan moeten de frequentiefuncties in het universum binomiale verdelingen zijn.

We zullen nu eens zien, of de binomiale door gauss- of poissonverdelingen te vervangen zijn. De benadering door een gaussverdeling eist een grote n . Dit komt alleen voor in de laatste kolom van tabel 2, die voor het jaar. Daar is $n = 365,25$. Inderdaad zien we op het oog al, dat de frequentieverdeling van de aantallen onweersdagen per jaar vrij goed door een gaussverdeling weergegeven kan worden.

Bij de maandelijks aantallen is hiervan alleen sprake bij de maanden Mei – Augustus, want de onweerskansen zijn dan ongeveer 0,5. Immers, wanneer p en q niet veel verschillen, wordt de gaussverdeling eerder benaderd.

Voor de overige maanden komt ook de poissoncurve niet in aanmerking, daar de kans op het optreden van onweer zelfs op een winterdag nauwelijks onder 0,1 daalt.

Ten bewijze van deze beweringen geven we in tabel 3 naast elkaar de frequentieverdeling van de onweersdagen en de daaraan aangepaste bernoulli- en poissonverdelingen en wel voor de maanden Januari, Februari, Maart, October, November en December.

We zien uit de vergelijking tussen B- en P-kolommen, dat vervanging van de bernoulliverdeling door een poissonverdeling aanleiding geeft tot afwijkingen, die niet zonder meer verwaarloosd mogen worden. Daar bovendien het rekenwerk bij de bernoulliverdeling nauwelijks groter is dan bij de poissonverdeling, mag de vraag gesteld worden, waarom we dan nog gebruik zouden maken van deze laatste verdeling. Bij de beantwoording moeten we letten op het verschil in de gebruikte grootheden in de twee formules. In de éne (Bernoulli) n en p , in

Tabel 2

Aantal onweersdagen in de maanden	J	F	M	A	M	J	J	A	S	O	N	D	Jaarlijkse aantal onweersdagen	
0	4	14									3	11		
1	11	8	6			1				2	4	7		
2	14	8	8			1		1	3	6	9	11	60—4	1
3	12	8	5	3		—		1	1	4	6	8	65—9	—
4	4	10	7	3		—		—	2	2	7	3	70—4	—
5	6	2	12	6	2	—		—	3	2	6	5	75—9	—
6	3	1	7	7	—	1		—	3	—	5	6	80—4	1
7	1	1	1	4	3	1		—	5	9	10	2	85—9	5
8	—	3	3	3	1	1	1	1	2	7	2	2	90—4	5
9	1	—	2	4	1	5	—	1	5	4	—	—	95—9	6
10		—	—	4	5	2	5	1	5	4	2	1	100—4	7
11		—	2	9	3	1	1	5	8	3	1		105—9	7
12		1	2	4	4	6	3	4	3	2	—		110—4	7
13			1	2	6	6	3	1	3	3	1		115—9	8
14				5	7	5	5	5	4	3			120—4	2
15				—	2	4	3	9	3	1			125—9	2
16				1	4	2	6	1	—	2			130—4	4
17				1	7	5	8	5	3	1			135—9	1
18					4	5	3	5	—	1				
19					2	4	5	4	1				Totaal	5967
20					—	1	2	6	1				Jaargem.	107
21					2	2	1	3	—				Maandgem.	8,88
22					1	2	4	—	—				Daggem.	0,292
23					—	1	3	2	1					
24					1	—	3	1						
25					—	1		—						
26					—			—						
27					1			1						
Totaal	155	147	275	497	793	816	930	898	573	459	258	166		
Maandgem.	2,77	2,62	4,91	8,88	14,16	14,57	16,62	16,04	10,23	8,20	4,61	2,96	in de 56 jaren	
Daggem.	0,089	0,093	0,158	0,296	0,457	0,486	0,536	0,517	0,341	0,264	0,154	0,096	1889—1944	

de tweede (Poisson) hun product, nl. μ . Nu kan het voorkomen, dat wel μ bepaald kan worden, maar niet p en n afzonderlijk. In dat geval kunnen geen bekende waarden in de bernoullifunctie gesubstitueerd worden en moet de poissonfunctie gebruikt worden. Een voorbeeld volgt in 2.3.3.

Tabel 3 levert ons nog meer vergelijkingsmogelijkheden. We zien, dat de opgetreden frequentieverdeling der onweersdagen noch door een bernoulli-, noch door een poissonverdeling bevredigend wordt weergegeven. Alleen Januari

Tabel 3

Januari		Februari		Maart		October		November		December	
B	P	B	P	B	P	B	P	B	P	B	P
0	4	3,08	3,52	14	3,65	4,06	—	0,27	0,41	—	0,00
1	11	9,37	9,73	8	10,47	10,65	6	1,56	2,03	2	0,05
2	14	13,77	13,47	8	14,48	13,98	8	4,40	4,98	6	0,25
3	12	13,05	12,43	8	12,86	12,22	5	8,00	8,15	4	0,86
4	4	8,96	8,60	10	8,23	8,02	7	10,54	10,00	2	2,16
5	6	4,74	4,76	2	4,05	4,21	12	10,72	9,82	2	4,19
6	3	2,02	2,20	1	1,59	1,84	7	8,74	8,04	—	6,53
7	1	0,71	0,87	1	0,51	0,69	1	5,87	5,64	9	8,38
8	—	0,21	0,30	3	0,14	0,23	3	3,32	3,46	7	9,03
9	1	0,05	0,09	—	0,03	0,07	2	1,58	1,89	4	8,30
10	—	0,01	0,03	—	0,01	0,02	—	0,66	0,93	4	6,56
11	—	0,00	0,01	—	0,00	0,00	2	0,24	0,41	3	4,50
12				1			2	0,07	0,17	2	2,70
13							1	0,02	0,06	3	1,42
14								0,02	0,66	1,09	—
15										0,00	0,01
16								1	0,27	0,60	
17								2	0,10	0,31	
18								1	0,03	0,15	
								1	0,01	0,07	
Totaal											
56 55,97 55,99 56 56,02 55,99 56 55,99 56,01 56 56,00 55,98 56 55,99 56,00 56 55,96 56,02											

maakt een gunstige uitzondering. Een G^2 -toets zal geen significant verschil kunnen aantonen.

Voor de andere vijf maanden vallen dadelijk belangrijke afwijkingen tussen waargenomen en bernoulliverdeling in het oog, die zeker significant zullen zijn. In het algemeen kan men zeggen, dat de staartfrequenties door de theorie te laag worden aangegeven en de centrale waarden te hoog. Voorts hebben bijna alle 12 maandverdelingen een bimodaal (tweetoppig) uiterlijk.

Deze interessante eigenschappen zullen een diepergaand statistisch onderzoek lonend maken, al blijven het geringe aantal waarnemingsjaren en een zekere mate van onbetrouwbaarheid in de gegevens moeilijk te overwinnen hindernissen.

Zojuist hebben we bezwaren geopperd tegen de veronderstelling, dat de frequentieverdeling der maandelijke aantallen onweersdagen door een bernoulli- of poissonverdeling weer te geven zou zijn. Nemen we voor een ogenblik aan, dat deze bezwaren niet steekhoudend zijn, dan levert een hernieuwde beschouwing van tabel 3 de bevestiging van enkele theoretische resultaten (slot 1.3.2).

In de wintermaanden zijn de verdelingen buitengewoon scheef, wat te danken

is aan de kleinheid van p , resp. μ . In de verdelingen van de zomermaanden en in die voor het gehele jaar verdwijnt de scheefheid. Voor de zomermaanden vinden we de oorzaak daarvan in het geringe verschil tussen p en q , voor de jaarverdeling in de grootte van n ($= 365,25$).

Niet alleen een verzameling van frequenties, dus een frequentieverdeling, kan met behulp van de bernoulliverdeling onderzocht worden, ook een afzonderlijke frequentie kan daarmee statistisch beoordeeld worden.

Als voorbeeld nemen we de temperaturenreeks van 210 achtereenvolgende zomers. x zomers zullen aan de warme kant en $(210-x)$ aan de koude kant van de normaal liggen. Zijn de kansen op de positieve en negatieve afwijkingen in het universum gelijk (onze nulhypothese), dan gehoorzaamt x aan de binomiale verdeling. De waarschijnlijkste waarde van x is 105. We vinden voor De Bilt in de periode 1735–1944 een waarde voor x van 93.

De steekproevenetheorie leert, dat we de nulhypothese nu niet onmiddellijk mogen verwerpen. De kans op het optreden van 93 of minder warme zomers kan met gebruikmaking van de formule voor de standaarddeviatie $\sigma = \sqrt{npq}$ bepaald worden. Wegens de symmetrie ($p = q$) en grote n is vervanging van de binomiale verdeling door de gaussische toegestaan. $\sigma = \sqrt{210 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 7,25$.

De absolute afwijking $105 - 93 = 12$ komt overeen met een relatieve afwijking $T = 12 : 7,25 = 1,66$. Bij de gaussverdeling is de overschrijdingskans van een dergelijke relatieve afwijking $= 0,1$, d.i. de kans op een even grote of nog grotere afwijking van hetzelfde of tegengestelde teken.

Deze overschrijdingskans is te groot om de nulhypothese te verwerpen.

2.3.3 DE VERDELING VAN POISSON

In de vorige paragraaf mislukte de vervanging van een bernoulliverdeling door een poissonverdeling, omdat p niet klein genoeg was. In de klimatologie kunnen we echter gebeurtenissen vinden, die waarschijnlijkheden hebben van elke gewenste grootte. Deze mogelijkheid hebben we te danken aan de aanwezigheid van continu variërende grootheden. Want de kans, dat de maximumtemperatuur op een dag $> T_0$ zal zijn, is $p(T_0)$ en de kans op een dagelijkse neerslaghoeveelheid $> R_0$ is $p(R_0)$. $p(T_0)$ en $p(R_0)$ zijn continue functies van T_0 resp. R_0 , zodat bij elke p een T_0 en R_0 te vinden is en bij elke T_0 en R_0 een p .

Een parallel bij onweersdagen bestaat strikt genomen niet. Er zijn geen dagen, waarop de gebeurtenissen $\frac{1}{2}$ onweer, $\frac{2}{5}$ onweer e.d. optreden. Juist, omdat de eigenschap „onweersdag of niet” niet afhangt van de hoeveelheid onweer, was het onmogelijk een met onweer samenhangende grootte te vinden, die een voldoende kleine p had om het gebruik van de poissonverdeling te rechtvaardigen. (Natuurlijk vinden we wel een voldoende kleine p , wanneer we overgaan tot het

beschouwen van het voorkomen van onweer gedurende tijdseenheden, die kleiner zijn dan een etmaal).

BROOKS and CARRUTHERS (1944,5) geven het voorbeeld van de dagen met meer dan 30 mm neerslag in de periode Mei – September voor de 61 jaren 1871–1931 (Londen). Er kwamen 28 van zulke dagen voor in het gehele tijdvak.

De overeenstemming tussen waargenomen en poissonverdeling is goed. Nog beter is die tussen de beide bernoulli- en de poissonverdelingen.

x	y	Poisson	Bernoulli	
			$n = 150$	$n = 120$
0	41	38,55	38,52	38,52
1	16	17,69	17,74	17,75
2	2	4,06	4,06	4,05
3	—	0,62	0,61	0,61
4	2	0,07	0,07	0,07
	61	60,99	61,00	61,00

x = aantal dagen met meer dan 30 mm neerslag, per zomer.

y = aantal zomers met x achtereenvolgens = 0, 1, 2, 3 en 4.

De eerste bernoulliverdeling heeft $n = 150$ en $p = \frac{28}{61 \times 150} = 0,00306$; de tweede $n = 120$ en $p = \frac{28}{61 \times 120} = 0,00383$. Het gebruik maken van de eerste bernoulliverdeling is correct, wanneer dagen met meer dan 30 mm neerslag onafhankelijk van elkaar zijn. Vertonen dergelijke dagen persistentie, dan moet voor n het effectieve aantal onafhankelijke dagen (1.12.2) gesubstitueerd worden. In de tweede bernoulliverdeling namen we de lage schatting $n = 120$. De onderlinge verschillen zijn niet noemenswaard.

2.4 CENTRUM EN GEMIDDELDE (universum- en steekproefgemiddelde)

Moeilijkheden van begripsmatige of rekenkundige aard zullen zich hier nauwelijks voordoen.

Heeft men te maken met waarnemingsuitkomsten, die ver uiteen liggen en die bijgevolg zelden gelijke uitkomsten geven, dan is de voor de hand liggende methode, alle getallen op de gewone wijze op te tellen. Hun som gedeeld door het aantal waarnemingen levert het gemiddelde op.

Zijn er veel gelijke waarnemingsuitkomsten, dan loont het vaak de moeite de frequentieverdeling te bepalen. We hebben dit gedaan met de zomertemperaturen van 1735 – 1944 uit tabel A van LABRIJN (1945).

Het resultaat is te zien in onze tabel 4.

Het uittellen van de frequenties stelt ons onmiddellijk in staat een grafiek te maken en zo een beter inzicht te verkrijgen in de verdeling der zomertemperaturen.

Tabel 4

14,0°C	1	15,0°C	4	16,0°C	13	17,0°C	13	18,0°C	1
1	1	1	7	1	9	1	9	1	—
2	1	2	6	2	9	2	7	2	1
3	1	3	10	3	6	3	2	3	—
4	—	4	7	4	4	4	2	4	3
5	—	5	8	5	8	5	—	5	1
6	2	6	9	6	8	6	2	6	1
7	2	7	6	7	5	7	—	7	1
8	3	8	10	8	3	8	1	8	—
9	1	9	16	9	5	9	1	9	—

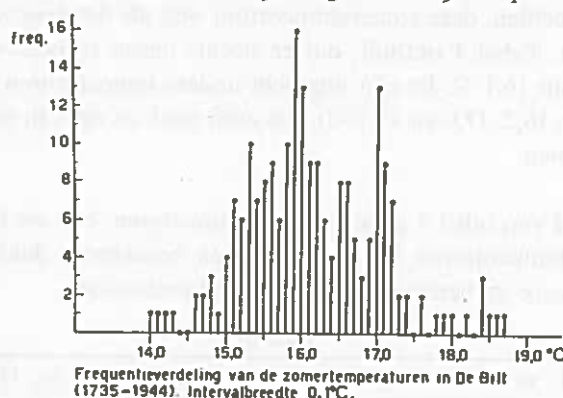


fig. 11

Fig. 11 bevat de grafische voorstelling van tabel 4. Het verrassende is het aanwezig zijn van twee hoge toppen. Twee lagere toppen, resp. bij 15,3 en 16,5 – 16,6 °C. vallen minder op, zodat we hun aanwezigheid waarschijnlijk wel aan het toeval of het steekproefeffect mogen toeschrijven. Gaarne zouden we dit ook met een van de grote toppen willen doen, daar de bimodale (tweetoppige) frequentiefuncties in de geofysica tot nog toe uitzonderingen zijn. Om de invloed van het steekproefeffect te verminderen gaan we het aantal elementen per interval vergroten door de breedte van de intervallen te verdubbelen. In plaats van 0,1°C wordt deze nu 0,2°C. Zo ontstaat fig. 12. Ofschoon elk interval nu gemiddeld tweemaal zo veel elementen telt als in het vorige geval, blijven de twee toppen duidelijk gescheiden. De bimodaliteit van de frequentieverdeling, die ook in de temperaturen van andere maanden en stations gevonden wordt (2.5), biedt de klimatoloog een interessant en nauwelijks ontgonnen terrein.

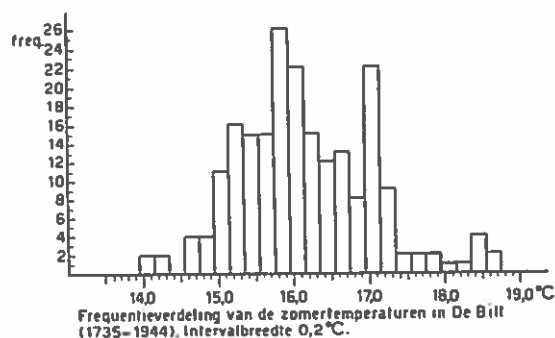


fig. 12

Het gemiddelde van de zomertemperaturen uit het tijdvak 1735 – 1944 is 16,1°C. Hoewel dit gemiddelde steeds de „normaal” genoemd wordt, moet men er zich voor wachten, deze zomertemperatuur ook als de meest waarschijnlijke te beschouwen. Tabel 4 onthult, dat er slechts negen zomers waren met een temperatuur van 16,1°C. Er zijn nog acht andere temperaturen (nl. 15,3, 15,6, 15,8, 15,9, 16,0, 16,2, 17,0 en 17,1°C), die even vaak of zelfs in nog meer jaren zijn voorgekomen.

Aan de hand van tabel 5 gaan we nu demonstreren, hoe we het gemiddelde van de zomertemperaturen het snelst kunnen berekenen, daarbij tegelijk de basis leggend voor de berekening van de standaarddeviatie.

Tabel 5

0a	0b	0c	0d	1a	1b	1c	1d	2a	2b	2c	2d	3a	3b	3c	3d	4a	4b	4c	4d
0	1	—	—	0	4	—	—	0	13	—	—	0	13	—	—	0	1	—	—
1	1	1	1	1	7	7	7	1	9	9	9	1	9	9	9	1	—	—	—
2	1	2	4	2	6	12	24	2	9	18	36	2	7	14	28	2	1	2	4
3	1	3	9	3	10	30	90	3	6	18	48	3	2	6	18	3	—	—	—
4	—	—	—	4	7	28	112	4	4	16	64	4	2	8	32	4	3	12	48
5	—	—	—	5	8	40	200	5	8	40	200	5	—	—	—	5	1	5	25
6	2	12	72	6	9	54	324	6	8	48	288	6	2	12	72	6	1	6	36
7	2	14	98	7	6	42	294	7	5	35	245	7	—	—	—	7	1	7	49
8	3	24	192	8	10	80	640	8	3	24	192	8	1	8	64	8	—	—	—
9	1	9	81	9	16	144	1296	9	5	45	405	9	1	9	81	9	—	—	—
12 65 457				83 437 2987				70 253 1487				37 66 304				8 32 162			

Ter vereenvoudiging van het rekenwerk is in de a-kolommen niet de waargenomen temperatuur genoteerd, maar zijn aantal tienden van °C = a_i . (a_i is dus = het cijfer achter de komma). In de b-kolommen bevinden zich de frequenties = b_i , in de c-kolommen hun producten $c_i = a_i b_i$. Tenslotte wordt dit

product in de d-kolommen nog eens vermenigvuldigd met a_i , zodat $d_i = a_i c_i = a_i^2 b_i$. De gegevens uit de d-kolommen hebben we echter niet nodig bij de berekening van het gemiddelde; zij komen te pas bij het bepalen van de standaarddeviatie in 2.6.

De vijf partiële sommen van de b-, c- en d-kolommen worden onder elkaar geplaatst en opgeteld.

Tabel 6

	b	c	d	e	f	g
0	12	65	457	—	—	—
1	83	437	2987	83	83	437
2	70	253	1487	140	280	506
3	37	66	304	111	333	198
4	8	32	162	32	128	128
	210	853	5397	366	824	1269

De vijf b-sommen worden achtereenvolgens met 0, 1, 2, 3 en 4 vermenigvuldigd. Zo ontstaat kolom e.

Nu geldt: $\frac{\sum c_i + 10 \sum e_i}{\sum b_i} = 10 (t_{gem.} - 14)$

Substitutie geeft:

$$\frac{853 + 3660}{210} = \frac{4513}{210} = 21 = 10 (t_{gem.} - 14)$$

$$t_{gem.} = 14 + 2,1 = 16,1$$

De gemiddelde zomertemperatuur in het tijdvak 1735–1944 is dus gelijk aan 16,1°C.

Omdat dit een naar beneden afgeronde uitkomst is, mogen we zomers met een temperatuur van 16,1°C of minder beschouwen als koude zomers. Warme zomers zijn die met een temperatuur van 16,2°C of meer.

Van de 210 zomers zijn er 93 te warm geweest en 117 te koud. Nemen we aan, dat het universum geen voorkeur heeft voor warme of koude zomers (de nulhypothese), dan heeft deze steekproef een $P = 10\%$ (slot van 2.3.2). De nulhypothese mag dus niet verworpen worden.

2.5 HET AANTAL VRIJHEIDSGRADEN

Als men zich aan de in 1.5 gegeven definitie houdt, is het vrij gemakkelijk steeds het juiste aantal vrijheidsgraden te bepalen.

Wanneer de theorie een verdeling van GAUSS verwacht met bekend centrum en standaarddeviatie, dan kunnen de steekproefresultaten daarmee kwantitatief vergeleken worden door gebruik te maken van de G^2 -toets.

Er behoeft geen enkele parameter uit het steekproefmateriaal geschat te worden, dus $b = 0$.

De enige lineaire voorwaarde is, dat de totale frequenties gelijk zijn (het oppervlak, ingesloten door de gausskromme en de horizontale as moet gelijk zijn aan dat, ingesloten door de grafische voorstelling van de frequentieverdeling). Dit geeft $a = 1$. Hieruit volgt $\nu = N-1$.

Is de theorie slechts in staat te voorspellen, dat er een gaussverdeling optreedt met nog onbekende waarden voor centrum en standaarddeviatie, dan dienen deze twee parameters geschat te worden om toepassing van de G^2 -toets mogelijk te maken. Dus $b = 2$ en a is weer $= 1$, $\nu = N-3$.

BROOKS and CARRUTHERS (1944, 29) vergaten het aantal vrijheidsgraden met het aantal geschatte parameters te verminderen.

Zij hadden eerst de frequentietabel opgesteld van de Augustus-temperaturen te Greenwich (1841-1940). De temperaturen waren over 8 cellen verdeeld, dus $N = 8$; $a = 1$. De G^2 -som bleek $= 14,89$ te zijn. Zij telden 7 vrijheidsgraden (wat een P van 0,04 opleverde) en verbonden aan dit resultaat de conclusie, dat de waarnemingen niet gaussisch verdeeld waren.

Daarna brachten zij echter merkwaardige overwegingen naar voren. „Het feit, dat de afwijkingen tussen de gausskromme en het frequentiepolygoon in het midden van de kromme voorkomen en niet aan de uiteinden, suggereert, dat ze van toevallige aard zijn en in een langere reeks zouden verdwijnen. De beste toets is, de gehele periode in twee gelijke delen te splitsen en de twee frequentieverdelingen te vergelijken. Dit toont, dat de afwijkingen bijna geheel ontstaan zijn in de eerste 50 jaren.”

Deze redenering is op wel haast alle punten aanvechtbaar. Als men een statistische toets op de juiste wijze toepast, moet men zich bij de uitspraak neerleggen en deze niet door drogredenen verkrachten.

Juist het feit, dat de grootste afwijkingen in het midden voorkomen, wekt grote twijfel aan de toevalligheid van hun karakter. De lage frequenties, dus de staartfrequenties leveren geringe bijdragen tot de G^2 -som. Men kan dus niet de aanmerking maken, dat de cellen aan de uiteinden te weinig elementen bevatten en dat de hoge G^2 -som ontstaan is door het benaderingskarakter van de G^2 -methode.

Dat de tweede helft van de waarnemingen zich niet zo uitzonderlijk gedraagt als de eerste helft, behoeft nog geen bewijs te zijn, dat de reeks op de lange duur hetzelfde gedrag zal aannemen als deze tweede helft. Het steekproefeffect kan immers ook nog verklaren, dat de tweede helft een gaussverdeling schijnt te volgen.

De eis, die BROOKS and CARRUTHERS impliciet stellen, nl. dat de Augustus-temperaturen de gaussverdeling volgen, behoeft in het geheel geen plausibele eis te zijn.

BIDER (1948) vond in de reeks van temperaturen te Basel in de zomer ook tweetoppigheid. Dit verschijnsel doet zich eveneens voor bij de zomertemperaturen in Nederland (2.4).

Het betoog van BROOKS en CARRUTHERS wordt helemaal aan de kaak gesteld, wanneer we de G^2 -toets op de juiste wijze uitvoeren.

De G^2 -som blijft natuurlijk gelijk (14,89). Het aantal vrijheidsgraden moet echter met 2 verminderd worden, als men een gevonden frequentieverdeling aanpast aan een gaussverdeling. Dus $v = 5$. De overschrijdingskans is nu 0,01, wat de mogelijkheid van het gaussisch verdeeld zijn van de Augustus-temperaturen te Greenwich zo goed als geheel uitsluit.

2.6 STANDAARDDEVIATIE EN VARIANTIE

De standaarddeviatie van de zomertemperaturen (1735–1944) kan eveneens met behulp van tabel 6 (p. 69) bepaald worden.

We moeten het volgende recept gebruiken:

Laat de f-kolom op dezelfde manier uit de e-kolom ontstaan als de e-kolom uit de b-kolom, d.w.z. vermenigvuldig de getallen uit de e-kolom respectievelijk met 0, 1, 2, 3 en 4. Op dezelfde wijze ontstaat de g-kolom uit de c-kolom. Het is niet moeilijk in te zien, dat de som van de kwadraten van alle afwijkingen ten opzichte van het aangenomen nulpunt ($14,0^\circ\text{C}$) gelijk is aan: $\sum d_i + 100 \sum f_i + 20 \sum g_i = 5.397 + 82.400 + 25.380 = 113.177$.

De bedoeling is echter de kwadratensom van de afwijkingen t.o.v. het steekproefgemiddelde te bepalen. Deze gevraagde som is gemakkelijk – met een eenvoudige correctie – uit de reeds berekende som te herleiden.

Stel de afwijkingen t.o.v. het fictieve nulpunt ($14,0^\circ\text{C}$) = v_i ($i = 1, \dots, 210$).

De afwijkingen t.o.v. het steekproefgemiddelde zijn dan = $v_i - \frac{\sum v_i}{n}$ (n is hier = 210).

De gevraagde kwadratensom is

$$\sum_{i=1}^n \left(v_i - \frac{\sum v_i}{n} \right)^2 = \sum_{i=1}^n v_i^2 - 2 \frac{\sum v_i}{n} \sum v_i + \frac{(\sum v_i)^2}{n} = \sum_{i=1}^n v_i^2 - \frac{(\sum v_i)^2}{n}$$

Deze formule veroorlooft ons, te rekenen met afwijkingen t.o.v. een willekeurig „gemiddelde” (we leiden de willekeur zo, dat het rekenkundige werk minimaal is).

$$\sum_{i=1}^{210} v_i^2 = 113.177$$

$$\sum v_i = 4.513$$

$$\frac{(\sum v_i)^2}{210} = 96.987$$

$$\sum_{i=1}^{210} (x_i - m)^2 = 16.190$$

$$s = \sqrt{\frac{16.190}{209}} = 8,8 \text{ tienden van } ^\circ C = 0,88 ^\circ C$$

2.6.1 DE FREQUENTIEFUNCTIE VAN DE STANDAARDDEVIATIE VAN DE ENKELVOUDIGE WAARNEMING

We gaan de standaarddeviaties bepalen van de Januari-temperaturen (Zwanenburg-De Bilt) in elk der 20 decennia uit de periode 1741–1940. In het eerste decennium (1741–1750) is de som der kwadraten van de afwijkingen van de 10 Januari-temperaturen t.o.v. hun decenniumgemiddelde = 18,87. Derhalve is $s_1 = \sqrt{\frac{18,87}{9}} = 1,448 ^\circ C$. Op soortgelijke wijze kunnen de standaarddeviaties uit de overige 19 decennia bepaald worden. In tabel 7 zijn ze aangegeven, tezamen met $(n-1) \frac{s^2}{\sigma^2}$. Daar σ , de standaarddeviatie van het universum, niet bekend is, hebben we hiervoor de standaarddeviatie s_{200} uit het tijdvak 1741–1940 gesubstitueerd.

De grootheid $(n-1) \frac{s^2}{\sigma^2}$ volgt de G^2 -verdeling. Zo'n verdeling heeft de eigenschap, dat het centrum der elementen gelijk is aan het aantal vrijheidsgraden. De

Tabel 7

	s	$(n-1) \frac{s^2}{\sigma^2}$
1741–50	1,448	2,62
1751–60	2,674	8,95
1761–70	3,366	14,17
1771–80	2,500	7,82
1781–90	2,694	9,08
1791–00	3,164	12,52
1801–10	2,735	9,35
1811–20	2,854	10,19
1821–30	3,711	17,23
1831–40	3,361	14,13
1841–50	2,887	10,42
1851–60	1,793	4,02
1861–70	2,445	7,48
1871–80	2,627	8,63
1881–90	2,352	6,92
1891–00	2,447	7,49
1901–10	1,884	4,44
1911–20	1,960	4,81
1921–30	2,368	7,02
1931–40	3,012	11,35

20 steekproefdecennia geven voor $(n-1) \frac{s^2}{\sigma^2}$ een gemiddelde van 8,932. Het aantal vrijheidsgraden is $n-1 = 9$. Ter verdere controle kunnen we de hoogste en de laagste waarde voor de steekproefstandaarddeviatie beschouwen. $(n-1) \frac{s^2}{\sigma^2}$ is voor de maximale en minimale standaarddeviatie gelijk aan 17,23 en 2,62. Slaan we een G^2 -tabel op, dan zien we, dat bij deze waarden overschrijdingskansen behoren van 5% en 98%. Dit betekent, dat bij uitbreiding van het aantal steekproeven 5% van dat aantal een grotere $(n-1) \frac{s^2}{\sigma^2}$ waarde zal krijgen dan het maximum van 17,23 en 2% van dat aantal een kleinere waarde voor $(n-1) \frac{s^2}{\sigma^2}$ dan het minimum van onze 20 decennia. Beide percentages zijn van een aannemelijke

grootte. Dit onderzoek kan ons derhalve geen aanleiding geven, de waarnemingen van de temperatuur of de statistische theorie te wantrouwen.

In tabel 8 zijn van alle maanden de grootste en kleinste standaarddeviaties opgesteld, benevens de corresponderende waarden van $(n-1) \frac{s^2}{\sigma^2}$ en de overschrijdingskansen.

Tabel 8

	Minimaal			Maximaal		
	s	$(n-1) \frac{s^2}{\sigma^2}$	P	s	$(n-1) \frac{s^2}{\sigma^2}$	P
Jan.	1,448	2,62	98	3,711	17,23	5
Feb.	1,776	5,43	80	3,217	17,82	4
Mar.	1,249	4,05	90	2,643	18,17	3
Apr.	0,938	3,66	93	1,867	14,48	10
Mei	0,933	3,75	92	2,117	19,32	2
Jun.	0,593	1,81	99,5	1,737	15,52	8
Jul.	0,766	2,93	97	1,827	16,69	5
Aug.	0,613	2,47	98	1,754	20,25	2
Sep.	0,663	3,03	96	1,472	14,91	9
Oct.	0,772	2,89	97	1,810	15,90	7
Nov.	0,695	1,69	99,5	2,116	15,73	7
Dec.	1,716	4,41	88	3,606	19,47	2

Alle extreme standaarddeviaties hebben een overschrijdingskans, die van een redelijke grootte is, als we rekening houden met het feit, dat de behandelde standaarddeviaties steeds de hoogste of laagste zijn uit een groep van 20 spreidingsgrootheden.

Tenslotte maken we de frequentietabel op van de 240 waarden voor $(n-1) \frac{s^2}{\sigma^2}$, behorende bij de 20 decennia 1741–1940 en de 12 maanden Jan.–Dec. De theorie eist, dat de frequenties gehoorzamen aan een G^2 -verdeling met $(n-1) = 9$ vrijheidsgraden.

Bij de vergelijking van de theoretische en waargenomen frequenties worden we weer voor de keuze geplaatst, de intervalgrenzen òf bij de waargenomen, òf bij de theoretische verdeling vast te stellen. In 2.3.1 hebben we beide wegen gevolgd. Daar bleek, dat de tweede methode meer voordelen geeft, mits de waargenomen waarden tot op een voldoende aantal decimalen berekend zijn. Dit is hier het geval, zodat we nu de intervalgrenzen gaan bepalen aan de hand van de theoretische verdeling.

In tabel IV van FISHER and YATES (1949) bevinden zich de overschrijdingskansen voor de G^2 -verdeling. Voor 9 vrijheidsgraden heeft $G^2 = 2,088$ een $P = 0,99$. Dit betekent, dat 1% der waarnemingen een G^2 -waarde zal vertonen, die tussen 0 en 2,088 ligt.

$G^2 = 2,532$ heeft een $P = 0,98$. Ook 1% der G^2 -waarden ligt dus tussen 2,088 en 2,532.

$G^2 = 3,325$ heeft een $P = 0,95$. Het volgende interval heeft derhalve de grenzen 2,532 en 3,325. Het moet 3% der waarnemingen omvatten. We kunnen zo doorgaan tot $G^2 = 27,877$ met een $P = 0,001$. G^2 -waarden $> 27,877$ hebben daarom een frequentie $= 10/100$.

In tabel 9 zijn de intervalgrenzen met de daarbij behorende P 's aangegeven en daarna voor elke maand de aantallen decennia, waarvan de $(n-1) \frac{s^2}{\sigma^2}$ -waarde in het desbetreffende interval viel. De volgende kolom geeft het totaal der 12 maanden weer. In de laatste kolom staan de theoretische aantallen. Deze zijn gelijk aan de op bovenstaande wijze afgeleide percentages, vermenigvuldigd met het totale aantal van 240.

Tabel 9

G^2	P	J	F	M	A	M	J	J	A	S	O	N	D		
0	1						1				1		2	2,4	
2,088	0,99						—		1		—		1	2,4	
2,532	0,98	1					2	2	1	1	1	1	9	7,2	
3,325	0,95	1		1	1	1	2	1	—	1	3	—	11	12	
4,168	0,90	2		3	2	6	2	—	1	2	2	4	5	29	24
5,380	0,8	—	4	3	2	1	1	1	4	1	1	2	3	23	24
6,393	0,7	5	6	4	8	4	4	5	4	2	3	3	—	48	48
8,343	0,5	6	4	6	1	3	1	3	5	5	6	2	6	48	48
10,656	0,3	1	2	—	3	—	1	2	—	5	2	3	2	21	24
12,242	0,2	3	2	1	3	3	5	4	2	2	1	2	3	31	24
14,684	0,1	—	—	1	—	—	1	2	1	1	1	2	—	9	12
16,919	0,05	1	2	1		2			—				1	7	7,2
19,679	0,02								1					1	2,4
21,666	0,01														2,16
27,877	0,001														0,24
∞	0														
														240	240

De goede overeenstemming tussen de aantallen in de laatste twee kolommen kan door toepassing van de G^2 -toets ook objectief bevestigd worden.

2.6.2 DE FREQUENTIEFUNCTIE VAN DE STANDAARDDEVIATIE IN GROTE STEEKPROEVEN

Voor grote n kan de G^2 -verdeling van de vorige onderparagraaf door de gaussverdeling vervangen worden. Deze heeft dan een universumstandaarddeviatie

$$= \frac{\sigma}{\sqrt{2n-2}} \text{ en het centrum in } x = \sigma.$$

Als voorbeeld nemen we de steekproefstandaarddeviatie in de perioden van 40 jaren gedurende 1741–1940. In de laatste twee eeuwen hebben we echter slechts vijf perioden van 40 jaren en dit aantal is te klein om de statistische theorie te illustreren. We kunnen ons materiaal vertwaalfvoudigen door alle maanden in het onderzoek te betrekken. De jaarlijkse gang, i.c. de ongelijkheid der 12 maandelijks universumstandaarddeviaties σ elimineren we door eerst voor elke maand de 5 steekproefstandaarddeviaties door de bijbehorende universumstandaarddeviaties σ te delen. Zo krijgen we dan 60 quotiënten, die onderling vergelijkbaar zijn. Hun centrum is 1 en hun standaarddeviatie is $\frac{1}{\sqrt{2n-2}} = \frac{1}{\sqrt{78}} = 0,113$.

In tabel 10 zijn deze 60 quotiënten in intervallen ingedeeld. De grenzen van de intervallen zijn in de eerste regel aangegeven.

Tabel 10

0,80	0,85	0,90	0,95	1,00	1,05	1,10	1,15	1,20
1	6	6	7	11	14	6	5	3

Het kost alleen maar vrij veel rekenwerk om de theoretische verdeling – ter vergelijking met de waargenomen verdeling – ook in tabel 10 op te nemen. We doen dit niet, omdat de kleine aantallen het nemen van een definitieve beslissing toch niet toelaten.

Het grootste quotiënt ($= 1,227$) kwam voor in de periode 1781–1820 en wel voor de maand April. De afwijking ($1,227-1$) t.o.v. het centrum is tweemaal zo groot als de standaarddeviatie ($0,113$) en heeft derhalve een eenzijdige overschrijdingskans P van $2,3\%$ (bepaald uit het diagram van fig. 5, p. 21).

Het kleinste quotiënt ($= 0,788$) viel in de periode 1901–1940 en heeft betrekking op de maand Mei. Het heeft een $P = 3,1\%$. Beide overschrijdingskansen zijn hoger dan we verwachten mochten, nl. ca $1/60 = 1,7\%$.

Verder zullen we alleen het te verwachten aantal in het interval $0,90-1,10$ bepalen. De absolute afwijking $1,10$ staat gelijk met de relatieve afwijking

$$T = \frac{1,10-1}{\sigma} = \frac{1,10-1}{0,113} = 0,883.$$

Deze heeft een dubbelzijdige overschrijdingskans van 38% . Het interval

0,90–1,10 moet derhalve 62% der waarnemingen of wel $0,62 \times 60 = 37,2$ getallen bevatten.

In tabel 10 vinden we een aantal van $7 + 11 + 14 + 6 = 38$. We mogen hieruit wel de conclusie trekken, dat een waarde van 40 voor n in dit verband „groot” genoemd mag worden.

2.7 DE FREQUENTIEFUNCTIE VAN DE STEEKPROEFGEMIDDELDEN

2.7.1 DE FREQUENTIEFUNCTIE VAN DE STEEKPROEFGEMIDDELDEN IN GROTE STEEKPROEVEN

Ook nu mogen we steekproeven met 40 elementen tot de grote steekproeven rekenen. Daardoor kunnen we als benadering van de verdeling der steekproefgemiddelden de gaussverdeling gebruiken. Deze heeft dan als centrum dat van het oorspronkelijke universum en als standaarddeviatie die van het universum, gedeeld door $\sqrt{n}$.

We bepalen voor elke maand de gemiddelde temperaturen m_{40} voor de vijf perioden 1741–80, 1781–1820, 1821–60, 1861–1900 en 1901–40. De uitkomsten zijn niet onderling vergelijkbaar, daar elke maand een ander centrum en een andere standaarddeviatie heeft. Dit nadeel kan verholpen worden door elk gemiddelde te beschouwen t.o.v. zijn centrum en de afwijking daarna te delen door $\sigma/\sqrt{n}$. Voor μ en σ moeten we weer genoeg nemen met de beste schattingen, nl. het gemiddelde en de standaarddeviatie, berekend over het tijdvak 1741–1940. De aldus genormeerde steekproefgemiddelden x hebben een centrum = 0 en een standaarddeviatie = 1.

De waargenomen verdeling der genormeerde afwijkingen x is in tabel 11 opgenomen, de theoretische verdeling eveneens. Deze laatste kan volgens 2.3.1 bepaald worden.

Tabel 11

	waargenomen frequentie	theoretische frequentie
$x < -1,95$	5	1,6
$-1,95 < x < -0,95$	8	8,7
$-0,95 < x < -0,05$	15	18,5
$-0,05 < x < 0,95$	21	20,9
$0,95 < x < 1,95$	9	8,7
$1,95 < x$	2	1,6
	60	60,0

Voor de standaarddeviatie der 60 genormeerde afwijkingen vinden we 1,17. De theorie deed een waarde van 1 verwachten.

In 1.6.2 leerden we, dat de steekproefstandaarddeviaties zelf een universum-

standaarddeviatie hebben $= \frac{\sigma}{\sqrt{2n-2}}$. In ons geval dus $\frac{1}{\sqrt{118}} = 0,092$. De absolute afwijking $1,17-1 = 0,17$ is 1,8 maal zo groot als de bijbehorende standaarddeviatie. Daaraan is een dubbelzijdige overschrijdingskans $P = 7\%$ verbonden.

De drie genormeerde steekproefgemiddelden met de grootste afwijkingen zijn $-2,6$ (December, 1781-1820), $+2,7$ (October, 1821-1860) en $+2,8$ (Januari, 1901-1940). De enkelzijdige overschrijdingskansen zijn resp. $0,47\%$, $0,35\%$ en $0,26\%$. Deze lage waarden staan in scherp contrast met de hoge waarden, die we in de vorige onderparagraaf 2.6.2 vonden bij het onderzoek naar de extreme standaarddeviaties in dezelfde tijdvakken.

Een meer gedetailleerd onderzoek zal moeten uitmaken, of er een oorzaak aan te wijzen is, waarom de periodengemiddelden meer spreiden dan verwacht werd en de standaarddeviaties in die perioden van 40 jaren daarentegen juist minder. Deze uitkomsten zijn koren op de molens der aanhangers van de Brückner-periode van ca 35 jaar. De omvang van deze verhandeling laat niet toe, dieper op dit onderwerp in te gaan.

Is er slechts één steekproef, dan kan met de methode van deze onderparagraaf beslist worden over de vraag, of de steekproef afkomstig kan zijn uit een universum met een zeker centrum.

De steekproef levert ons zijn gemiddelde m en standaarddeviatie s . De laatste wordt aan de universumstandaarddeviatie σ gelijkgesteld. Geeft één of andere hypothese als universumcentrum de waarde μ , dan kan $\frac{m-\mu}{\sigma/\sqrt{n}}$ berekend worden. Daaruit volgt een overschrijdingskans, waarna de hypothese al of niet verworpen wordt.

2.7.2 DE FREQUENTIEFUNCTIE VAN DE STEEKPROEFGEMIDDELDEN IN KLEINE STEEKPROEVEN

Elk van de nu te beschouwen 240 steekproeven bezit als elementen de 10 temperaturen van een zekere maand in één der 20 decennia van 1741-1940. Na berekening van m en s kunnen we het quotiënt $\frac{m-\mu}{s/\sqrt{n}}$ bepalen. Voor μ nemen we weer het gemiddelde uit de periode 1741-1940.

De verdienste van de methode van STUDENT is, dat σ niet geschat hoeft te worden. Men moet alleen de steekproefstandaarddeviatie gebruiken.

De 240 quotiënten $\frac{m-\mu}{s/\sqrt{n}}$ volgen de t -verdeling van STUDENT met $n-1 = 9$ vrijheidsgraden. Om dit te controleren verdelen we de 240 getallen over intervallen, waarvan de grenzen bepaald worden aan de hand van tabel III van FISHER and

YATES (1949). Daar vinden we bij 9 vrijheidsgraden, dat $t = 0,129$ een overschrijdingskans van 0,9 heeft. 10% der quotiënten verwachten we derhalve in het interval $0 \leq t < 0,129$ en wel 5% negatieve en 5% positieve quotiënten. $t = 0,261$ heeft een $P = 0,8$. In interval $0,129 \leq t < 0,261$ mogen we daarom ook 10% (weer 5% negatieve en 5% positieve) der quotiënten verwachten. Op gelijke wijze worden de grenzen der overige intervallen vastgelegd. In tabel 12 is de frequentieverdeling van de STUDENT-quotiënten opgesteld. Negatieve en positieve waarden zijn apart gehouden. De overeenkomst tussen waargenomen en theoretische frequenties is redelijk. Een minder subjectieve uitspraak volgt in 2.8, wanneer de G^2 -toets is toegepast.

Behalve voor dergelijke massale onderzoeken kan de toets van STUDENT natuurlijk ook voor een enkele steekproef toegepast worden, juist zoals aan het slot van 2.7.1 aangegeven is. De vervanging van σ door s behoeft nu niet plaats te vinden.

Stellen we bv. als nulhypothese, dat de jaarlijkse neerslaghoeveelheden in Delft en Scheveningen niet verschillen (geen kusteffect aanwezig). We onderstellen dus

$\mu = 0$. De overige grootheden in $t = \frac{m - \mu}{s/\sqrt{n}}$ worden in tabel 13 berekend.

Het resultaat wordt $t = 4,83$. Daar er $n - 1 = 9$ vrijheidsgraden zijn, heeft deze t -waarde een overschrijdingskans, die iets kleiner is dan 0,001. De geringe P

Tabel 13

	Jaarsom neerslag in mm			Jaarsom neerslag in dm		
	Delft	Scheveningen	D—S	(D—S) ²	D S	D—S
1931	801	705	96	9.216	8 7	1
32	773	747	26	676	8 7	1
33	697	575	122	14.884	7 6	1
34	633	541	92	8.464	6 5	1
35	858	869	— 11	121	9 9	0
36	829	718	111	12.321	8 7	1
37	942	756	186	34.596	9 8	1
38	658	563	95	9.025	7 6	1
39	901	868	33	1.089	9 9	0
40	827	710	117	13.789	8 7	1
			867	104.181		8
$(n-1) s^2 = 104.181 - \frac{867^2}{10} = 29.012$						
$m = 86,7$			$n = 10$	$(n-1) s^2 = 8 - \frac{8^2}{10} = 1,6$		
			$\mu = 0$	$m = 0,8$		
$t = \frac{m - \mu}{s} \sqrt{10} = 4,83$						
$t = \frac{m - \mu}{s} \sqrt{10} = 6$						

dwingt ons de nulhypothese te verwerpen. We zijn er zeker van, dat in Delft meer neerslag valt dan in Scheveningen.

Een enkel woord over de afronding moge hier nog volgen. Gebruiken we mm als eenheid, dan moeten we met grote getallen werken. De spreidingsbreedte is niet minder dan 197. Gaan we over op dm , dan is de spreidingsbreedte 1, dus veel kleiner dan het optimale aantal van 6 (zie 2.11.1). Het verschil in de twee t -waarden (4,83 en 6) is dan ook te groot. Nemen we als eenheid 20 mm, dan is de spreidingsbreedte slechts 10. t wordt nu = 4,79, minder dan 1% afwijkend van de juiste waarde.

2.7.3 SIGNIFICANTIE VAN HET VERSCHIL TUSSEN TWEE STEEKPROEFGEMIDDELDEN (STUDENT'S t -TOETS)

In het slot van 2.7.2 hebben we eigenlijk ook het verschil tussen twee steekproefgemiddelden onderzocht. Daar hebben we nl. paarsgewijs de verschillen bepaald tussen twee elementen, die kennelijk bij elkaar horen. Door deze kunstgreep toe te passen ontstond de steekproef der verschillen. Zo worden de oorspronkelijke waarnemingen wel niet ten volle gebruikt, maar deden we dit wel, dan hadden we met afhankelijke grootheden gewerkt. Immers, de jaarsommen van de neerslag in Scheveningen en Delft lopen tot op zekere hoogte parallel. De gevonden verschillen mogen we wel onafhankelijk van elkaar noemen.

De nu te beschrijven methode maakt gebruik van alle inlichtingen, maar deze moeten dan ook onderling onafhankelijk zijn. Dit kunnen we onderstellen van de 30 Januari-temperaturen y_j uit de periode 1871–1900 en de Januari-temperaturen x_i van 1901–1930. $n_1 = n_2 = 30$. We kunnen berekenen, dat $m_1 = 2,65^\circ\text{C}$ en $m_2 = 2,02^\circ\text{C}$ en ook, dat

$$s^2 = \frac{\sum_{i=1}^{n_1} (x_i - m_1)^2 + \sum_{j=1}^{n_2} (y_j - m_2)^2}{n_1 + n_2 - 2} = 0,65$$

$$t = \frac{m_1 - m_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = 3,045 \quad v = n_1 + n_2 - 2 = 58 \quad P = 0,005$$

Deze waarde ligt ver beneden de gebruikelijke betrouwbaarheidsdrempel van 0,05, zodat we de temperatuurstijging in het tijdvak 1871–1930 significant moeten noemen.

Daar het selectieeffect (3.2) hier een rol gespeeld kan hebben, is het verstandig, het verschil in de gemiddelde Januari-temperatuur van de twee klimaatperioden niet zonder meer als reëel te beschouwen.

Ten slotte geven we het antwoord op de vraag, die aan het eind van 2.1b gesteld werd. We hebben deze gegevens nodig: $m_{10} = 108$, $s_{10} = 9,3$ (betrekking hebbend op de periode 1901–1910), $m_{28} = 83$, $s_{28} = 21,2$ (betrekking hebbend op de periode 1911–1938).

Verschilt nu m_{10} significant van m_{28} ? Het gewogen gemiddelde s van s_{10} en s_{28} bepalen we uit $s^2 = [(n_{10}-1) s_{10}^2 + (n_{28}-1) s_{28}^2] / (n_1 + n_2 - 2) = (9 \times 9,3^2 + 27 \times 21,2^2) / 36 = 359$.

$$t = \frac{m_{10} - m_{28}}{s} \sqrt{\frac{n_{10} n_{28}}{n_{10} + n_{28}}} = \frac{25}{\sqrt{359}} \sqrt{\frac{10 \times 28}{38}} = 3,58 \quad v = 36 \quad P = 0,001$$

Ook deze P -waarde ligt ver beneden de 5%-drempel, zodat we het steekproef-effect bij onze verdere onderzoeken van dit geval mogen uitsluiten.

2.8 DE G^2 -TOETS

Deze toets is in het voorgaande reeds herhaaldelijk toegepast. Men zie 2.3.1. Als laatste voorbeeld zullen we nu de twee totaalkolommen van tabel 12 (p. 78) aan de G^2 -toets onderwerpen. Om te voldoen aan de eis, dat het theoretische aantal in elk hokje > 10 is, moeten we de laatste 5 hokjes samenvoegen. Dan ontstaat tabel 14.

Tabel 14

waargenomen aantal		theoretisch aantal	afwijkingen		kwadraten der afwijkingen	
—	+	in beide	—	+	—	+
15	14	12	3	2	9	4
3	11	12	— 9	— 1	81	1
12	11	12	0	— 1	0	1
13	8	12	1	— 4	1	16
10	17	12	— 2	5	4	25
10	11	12	— 2	— 1	4	1
11	13	12	— 1	1	1	1
14	9	12	2	— 3	4	9
6	13	12	— 6	1	36	1
19	20	12	7	8	49	64
113	127	120	+13	+17	189	123
			—20	—10		
			$G^2 = \frac{189}{12} = 15,8$		$G^2 = \frac{123}{12} = 10,2$	
			$v = 10$		$v = 10$	
			$P = 10\%$		$P = 40\%$	

We mogen derhalve niet beweren, dat de gevonden verdelingen significant afwijken van de theoretische.

Er is een verschil met G^2 -toetsen, die vroeger gebruikt zijn: de som der afwijkingen is hier $\neq 0$. In de meeste gevallen weten we niet, hoeveel elementen in elk interval verwacht worden. Wel weten we van de theoretische verdeling bv. welk percentage elk hokje omvatten zal. Met behulp van het totale aantal waarnemingen weten we dan, hoeveel elementen elk hokje in de theoretische verdeling krijgt, maar het gevolg is, dat de algebraïsche som der verschillen tussen waargenomen en theoretische frequentie $= 0$ is. Derhalve is het aantal vrijheidsgraden één minder dan het aantal hokjes, omdat de frequentie in het laatste hokje niet meer willekeurig te kiezen is.

Bezien we de eerste kolom van tabel 14, dan zien we, dat de frequentie in het laatste hokje wel vrij te kiezen is, want de som van de positieve afwijkingen is niet gelijk aan die van de negatieve afwijkingen. Het aantal vrijheidsgraden v is steeds gelijk aan het aantal intervallen, waarin de frequenties vrij te kiezen zijn, dus in dit geval gelijk aan het totale aantal intervallen n .

De waarschuwing, dat de G^2 -toets alleen betrekking heeft op frequenties, is overbodig, wanneer men steeds de formule $G^2 = \sum \frac{(x_i - X_i)^2}{X_i}$ voor ogen houdt.

Heeft men deze formule bv. op % of cm toegepast, dan wordt G^2 tienmaal zo groot, als men van de genoemde grootheden op $\frac{0}{100}$ of mm overgaat. Daar het aantal vrijheidsgraden gelijk blijft, kan het resultaat, dat eerst insignificant was, vaak alleen door transformatie van de eenheid significant worden. Dit is absurd en daaruit volgt al, dat de G^2 -toets slechts op aantallen of frequenties toegepast mag worden. (CROSSLEY 1948, VAN DER BIJL 1948, DAW 1948).

De G^2 -verdeling kwamen we ook tegen bij het bespreken van de frequentiefunctie van de standaarddeviatie (1.6.1 en 2.6.1). De grootheid $(n-1) \frac{s^2}{\sigma^2}$ volgt de G^2 -verdeling. De 20×12 waarden voor $(n-1) \frac{s^2}{\sigma^2}$ bleken inderdaad aan een G^2 -verdeling te gehoorzamen.

Andere eigenschappen van de G^2 -verdeling leren we uit het volgende kennen. Voor elke maand leverden de standaarddeviaties s van de 20 decennia (1741–1940) evenzovele waarden voor $(n-1) \frac{s^2}{\sigma^2}$ (voor Januari is dit in tabel 7 gebeurd).

Het gemiddelde, $\bar{G}^2$, van deze 20 waarden moet in de buurt van $n - 1 = v = 9$ liggen. Wij vonden voor

	J	F	M	A	M	J	J	A	S	O	N	D
$\bar{G}^2$	8,932	9,444	8,636	8,274	8,678	8,098	9,346	8,655	9,128	8,068	8,580	9,024
$\bar{G}^2 - 9$	-0,068	0,444	-0,364	-0,726	-0,322	-0,902	0,346	-0,345	0,128	-0,932	-0,420	0,024

De G^2 -verdeling heeft het centrum in ν en de standaarddeviatie $\sigma(G^2) = \sqrt{2\nu}$. Voor elke maand zou uit de twintig waarden voor $(n-1) \frac{s^2}{\sigma^2} = G^2$ de steekproefstandaarddeviatie $s(G^2)$ berekend kunnen worden, die dan niet veel van $\sqrt{2\nu}$ mag afwijken.

Wij zullen ingewikkelder paden bewandelen. Het gemiddelde van twintig G^2 -waarden heeft eveneens een centrum $\nu = 9$, maar een standaarddeviatie

$$\sigma(\bar{G}^2) = \frac{\sqrt{2\nu}}{\sqrt{20}} = \sqrt{\frac{18}{20}} = 0,95.$$

Alle afwijkingen $|\bar{G}^2 - 9|$ zijn kleiner dan deze standaarddeviatie, wat op zichzelf al vreemd is, daar bij twaalf afwijkingen er ca vier te verwachten zijn, groter dan de standaarddeviatie. Het gemiddelde, $\bar{G}^2 - 9$, van de twaalf afwijkingen $\bar{G}^2 - 9$ heeft een centrum 0 en een standaarddeviatie $\sigma(\bar{G}^2) = \frac{0,95}{\sqrt{12}} = 0,27$.

We berekenen uit de twaalf afwijkingen een gemiddelde $\bar{G}^2 - 9 = -0,261$ en een standaarddeviatie $s(\bar{G}^2) = 0,45$. Het gemiddelde, $\bar{G}^2$, wijkt derhalve minder van zijn centrum af dan zijn standaarddeviatie $\sigma(\bar{G}^2) = 0,27$. Tussen de steekproefstandaarddeviatie (0,45) der twaalf afwijkingen en hun universumstandaarddeviatie (0,95) is wel een groot verschil.

De overschrijdingskans wordt op de manier van 2.6.1 bepaald. Het blijkt, dat

$$(n-1) \left\{ \frac{s(\bar{G}^2)}{\sigma(\bar{G}^2)} \right\}^2 = 2,47; \nu = n-1 = 11 \text{ en } P = 0,994.$$

Er is slechts een kans van 0,006 op een kleinere waarde voor $(n-1) \left\{ \frac{s(\bar{G}^2)}{\sigma(\bar{G}^2)} \right\}^2$. Het verschil tussen $s(\bar{G}^2)$ en $\sigma(\bar{G}^2)$ is derhalve significant.

We kunnen wel inzien, waarom de steekproefstandaarddeviatie $s(\bar{G}^2)$ kleiner moet zijn dan de berekende standaarddeviatie $\sigma(\bar{G}^2)$. De geringe persistentie in de maandtemperaturen is de oorzaak, dat de twaalf maanden zich niet onafhankelijk gedragen, waardoor tenslotte de spreiding in de twaalf afwijkingen $\bar{G}^2 - 9$ kleiner moet zijn dan de verwachte waarde. Maar het is moeilijk in te zien, dat de geringe persistentie zulk een groot verschil tussen $s(\bar{G}^2)$ en $\sigma(\bar{G}^2)$ veroorzaakt kan hebben.

2.9 DE z -TOETS VAN FISHER

Het eerste voorbeeld ontleen we aan 2.6.1, waar we de twintig decennium-temperaturen 1741–1940 (Zwanenburg–De Bilt) reeds behandeld hebben.

Voor s_1 nemen we de hoogste standaarddeviatie, in een decennium waargenomen en wel voor de maand Januari. Voor s_2 nemen we de standaarddeviatie berekend over de volle 200 jaren. $F = e^{2z} = \frac{s_1^2}{s_2^2} = 1,91$, nl. $1/9$ van de waarde 17,23 voor $(n-1) \frac{s^2}{\sigma^2}$ (zie tabel 7). Verder is $\nu_1 = 9$ en $\nu_2 = 199$.

Tabel V van FISHER and YATES (1949) bevat de z - en F -waarden, behorende bij de overschrijdingskansen 20, 10, 5, 1 en 0,1%. We vinden bij deze kansen achtereenvolgens de door interpoleren bepaalde 9 F 199-waarden 1,37, 1,65, 1,92, 2,47 en 3,20.¹⁾ De laatste decimaal is niet nauwkeurig. De berekende waarde 1,91 heeft derhalve een overschrijdingskans van 5%. In tabel 8 werd dezelfde P gevonden.

Verbazing mag dit resultaat niet wekken, daar de z -verdeling voor grote ν_2 door de G^2 -verdeling benaderd mag worden en voor $\nu_2 = \infty$ zelfs gelijk is aan de G^2 -verdeling.

Nu nemen we de kleinste standaarddeviatie in ogenschouw. Daar steeds $s_1 > s_2$ moet zijn, heeft s_1 in dit geval betrekking op de spreiding over de 200 jaren en s_2 op de minimale spreiding, ooit in een Januari-decennium waargenomen. $\frac{s_2^2}{s_1^2} = 1/9 \times 2,62 = 0,29$ (zie tabel 8).

$$e^{2z} = F = \frac{s_1^2}{s_2^2} = 3,43 \text{ met } \nu_1 = 199 \text{ en } \nu_2 = 9.$$

Uit tabel V van FISHER and YATES leren we door interpolatie, dat deze F -waarde een overschrijdingskans van ca 2% heeft. Dit staat gelijk met een onderschrijdingskans van ca 98%, wat in overeenstemming is met de P -waarde uit tabel 8 (onderschrijdingskans = 1 – overschrijdingskans).

Op dezelfde wijze kunnen ook de overige waarden bewerkt worden.

Interessanter is het geval, waarbij we voor s_1 en s_2 de hoogste en laagste Januari-standaarddeviatie nemen. $F = e^{2z} = \frac{s_1^2}{s_2^2} = \frac{3,711^2}{1,448^2} = 6,57$ en $\nu_1 = \nu_2 = 9$. P ligt nu tussen 1 en 0,1%. Interpolatie kan geen nauwkeuriger waarde geven dan 0,6 – 0,8%.

¹⁾ $\nu_1 F \nu_2$ geeft een waarde voor F aan, behorende bij ν_1 vrijheidsgraden in de teller en ν_2 in de noemer.

Een formele beoordeling van deze lage overschrijdingskans leidt tot verwerping van de nulhypothese, d.w.z. wij zouden dan moeten aannemen, dat de standaarddeviaties s_1 en s_2 behoren bij steekproeven, die *niet* afkomstig zijn uit eenzelfde universum. Een juistere beoordeling houdt rekening met het feit, dat deze twee steekproeven niet willekeurig genoemd mogen worden en dat de z -toets daarom met voorzichtigheid gehanteerd moet worden. Het in 3.2 te behandelen selectieeffect speelt ook hier een rol.

Overschrijdingskansen, berekend met behulp van statistische methoden hebben in het algemeen steeds betrekking op willekeurige, niet uitgezochte steekproeven. Onze twee steekproeven zijn juist vanwege hun maximale resp. minimale spreiding geselecteerd.

Vermenigvuldiging van P met de zgn. selectiefactor (3.2) heft het selectie-effect op, waarna er geen bezwaren meer zijn om de overschrijdingskansen rechtstreeks met de betrouwbaarheidsdrempel te vergelijken.

De selectiefactor is gelijk aan het aantal verschillende paren, dat uit tien afzonderlijke elementen gekozen kan worden, dus $= \frac{1}{2} \cdot 10 \cdot 9 = 45$.

De voor het selectieeffect gecorrigeerde overschrijdingskans is $45 \times (0,6 \text{ à } 0,8) = 27 \text{ à } 36\%$. Deze waarde ligt ruimschoots aan de toevallige kant van de betrouwbaarheidsdrempel van 5% en ook hier is dus generlei reden om aan te nemen, dat de steekproeven uit een verschillend universum afkomstig zijn.

In het volgende tabelletje zijn de verhoudingen tussen hoogste en laagste varianties voor de twaalf maanden aangegeven.

Maand	J	F	M	A	M	J	J	A	S	O	N	D
$F = \frac{s_{\max}^2}{s_{\min}^2}$	6,57	3,28	4,49	3,96	5,15	8,57	5,70	8,20	4,92	5,50	9,31	4,41

Bij de overschrijdingskansen $P = 5, 1$ en $0,1\%$ behoren de $9F9$ -waarden: 3,2, 5,4 en 10,1. Geen enkele F -waarde uit deze tabel bereikt het $0,1\%$ -niveau. Dit betekent, dat alle variantiequotiënten plausible waarden aangenomen hebben. We moeten immers de corresponderende overschrijdingskansen vermenigvuldigen met de selectiefactor (die $= 45$ is), voor we mogen onderzoeken, aan welke zijde van de betrouwbaarheidsdrempel de gevonden P 's liggen.

Als laatste voorbeeld moeten we nog de standaarddeviaties van 2.1b behandelen. Voor het tijdvak 1899–1938 zijn de standaarddeviaties van de data van de laatste „killing frost” in Vancouver en Bismarck resp. 22,5 en 9,7. Hun quotiënt is $= 2,32$. Omdat $v_1 = v_2 = 39$ is, is $P = 0,5\%$.

Daar in 2.7.3 aangetoond is, dat de reeks niet homogeen is, verdient het de voorkeur, alleen het tijdvak 1911–38 te beschouwen. Voor die periode waren de

standaarddeviaties 21,2 en 9,0. Hun quotiënt is $= 2,36$. $v_1 = v_2 = 27$. $P = 2\%$.

Daarmee is statistisch bewezen, dat de laatste „killing frost“-data in het kustgebied meer spreiden dan in het binnenland. CONRAD bereikte op grond van onvoldoend feitenmateriaal juist de tegenovergestelde conclusie.

2.10 DE VARIANTIEANALYSE

2.10.1 DE ENKELVOUDIGE VARIANTIEANALYSE

Ons eerste voorbeeld ontleen we aan een artikel van SCHERHAG (1948) over de samenhang tussen de zonnevlekken en de wintertemperaturen van Berlijn. Op p. 240-1 (Tabelle 4) van dit artikel vindt men de Berlijnse wintertemperaturen (1729-1948, enkele jaren ontbreken), ingeplant in het zonnevlekken-schema van SCHERHAG. Zo'n schema wordt als volgt opgebouwd.

Elke dag pogen verscheidene over de gehele wereld verspreide waarnemers het aantal zonnevlekken te bepalen. Uit de gelukte en betrouwbare pogingen wordt zo goed mogelijk een dagelijks gemiddelde berekend. De 365 daggemiddelden leveren een jaargemiddelde. Dit jaargemiddelde vertoont een periode van ca 11 jaren. Het schema wordt zo ingericht, dat de kolom onder het hoofd „Max + k“ slechts die jaren bevat, die k jaren na een zonnevlekkenmaximum-jaar komen. Luidt het hoofd „Min - k“, dan bevat de kolom de jaren, die k jaar geleden zijn voor een zonnevlekkenminimum. De onder elkaar staande jaren in een kolom liggen dus ca 11 jaren uiteen. Elke rij bevat de opvolgende jaren van een zonnevlekken-cyclus. Op deze wijze krijgt elk jaar een bepaalde plaats in het schema.

SCHERHAG heeft nu in zijn schema alle jaren vervangen door de corresponderende wintertemperaturen. De kolomgemiddelden bleken min of meer te spreiden. Deze spreiding is te danken aan 1) het steekproefeffect en eventueel 2) de variërende zonnevlekkeninvloed.

Het is echter beslist onjuist, de onderlinge verschillen tussen de kolomgemiddelden zonder diepgaand onderzoek als reëel te beschouwen, zoals SCHERHAG gedaan heeft. Want daaruit zou moeten volgen, dat de temperatuur van de winter afhangt van het aantal jaren, dat hem scheidt van een zonnevlekken-extremum. Dit uiterst belangrijke resultaat eist een grondige fundering.

Wanneer het steekproefeffect nog niet tot verwaarloosbare proporties is teruggebracht, is het zeer gevaarlijk de opgetreden verschillen toe te schrijven aan de tweede oorzaak, de zonnevlekkeninvloed.

De aangewezen methode om verschillen in kolomgemiddelden statistisch te onderzoeken is de variantieanalyse.

We moeten derhalve de variantie binnen de kolommen en die tussen de kolommen gaan bepalen.

$$s^2_{\text{bik}} = \frac{1}{n-k} \sum \sum (x_{ij} - x_{.j})^2$$

$$s^2_{\text{tuk}} = \frac{1}{k-1} \sum r_j (x_{.j} - x_{..})^2$$

Rechtstreekse berekening van deze grootheden is omslachtig. De kortste methode maakt gebruik van het volgende recept.

Kwadrater elk kolomtotaal en deel het door het in die kolom aanwezige aantal elementen. De som van al deze quotiënten is B .

De som der kolomtotalen, dus de som van alle n elementen is A . $\frac{A^2}{n}$ wordt bepaald. Deze grootheid en B zijn in tabel 15 berekend.

Tabel 15

	Min.					Max.					
	-1½	-½	+½	+1½	+2½	-½	+½	+1½	+2½	+3½	-2½
Kolomtotalen	3,2	4,6	10,2	2,1	10,9	-7,3	-16,2	-2,4	16,3	-2,8	10,0
	3,2 ²	4,6 ²	10,2 ²	2,1 ²	10,9 ²	7,3 ²	16,2 ²	2,4 ²	16,3 ²	2,8 ²	10,0 ²
	14	19	20	20	32	20	20	18	19	18	14
	0,731	1,114	5,202	0,221	3,713	2,664	13,122	0,320	13,984	0,436	7,143

De som van deze elf quotiënten is $48,36 = B$. De som der kolomtotalen $A = 28,6$.

$$\frac{A^2}{n} = \frac{28,6^2}{214} = 3,82.$$

Tot slot van de voorbereidende werkzaamheden hebben we nog D nodig, d.i. de som van de kwadraten van alle n elementen. D is in ons voorbeeld 988,70.

Nu kunnen we de variantie werkelijk gaan analyseren met behulp van dit schema.

Schema voor de enkelvoudige variantieanalyse

	Kwadratensom	Aantal vrijheidsgraden	Variantie
Tussen kolommen	$B - \frac{A^2}{n}$	$k - 1$	$\frac{1}{k-1} (B - \frac{A^2}{n})$
Binnen kolommen	$D - B$	$n - k$	$\frac{1}{n-k} (D - B)$

De verhouding der varianties is $\frac{n-k}{k-1} \times \frac{B - A^2/n}{D - B}$. Substitutie van A , B , D ,

$$n (= 214) \text{ en } k (= 11) \text{ geeft: } \frac{203}{10} \times \frac{48,36 - 3,82}{988,70 - 48,36} = 0,962.$$

Dit resultaat betekent, dat de spreiding tussen de kolommen kleiner is dan de spreiding binnen de kolommen. Het doel van het onderzoek is na te gaan, of er een werkelijke kolommeninvloed is, m.a.w. of de kolomgemiddelden ook in het universum verschillen. Is de zonnevlekkeninvloed reëel, dan moet de universumstandaarddeviatie tussen de kolommen σ_{tuk} groter zijn dan de universumstandaarddeviatie binnen de kolommen σ_{bik} . Zijn de wintertemperaturen onafhankelijk van de zonnevlekken, dan is $\sigma_{tuk} = \sigma_{bik}$. Wij kennen geen van beide, wel echter hun schattingen, nl. de steekproefstandaarddeviaties. Het rekenresultaat, $s_{tuk} < s_{bik}$, geeft geen enkel recht te beweren, dat $\sigma_{tuk} > \sigma_{bik}$. Berlijn zal dus nog vele winters moeten doormaken, al eer de variantieanalyse kan aantonen, dat de zonnevlekken de wintertemperaturen al of niet beïnvloeden.

De in deze onderparagraaf gevolgde wijze van behandeling is de eenvoudigste, omdat hij kan aansluiten op de tabel 4 van SCHERHAG.

Zijn echter de elementen nog niet gerangschikt in een zonnevlekken-schema, dan geeft de nu te bespreken methode voordelen.

De tweede kolom in tabel 16 heeft tot hoofd: $\text{Min} - 1\frac{1}{2}$. Dit betekent, dat de in deze kolom voorkomende winters anderhalf jaar liggen voor het midden van een zonnevlekkenminimumjaar. $\text{Max} + 1\frac{1}{2}$ betekent, dat de winters van de achtste kolom een half jaar optraden na het midden van een zonnevlekkenmaximumjaar.

SCHERHAG's zonnevlekken-schema vertelt, welke jaren in elke kolom moeten komen. De lijst met de wintertemperaturen van 1729–1948 stelt ons in staat, de frequenties van de temperaturen in de kolommen te bepalen. Als bijproduct krijgen we zonder veel moeite de frequentieverdeling van de 214 wintertemperaturen (kolom onder: Totaal).

Uit de frequentieverdelingen in de kolommen laten zich het aantal elementen, het eerste en het tweede moment van elke kolom afleiden. Het eerste moment onder aan de kolom is gelijk aan de som der in die kolom aanwezige wintertemperaturen.

De rij „eerste moment” is dus vergelijkbaar met de eerste rij van tabel 15.

De rij „tweede moment” levert D , de som van de kwadraten van de 214 temperaturen.

De kolommen „eerste en tweede moment” vervullen hier alleen de functie van controle op het eindresultaat van de overeenkomstige rijen.

Substitutie van de in tabel 16 gevonden waarden van A , B en D geeft een variantiequotiënt $= \frac{203}{10} \times \frac{50 - 4}{993 - 50} = 0,990$. Dit betekent weer $s_{tuk} < s_{bik}$. We zijn derhalve niet in staat te bewijzen, dat $\sigma_{tuk} > \sigma_{bik}$.

Tabel 16

Temp. in °C.	Min.					Max.					Total	Temp. Eerste Tweede in °C. moment	
	-1½	-½	+½	+1½	+2½	-½	+½	+1½	+2½	+3½			
4		1	1	1	3		2	2	2	2	5	4	20
3	4	1	3	2	8	1	2	3	4	1	23	3	69
2	1	3	4	3	4	4	2	3	4	2	38	2	76
1	—	4	3	4	9	3	4	3	6	1	33	1	33
0	3	4	3	2	2	2	3	2	3	5	39	0	0
—1	3	4	1	4	2	4	3	3	3	4	34	—1	—34
—2	1	—	4	1	3	3	1	3	1	—	17	—2	—34
—3	2	—	—	2	3	2	—	—	—	—	11	—3	—33
—4		1	—	1		—	2	—	2	1	6	—4	—24
—5		1	1			1	2	2			7	—5	—35
—6							—				0	—6	0
—7							1				1	—7	—7
Aantal in elke kolom											214 = "	+198	993
												—167	
												31	
Eerste moment											31 = A		
Tweede moment											993 = D		
											961 = A ²		
											214 = "		
											4		

B, de som van deze elf quotiënten = 50.

Het afronden tot op hele graden C heeft het tabeller- en rekenwerk vergemakkelijkt. Vergelijking met de niet afgeronde bewerking toont, dat er slechts onbelangrijke fouten optreden. De eindcijfers, de variantiequotiënten, zijn 0,962 en 0,990. Beide uitkomsten voeren tot dezelfde negatieve conclusie.

De enige voorwaarde gesteld aan het toepassen van de z -toets was het normaal verdeeld zijn van de elementen in de kolommen. De frequentieverdeling in tabel 16 toont, dat de temperaturen inderdaad in grote trekken de verdeling volgens GAUSS gehoorzamen.

2.10.2 DE TWEEVOUDIGE VARIANTIEANALYSE

De Juli-luchtdruk in De Bilt (1849–1949) zal nu behandeld worden, omdat we deze in 3.1 nodig hebben bij de critisering van een onderzoek van BAUR.

Ter wille van de vergelijkbaarheid moeten we gebruik maken van het zonnevlekken-schema, zoals Prof. BAUR dat in HANN-SÜRING (Lehrbuch der Meteorologie, vijfde ed., p. 961) opgesteld heeft. In 75 van de 196 jaren verschillen de schema's van BAUR en SCHERHAG van elkaar (VAN DER BIJL 1951). De oorzaak moeten we zoeken in het gebruiken van twee soorten zonnevlekgetallen. BAUR neemt de echte jaarlijkse getallen, SCHERHAG middelt eerst die getallen over drie opeenvolgende jaren. Daardoor verspringen de extremen vaak op een ander jaar.

In tabel 17 zijn de met 1000 mbar verminderde gemiddelde luchtdrukwaarden te De Bilt in de 101 Julimaanden van 1849–1949 weergegeven. Het eerste element, Juli 1849, staat in de eerste rij en in de kolom „Max + 1”. Dit betekent dus, dat Juli 1849 (met een gemiddelde luchtdruk van 1015,3 mb) één jaar na het zonnevlekkenmaximum van 1848 lag. De jaren 1856 en 1860 hadden een zonnevlekkenminimum resp. een maximum. De eerste waarnemingen in de „Min”- en „Max”-kolom komen derhalve 7 resp. 11 jaar na 1849. Niet steeds is de afstand in jaren tussen de zonnevlekkenextrema constant. Dit verklaart het voorkomen van de lege plekken in tabel 17. (De grafische voorstelling bevindt zich in fig. 14 op p. 121).

Bij de tweevoudige variantieanalyse mogen echter geen lege plekken aanwezig zijn. Alle rijen, dus ook alle kolommen moeten evenveel elementen tellen. De eenvoudigste manier om een complete rechthoek te verkrijgen is het weglaten van de onvolledige rijen of kolommen. Het grootste aantal elementen houden we nog over, wanneer we de kolommen Min + 1, Min + 2, Max + 2, Max + 3 en Max + 4 schrappen, benevens het jaar 1849. Het restant wordt met 10 mbar verminderd en afgerond op hele mbar, waarna tabel 18 ontstaat. De rij- en kolomtotalen worden opgemaakt.

Tabel 17

	-3	-2	-1	Min.	+1	+2	-2	-1	Max.	+1	+2	+3	+4
1849—1852										15,3	15,5	11,5	17,5
1853—1863	14,6	16,2	13,5	16,2	16,3		13,7	20,3	15,8	10,1	14,3	20,3	
1864—1874	16,6	15,5	13,3	13,0			17,7	20,0	16,0	12,7	14,9	16,2	17,1
1875—1885	14,6	18,9	13,7	15,9	09,6	13,7	16,6	13,2	11,3	15,8	22,0		
1886—1897	14,3	18,1	08,4	13,5	12,2	13,7	16,4	13,0	13,2	12,6	16,6	15,9	
1898—1909	17,6	17,5	15,5	17,0	16,3	13,7	18,4	17,6	17,6	17,3	16,7	12,8	
1910—1919	11,3	21,6	15,0	16,0	11,9		12,7	16,7	17,6	15,0	14,9		
1920—1929	14,4	18,1	13,5	16,0	12,3	13,7	15,4	12,7	18,4	17,5			
1930—1940	09,9	10,0	12,3	17,9	16,4		18,9	11,1	15,3	15,0	13,0	13,1	
1941—1949	15,5	14,4	16,0	13,7			15,8	17,6	15,5	15,0	19,6		

Gemiddelde barometerstand te De Bilt in de Julimaanden 1849—1949 ingepast in het zonnevlekkenschema van Prof. F. BAUR. Alle waarden zijn met 1000 mbar verminderd.

Tabel 18

-3	-2	-1	Min.	-2	-1	Max.	+1	Totaal
5	6	4	6	4	10	6	0	41
7	6	3	3	8	10	6	3	46
5	9	4	6	7	3	1	6	41
4	8	—2	4	6	3	3	3	29
8	8	6	7	8	8	8	7	60
1	12	5	6	3	7	8	5	47
4	8	4	6	5	3	8	8	46
0	0	2	8	9	1	5	5	30
6	4	6	4	6	8	6	5	45
40	61	32	50	56	53	51	42	385 = A

Zonnevlekkenschema, waarin afgeronde en met 1010 mb verminderde gemiddelde luchtdruk in enige Julimaanden van 1849—1949 (De Bilt).

Nu worden alle getallen van tabel 18 (behalve A) gekwadrateerd, waarmee tabel 19 ontstaat. De twee kolommen rechts in deze tabel 19 bevatten de kwadraten van de sommen van de kolommen en rijen van tabel 18.

We moeten de kwadratsommen $k \sum (x_{i.} - x_{..})^2$, $r \sum (x_{.j} - x_{..})^2$ en $\sum \sum (x_{ij} - x_{i.} - x_{.j} + x_{..})^2$ bepalen. Rechtstreekse berekening is weer omslachtig. In het nu volgende schema voor de tweevoudige variantieanalyse staan de gevraagde kwadratsommen, echter berekend met behulp van de grootheden A, B, C, D, k en r.

Tabel 19

De gekwadrateerde elementen van tabel 18								Kwadraten v. d. rij- en kolomtotalen van tabel 18	
-3	-2	-1	Min.	-2	-1	Max.	+1		
25	36	16	36	16	100	36	0	1681	1600
49	36	9	9	64	100	36	9	2116	3721
25	81	16	36	49	9	1	36	1681	1024
16	64	4	16	36	9	9	9	841	2500
64	64	36	49	64	64	64	49	3600	3136
1	144	25	36	9	49	64	25	2209	2809
16	64	16	36	25	9	64	64	2116	2601
0	0	4	64	81	1	25	25	900	1764
36	16	36	16	36	64	36	25	2025	
232	505	162	298	380	405	335	242	2559	17169
								= D	= kC
									= rB

A is de som van alle oorspronkelijke elementen (hier = 385), rB is de som van de k (= 8) kwadraten van de kolomtotalen (hier = 19155), kC is de som van de r (= 9) kwadraten van de rijtotalen (hier = 17169), D is de som van de kwadraten van alle n (= kr = 72) oorspronkelijke elementen (hier = 2559).

Schema voor de tweevoudige variantieanalyse

	Kwadratensom	Aantal vrijheidsgraden	Variantie
Tussen kolommen	$B - \frac{A^2}{n}$	$k - 1$	$\frac{1}{k-1} (B - \frac{A^2}{n})$
Tussen rijen	$C - \frac{A^2}{n}$	$r - 1$	$\frac{1}{r-1} (C - \frac{A^2}{n})$
Residu	$D + \frac{A^2}{n} - B - C$	$(k-1)(r-1)$	$\frac{1}{(k-1)(r-1)} (D + \frac{A^2}{n} - B - C)$

De eerste twee varianties delen we door de derde. Zo krijgen we het variantiequotiënt voor de rijen, nl.

$$\frac{(C - \frac{A^2}{n})(r-1)}{D + \frac{A^2}{n} - B - C} \quad (\text{in ons geval} = 1,78) \text{ en}$$

het variantiequotiënt voor de kolommen, nl.

$$\frac{(B - \frac{A^2}{n})(k-1)}{D + \frac{A^2}{n} - B - C} \quad (\text{in ons geval} = 1,62).$$

In het schema voor de tweevoudige variantieanalyse lezen we, dat de teller van het variantiequotiënt voor de rijen $(r-1)$, dus 8 vrijheidsgraden bezit en de noemer $(k-1)(r-1) = 56$. We weten, dat $s_{tur} > s_{residu}$. Mag hieruit nu geconcludeerd worden, dat ook $\sigma_{tur} > \sigma_{residu}$? Of zou het toeval een dergelijk groot verschil tussen de twee schattingen voor de spreiding in het universum kunnen veroorzaken? Wat is de kans, dat een steekproef een waarde voor $\frac{s_{tur}}{s_{residu}} \geq 1,78$ levert, wanneer $\sigma_{tur} = \sigma_{residu}$? Om deze vraag te beantwoorden moeten we de F -tabellen nagaan.

Uit deze blijkt, dat bij 8 vrijheidsgraden in de teller en 56 in de noemer $F = 1,44$ een overschrijdingskans $P = 20\%$ en $F = 1,78$ een $P = 10\%$ heeft. We vonden voor de rijen, dat $8 F 56 = 1,78$ was. Deze $8 F 56$ heeft derhalve een overschrijdingskans, die gelijk is aan 10% . We mogen dus niet uit $s_{tur} > s_{residu}$ besluiten, dat $\sigma_{tur} > \sigma_{residu}$. M.a.w. de variantieanalyse kan niet aantonen, dat de Juli-luchtdruk in de ene zonnevlekkencyclus werkelijk verschilt van de druk in een andere cyclus.

Wat de kolommen betreft, we berekenden, dat het variantiequotiënt $7 F 56 = 1,62$ was. Bij een overschrijdingskans van 20% (resp. 10%) behoort een waarde van $7 F 56 = 1,45$ (resp. $1,82$), zo leren ons de F -tabellen. De berekende waarde van $1,62$ heeft dus een P tussen 10 en 20% in. Ook een duidelijk kolomeffect ontbreekt derhalve. We kunnen niet aantonen, dat de plaats van het jaar in de zonnevlekkencyclus de Juli-luchtdruk in De Bilt beïnvloedt.

Men zal de vraag opwerpen, waarom nu wel overschrijdingskansen bepaald zijn en bij de enkelvoudige variantieanalyse van de Berlijnse wintertemperaturen niet. De oorzaak van deze asymmetrie wordt duidelijk, wanneer we nogmaals voor ogen houden, wat het doel van de variantieanalyse is. En dit is het vaststellen, of de spreiding in de kolommen (eventueel rijen, enz.) significant groter is dan de toevalspreiding.

In het geval van de Juli-luchtdruk vonden we, dat zowel de kolom- als de rij-spreiding groter waren dan de toevalspreiding, maar het onderzoek naar de overschrijdingskansen leerde, dat de verschillen niet significant waren.

In het geval van de Berlijnse wintertemperaturen bleek de kolomspreiding zelfs kleiner te zijn dan de toevalspreiding. Het enige, wat een significantie-onderzoek hier zou kunnen leren, is, dat de kolomspreiding significant kleiner is dan de toevalspreiding en dit is in normale gevallen theoretisch onmogelijk. Daar ons doel – het aantonen van een significant grotere kolomspreiding – onbereikbaar is, indien de kolomspreiding kleiner is dan de toevalspreiding, heeft het geen zin overschrijdingskansen te bepalen.

2.11 CORRELATIE EN REGRESSIE

2.11.1 ENKELVOUDIGE CORRELATIE

De correlatiecoëfficiënt is pas in de 19^{de} eeuw door BRAVAIS (de formule in 1846, herontdekt door GALTON in 1888) en EDGEWORTH (de benaming in 1892) ingevoerd. Dat de correlatiecoëfficiënt inderdaad in een behoefte voorziet, moge blijken uit het volgende:

DERHAM (1714–1716) publiceerde de jaarlijkse regenhoeveelheden, gevallen in Upminster (Essex) in de jaren 1697–1714. De redacteur van de *Philosophical Transactions* liet deze cijfers volgen door de gegevens van Parijs voor het tijdvak 1689–1711. In het postscriptum wordt er dan nog even op gewezen, dat de jaren niet precies samenvallen vanwege „the Diversity of Stile”. Het verschil bedraagt overigens slechts 13 dagen. Tot besluit wordt de opmerking gemaakt: „We have forbore to make any Remarks upon this Comparison, leaving it to the Consideration of the Curious Reader”.

Als wij de rol van „nieuwsgierige lezer” mogen spelen, dan kunnen we, gewapend met het sindsdien ontdekte begrip correlatiecoëfficiënt, de parallelle gang tussen de regenhoeveelheden van de twee meteorologische stations in een getal vastleggen.

De twee waarnemingsperioden hebben de jaren 1697–1711 gemeen. Over dit tijdvak van 15 jaren gaan we de correlatiecoëfficiënt bepalen. Dit gebeurt in tabel 20. De bewerkingen spreken grotendeels voor zichzelf. Om de berekeningen te vereenvoudigen wordt van de gemeten regenhoeveelheden eerst een dusdanig groot getal afgetrokken, dat de nieuwe cijfers zo klein mogelijk zijn.

Toch blijven x_i en y_i vrij grote getallen. De spreidingsbreedte, het verschil tussen hoogste en laagste waarneming, is in x_i $116-1 = 115$, in y_i $128-8 = 120$.

We gaan nog meer afronden en zullen dan zien, dat het eindresultaat, de waarde van r , weinig varieert.

Vermenigvuldigen we $(U_i - 15)$ en $(P_i - 14)$ niet met 10, maar met 1, dan is na afronding de spreidingsbreedte in x_i $12-0 = 12$ en in y_i $13-1 = 12$. De nieuwe r wordt

$$r = \frac{503 - \frac{73 \times 93}{15}}{\sqrt{\left(557 - \frac{73^2}{15}\right) \left(719 - \frac{93^2}{15}\right)}} = 0,297$$

Deling van de oorspronkelijke gegevens $(U_i - 15)$ en $(P_i - 14)$ door 2 en afronding geeft tot spreidingsbreedten 6, zowel in x_i als in y_i . Nu is $r = 0,194$.

Deling door 5, gevolgd door afronding laat een spreidingsbreedte van slechts 2 in x_i en 3 in y_i over. Men zie tabel 21.

Tabel 20

	<i>i</i>	<i>U_i</i>	<i>P_i</i>	<i>x_i</i>	<i>y_i</i>	<i>x_i y_i</i>	<i>x_i²</i>	<i>y_i²</i>
1697	1	15,52	21,60	5	76	380	25	5.776
98	2	24,46	23,20	95	92	8.740	9.025	8.464
99	3	15,11	19,93	1	59	59	1	3.481
1700	4	19,03	21,38	40	74	2.960	1.600	5.476
01	5	18,69	22,78	37	88	3.256	1.369	7.744
02	6	20,38	17,42	54	34	1.836	2.916	1.156
03	7	23,99	18,51	90	45	4.050	8.100	2.025
04	8	15,81	21,20	8	72	576	64	5.184
05	9	16,93	14,82	19	8	152	361	64
06	10	24,29	16,32	93	23	2.139	8.649	529
07	11	16,31	19,11	13	51	663	169	2.601
08	12	19,22	19,51	42	55	2.310	1.764	3.025
09	13	26,56	23,21	116	92	10.672	13.456	8.464
1710	14	18,37	17,10	34	31	1.054	1.156	961
11	15	23,60	26,84	86	128	11.008	7.396	16.384
				733	928	49.855	56.051	71.334

U_i = regenhoeveelheid in inches te Upminster in jaar $i + 1696$.

P_i = idem in Paris.

$x_i = 10 (U_i - 15)$

$y_i = 10 (P_i - 14)$

x_i en y_i zijn afgerond op gehele waarden.

$$r = \frac{\sum_{i=1}^n (x_i - m_1) (y_i - m_2)}{\sqrt{\sum_{i=1}^n (x_i - m_1)^2 \sum_{i=1}^n (y_i - m_2)^2}}$$

$$\sum_{i=1}^n (x_i - m_1) (y_i - m_2) = \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum x_i \sum y_i = 49.855 - \frac{733 \times 928}{15} = 4.507$$

$$\sum_{i=1}^n (x_i - m_1)^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n} = 56.051 - \frac{733^2}{15} = 20.232$$

$$\sum_{i=1}^n (y_i - m_2)^2 = \sum_{i=1}^n y_i^2 - \frac{(\sum y_i)^2}{n} = 71.334 - \frac{928^2}{15} = 13.922$$

$$r = \frac{4.507}{\sqrt{20.232 \times 13.922}} = 0,269$$

Tabel 21

	i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	
1697	1	—	2	—	—	4	$x_i = \frac{1}{5} (U_i - 15)$
98	2	2	2	4	4	4	
99	3	—	1	—	—	1	$y_i = \frac{1}{5} (P_i - 14)$
1700	4	1	1	1	1	1	
01	5	1	2	2	1	4	x_i en y_i zijn afgerond op gehele waarden.
02	6	1	1	1	1	1	
03	7	2	1	2	4	1	
04	8	—	1	—	—	1	
05	9	—	—	—	—	—	$r = \frac{22 - \frac{15 \times 19}{15}}{\sqrt{(25 - \frac{15^2}{15})(33 - \frac{19^2}{15})}}$
06	10	2	—	—	4	—	$= \frac{3}{\sqrt{90}} = 0,316$
07	11	—	1	—	—	1	
08	12	1	1	1	1	1	
09	13	2	2	4	4	4	
1710	14	1	1	1	1	1	
11	15	2	3	6	4	9	
		15	19	22	25	33	

Naarmate we meer afronden, vinden we voor r achtereenvolgens 0,269, 0,297, 0,194 en 0,316. Volgens tabel VI van FISHER and YATES heeft een steekproefwaarde van $r = 0,4409$ bij $n-2 = 13$ vrijheidsgraden nog een overschrijdingskans $P = 10\%$, uitgaande van een universumwaarde $\rho = 0$. Het geringe aantal waarnemingen verhindert, dat we de samenhang tussen de neerslaghoeveelheden in Parijs en Upminster significant kunnen noemen.

Een enkele opmerking over de theorie van het afronden is hier wel op zijn plaats. WALKER (1926) was de eerste in de meteorologische literatuur, die er aandacht aan besteedde. Zijn conclusie was, dat de standaarddeviatie ca 4 afrondingseenheden moest bedragen. HAMAKER (1948, 190) kwam tot de regel: „Het afrondingsinterval moet zijn van de orde van de helft van de standaarddeviatie of 1/6 van de spreidingsbreedte, mits deze laatste uit 5 tot 10 waarnemingen is bepaald”. De afrondingseenheid van HAMAKER is tweemaal zo groot als die van WALKER. Bij een aantal waarnemingen van omstreeks 10 kan de afrondingseenheid eenvoudiger bepaald worden via de spreidingsbreedte dan met behulp van de standaarddeviatie. Zo was in ons voorbeeld de spreidingsbreedte achtereenvolgens gelijk aan 115, 12, 6 en 2 afrondingseenheden. Het derde geval voldoet aan HAMAKER's eis, maar het resultaat van het vierde geval toont, dat verdergaande afronding onder dringende omstandigheden ook nog bruikbaar is.

We keren terug naar ons uitgangspunt: de oudste correlatie tussen meteorologische grootheden. Het heeft zin te onderzoeken, hoe die correlatie in andere tijdvakken is. Berekend is de correlatiecoëfficiënt tussen de jaarlijkse neerslag-sommen in Parijs en Greenwich in de periode 1874–1934. Greenwich ligt slechts 20 km van Upminster verwijderd en is derhalve een waardige plaatsvervanger. De keuze van begin- en eindjaar werd bepaald door de gegevens in de „World Weather Records”. Het resultaat was $r = 0,543$. Deze correlatiecoëfficiënt heeft $\nu = n - 2 = 59$ vrijheidsgraden. Tabel VI van FISHER and YATES leert ons, dat $P(r = 0,408, \nu = 60; p = 0) = 0,001$. Onze correlatiecoëfficiënt ligt ver buiten de grens van 0,408. Zijn overschrijdingskans is dus veel minder dan 0,001. De correlatie tussen de jaarlijkse regensommen in Parijs en Greenwich (Upminster) is in het laatste tijdvak significant.

Nemen we eens aan, dat de universumwaarde ρ van de correlatiecoëfficiënt gelijk is aan de zojuist gevonden steekproefwaarde 0,543. Wijkt de „oudste” meteorologische correlatiecoëfficiënt significant van deze waarde af?

FISHER's transformatie levert $\zeta = \frac{1}{2} \ln (1 + 0,543) - \frac{1}{2} \ln (1 - 0,543) = 0,608$ en $z = \frac{1}{2} \ln (1 + 0,269) - \frac{1}{2} \ln (1 - 0,269) = 0,276$.

$$(z - \zeta) = -0,332 \quad \sigma = \frac{1}{\sqrt{\nu-1}} = \frac{1}{\sqrt{12}} = 0,289$$

$$\mu = \frac{\rho}{2\nu+2} = \frac{0,543}{28} = 0,019$$

$$\frac{(z - \zeta) - \mu}{\sigma} = \frac{-0,351}{0,289} = -1,22. \text{ Een dergelijke relatieve afwijking heeft een}$$

dubbelzijdige overschrijdingskans van 22% (bepaald uit fig. 5, p. 21) en is derhalve niet significant.

2.11.1.1 SIGNIFICANTIE VAN HET VERSCHIL TUSSEN TWEE CORRELATIE-COEFFICIENTEN

In de hieraan voorafgaande regels stelden we een steekproefwaarde gelijk aan zijn universumwaarde. Deze minder juiste handelwijze vervangen we nu door een betere methode. Als voorbeeld kiezen we niet het vorige geval, waarbij de gecorreleerde grootheden in de twee steekproeven dezelfde waren en de twee perioden, waarover gecorreleerd werd, verschilden. We gaan nu twee correlatiecoëfficiënten vergelijken, waarbij het tijdvak gelijk is en de gecorreleerde grootheden verschillen.

Voor de periode 1885–1930 correleerde BAUR (zie 3.1) de wintertemperatuur van Centraal Europa met de daaraan voorafgaande temperatuur van NW Siberië in November en het temperatuurverschil tussen November en October.

Het resultaat van de berekening was:

$$r_1 = r_{\text{Nov.}} = 0,42 \quad r_2 = r_{\text{Nov.-Oct.}} = 0,63$$

$$v_1 = v_2 = n - 2 = 44$$

$$z_1 = \frac{1}{2} \ln(1 + r_1) - \frac{1}{2} \ln(1 - r_1) = 0,448 \quad z_2 = 0,742$$

$$\sigma = \sqrt{\frac{1}{v_1 - 1} + \frac{1}{v_2 - 1}} = \sqrt{\frac{2}{43}}$$

$$\text{De relatieve afwijking is } \frac{z_1 - z_2}{\sigma} = \frac{-0,294}{\sqrt{\frac{2}{43}}} = -1,36 \text{ en } P = 17\%.$$

De overschrijdingskans is te groot om het verschil tussen de correlatiecoëfficiënten significant te kunnen noemen.

2.11.2 ENKELVOUDIGE REGRESSIE

We onderzoeken de regressie tussen de Augustus- en de jaartemperaturen te De Bilt. Nemen we aan, dat de temperatuur van Augustus onafhankelijk is van die der andere maanden, dan is β , de regressiecoëfficiënt, gemakkelijk te bepalen. Zij x de Augustus-temperatuur en y de jaartemperatuur, dan is: $12y = x + T_{\text{Jan.}} + T_{\text{Feb.}} + \dots + T_{\text{Dec.}}$. We beschouwen alleen de afwijkingen ten opzichte van de respectieve centra, zodat $\mu_i = 0$ is.

$$\text{Nu geldt: } E(xy) = E\left\{x \cdot \frac{1}{12} (x + T_{\text{Jan.}} + \dots + T_{\text{Dec.}})\right\} = \frac{1}{12} E(x^2), \text{ omdat}$$

$E(x \cdot T_{\text{Jan.}}) = \dots = E(x \cdot T_{\text{Dec.}}) = 0$. Immers, de Augustus-temperatuur hangt niet af van de temperatuur in de overige maanden.

$$\beta = \frac{E(xy)}{E(x^2)} = \frac{\frac{1}{12} E(x^2)}{E(x^2)} = \frac{1}{12} = 0,0833$$

Voor het steekproefmateriaal nemen we de jaren 1901–1940. De berekening van de regressiecoëfficiënt b gaat in het begin op dezelfde wijze als de berekening van een correlatiecoëfficiënt. Ook nu wordt eerst een „voordelig” nulpunt opgezocht, zowel voor x als voor y , met het doel het cijferwerk minimaal te maken. Bepaald worden vervolgens $\sum xy$, $\sum x$, $\sum y$ en $\sum x^2$; bij de correlatiecoëfficiënt hebben we ook $\sum y^2$ nodig.

$$b = \frac{\sum_{i=1}^n (x_i - m_1)(y_i - m_2)}{\sum_{i=1}^n (x_i - m_1)^2}$$

De teller is gelijk aan $\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}$ en de noemer gelijk aan

$$\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}.$$

We vonden $b = \frac{10,125}{50,71} = 0,1997$. Bij het onderzoek naar de significantie van het grote verschil tussen b en β hebben we s_y^2 nodig.

$$(n-1) u^2 = (n-1) s_y^2 - b^2 (n-1) s_x^2$$

$$39 u^2 = 11,08 - 0,1997^2 \times 50,71 = 9,06$$

$$t = \frac{1}{u} (b - \beta) s_x \sqrt{n-2} = \left(\sqrt{\frac{9,06}{39}} \right)^{-1} \times 0,1164 \sqrt{\frac{50,71}{39}} \sqrt{38} =$$

$$= 0,1164 \sqrt{\frac{50,71 \times 38}{9,06}} = 1,698 \quad v = n-2 = 38 \quad P = \text{ca } 10\%$$

Het belangrijke verschil tussen β en b mogen we daarom niet significant noemen. In werkelijkheid is de overschrijdingskans nog hoger, daar we uitgingen van de onjuiste veronderstelling, dat de Augustus-temperatuur geen invloed heeft op de temperaturen der overige maanden. Natuurlijk is de correlatie tussen de Juli- en Augustus- en tussen de Augustus- en September-temperaturen niet nul. Beide correlatiecoëfficiënten zijn ca 0,4 = 40% (zie 2.12.1). β wordt daardoor $2 \times 40 = 80\%$ groter en dus ca 0,15. Het verschil tussen b en β is nu kleiner, nl. 0,05. Dit levert een $t = 0,73$ en een dubbelzijdige overschrijdingskans P van ca 45%.

2.11.3 MEERVOUDIGE REGRESSIE

VARANGOT (1948) voerde de meervoudige regressierekening in bij de herleiding van klimatologische reeksen. Hij vatte het probleem aan van de schatting van het etmaalgemiddelde van de temperatuur uit de termijnwaarnemingen van 8^h, 14^h en 19^h.

Elke oplossing moet gebruik maken van een tijdvak, waarin volledige 24-uurwaarnemingen bekend zijn. In deze periode kan zowel het overdaggemiddelde $\frac{1}{3} (T_8 + T_{14} + T_{19})$ als het etmaalgemiddelde $\frac{1}{24} (T_0 + T_1 + \dots + T_{23})$ berekend worden.

De eenvoudigste oplossing is die, waarbij het verschil van deze gemiddelden bepaald wordt in tijdvakken, waarin beide bekend zijn. Uit dit verschil en het

overdaggemiddelde is dan een schatting voor het etmaalgemiddelde af te leiden, wanneer dit etmaalgemiddelde zelf niet bekend is.

Men voelt aan, dat een methode, die gebruik maakt van de 8^h , 14^h en 19^h -gegevens zelf, betere schattingen levert dan de zo juist genoemde, waarbij alleen het gemiddelde van de drie waarnemingen een rol speelt.

VARANGOT legde de gegevens van de Julimaanden uit de periode 1921-1940 aan zijn berekeningen ten grondslag. Het resultaat was de regressievergelijking: $X''_0 = 0,755 x_1 - 0,089 x_2 + 0,316 x_3 + 0,15 ^\circ\text{C}$.

Hierin betekenen:

X''_0 = het met behulp van deze regressievergelijking geschatte etmaalgemiddelde

x_0 = maandgemiddelde van het etmaalgemiddelde der temperatuur

x_1 = " " de temperatuur om 8^h

x_2 = " " " " 14^h

x_3 = " " " " 19^h

Een indruk van de kwaliteit van de schatting uit de regressievergelijking geeft $\Sigma(x_0 - X''_0)^2$, gesommeerd over de 20 jaren van het basistijdvak. De som was 0,289.

De eenvoudigste methode vond als verschil tussen etmaal- en overdaggemiddelde $-1,7^\circ\text{C}$. Dit getal vermeerderd met het overdaggemiddelde geeft eveneens een schatting X'_0 voor het etmaalgemiddelde x_0 . De som $\Sigma(x_0 - X'_0)^2$ is nu 1,752, ruim zes maal zo groot als de analoge kwadratensom bij de regressiemethode.

Gebruiken we de regressievergelijking buiten het basistijdvak 1921-1940, dan moeten we minder goede resultaten verwachten. Immers, de regressiecoëfficiënten zijn zo berekend, dat de kwadratensom in het basistijdvak minimaal is.

Een ander tijdvak heeft andere regressiecoëfficiënten, die de minimale kwadratensom leveren. Gebruikt men in dat andere tijdvak de regressiecoëfficiënten van het basistijdvak, dan zal de kwadratensom niet minimaal kunnen zijn.

Passen we de vergelijkingen van 1921-1940 toe op de gegevens van de 20 jaren 1909-1920 en 1941-1948, dan zijn de kwadratensommen $\Sigma(x_0 - X''_0)^2 = 0,70$ en $\Sigma(x_0 - X'_0)^2 = 2,01$.

We keren weer terug naar het artikel van VARANGOT. Deze vestigt de aandacht op het negatieve teken van de coëfficiënt van x_2 en slaagt er in een aannemelijke verklaring te vinden. Als statisticus verzuimt VARANGOT het onderzoek naar de significantie niet: "Overigens kan men het negatief zijn van de coëfficiënt van x_2 niet zonder meer als significant beschouwen: de standaarddeviatie van deze coëfficiënt bedraagt namelijk 0,070 (van de coëfficiënten van x_1 en x_3 zijn de standaarddeviaties resp. 0,105 en 0,117)".

Dat VARANGOT niet ten onrechte tot voorzichtigheid maant, blijkt uit het regressieonderzoek van de Juli-temperaturen voor 1901-1920. Dit gaan we nu behandelen.

Het algemene element van het $(m + 1) = 4$ -dimensionale universum is $(x_{0i}, x_{1i}, x_{2i}, x_{3i})$ met $i = 1, 2, \dots, 20$.

x_{0i} = maandgemiddelde van het etmaalgemiddelde van de temperatuur in jaar $1900 + i$

x_{1i} = maandgemiddelde van de temperatuur om 8^h in jaar $1900 + i$

x_{2i} = " " " " " 14^h " " $1900 + i$

x_{3i} = " " " " " 19^h " " $1900 + i$

Alle waarnemingen zijn beschouwd als afwijkingen ten opzichte van hun gemiddelde van 1901-1920. Zij hebben betrekking op de maand Juli en het station De Bilt.

De drie regressiecoëfficiënten worden berekend uit de drie normaalvergelijkingen. Hierin komen de sommen $\sum_{i=1}^{20} x_{hi} x_{ji}$ voor ($h = 0, 1, 2, 3; j = 1, 2, 3$).

Substitutie van deze sommen levert de normaalvergelijkingen:

$$44,27 = 45,56 b_1 + 61,95 b_2 + 51,71 b_3$$

$$61,50 = 61,95 b_1 + 88,45 b_2 + 72,00 b_3$$

$$51,16 = 51,71 b_1 + 72,00 b_2 + 60,61 b_3$$

De oplossing is: $b_1 = 0,365$ $b_2 = 0,183$ $b_3 = 0,315$

Inderdaad heeft x_2 nu een positieve coëfficiënt, waarvan de mogelijkheid door VARANGOT aangetoond is.

De meest onverwachte conclusie, die VARANGOT uit zijn onderzoek trok, was deze: „Men vindt een nauwkeuriger schatting X'''_0 voor x_0 door alleen van x_1 gebruik te maken dan door de „klassieke’ methode toe te passen, waarbij men x_1 en x_2 en x_3 in de berekening betreft”. VARANGOT werd tot deze gevolgtrekking geleid door de hoge regressiecoëfficiënt van x_1 .

Volgens 1.11.2 is $\Sigma (x_0 - X'''_0)^2 = \Sigma x_{0i}^2 - \frac{(\Sigma x_{0i} x_{1i})^2}{\Sigma x_{1i}^2} = 0,435$ (voor 1921-1940). Inderdaad is deze kwadraten som nog niet het vierde gedeelte van $\Sigma (x_0 - X'_0)^2 = 1,752$, de kwadraten som bij de „klassieke’ methode.

De grote rol, die x_1 in 1921-1940 speelde ($b_1 = 0,755$), is in de jaren 1901-1920 een middelmatige. De zojuist geciteerde conclusie lijkt nu in gevaar te komen.

Bepalen we echter voor 1901-1920 de som $\Sigma (x_0 - X'''_0)^2 = 43,94 - \frac{(44,27)^2}{45,56} =$

$0,92$, de som $\Sigma (x_0 - X''_0)^2 = 43,94 - 0,365 \times 44,27 - 0,183 \times 61,50 - 0,315 \times 51,16 = 0,41$ en de som $\Sigma (x_0 - X'_0)^2 = 2,07$, dan blijkt opnieuw, dat de 8^h -temperatuur een betere schatting voor het etmaalgemiddelde geeft dan het overdaggemiddelde.

Daar de regressiecoëfficiënten voor de twee tijdvakken 1901–1920 en 1921–1940 nogal uiteenlopen, is een onderzoek naar de significantie der verschillen op zijn plaats. De absolute verschillen zijn achtereenvolgens 0,390, 0,272 en 0,001. Stellen we de regressiecoëfficiënten van het universum gelijk aan die van de periode 1921–1940, dan behoeven we de absolute verschillen slechts door de universumstandaarddeviaties te delen om de relatieve verschillen T (1.3.1) te krijgen. VARANGOT vond voor de steekproefstandaarddeviaties resp. 0,105, 0,070 en 0,117. Stellen we ook deze gelijk aan hun universumwaarden, dan geeft deling de relatieve verschillen 3,7, 3,9 en 0,01. Volgens de gaussverdeling zijn de bijbehorende overschrijdingskansen 0,04 %, 0,01 % en 99 %.

Het delen van de absolute verschillen der regressiecoëfficiënten door de standaarddeviaties der regressiecoëfficiënten is minder juist. Het is beter te delen door de standaarddeviaties der verschillen. Deze is gemakkelijk te berekenen, want bij gelijksoortige en ongecorreleerde grootheden is $\sigma_{a-b}^2 = \sigma_a^2 + \sigma_b^2 = 2\sigma_a^2$. De standaarddeviaties van de verschillen der regressiecoëfficiënten zijn dus $\sqrt{2}$ maal zo groot, dus resp. 0,148, 0,100 en 0,165. De verschillen zijn resp. 0,390, 0,272 en 0,001. De eerste twee zijn 2,6 en 2,7 maal zo groot als de standaarddeviaties der verschillen. Hun overschrijdingskansen zijn < 1 %. Zou de toegepaste statistische methode betrouwbaar zijn, m.a.w. zouden de regressiecoëfficiënten inderdaad reeds gaussisch verdeeld zijn, dan moeten de veranderingen in het gedrag van de Juli-temperatuur als reëel beschouwd worden.

Laten we, alvorens een uitspraak te doen, eerst eens de exacte methode toepassen. We kunnen de regressiecoëfficiënten van het tijdvak 1921–1940 gelijkstellen aan de coëfficiënten van het universum. We onderzoeken dan, of de b_j -waarden van het andere tijdvak (1901–1920) significant afwijken van die uit het eerst genoemde tijdvak.

Zo krijgen we door substitutie in de formule $t_j = \frac{(b_j - \beta_j) \sqrt{n-m-1}}{\sqrt{\sum_{i=1}^n (x_{0i} - X_{0i})^2}} \frac{\sqrt{D}}{\sqrt{D_{1j}}}$

$$\begin{aligned} t_1 &= \frac{0,390 \sqrt{20-3-1}}{\sqrt{0,41}} \frac{\sqrt{238,8}}{\sqrt{177,0}} = 2,83 & \nu &= 16 & P &= 1,3\% \\ t_2 &= \frac{0,272 \sqrt{20-3-1}}{\sqrt{0,41}} \frac{\sqrt{238,8}}{\sqrt{87,47}} = 2,81 & \nu &= 16 & P &= 1,4\% \\ t_3 &= \frac{0,001 \sqrt{20-3-1}}{\sqrt{0,41}} \frac{\sqrt{238,8}}{\sqrt{192,0}} = 0,007 & \nu &= 16 & P &= 99,4\% \end{aligned}$$

Deze uitkomsten zijn vergelijkbaar met die van het onderzoek, waarbij we de absolute verschillen gedeeld hebben door de standaarddeviaties van de regressiecoëfficiënten zelf. Dit gaf de overschrijdingskansen 0,04 %, 0,01 % en 99 %.

Deze kansen zijn veel kleiner dan de volgens de exacte methode bepaalde. Onze conclusie moet dan ook zijn, dat de benaderende methode niet toegepast mag worden op steekproeven van 20 elementen.

Konden we bij de benaderende methode de standaarddeviatie der regressiecoëfficiënten vervangen door de standaarddeviatie van de verschillen der regressiecoëfficiënten, deze verbetering kan nu niet zo eenvoudig aangebracht worden. Doch ongetwijfeld zullen we geen grote fout maken, als we t_j door $\sqrt{2}$ delen. t_1 en t_2 worden dan ca 2, waarmee een overschrijdingskans $> 5\%$ verbonden is. Nadat we op deze wijze eindelijk tot de beste toets gekomen zijn, is het resultaat, dat de regressiecoëfficiënten niet significant verschillen.

2.11.4 MEERVOUDIGE CORRELATIE

We houden ons aan het voorbeeld van de Juli-temperaturen van de periode 1901–1920.

De multiële correlatiecoëfficiënt $R_{0,123}$ is de correlatiecoëfficiënt tussen de x_0 en X''_0 van dat voorbeeld. Nu de regressiecoëfficiënten bekend zijn, kan $R_{0,123}$ het gemakkelijkst uit de formule

$$R_{0,123}^2 = 1 - \frac{\sum_{i=1}^{20} (x_{0i} - X_{0i})^2}{\sum_{i=1}^{20} x_{0i}^2} \quad \text{berekend worden.}$$

$$R_{0,123} = \sqrt{1 - \frac{0,41}{43,94}} = 0,9954$$

Kunnen we echter niet op een regressieonderzoek steunen, dan is $R_{0,123}$ het eindresultaat van correlatieberekeningen.

De opbouw van het correlatieonderzoek begint met de bepaling van alle mogelijke totale correlatiecoëfficiënten.

$$\begin{aligned} \text{Dit zijn: } r_{01} &= 0,9894 & r_{12} &= 0,9759 \\ r_{02} &= 0,9865 & r_{13} &= 0,9840 \\ r_{03} &= 0,9914 & r_{23} &= 0,9834 \end{aligned}$$

Hieruit leiden we de 3 partiële correlatiecoëfficiënten van de 1^{ste} orde af met behulp van de algemene formule op p. 38, nl.

$$r_{01,23 \dots p(p+1)} = \frac{r_{01,23 \dots p} - r_{0(p+1),23 \dots p} r_{1(p+1),23 \dots p}}{\sqrt{(1 - r_{0(p+1),23 \dots p}^2) (1 - r_{1(p+1),23 \dots p}^2)}}^{1/2}$$

$$r_{02,1} = \frac{r_{02} - r_{01} r_{12}}{\sqrt{(1 - r_{01}^2) (1 - r_{12}^2)}} = 0,6608$$

$$r_{03.1} = \frac{r_{03} - r_{01} r_{13}}{\sqrt{(1 - r_{01}^2)(1 - r_{13}^2)}} = 0,6892$$

$$r_{23.1} = \frac{r_{23} - r_{21} r_{13}}{\sqrt{(1 - r_{21}^2)(1 - r_{13}^2)}} = 0,5944$$

We zien, dat bij deze coëfficiënten, waarin de invloed van de ochtendtemperatuur geëlimineerd is, de correlatie aanmerkelijk kleiner is dan bij de totale correlatiecoëfficiënten. Deze vermindering zet zich door, wanneer we nog een temperatuur uitschakelen.

$$r_{03.12} = \frac{r_{03.1} - r_{02.1} r_{23.1}}{\sqrt{(1 - r_{02.1}^2)(1 - r_{23.1}^2)}} = 0,4911$$

Nu zijn we zover, dat we de multipele correlatiecoëfficiënt kunnen bepalen volgens $1 - R_{0.123}^2 = (1 - r_{01}^2)(1 - r_{02.1}^2)(1 - r_{03.12}^2)$.

We vinden $R_{0.123} = 0,9955$.

Dit resultaat is in goede overeenstemming met de uitkomst bij het regressie-onderzoek, in het begin van deze onderparagraaf verkregen.

Tot slot tonen we aan, dat de correlatie- en regressiemethoden – hoe nauw ook met elkaar verbonden – niet tot eenzelfde conclusie behoeven te voeren. Dit blijkt uit de kolommen u en v van het volgende staatje

	$R_{0.123} = u$	$\Sigma (x_{0i} - X''_{0i})^2 = v$	Σx_{0i}^2
1901–20	0,9954	0,412	43,94
1921–40	0,9916	0,289	17,36

In het eerste tijdvak hebben we de hoogste correlatie, in het tweede de beste regressie.

De paradox verdwijnt, als we het verband tussen u en v nog eens neerschrijven:

$$u^2 = 1 - \frac{v}{\Sigma x_{0i}^2}$$

2.11.4.1 DE FREQUENTIEFUNCTIE VAN DE MULTIPELE CORRELATIECOEFFICIENT R , WANNEER DE UNIVERSUMWAARDE $P = 0$ IS

Het heeft geen zin de multipele correlatiecoëfficiënten $R_{0.123}$ van 2.11.4 te toetsen, daar we er zeker van zijn, dat de universumwaarde $P_{0.123} \neq 0$ is, terwijl de werkelijke waarde van $P_{0.123}$ in een zeer vage duisternis ligt.

De door ons geconstrueerde grafiek in fig. 7 op p. 41 kan wel diensten bewijzen bij de beschouwing van de multipele correlatiecoëfficiënten, die VISSER verkreeg bij het opstellen van de regressievergelijkingen voor neerslag en temperatuur in Nederland (Med. en Verh. 51, p. 117–8). N is daar steeds = 39, m is 5 of 6.

VISSER schrijft op p. 119: „De multiële correlatiecoëfficiënt voor de neerslag loopt uiteen van 0,703 tot 0,847, de overschrijdingskans van 0,087 tot 0,046 %; voor de temperatuur resp. van 0,693 tot 0,845 en 0,090 – 0,048 %. De vergelijkingen voldoen alle aan de hoge gestelde eisen”.

Op p. 83 vermeldt VISSER: „De overschrijdingskans werd door grafische interpolatie ontleend aan de tabellen, die FISHER geeft. De verkregen waarde is dus niet geheel nauwkeurig maar voor ons doel voldoende”.

Inderdaad toont raadpleging van onze grafiek op p. 41, dat de interpolatie onnauwkeurig was. Wij vinden in plaats van 0,087 tot 0,046 % en 0,090-0,048 % de waarden 0,024 % tot $2,9 \cdot 10^{-6}$ % en 0,042 %– $8,5 \cdot 10^{-6}$ %. Daaruit blijkt, dat in werkelijkheid nog beter aan de hoge eisen voldaan is.

Staat de significantie derhalve vast, de vraag naar de realiteit van het correlatieonderzoek mag niet bevestigend beantwoord worden, alvorens de rol van het selectieeffect in het correlatieonderzoek aan een grondige beschouwing is onderworpen.

2.12 PERSISTENTIE

Aan deze gevaarlijke vijand van de statisticus wordt in 3.3 een uitvoerige bespreking gewijd. In de nu volgende regels worden enkele bestrijdingsmiddelen behandeld.

2.12.1 AUTOCORRELATIECOEFFICIENT VAN DE k^{de} ORDE

Waarnemingen, die elke minuut verricht of afgelezen kunnen worden, bestaan nauwelijks in de klimatologie. We zien er daarom vanaf om correlatiecoëfficiënten tussen dergelijke waarnemingen te berekenen.

Wel bestaan uurlijkse waarnemingen. We kunnen bv. de temperaturen van 8^h en 9^h op de 31 dagen van Augustus 1948 te De Bilt correleren, evenzo die van 8^h en 10^h . Het resultaat is: $r(8,9) = 0,91$ en $r(8,10) = 0,79$. De grootte van $r(8,x)$ neemt af met toenemende x . Uit een soortgelijk onderzoek als dat van 2.11.3 ontleen we voor dezelfde maand $r(8,14) = 0,57$. De daling zet zich dus voort, maar gaat blijkens $r(8,19) = 0,77$ later weer in stijging over. Dit gedrag is niet karakteristiek voor elke maand. Blijkbaar was in Augustus 1948 de band tussen ochtend- en avondtemperatuur sterker dan die tussen ochtend- en middagtemperatuur. Een afwisselende, niet alle dagen even grote bewolking verklaart deze anomalie.

Van de uurlijkse waarnemingen stappen we over op de daggemiddelden. De autocorrelatiecoëfficiënt van de eerste orde van de etmaalgemiddelden van de

temperatuur in Augustus 1948 te De Bilt is 0,70. Zoals verwacht mag worden, is de autocorrelatiecoëfficiënt van de eerste orde van de 8^h temperaturen (eveneens De Bilt, Aug. 1948) niet zo groot: $r(8,32) = 0,57$.

De correlaties tussen de maandgemiddelden zijn ook berekend. Als voorbeeld geven we de correlaties van Augustus met de 12 daarop volgende maanden voor het tijdvak 1741–1940.

i	0	1	2	3	4	5	6	7	8	9	10	11	12
$r(\text{Aug.}, \text{Aug.} + i)$	1,00	+ 0,38	+ 0,25	+ 0,07	+ 0,15	- 0,08	- 0,02	+ 0,08	+ 0,03	- 0,01	+ 0,05	+ 0,12	+ 0,11

Afgezien van steekproefafwijkingen blijkt hieruit, dat de correlatie met het groter worden van de afstand afneemt.

Tot slot zijn de correlatiecoëfficiënten tussen de jaargemiddelden van de temperatuur te De Bilt bepaald.

Bij de autocorrelatiecoëfficiënt van de eerste orde der jaartemperaturen zijn de 40 jaargemiddelden van 1901–1940 gecorreleerd met de 40 jaargemiddelden van 1902–1941. Het resultaat was: $r_1 = + 0,010$.

Bij de autocorrelatiecoëfficiënt van de tweede orde der jaartemperaturen zijn de 40 jaartemperaturen van 1901–1940 gecorreleerd met de 40 jaartemperaturen van 1903–1942. Nu was de uitkomst: $r_2 = + 0,012$.

Beide bedragen zijn zo laag, dat het verantwoord is te zeggen, dat de jaargemiddelden van de temperatuur in De Bilt onafhankelijk van elkaar zijn. De jaartemperaturen zijn niet persistent.

2.12.2 HET EFFECTIEVE AANTAL ONAFHANKELIJKE WAARDEN $\gamma(g)$

In deze onderparagraaf zullen we een voorbeeld van BARTELS (1947, 91) volgen. Hij berekende de schatting $c(31)$ voor het effectieve aantal onafhankelijke waarden in de 31 dagelijkse temperatuurgemiddelden op de Sonnblick in Januari. De uitvoerige berekeningen, die STEINHAUSER (1938) met de meteorologische gegevens van de Sonnblick verricht heeft, omvatten ook de standaarddeviatie van de etmaalgemiddelden van de temperatuur in Januari. Voor het tijdvak 1887–1934 is deze $s = 5,06$ °C. Zou het dagelijkse temperatuurgemiddelde generlei persistentie vertonen, dan moet het maandgemiddelde van de temperatuur in Januari op de Sonnblick een standaarddeviatie hebben $= 5,06/\sqrt{31} = 0,91$ °C. De 48 Januari-gemiddelden geven echter een standaarddeviatie $s(31) = 2,21$ °C. Uit de formule $c(g) = \frac{s^2}{s^2(g)}$ volgt dan, dat $c(31) =$

$$\left(\frac{5,06}{2,21}\right)^2 = 5,2.$$

2.12.3 HET AEQUIVALENTE HERHALINGSGETAL $\omega(g)$

De beste schatting $o(g)$ voor de grootheid $\omega(g)$ vinden we uit de formule

$$o(g) = \frac{g}{c(g)} = g \frac{s^2(g)}{s^2}$$

Het equivalente herhalingsgetal van het dagelijkse temperatuurgemiddelde op de Sonnblick was in de Januarimaanden (1887–1934) dus gelijk aan $\frac{31}{5,2} = 6,0$.

De juiste betekenis ervan is de volgende:

Bepalen we de temperatuurgemiddelden over hexaden (perioden van zes dagen), dan zal de standaarddeviatie van de hexadentemperaturen even groot zijn als de standaarddeviatie, berekend uit de 31 etmaalgemiddelden.

De persistentie heeft de neiging de standaarddeviatie van de gemiddelden te vergroten. Bestond er geen persistentie in de dagelijkse temperatuurgemiddelden, dan zouden de hexadengemiddelden dichter om het totaalgemiddelde schommelen.

Als de autocorrelatiecoëfficiënten bekend zijn, kan $\omega(g)$ ook berekend worden met behulp van de formule $\omega(g) = 1 + 2 \left(\frac{g-1}{g} \rho_1 + \frac{g-2}{g} \rho_2 + \dots + \frac{1}{g} \rho_{g-1} \right)$.

De coëfficiënten ρ_i zijn niet bekend, maar we zullen twee series aangenomen waarden substitueren.

Eerst de serie $\rho_i = 1 - \frac{i}{10}$, daarna $\rho_i = 1 - \frac{i}{5}$. De eerste serie levert $\omega(31) = 8,94$; de tweede $\omega(31) = 4,74$. De berekende schatting $o(31) = 6,0$ ligt tussen de waarden 8,94 en 4,74 in. Blijkbaar moeten de ware waarden van ρ_i voorname-lijk gezocht worden in het interval tussen $1 - \frac{i}{5}$ en $1 - \frac{i}{10}$.

HOOFDSTUK 3

ENIGE FOUTENBRONNEN BIJ HET STATISTISCH
ONDERZOEK

The detection of errors is the first and surest step towards the
discovery of the truth. HALLEY 1704

3.0 INLEIDING

Wanneer de geofysicus uit zijn omvangrijk waarnemingsmateriaal regels en wetten wil afleiden, kan hij in vele gevallen moeilijk anders doen dan statistische methoden toepassen. Om het gebruik van deze methoden te stimuleren en te vergemakkelijken heeft de theoretische statisticus ze in de vorm van recepten beschikbaar gesteld.

Dit heeft het voordeel, dat meer en meer geofysici hun materiaal kritisch-statistisch gaan bewerken met als gunstig gevolg een daling in het aantal veronderstellingen, die op zeer lichtvaardige gronden en uiterst zwakke aanwijzingen jaarlijks in de geofysische tijdschriften gepubliceerd worden.

Het nadeel van de in receptvorm gegoten statistische methoden is echter de welhaast onvermijdelijke mechanische, gedachtenloze toepassing. Soms, bv. in de eenvoudigste gevallen, kan dit in het geheel geen kwaad. In andere gevallen, vooral wanneer het erom gaat, oorzakelijke samenhangen op te sporen, lopen de statistische methoden gevaar verkracht te worden.

Het is een onmogelijke eis van elke gebruiker van statistische methoden te vergen, dat hij de gehele mathematisch-statistische theorie beheerst.

Gelukkig is er een andere oplossing, die het aantal misverstanden en fouten op dit speciale gebied kan reduceren. We zullen trachten deze in het kort te schilderen.

Om verkeerde toepassingen van de statistische methoden te voorkomen kan men als eis stellen, dat er „statistisch gedacht” moet worden. Anderen stellen de eis, dat „het gezonde verstand” gebruikt moet worden.

De eerste formulering van de eis, vervat in algemene bewoordingen, kan men herhaaldelijk vinden in het Nederlandse tijdschrift „Statistica” (bv. NASS 1950, 232). Naar onze mening heeft de geofysicus onder „statistisch denken” te verstaan, het herkennen en het vermijden van een aantal soorten van fouten, die in het vervolg van dit hoofdstuk uitvoerig behandeld zullen worden.

De tweede formulering was het telkens weer – te pas en te onpas – terugkerende thema van de Meteorologen-Tagung te Hamburg in October 1950 (zie de

Diskussionsbemerkung zu Prof. KOPPE in het Sonderheft van de Annalen der Meteorologie van 1951, p. 94). Ook BEVERIDGE (1950) zegt: „The use of statistics does not lessen the necessity for using common sense”.

Het gezonde verstand kan in eenvoudige gevallen beslissen, maar soms kan dit criterium niet de doorslag geven. Het kan voorkomen, dat, wat de een „bon sens” vindt, door de ander „nonsense” genoemd wordt. Een dergelijke tweeslachtigheid wordt vermeden, wanneer men gebruik gaat maken van een nadere precisering van het begrip „gezond verstand”. Of liever van vergrijpen tegen het gezonde verstand.

Deze vergrijpen kunnen in enige categorieën ingedeeld worden. Voorlopig is hun aantal beperkt tot vijf, maar het is best mogelijk, dat andere fouten in de toekomst zo vaak zullen optreden, dat het de moeite loont, meer categorieën aan de reeds bestaande toe te voegen.

De vergrijpen ontstaan, wanneer men de invloed van het volgende vijftal effecten verwaarloost: 1) het steekproefeffect, 2) het selectieeffect, 3) het persistentieeffect, 4) het pseudocorrelatieeffect en 5) het krommingseffect.

Men kan deze vijf effecten in twee vrij duidelijk gescheiden groepen indelen. Het steekproefeffect, het selectieeffect en het pseudocorrelatieeffect behoren thuis in de groep, waarmee de „geofysicus – niet statisticus” moeite zal hebben, omdat ze volkomen buiten zijn terrein liggen. Daarentegen zal de „statisticus – niet geofysicus” door de overige twee effecten verrast kunnen worden, omdat deze in wezen door het bijzondere karakter van sommige geofysische verschijnselen veroorzaakt worden.

Deze vijf vijanden van de geofysicus-statisticus kunnen zijn werk volkomen bederven en de moeizaam verkregen berekeningen tot waardeloze rekenresultaten degraderen.

Wanneer de geofysicus bij elk onderzoek, ja zelfs bij elk onderzoekje zich afvraagt, of hij voldoende rekening gehouden heeft met het vijftal zo juist opgesomde effecten, dan zullen de vaktijdschriften minder artikelen kunnen publiceren, waarop de statisticus aanmerkingen moet maken.

Na deze algemene beschouwingen volgt eerst een korte bespreking van de vijf effecten, daarna komt pas de uitvoerige bespreking.

3.0.1 HET STEEKPROEFEEFFECT

Het eerste, het steekproefeffect is het belangrijkste. Het komt overal daar voor, waar we conclusies, getrokken uit het waarnemingsmateriaal (= de steekproef), willen generaliseren, m.a.w. geldig maken voor niet gemeten objecten (= het universum). En dit generaliseren is nu juist het doel van welhaast alle wetenschappelijke onderzoeken.

De beste remedie tegen het steekproefeffect verkrijgen we, als we het onbekende universumresultaat niet exact gelijk stellen aan het steekproefresultaat, maar het in een interval van waarden, aan weerszijden van het steekproefresultaat, aanwezig denken.

3.0.2 HET SELECTIEEFFECT

Het is zeer moeilijk te omschrijven, wat men onder het tweede effect – het selectieeffect – heeft te verstaan.

Bij de gewone statistische bewerking van gegevens zal men dit effect nauwelijks ontmoeten, ook niet bij het toetsen van een van tevoren nauwkeurig door de theorie vastgelegde hypothese. Pas wanneer men verband gaat zoeken tussen twee verschillende grootheden, loopt men gevaar, de dupe te worden van het selectieeffect.

Dit gebeurt vooral, wanneer men het vraagstuk niet objectief te lijf gaat en zich dus bij elke stap bewust of onbewust laat leiden door niet geverifieerde onderstellingen over de werkelijkheid.

Dergelijke vooropgezette meningen spelen vaak een rol. Welbekende voorbeelden uit de geofysica zijn: a) het verband tussen zonnevlekken en sommige biologische grootheden (PETERSEN 1947), b) het verband tussen Aran en vele biologische grootheden (CURRY 1946), c) het verband tussen de planeten en de meteorologische grootheden (JOHNSON 1946). Andere takken van wetenschap zouden deze opsomming ontstellend lang kunnen maken.

Dat een dergelijke wijze van wetenschappelijk onderzoek niet gezond is, blijkt uit het niet zeldzaam voorkomen van een idee-fixe. En de oorsprong daarvan moet gezocht worden in het op een niet kritische wijze aanvaarden van ongeverifieerde onderstellingen over de werkelijkheid. ROUSSEAU zegt terecht in zijn „Emile” (livre 4): „Les idées générales et abstraites sont la source des plus grandes erreurs des hommes.”

Aan de hand van talrijke voorbeelden zal in 3.2 getracht worden duidelijk te maken, wat het selectieeffect is en hoe het geëlimineerd kan worden.

3.0.3 HET PERSISTENTIEEFFECT

Dit doet zich voor bij gebeurtenissen, die niet los van elkaar staan. De meeste statistische wetten berusten op de veronderstelling, dat de enkelvoudige gebeurtenissen geheel onafhankelijk van elkaar zijn.

Degene, die dergelijke statistische methoden zonder enige beperking toepast op een verzameling van onderling afhankelijke gebeurtenissen, maakt zich schuldig aan het verwaarlozen van het persistentieeffect.

Als de tijdsruimte tussen de te beschouwen gebeurtenissen maar voldoende

groot is, zullen deze gebeurtenissen als geheel onderling onafhankelijk behandeld mogen worden.

Om het persistentieeffect uit te schakelen behoeft men dus niets anders te doen dan zoveel tussenliggende gebeurtenissen te schrappen, tot de overblijvende inderdaad onafhankelijk blijken te zijn.

Deze methode is theoretisch juist en het enige bezwaar, dat geopperd kan worden, is het volgende: het waarnemingsmateriaal kan door deze aderslatingen zo dun worden, dat de steekproef, gevormd door de resterende waarnemingen, slechts zeer onbetrouwbare parameterwaarden oplevert. Het steekproefeffect kan dan een rol gaan spelen.

3.0.4 HET PSEUDOCORRELATIEFFECT

Dit effect heeft ook heel wat slachtoffers gemaakt, vooral onder hen, die jacht maakten op hoge correlaties.

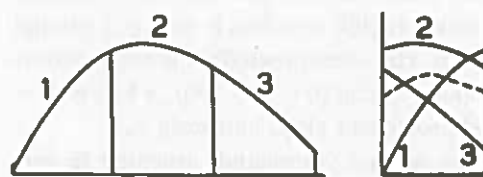
Correleert men twee grootheden, die een of meer termen gemeen hebben, dan zal de correlatiecoëfficiënt een universumwaarde hebben, die $\neq 0$ is. De gevonden steekproefwaarde zal in heel veel gevallen zo ver van 0 afwijken, dat de correlatie significant genoemd moet worden. Dit mag natuurlijk in het geheel geen verwondering teweeg brengen. Toch zijn enkele onderzoekers er zo door verast, dat ze een verkeerde conclusie trokken.

Bij een significante correlatie tussen a_i en $(a_i - b_i)$ – één van de eenvoudigste gevallen – concludeert men dan tot de aanwezigheid van een verband tussen a_i en b_i , wat in het algemeen onjuist is. De hoge waarde van de correlatiecoëfficiënt wijst slechts op een samenhang tussen a_i en $(a_i - b_i)$. Voor $a_i \neq b_i$ is dit verband triviaal.

3.0.5 HET KROMMINGSEFFECT

Het krommingseffect treedt zelden op. Het werd omstreeks 1922 door BARTELS ontdekt. In 1.1 werd er reeds een voorbeeld van gegeven.

Het krommingseffect berust op de volgende eigenschap van convexe curven: men verdeelt de abscissenas van een convexe kromme in een willekeurig aantal delen van gelijke lengte. Vervolgens plaatst men de stukjes zo, dat de begin- en eindpunten van de abscissensegmenten samenvallen. Voor elke abscissenwaarde



in het gemeenschappelijke interval bepaalt men nu de gemiddelde ordinaten van alle stukjes van de kromme. Het zal dan blijken, dat de op deze wijze gevonden „gemiddelde curve” een glad convex uiterlijk heeft.

3.1 HET STEEKPROEFEFFECT

Reeds enkele malen zijn de begrippen steekproef en universum gedefinieerd en toegelicht. De mededeling, dat de parameter berekend uit de steekproefgegevens in het algemeen niet gelijk zal zijn aan de overeenkomstige parameter van het universum, zal dan ook geen verbazing meer kunnen verwekken. De oorzaak voor dit verschil is nu bekend: universum en steekproef hebben immers een verschillende samenstelling.

Vroeger was men maar al te vaak geneigd het verschil tussen steekproef- en universumparameter aan de één of andere, nauwkeurig afgebakende oorzaak toe te schrijven. Men vergat dan, dat de grootte van de steekproefparameter ook afhankelijk is van de wijze van samenstelling van de steekproef, zodat het niet altijd noodzakelijk is een bepaalde concrete oorzaak verantwoordelijk te stellen voor het gevonden verschil.

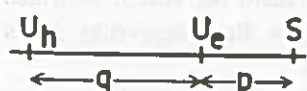
Wanneer iemand tweemaal dezelfde optelling verricht en de resultaten zijn ongelijk, dan zegt men korter, dat een rekenfout daarvan de oorzaak is geweest, in plaats van een lang verhaal, dat bv. de ene optelling foutloos verlopen is, maar dat de andere maal ergens in het midden van de kolom de som van $24 + 7$ met 32 aangegeven werd.

Of een ander voorbeeld: vindt men in een boek een foutief woord en weet men, wat er eigenlijk had moeten staan, dan kan men deze discrepantie verklaren door te zeggen, dat de zetter een fout gemaakt heeft door bv. een p in plaats van een v te nemen, enz. enz. Veel korter is natuurlijk de verklaring, dat een zetfout de schuld is van de onregelmatigheid.

Op soortgelijke wijze komen we er toe om het verschil tussen steekproef- en universumparameter kortweg te beschouwen als een steekproeffout (sampling error). Met andere woorden: we kunnen dat verschil toeschrijven aan het steekproefeffect. Konden we dit niet doen, dan waren we gedwongen steeds een uitvoerige beschrijving te geven van het feit, waarom de waarden voor steekproef- en universumparameter in het algemeen verschillen.

Vele statistici gaan weleens zo ver te beweren, dat een bepaald verschil een gevolg is van het steekproefeffect. Dit mag men echter nooit beweren. Immers, het is in het gunstigste geval slechts mogelijk aan te tonen, dat het verschil veroorzaakt kan zijn door het steekproefeffect. Steeds bestaat daarnaast de mogelijkheid van een concrete invloed, die verantwoordelijk te stellen is voor $q\%$ van het gevonden verschil ($0 \leq q < 100$ of $q > 100$). Het steekproefeffect is verantwoordelijk te stellen voor $p\%$ van het gevonden verschil ($0 < p \leq 100$). p kan niet = 0 zijn en q niet = 100, omdat het steekproefeffect altijd aanwezig is.

Aan de hand van een tekening zullen we het voorgaande trachten te verduidelijken.



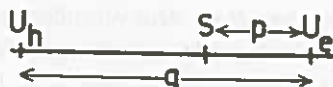
U_h stelt de parameterwaarde voor, geleverd door een hypothetisch universum, nl. het universum, dat uit de door ons opgestelde theorie of veronderstellingen volgt.

U_e is de parameterwaarde in het echte universum: dit is het universum, dat we zouden krijgen, als we de steekproef naar willekeur konden uitbreiden.

U_e is en blijft onbekend, U_e kan met U_h samenvallen; in dat geval zijn onze hypothesen geheel en al juist.

S is de steekproefwaarde van de desbetreffende parameter. De waarschijnlijkheid, dat S met U_h of U_e samenvalt, is zo gering, dat we deze mogelijkheden verwaarlozen. Vandaar $p > 0$ en $q \neq 100$.

q kan ook groter zijn dan 100, namelijk in het geval, dat U_e buiten het interval $< U_h, S >$ ligt.



Blijft het verschil $S - U_h$ binnen redelijke grenzen – de grenzen worden langs statistische weg bepaald – dan is het ongemotiveerd van een significant verschil $S - U_h$ te spreken. M.a.w. er is dan geen enkele reden om $U_e \neq U_h$ te beschouwen. Evenmin is er reden om dit als bewijs van $U_e = U_h$ te laten gelden.

De juiste conclusie luidt: het steekproefmateriaal geeft geen aanleiding om aan de juistheid van onze hypothese te twijfelen. Korter: men mag niet zeggen $U_e \neq U_h$.

De grootte van p kan nooit berekend worden, daarom blijft ook q onbekend. Op dit laatste is één uitzondering mogelijk, nl. wanneer het verschil tussen de steekproefparameter S en de parameter U_h van het hypothetische universum zo groot is, dat het niet uitsluitend aan het steekproefeffect geweten kan worden. Dan moet daarvoor een bepaalde oorzaak aanwezig zijn, dus $q > 0$ en $U_h \neq U_e$. Heeft men kunnen berekenen, dat het steekproefeffect slechts heel gering kan zijn (bv. bij een serie laboratoriummetingen) en is daarentegen het verschil tussen steekproef- en universumparameters ($S - U_h$) zeer groot, dan ligt p zo dicht bij 0, dat het steekproefeffect verwaarloosd mag worden ($S = U_e$). Het verschil tussen de beide parameters moet dan aan één of meerdere concrete oorzaken worden toegeschreven ($S - U_h = U_e - U_h$).

De invloed van het steekproefeffect is reeds eenmaal gedemonstreerd aan de hand van een voorbeeld in 2.1b. Daar werd aangetoond, dat een conclusie van CONRAD wel geldig was voor het tijdvak 1901–1910, maar zijn geldigheid verloor voor het tijdvak 1899–1938.

Het steekproefeffect speelt ook een grote rol in het felle duel tussen de Duitse

meteorologen RODEWALD en BAUR. Moge men de felheid betreuren, waarmee zij elkaar aanvallen, er is toch ook een lichtzijde aan dit tweegevecht. Er is namelijk een lachende derde. En dat is de statistica!

Beide heren beschuldigen elkaar van het op onvoldoende wijze analyseren van het beschikbare waarnemingsmateriaal en het te snel generaliseren van enkele markante verschijnselen, die optraden bij een gering aantal gevallen.

De belangstellende toeschouwer krijgt tijdens het verloop van het duel op duidelijke wijze een beeld voor ogen gevoerd van de macht van het steekproef-effect. De toeschouwer bemerkt bovendien, hoe noodzakelijk het is, dat aan sommige geofysici een statistische bril wordt voorgeschreven!

De geschiedenis van het duel is als volgt:

RODEWALD (1948) had bemerkt, dat de vijf strenge winters van de laatste 20 jaar voorafgegaan werden door grote drukveranderingen. Hij construeerde de verdeling van de maandelijkse luchtdrukafwijkingen voor October, November en December en wel gemiddeld over de vijf jaren 1928, 1939, 1940, 1941 en 1946. Vervolgens werden de veranderingen in de gemiddelde luchtdrukverdelingen bepaald tussen de opvolgende maanden. De gegevens van October werden afgetrokken van die van November, de luchtdrukwaarden van November afgetrokken van de overeenkomstige in December. Zo ontstonden kaarten, die aangaven de gemiddelde veranderingen van de afwijkingen van de luchtdrukverdeling boven Europa en het Oostelijk gedeelte van de Atlantische Oceaan van October op November en van November op December voor de vijf strenge winters uit de laatste 20 jaren. Zouden deze winters onafhankelijk van hun voorgeschiedenis tot stand gekomen zijn, dan moest men op beide kaarten een min of meer vlak verloop van toevallige afwijkingen aantreffen.

Het werk van RODEWALD bleek onverwachte resultaten te geven. Op beide kaarten kwamen suggestieve gedeelten voor. De eerste ($\Delta P_{\text{Nov.}} - \Delta P_{\text{Oct.}}$) toonde een sterke luchtdrukdaling boven de Noordzee en Groot Brittannië, de tweede ($\Delta P_{\text{Dec.}} - \Delta P_{\text{Nov.}}$) een even grote luchtdrukstijging in dezelfde omstreken. Het is zeer verleidelijk deze karakteristieke veranderingen te gebruiken voor weersverwachtingen op lange termijn, i.c. wintervoorspellingen.

BAUR (1950) bleek de verleiding te kunnen weerstaan. Hij vroeg zich onmiddellijk af: kan RODEWALD's resultaat het boze werk geweest zijn van het steekproefeffect?

Deze vraag kan ontkennend beantwoord worden, wanneer men kan aantonen, dat alle grote luchtdrukdalingen¹⁾ van October op November ($\Delta P_{\text{Nov.}} - \Delta P_{\text{Oct.}} < 0$) en sterke drukstijgingen¹⁾ van November op December ($\Delta P_{\text{Dec.}} - \Delta P_{\text{Nov.}}$

¹⁾ Steeds beschouwd als afwijking van de normale drukverandering.

>> 0) in het gebied van de Noordzee steeds gevolgd werden door een koude of strenge winter.

Zoals BAUR terecht opmerkt, mag men zich niet tevreden stellen met het vermelden van die argumenten, die pleiten voor de hypothese van RODEWALD. Het klimatologisch arsenaal van waarnemingen geeft de gelegenheid de hypothese ook in andere jaren te toetsen en men moet er daarom ook gebruik van maken.

RODEWALD heeft in zijn artikel van 1948 feitelijk niet veel meer dan beschrijvende statistiek geleverd. De door hem gevonden samenhang gold wel voor de periode 1928–48, maar het gaat niet aan, deze samenhang zonder enig nader onderzoek als causaal te beschouwen. Dit mag eerst, wanneer toevalseffecten uitgesloten geacht mogen worden. Pas dan kan het gevonden verband ook buiten het onderzochte tijdvak geldig verklaard worden en o.a. voor voorspellingsdoel-einden toegepast worden. (Zie ook 4.4).

BAUR toonde vervolgens overduidelijk aan, dat er van de hypothese van RODEWALD voor 1928 weinig terecht kwam – tenminste van de hypothese van RODEWALD, zoals deze door BAUR opgevat werd. Want de controverse bleek nog vergroot te zijn door een misverstand.

Prof. BAUR veronderstelde namelijk, dat RODEWALD het karakter van de te verwachten winter wilde voorspellen aan de hand van het al of niet aanwezig zijn van anomalieën in de maandelijkse luchtdrukverandering in de naaste omgeving van Schotland. In dat geval zou het juist zijn, Edinburgh – met zijn lange rij van meteorologische waarnemingen vanaf 1771! – als representatief punt te beschouwen. BAUR berekende de correlatiecoëfficiënt tussen het luchtdrukverschil November – October in Edinburgh en de wintertemperatuur in Centraal Europa voor het tijdvak 1771/72 – 1948/49. Uitkomst: $+0,05$. Deze lage waarde geeft geen aanleiding een lineair verband aanwezig te onderstellen.

Vervolgens werd de correlatiecoëfficiënt met het luchtdrukverschil December – November bepaald. Deze was voor hetzelfde tijdvak $-0,36$. De significantie van dit resultaat noemt BAUR triviaal, een woord, waartegen RODEWALD in dit verband bezwaar heeft. M.i. echter ten onrechte, want het is gemakkelijk in te zien, dat de correlatiecoëfficiënt negatief moet zijn. Hier spelen persistentie en pseudocorrelatie een rol. De volledige verklaring volgt in 3.4.

Met recht concludeerde BAUR uit deze correlatiewaarden, dat het luchtdrukverloop in Edinburgh in de herfst de wintertemperatuur in Centraal Europa maar weinig beïnvloedt.

RODEWALD riposteerde hierop met de opmerking, dat men dit ook al kon zien aan zijn kaarten met het drukverloop, gemiddeld over de jaren 1885, 1892, 1894 en 1900. (Deze jaren werden gevolgd door een strenge winter). Edinburgh ligt in November op en in December vlak bij de nullijn van de isallobarische centra.

Edinburgh mag dus in geen geval representatief genoemd worden. De resultaten van Prof. BAUR zijn daarmee in overeenstemming.

De kaart met de gemiddelde verandering van de luchtdruk voor de strenge winters 1886, 1893, 1895 en 1901 vertoont dezelfde isallobarische centra als de kaart, waarop de barische voorbereidingen voor de strenge winters van de laatste twee decennia te zien zijn. Deze overeenkomst werd door RODEWALD als een wetmatigheid beschouwd. Maar ook dit is prematuur, want de zachte winters 1893/94 en 1926/27 kwamen na een drukverloop, analoog aan dat voor strenge winters (BAUR 1950).

Na op deze wijze de aanval van BAUR enigszins gepareerd te hebben, ging RODEWALD vervolgens zelf tot de aanval over. Hij toonde aan, dat BAUR zelf in enige gevallen geen rekening gehouden heeft met het steekproefeffect. Het meest illustratieve voorbeeld is dat van het verband tussen de temperatuur van de eerste Decemberdekade in Berlijn en de temperatuur van de daaropvolgende winter in Centraal Europa.

RODEWALD zegt daarover (1950, 232): „Wenn nun wenig später eine BAURsche 100-Prozentigkeit folgt, nämlich das Auftreten milder Mitteleuropa-Winter nach einem um wenigstens 2,5° zu warmen 1. Dezember-Drittel in Berlin, so muss der Leser annehmen, dass dieser nach BAUR ‚eindrucksvolle Beleg‘ das vorhandene Material vollständig erfasse und ausschöpfe. Der Gedanke, dass in einem Aufsatz, der das Prinzip der Vollständigkeit in der Statistik verfißt, eben gegen dieses Prinzip verstossen werden könnte, wird ihm nicht kommen”.

Prof. BAUR vond, dat zijn regel na 1848 steeds uitkwam. RODEWALD ontdekte, dat het eerste jaar vóór dit tijdvak reeds een uitzondering bracht. De gemiddelde temperatuurafwijking te Berlijn van 1–10 December 1847 bedroeg nl. + 4,1°C. Er volgde een strenge winter: de gemiddelde temperatuur in Januari en Februari was in Centraal Europa 2,4°C beneden de normaal. RODEWALD onderzocht alle jaren van 1766–1847. Er bleken nog meer uitzonderingen op de regel van Prof. BAUR te zijn. Men kan ze vinden in Tabelle 5 (van RODEWALD 1950, 232). RODEWALD trekt deze conclusies: „Men ziet, dat er in dit tijdvak geen sprake kan zijn van een indrukwekkend optreden van de ‚herhalingsneiging‘ na uitgesproken warme of koude eerste Decemberdekaden in Berlijn. Koude, zachte en normale winters (omvattend de maanden Januari en Februari) komen in ongeveer gelijke aantallen voor, onverschillig, hoe de voorafgaande eerste Decemberdekade zich gedroeg.”

Behalve in deze, door RODEWALD naar voren gebrachte gevallen verwaarloost Prof. BAUR daarvóór reeds eenmaal het steekproefeffect, en wel onmiddellijk na de filippica tegen RODEWALD.

BAUR (1950, 79) correleert eerst de temperatuur in NW Siberië in de maand November met de temperatuur van de daaropvolgende winter in Centraal Europa en vindt $r_{\text{Nov.}} = +0,42$ (1885–1930). Daarna correleert hij het verschil tussen de November- en October-temperatuur in NW Siberië met de daaropvolgende wintertemperatuur in Centraal Europa en vindt $r_{\text{Nov.-Oct.}} = +0,63$ (eveneens 1885–1930). BAUR's conclusie luidt: „Men ziet, dat de correlatie met de verandering in temperatuur van October op November groter is dan met de November-temperatuur zelf”. Strikt genomen geldt dit alleen voor het beschouwde tijdvak 1885–1930. Er is geen enkele aanleiding om het verschil in de correlatiecoëfficiënten niet aan het steekproefeffect toe te schrijven, want in 2.11.1.1 vonden we een overschrijdingskans van 17% bij het toetsen van de hypothese, dat de twee correlatiecoëfficiënten uit hetzelfde universum afkomstig zijn.

Desondanks telt BAUR dit verschil ($0,63 - 0,42$) als significant, getuige zijn slotzin: „Es kommt also nicht allein auf den Kaltluftvorrat als solchen, sondern gewissermassen auf die (zeitliche) Phase an, in der er sich befindet”.

De reactie van HUSSLEIN (1949) op het eerste artikel van RODEWALD (1948) mag hier niet onvermeld blijven, echter wel onbesproken.

Men leze zelf eens na, hoe HUSSLEIN op waarlijk voorbeeldige wijze aantoont, dat de hypothese van RODEWALD niet door alle gevallen bevestigd wordt.

Dat verwaarlozing van het steekproefeffect geen zelden voorkomend verschijnsel is, bewijst de nu volgende bespreking van het artikel „Zurückführung des Grosswetters auf solare Erscheinungen” (BAUR 1949).

Beschouwen we eerst eens de slotsom, waartoe BAUR in zijn résumé komt (p. 358): „Tijdens de zonnevlekkenextrema neemt de planetarische circulatie van de atmosfeer af, de moessoncirculatie echter toe, de subtropische hogedruk gordels hebben minder betekenis en de tegenstelling tussen winter- en zomerluchtdruk op het Aziatische continent is groter. Ongeveer $1\frac{1}{2}$ tot $2\frac{1}{2}$ jaar voor de zonnevlekkenextrema neemt de planetarische circulatie toe, de subtropische hogedruk gordels tonen meer invloed en zijn naar de polen verschoven, het verschil tussen winter- en zomerluchtdruk op het vasteland is verminderd, de neiging tot strenge winters is geringer”.

Prof. BAUR verklaart later nog: „Deze kenmerken treden niet in elke cyclus op, maar toch in een overwegende meerderheid. Zo was bv. in de 13 zonnevlekkencycli van het minimum 1798 tot het minimum 1944 de zomer (1 Juli–31 Augustus) twee jaren voor het minimum in 8 gevallen te droog en slechts in 5 gevallen natter dan normaal; de zomer (Juli + Augustus) twee jaren voor het maximum was in 10 gevallen te droog en slechts in 3 gevallen te nat” (p. 360–2).

Ongetwijfeld zijn dit de markantste kenmerken van de atmosfeer, zoals Prof.

BAUR zich die voorstelt. Maar aan de andere kant zijn deze eigenschappen nu toch weer niet zo markant, dat de vraag „toeval of realiteit?” overbodig schijnt. BAUR stelt zichzelf ook die vraag en geeft een bewijs, waaruit de realiteit van de zonnevlekkeninvloed zou moeten blijken. In 3.2 worden enkele bezwaren tegen dit bewijs aangevoerd.

BAUR bepaalde het gemiddelde verloop in een zonnecyclus van een tiental meteorologische grootheden over de gehele wereld verspreid. Het gemiddelde werd berekend over 5 tot 18 zonnevlekkencycli.

De tabel op p. 961 van HANN-SÜRING: Lehrbuch der Meteorologie (vijfde uitg.) vermeldt, hoe de jaren na 1752 in BAUR's zonnevlekkenschema zijn ingedeeld. Om de gemiddelde gang van meteorologische elementen in de zonnevlekkencyclus te krijgen, behoeft men niets anders te doen dan de jaren in het zojuist genoemde schema te vervangen door de waarden van de te onderzoeken meteorologische grootheden in de desbetreffende jaren.

Zijn resultaten geeft BAUR (1949, 361) in een grafiek weer en in getallenwaarden in HANN-SÜRING (p. 962, vijfde uitg.). Twee uitkomsten zullen we eens nader bekijken.

	Jaren voor			zonne- vlekken		na		voor			zonne- vlekken		na	
	3	2	1	min	1	2		2	1		max	1	2	
A	-0,27	+1,11	-0,26	+0,16	-0,77	-0,80		+0,38	-0,50	+0,15	+0,67	+0,67	0,00	
B	+5,8	-9,5	+8,1	-4,1	+5,9	+1,0		-25,8	+4,0	+15,9	+2,7	-3,9		

A: afwijking van de luchtdruk te Berlijn in mb in Juli + Augustus, 1875-1944, ten opzichte van het gemiddelde over die periode.

B: afwijking van de hoeveelheid neerslag in mm voor Centraal Europa in Juli + Augustus, 1803-1943, ten opzichte van het gemiddelde over die periode.

Wij hebben dezelfde bewerkingen toegepast op de waarnemingen in Nederland en daarbij het tijdperk uitgebreid.

	Jaren voor			zonne- vlekken		na		voor			zonne- vlekken		na	
	3	2	1	min	1	2		2	1		max	1	2	
A'	-0,17	+0,95	-1,37	+0,15	-1,30	-1,58		+0,38	+0,47	-0,18	+0,18	+0,67		
B'	+4,2	-6,7	-6,1	+18,9	+16,2	-8,1		-26,1	+14,4	+7,2	-11,7	-5,9		

A': afwijking van de luchtdruk in de maanden Juli en Augustus te Utrecht-De Bilt in mb voor het tijdvak 1849-1949, ten opzichte van het gemiddelde over die periode.

B': afwijking van de neerslaghoeveelheden in Juli en Augustus te Zwanenburg-Hoofddorp in mm voor het tijdvak 1752-1949, ten opzichte van het gemiddelde voor die periode.

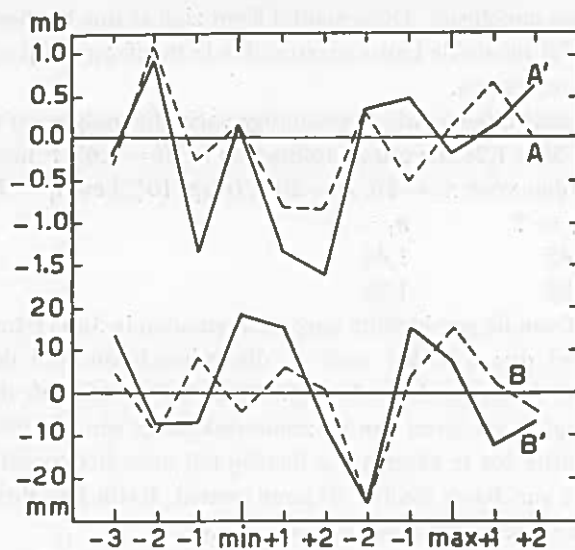


fig. 13

Ter illustratie worden A, B, A' en B' in grafiek weergegeven (fig. 13). Naast grote overeenkomsten zien we ook grote verschillen. De overeenkomsten vinden hun oorzaak in het gedeeltelijk samenvallen van de tijdvakken (A: 70% van A', B: 70% van B') en de relatief geringe geografische afstand. Deze zelfde oorzaken kunnen ook de verschillen verklaren. Immers, nemen we aan, dat Berlijn en De Bilt (resp. Hoofddorp) op gelijke wijze op de zonnestoringen reageren, (geografisch effect = 0), dan moeten we de verschillen tussen A en A' en tussen B en B' verklaren door het steekproefeffect. Dit betekent, dat BAUR's resultaten niet de ware invloed van de zonnevlekken voorstellen, maar versluierd door niet onbelangrijke toevalsverschijnselen. Zouden we daarentegen aannemen, dat er over voldoende zonnevlekkencycli gemiddeld is om het steekproefeffect te niet te doen, dan moeten we de verschillen verklaren door de ongelijke invloed, die de zon op de atmosfeer te Berlijn en De Bilt (resp. Hoofddorp) oefent.

Beide verklaringsmogelijkheden doen ernstige afbreuk aan de algemene geldigheid van de slotsom, waartoe Prof. BAUR in zijn résumé gekomen was en die op p. 117 geciteerd werd.

Accepteren we de hypothese, dat de geografische ligging geen of weinig verschillen teweeg brengt, dan moeten we het steekproefeffect nader onderzoeken.

Dit doen we door van alle maanden het gemiddelde verloop van de barometerstand te De Bilt tijdens de zonnevlekkencyclus te bepalen. Van de zomermaanden bleek Juli het meest in overeenstemming te zijn met de door Prof. BAUR getrokken algemene conclusies. Deze maand leent zich er dus het best toe om een significant verschil tussen de barometerstanden in de afzonderlijke jaren van de zonnecyclus te ontdekken.

De in 2.10.2 reeds uitgevoerde tweevoudige variantieanalyse gaf tot resultaat: voor de rijen $8F56 = 1,78$ en voor de kolommen $7F56 = 1,62$. F heeft de volgende drempelwaarden voor $n_2 = 56$, $P = 20\%$ (resp. 10%) en $n_1 = 7$ (resp. 8).

P	$n_1 = 7$	$n_1 = 8$
20%	1,45	1,44
10%	1,82	1,78

De steekproef van de gemiddelde barometerstanden in Juli (Utrecht-De Bilt, 1849-1949) levert dus waarden voor F , die ruimschoots aan de „toevallige kant” liggen van de drempelwaarde 5% . Deze steekproef geeft daarmee geen enkele aanleiding, de elf jaren van de zonnevlekkencyclus een significant verschillende Juli-luchtdruk toe te kennen. En daarbij telt onze steekproef nog wel 101 jaren, terwijl die van BAUR slechts 70 jaren omvat. BAUR had dus nog minder recht, dergelijke verreikende conclusies te trekken.

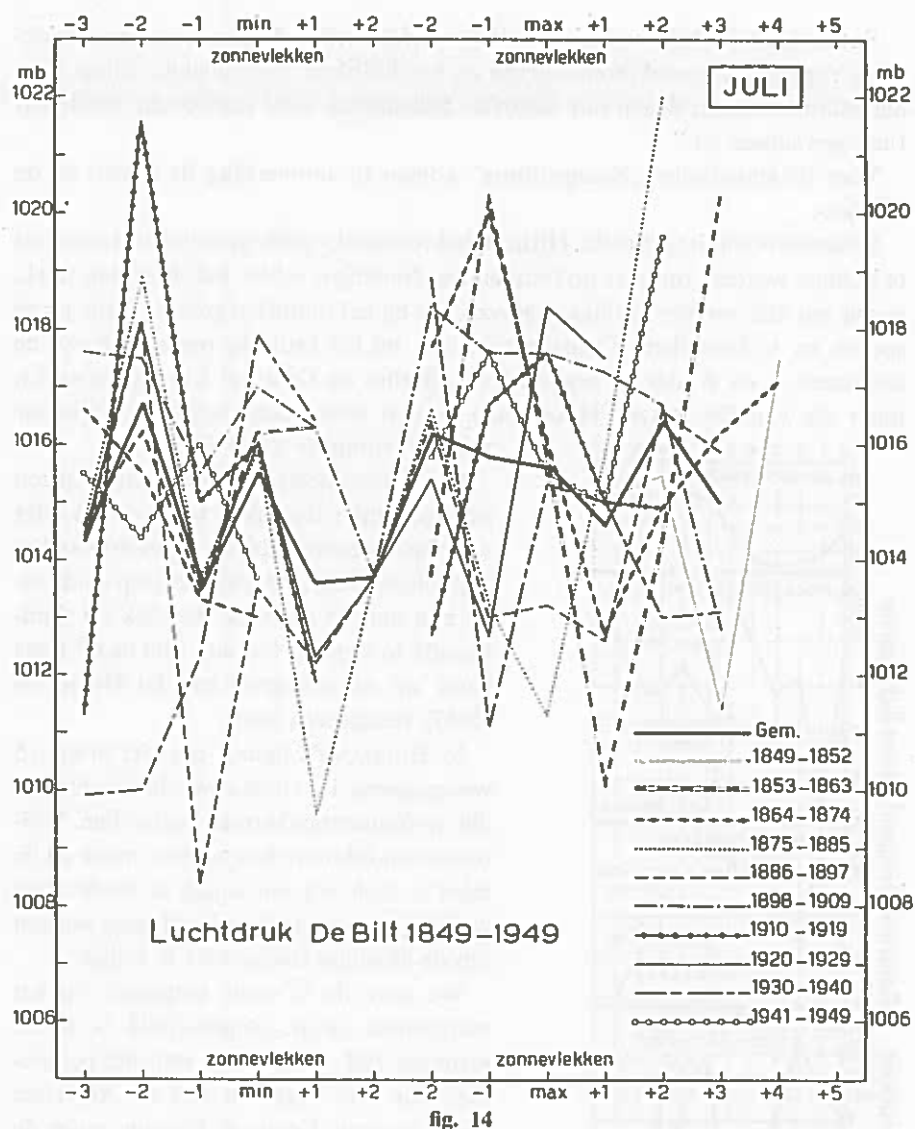
Voor hen, die bezwaren tegen of moeilijkheden met de toepassing van de variantieanalyse hebben, kan een grafische voorstelling verhelderend werken. Voor hen heeft deze „Statistica pauperum” dezelfde functie als de „Biblia pauperum” (een bijbel zonder woorden) voor de analfabeten uit de Middeleeuwen.

Fig. 14 geeft de luchtdrukwaarden in Juli (Utrecht-De Bilt, 1849-1949), ingepast in het zonnevlekkenschema van Prof. BAUR (tabel 17 op p. 91 bevat de getalswaarden). Ook de tekening laat zien, dat het hoogst voorbarig is te spreken van een invloed van de zon op de Juli-luchtdruk.

Daar de tijdsduur tussen zonnevlekkenminimum en -maximum niet altijd precies 5 jaar is, maar ook wel eens 3 of 4 jaar komen er in de kolommen van 1 en 2 jaar na het minimum minder maandgemiddelden voor dan in andere kolommen.

De wespentaille (kolom: min +2) is een zeer bijzondere merkwaardigheid. De kwestie, of het constant zijn van de Juli-luchtdruk in het tweede jaar na het zonnevlekkenminimum op realiteit of toeval berust, zal in 3.2 bij het selectie-effect besproken worden.

Een grafische voorstelling kan dus zowel schadelijk als nuttig zijn. Prof. BAUR las uit het gedrag van de krommen A en B (fig. 13) de invloed van de zonneerupties, kende dus aan elke afwijking in het gemiddelde verloop een reële oorzaak toe. Onze grafische voorstelling in fig. 14 gaf daarentegen een verward gedrag



van de luchtdruk in elke cyclus ten opzichte van het gemiddelde verloop. Deze schijnbare tegenstelling is een fraaie illustratie van Sir Normand LOCKYER's citaat: „La méthode des moyennes, c'est la méthode de ne jamais connaître le vrai”.

Het is daarom wenselijk, de krommen, die de gemiddelde gang van een of ander element voorstellen, niet alleen te publiceren, maar òf vergezeld te doen gaan van het verloop in elke cyclus, òf een statistische toets toe te passen.

FLOHN (1948) schrijft: „Es ist zweifellos gefährlich, sich nur auf einen subjektiven Vergleich von mehreren Kurven zu beschränken; gar zu leicht erliegt man Selbsttäuschungen, denen nur schärfste Selbstkritik oder statistische Nachprüfung gewachsen ist.”

Voor de statistische „Nachprüfung” komen in aanmerking de z -toets en de G^2 -toets.

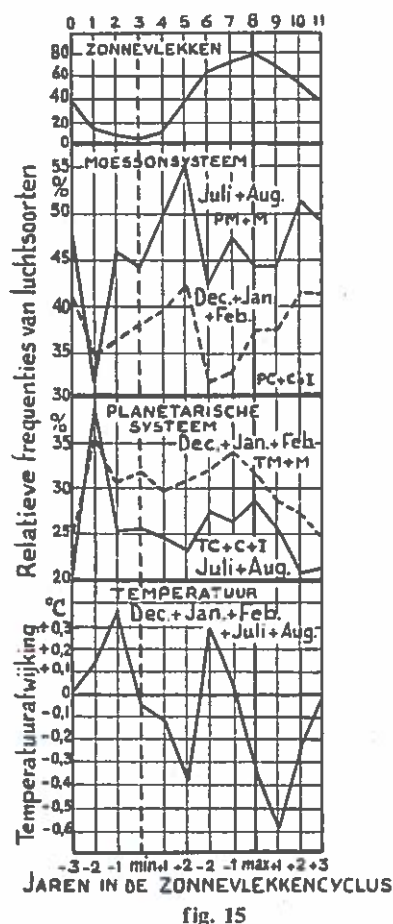
De eerste is zojuist gebruikt. Hij heeft het voordeel, op alle grootheden toegepast te kunnen worden, dus ook op frequenties. Nadelig is echter het meerdere werk, nodig om alle oorspronkelijke gegevens, die bij het middelen gebruikt zijn, op te sporen en te bewerken. Daarom zijn ook bij het kritische onderzoek van de krommen A en B niet de gegevens van Berlijn en Centraal Europa verwerkt, maar die van De Bilt en Hoofddorp, omdat deze gemakkelijker bereikbaar

waren (krommen A' en B').

De G^2 -toets heeft het nadeel, dat hij alleen op frequenties toegepast kan worden. Het voordeel is echter, dat de gegevens van het gepubliceerde gemiddelde verloop voldoende zijn om het resultaat dadelijk op significantie te kunnen toetsen. Met de G^2 -toets gaan we nu een geval na, dat HOFMANN (1947) beschreven heeft.

In HOFMANN's figuur, die hier in fig. 15 weergegeven is, vinden we vier krommen, die op frequenties betrekking hebben. Weliswaar op relatieve frequenties, maar uit de tekst is toch vrij eenvoudig te vinden met welke factor vermenigvuldigd moet worden om de absolute frequenties te krijgen.

We gaan de G^2 -toets toepassen op het markantste geval. Ongetwijfeld is dit de kromme $PM + M$. Deze stelt het percentage van alle dagen in Juli en Augustus voor, waarop Centraal Europa onder de invloed stond van maritieme (M) of polair-maritieme (PM) luchtsoorten in het tijdvak 1881–1941. HOFMANN's kromme is dus het gemiddelde van hoogstens 5 à 6 individuele krommen; in 61 jaren komen immers maximaal 5 of 6 zonnevlekkencycli voor.



Elk punt van de kromme – beter: elk hoekpunt van de polygoontrek – is dus ook het gemiddelde van de dagen met PM of M in – laten we zeggen – 5,5 jaar. De maanden Juli en Augustus tellen 62 dagen. Om de relatieve frequenties van de luchtsoorten, uitgedrukt in %, om te zetten in absolute frequenties, zou derhalve vermenigvuldigd moeten worden met de factor $A = \frac{5,5 \times 62}{100}$. Met recht: „zou”,

want de luchtsoorten op de 62 dagen zijn niet onafhankelijk van elkaar. Er is een persistentieeffect aanwezig. Bij de bespreking van dit effect in 1.12.3 hebben we ervaren, dat we het getal 62 moeten delen door een factor $\omega(62)$, die ongeveer overeenkomt met de gemiddelde duur van een $PM + M$ -periode. Op deze wijze wordt het aantal afhankelijke dagen vervangen door het effectieve aantal onafhankelijke dagen. De grootte van $\omega(62)$ kennen we niet exact, maar het is al voldoende, wanneer we weten tussen welke uiterste grenzen $\omega(62)$ in ligt. Het is nu zo goed als zeker, dat $2 < \omega(62) < 6$.

Hieruit volgt: $\frac{5,5 \times 62}{100 \times 6} < A < \frac{5,5 \times 62}{100 \times 2}$
 of: $0,57 < A < 1,70$

Het is niet nodig om de relatieve frequenties dadelijk in absolute frequenties om te zetten. Het vermenigvuldigen met A behoeft pas in de laatste stap te geschieden.

Hoe verleidelijk de kromme er ook uitziet en hoe gaarne we ook willen spreken over een invloed van de zonnevlekken op de aardse atmosfeer, de vraag: „toeval of realiteit?” mag niet zonder onderzoek met „realiteit” beantwoord worden.

We stellen daarom de nulhypothese, dat er van een zonnevlekkeninvloed helemaal geen sprake is. Is deze hypothese juist, dan zou de theoretische ($PM + M$)-kromme een horizontale lijn moeten zijn. De afwijkingen tussen theoretische en waargenomen waarden worden in de G^2 -toets verwerkt (1.8).

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	
0	48	3		9	Kolom <i>a</i> geeft het jaar in de zonnevlekkencyclus.
1	32		13	169	Het jaar 3 bv. komt overeen met het minimum in het schema
2	46	1		1	van Prof. BAUR. Kolom <i>b</i> stelt het uit de figuur 15 geschatte per-
3	44		1	1	centage voor.
4	50	5		25	<i>c</i> en <i>d</i> zijn de positieve en negatieve afwijkingen ten opzichte van
5	55	10		100	het voorlopige gemiddelde 45. Kolom <i>e</i> bevat de kwadraten van
6	42		3	9	deze afwijkingen.
7	48	3		9	
8	44		1	1	
9	44		1	1	
10	52	7		49	
	29	19	374		

$$\begin{aligned}\bar{M} &= 45 + \frac{29-19}{11} = 46 \\ \bar{M} \bar{G}^2 &= 374 - \frac{(29-19)^2}{11} = 365 \\ \bar{G}^2 &= \frac{365}{46} = 7,9\end{aligned}$$

$\bar{M}$ en $\bar{G}^2$ hebben betrekking op de relatieve frequenties.

Om de gevraagde G^2 te verkrijgen behoeven we $\bar{G}^2$ alleen maar met de factor A te vermenigvuldigen. Het resultaat is:

$$\begin{aligned}0,57 \times 7,9 &< G^2 < 1,70 \times 7,9 \\ \text{of} \quad 4,5 &< G^2 < 13,4\end{aligned}$$

Het aantal vrijheidsgraden is $11-1 = 10$.

Een (χ^2 , G^2 , P)-diagram of tabel leert ons, dat de overschrijdingskans tussen 0,92 en 0,20 ligt. Elke aanleiding om de nulhypothese te verwerpen ontbreekt. De op het oog zo „reëel” schijnende afwijkingen van de horizontale rechte $\bar{M} = 46\%$ mogen aan het toeval toegeschreven worden.

Tot slot van deze paragraaf nog enige voorbeelden, waarin het steekproef-effect een voorname rol gespeeld kan hebben.

Vroeger hebben wij een citaat van FLOHN (1948) opgenomen, waarin hij maant tot strenge zelfcritiek of statistische contrôle. Des te onbegrijpelijker¹⁾ is het, wanneer hij daarop laat volgen:

„Berechnet man z.B. die Häufigkeit aller Westlagen nach dem Homburger Kalender für die Zeitabschnitte 1881–1900, 1901–20 und 1921–40 und bezeichnet sie mit den Indizes 1, 2, 3, so ergeben sich zwischen den Teilreihen folgende Korrelationskoeffizienten:

	Winter	Frühling	Sommer	Herbst	Jahr
r_{12}	0,24	0,44	0,44	-0,05	0,29
r_{13}	0,18	0,59	0,44	-0,24	0,36
r_{23}	-0,40	0,19	0,13	-0,27	0,25

Trotzdem der Korrelationskoeffizient auf geringe zeitliche Verschiebungen bereits überaus scharf anspricht und daher als Gütemass wenig geeignet ist, ergeben sich doch im Frühling und im Sommer recht hohe positive Werte; die

¹⁾ Deze tegenstelling is te vergelijken met de tegenstelling tussen de opvolgende werken van Descartes (1637): „Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences” en „Le traité des Météores”. De Biographie Universelle (1814) zegt over Le traité des Météores: „Descartes, y donnant carrière à son imagination, entreprend d’expliquer tous les phénomènes météorologiques, même la formation de la foudre. Celui qui avait tant recommandé le doute, s’imagine qu’il suffit d’alléguer vaguement un mode possible d’un phénomène pour en avoir assigné la cause véritable, sans penser qu’en se hasardant ainsi à deviner par intuition, pour ainsi dire, les principes des faits, il y a l’infini à parier contre un qu’on tombera dans l’erreur.”

zeitliche Persistenz und die Klimaverwerfung um 1900 lässt sich damit objektiv zahlenmässig belegen ($r_{23} > r_{12} > r_{13}$). Derartige Kontrollrechnungen sind auf jeden Fall empfehlenswert".

Wanneer men mag aannemen, dat FLOHN de frequentie van de „westsituaties" („Westlagen") in de jaren n met hun frequentie in de jaren $n + 20$ gecorreleerd heeft ($n = 1881-1900$ voor r_{12} en $n = 1901-1920$ voor r_{23}) en op dezelfde manier r_{13} tussen de jaren n en $n + 40$ ($n = 1881-1900$), dan zijn alle correlatiecoëfficiënten berekend uit 20 paren waarden. Dat betekent, dat de correlatiecoëfficiënten $20-2 = 18$ vrijheidsgraden hebben.

Bij een universumwaarde $\rho = 0$ zal 1 % van de steekproeven ter grootte van 20 elementen een correlatiecoëfficiënt $> 0,56$ en 0,1 % van de steekproeven een correlatiecoëfficiënt $> 0,68$ hebben. Men kan dus twijfel uiten aan de realiteit van alle door FLOHN berekende correlatiecoëfficiënten. En pas goed aan de realiteit van de onderlinge verschillen tussen r_{23} , r_{12} en r_{13} . Het is een raadsel, hoe de tijdelijke persistentie (welke eigenlijk?) en de klimaatsprong omstreeks 1900 zich daarmee objectief numeriek laten bewijzen. De laatste, boven geciteerde zin, moet ontkend worden: „dergelijke contrôle-berekeningen zijn allerminst aan te bevelen".

Het zelfbedrog eist ook in de meteorologie zijn slachtoffers. De gewilligste daaronder passen allen de volgende methode toe.

— Men berekene het gemiddelde van alle bij éénzelfde X behorende waarden Y . Daarna zette men al die gemiddelde Y -waarden in een grafiek uit tegen de bijbehorende X -waarden.

Tenslotte verklare men zelfvoldaan: De grafische voorstelling vertoont duidelijk een nauw verband tussen de grootheden X en Y . —

Op deze manier kan men de invloed van een willekeurige X op een willekeurige Y bewijzen en met dit bewijs een bewijs leveren van het gezegde: „Met statistiek kan men alles bewijzen".

Dat inderdaad een bewijs van een of ander verband volgens deze methode mogelijk is, blijkt uit het volgende:

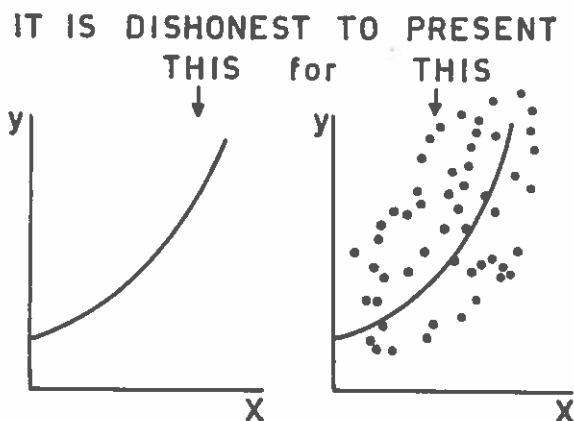
Slechts in de zeldzaamste gevallen behoort bij alle X -waarden dezelfde waarde van Y (de streep boven Y betekent, dat het gemiddelde over enkele waarden van Y bepaald is). De toevalsfluctuaties zorgen ervoor, dat in het algemeen $\bar{Y}_i \neq \bar{Y}_j$ (voor $i \neq j$). En zou het eens gebeuren, dat $\bar{Y}_i = \bar{Y}_j$ voor alle i en j , dan verdwijnt deze gelijkschakeling dadelijk, wanneer voor iedere Y_i nog een waarneming erbij genomen wordt. Want bijna zeker geldt $Y_i \neq \bar{Y}_i$. De nieuwe $\bar{Y}_i$'s zijn dan niet

meer aan elkaar gelijk. Door de toevalsspreiding of het steekproefeffect is $\frac{\partial \bar{Y}}{\partial X} \neq 0$ — ook als tussen X en Y helemaal geen verband aanwezig is.

Daarom mag uit ieder geval, waarvan de grafische voorstelling geen horizontale rechte, maar een gebroken lijn bevat, niet zonder meer geconcludeerd worden, dat er een verband tussen de ordinaat- en abscisgrootheden bestaat.

Het minste wat men in zulke gevallen ter statistische controle kan doen, of beter, moet doen, is het tekenen van alle afzonderlijke curven in een figuur (zoals bv. gebeurd is in fig. 14 op p. 121 van deze verhandeling). Wanneer een soortgelijke wirwar optreedt als in onze fig. 14, dan is twijfel aan de echtheid van het bewijs gerechtvaardigd.

In onderstaande tekening, ontleend aan MORONEY (1951, 29), ligt hetzelfde advies opgesloten als in de vorige alinea gegeven is.



Een andere controle methode is die van het schudden. Bij deze methode moet men de gehele waarnemingsreeks permuteren, zodat een eventueel verband door deze lukrake verwisseling totaal verloren gaat. Werkt men met de nieuwe, gepermuteerde reeks op dezelfde wijze verder als met de oorspronkelijke reeks gedaan werd, dan krijgt men uitkomsten, die gegarandeerd zinloos zijn. Zijn deze echter van gelijke grootte als in het oorspronkelijke eindresultaat, dan is het bestaan van een werkelijke zonnevlekkeninvloed zeer twijfelachtig. Het beste schudt men niet éénmaal, maar tien of meerdere malen om het toeval in zijn verschillende uitingsvormen op het spoor te komen (BARTELS 1948a, 216). Voor reeksen met 100 of minder waarnemingen kan men daartoe tabel 6 van BARTELS (1948a) gebruiken. Andere permutaties, die van die van BARTELS onafhankelijk zijn, kunnen zelfs bij tijdreeksen van 167 elementen gebruikt worden (VAN DER BUIJ 1951b).

De derde contrôle geeft de variantieanalyse. Met de in 1.10.1 aangegeven vorm gaat het slechts met niet-gladgestreken waarden. De vaak toegepaste vereffening $X'_n = 1/4 (X_{n-1} + 2 X_n + X_{n+1})$ veroorzaakt een kunstmatige autocorrelatie en verhindert de normale berekening. (Ook dat is weer een steun voor de stelling, dat men dergelijke vereffeningsprocédés zo weinig mogelijk moet gebruiken).

Het is – helaas – gemakkelijk, deze paragraaf uit te breiden met vele andere concrete voorbeelden. We doen dit niet, omdat het steekproefeffect zijn grote macht nu wel voldoende getoond zal hebben.

Veel van het voorafgaande zou overbodig geweest zijn, wanneer de in deze paragraaf bekritiseerde onderzoekers de wijze woorden ter harte genomen hadden, die BUYS BALLOT reeds in 1882 neerschreef:

— Evenzoo, en dit schijnt men te vergeten, kan het bij zoo groote ongelijkheden en storingen als waaraan de temperatuur onderworpen is, niet anders of de waarnemingen moeten in een zeker aantal kolommen geschikt, welk ook, voor de eene kolom een grooter som geven dan voor de andere. Men vleit zich dan te spoedig dat het aantal dier kolommen werkelijk eene periode bepaalt. HORNSTEIN, STEWART, MELDRUM, BLANFORT, vooral vele anderen die minder zorgvuldig het gevondene toetsen, die de reeks van waarnemingen niet lang genoeg nemen, ze niet in groepen afdeelen, om te zien of de verschillende groepen onderling ook in époque overeenstemmen, slaan daarom telkens periodische afwisselingen van temperatuur, barometerhoogten en regenhoeveelheden voor. Sommigen zoeken zelfs verband met de zonnevlekken, voor allerlei verschijnselen, hetwelk nog niet door de waarnemingen gewettigd wordt. —

3.2 HET SELECTIEEFFECT

If we look long enough, we will find something very peculiar and nonrandom about any given sequence, and can prove that the probability of this peculiarity arising by chance is very small. The difficulty is that randomness is not a property of a sequence, but of the process that produces them. LEVENE 1952

Tot degenen, die „zelfs een verband met de zonnevlekken gezocht” hebben, behoort ook SCHNEIDER. Hij onderzocht de invloed van de zonnevlekken op de bloeitijden van de paardenkastanje.

In zijn artikel schrijft SCHNEIDER (1950, 108):

„Da die Sonnenfleckenperioden und mit ihnen die Märztemperatur die Phasen wechseln, hat es keinen Sinn, ein Periodogramm der Märztemperatur für den gesammten Zeitraum aufzustellen. Es wurde deshalb nur für Frankfurt ein Periodogramm berechnet, da für die Länge der Frankfurter Reihe (ab 1835) die

Phase der Sonnenfleckenperiode annähernd konstant bleibt. Die Irrfahrtzeiträume fanden dabei natürlich keine Berücksichtigung".

De schrijver berekent een kans van 0,014% voor de mogelijkheid, dat de door hem gevonden periode van 11 jaren door het toeval veroorzaakt zou zijn.

Als men een hypothese opstelt (i.c. de realiteit van een periode van 11 jaren in de temperatuur van Maart te Frankfurt) en men wil deze hypothese bewijzen, terwijl men voor de hypothese ongunstige gevallen schrapt, – i.c. die van de dronkemansperioden ¹⁾ – dan is het natuurlijk vanzelfsprekend, dat de gunstige gevallen overblijven en er een zeer significant resultaat bereikt wordt. De betekenis van dit resultaat is echter hoogst twijfelachtig.

Slechts wanneer men tevoren weet, dat in bepaalde perioden de hypothese om theoretische of experimentele redenen onjuist is, mag men deze perioden buiten beschouwing laten. Maar wanneer men de ongunstige perioden eerst *a posteriori* herkent, kan het selectieeffect een rol spelen.

Een dikwijls voorkomende methode om dubieuze geofysische samenhangen langs statistische weg te bewijzen is de volgende:

Het tijdvak, waarin men de onderhavige samenhang aantreft, verlengt men zover, tot men op een geval stuit, waarin niet aan de samenhang voldaan is. Past men nu statistische toetsen toe, dan vindt men vaak een significante relatie.

De realiteit van dit verband tracht men aannemelijk te maken door een nulhypothese op te stellen, die de gevonden samenhang juist ontkent. Op grond van deze nulhypothese berekent men de kans voor het optreden van de waargenomen verschijnselen, opgevat als een willekeurige steekproef uit een universum, waarin het bestudeerde verband dus afwezig gedacht is. Dit betekent: indien er werkelijk geen verband aanwezig is en men zou over zeer vele steekproeven kunnen beschikken, dan zou slechts $p\%$ van de steekproeven zo'n – zuiver toevallige – overeenstemming tonen als de onderzochte steekproef. Meestal heeft p een zeer lage waarde. Nu moeten we één van beide: óf aannemen, dat de eerste de beste steekproef, die men beschouwde, een van de weinig voorkomende gevallen (in $p\%$) is, óf de nulhypothese verwerpen.

Meestal is p zo klein, dat de eerste mogelijkheid geschrapt kan worden ten gunste van de tweede. Het gevolg is, dat men moet ontkennen, dat het verband toevallig is. Op deze wijze heeft men langs statistische weg bewezen, dat er een causaal verband bestaat.

Het schijnt allemaal in orde, maar men heeft over het hoofd gezien, dat de steekproef niet de eerste de beste is, maar één, waarvan de grenzen door de te

¹⁾ = Irrfahrtperioden = random walk periods.

onderzoeken eigenschappen bepaald zijn. Kort gezegd: het selectieeffect heeft de significantie te voorschijn geroepen of tenminste aanzienlijk verhoogd.

Door te selecteren zondigt men tegen de hoofdwet van de statistica. De drie wijzen, waarop BOK (1946) deze eis weergegeven heeft, zijn reeds in de noot op p. 10 vermeld.

Dergelijke selectieve bepalingen van de steekproefgrenzen kunnen we vermoeden in enkele tabellen van Prof. BAUR, bv. in die op p. 963-6 en 970 van het Lehrbuch der Meteorologie (HANN-SÜRING, vijfde uitgave). Fig. 16 bevat deze grenzen.

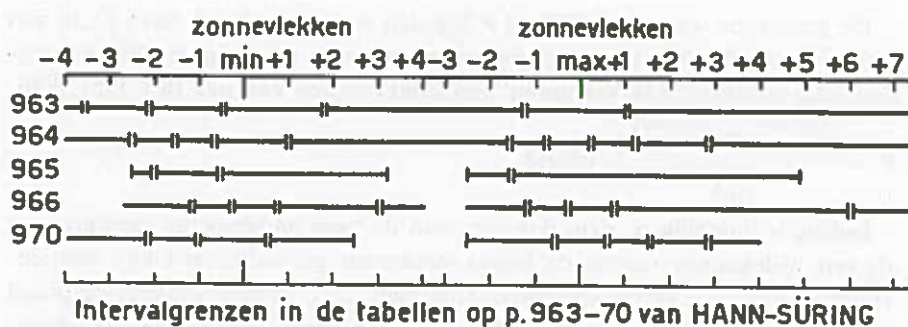


fig. 16

De enige regelmatigheid, die we in het vastleggen van deze grenzen kunnen ontdekken, vinden we in de tabel van p. 966. Deze heeft betrekking op de verdeling van de zeer warme maanden in Centraal Europa over de zonnevlekken-cyclus (Januari 1761-Februari 1947).

De grenzen zijn daar bij zonnevlekkenminimum en -maximum dezelfde, zodat tenminste de helft van de grenzen a priori vaststond. Maar overigens wekt fig. 16 sterk de indruk, dat alle grenzen zo gekozen zijn, dat naast elkaar gelegen intervallen grote en kleine aantallen van de verdeelde elementen (warme maanden, strenge winters enz.) bevatten. Daarmee verhoogt men echter de significantie op een ongeoorloofde manier. Neemt men grenzen, die onafhankelijk van het voorkomen van strenge winters of koude maanden zijn, dan zal misschien geen enkel resultaat significant zijn. Er blijft niets beters over dan af te wachten, tot het waarnemingsmateriaal zich ver- n -voudigd heeft. Voor n geldt: $n \gg 1$. BARTELS (1930, 42) heeft in een van zijn vaak moeilijk toegankelijke artikelen berekend: $n > 14$. Hij laat daarop volgen: „... ist die Periodensuche z.B. für Mitteleuropa bisher ziemlich aussichtslos”.

Dat Prof. BAUR inderdaad nu en dan met „optimale” grenzen werkt, blijkt uit zijn publicatie in de Bad Kissinger Wetterkarten van 1 en 2 October 1949. In

het interval 1,1 jaar na het zonnevlekkenmaximum tot 3,5 jaar voor het minimum kwamen van 1761 tot 1949 veertig winters voor, waaronder slechts één strenge. Het daaraan voorafgaande interval (0,5 jaar vòòr tot 1,0 jaar na het maximum) telt 31 winters, waaronder 9 strenge.

Om uit te maken, of het verschil tussen $9/31$ ($= 29\%$) en $1/40$ ($= 2,5\%$) significant is, berekenen we de kans, dat beide percentages afkomstig zijn uit een universum met een percentage = het gewogen gemiddelde van de beide steekproefpercentages. Dit gewogen gemiddelde is $\frac{31 \times 9/31 + 40 \times 1/40}{31 + 40} = \frac{10}{71}$.

De gevraagde waarschijnlijkheid P bepalen we aan de hand van § 12.16 van KENDALL (1947, 303). De daar beschreven methode wordt in de Duitse literatuur „das FISHERSche H-Verfahren” genoemd (zie ook VAN DER BIJL 1951, 200).

$$P = \frac{\binom{31}{9}\binom{40}{1} + \binom{31}{10}\binom{40}{0}}{\binom{71}{10}} = 0,0018$$

Helaas is duidelijk te zien, dat geen van de twee onderzochte steekproeven als een willekeurige (eerste de beste) steekproef gekwalificeerd kan worden. Voegt men aan de eerste steekproef (die met $2,5\%$ strenge winters) de direct aangrenzende linker- en rechterkolom toe, dan wordt het percentage verzesvoudigd (VAN DER BIJL 1949).

Slechts wanneer tevoren bekend zou zijn, dat de winters van het interval 1,1 jaar na het maximum tot 3,5 jaar voor het minimum zich anders zouden gedragen dan de winters van het interval 0,5 jaar voor tot 1,0 jaar na het maximum, heeft men het recht, de boven geschetste methode te volgen. Dat mag niet in gevallen als deze, waarin BAUR zijn hypothese opstelt op grond van zekere waarnemingen en vervolgens het „bewijs” levert op grond van dezelfde waarnemingen.

Om het selectieeffect te elimineren, moet de berekende kans met een correctiefactor vermenigvuldigd worden. Deze zgn. selectiefactor is gelijk aan het aantal van alle mogelijke groepen van ongeveer 40 winters. Dit getal is minstens 32, want zoveel kolommen zijn er (iedere kolom kan beginkolom van een groep zijn). Daarmee wordt de kans op het voorkomen van tegengestelde intervallen minstens $32 \times 0,0018 =$ minstens 0,05. De significantie is nu verdwenen.

Eigenlijk zou men de kans van alle mogelijke combinaties van alle mogelijke groepen moeten berekenen. Wanneer deze kansen niet op toevallige wijze in het interval 0 tot 1 verdeeld liggen, is het een of andere verband tussen zonnevlekken en wintertemperaturen aanwezig. Zijn de kansen daarentegen lukraak verdeeld tussen 0 en 1, dan mag men de realiteit van het verband ontkennen. Maar het gaat in zulke gevallen niet aan, de laagste kansen ($< 0,0027$ of zelfs $< 0,05$) uit te kiezen en deze als significant te bestempelen.

Een voorbeeld, waarin de methode van de selectiefactor goede resultaten levert, demonstreren wij aan de hand van gegevens van BARTELS (1948a, 212). Zijn tabel 1 van toevalsgetallen bevat rijen met elk 60 cijfers. BARTELS schrijft: „Ziffer 3 kommt in Spalte g 5 überhaupt nicht vor”. De kans, dat 60 op elkaar volgende cijfers geen 3 bevatten, is $(0,9)^{60} = 0,0018$. Volgens de grootte moet dit een significant resultaat zijn, maar wij mogen niet vergeten, dat het hier gaat om „kennis achteraf”. Wij wisten niet van te voren, dat juist rij g 5 geen drieën zou bevatten.

De selectiefactor berekenen we op de volgende manier: rij g 5 is uitgezocht uit 60 rijen, het cijfer 3 uit de tien cijfers 0–9. De selectiefactor is derhalve $60 \times 10 = 600$. De voor het selectieeffect gecorrigeerde kans wordt dus 1,08 (>1 , dit heeft echter niets te betekenen), in dit verband een zeer plausibel resultaat.

Deze paradox laat zich gemakkelijk oplossen, maar er zijn moeilijker gevallen. Sommige van hen zijn elders uitvoerig beschreven (VAN DER BIJL 1951).

Na deze excursie op meer exact gebied keren we terug naar de moerassige streken, die in het uitgestrekte gebied van de meteorologie helaas nog voorkomen.

We waren gebleven bij de eerste motivering van de voorspelling voor de winter van 1949–50. De tweede motivering vond Prof. BAUR in tellurische voorbereidingen (ΔP Sept. III > 2 mb en ΔP Sept. > 0 mb te Moskou, ΔP Sept. III > 0 mb te Berlijn $\rightarrow \Delta T$ winter $> 0,6^\circ\text{C}$ in Centraal Europa). Daar deze kennis eveneens geen voorkennis is, speelt het selectieeffect hier ook een niet onbelangrijke rol. Helaas is de grootte van de selectiefactor niet aan te geven – zelfs niet bij benadering.

Een werkelijk verrassende indruk van de macht van het selectieeffect geeft „Hoffmann's Erzählung” (HOFMANN 1951, 160; VAN DER BIJL 1951, 201).

Als laatste voorbeeld van dit hoofdstuk wordt nog eens teruggegrepen op een voorbeeld uit het vorige hoofdstuk (fig. 14 op p. 121).

De ideale wespentaille stemt tot nadenken: in ieder tweede jaar na een zonnevlekkenminimum had Juli een gemiddelde luchtdruk van 1013,7 mb. Weliswaar heeft dat in slechts vier jaar plaats gevonden, maar in de 97 andere jaren van de periode 1849–1949 vindt men slechts nog drie Julimaanden met dezelfde gemiddelde luchtdruk (tab. 17 op p. 91). De kans, dat $P_{\text{Juli}} = 1013,7$ mb, is dus $\frac{7}{101}$.

De kans, dat alle vier getallen in de rij „min + 2” = 1013,7 zijn, is $(\frac{7}{101})^4 = 0,000023$. Natuurlijk is het selectieeffect ook hier werkzaam geweest, maar de

selectiefactor is slechts 13 (er zijn 13 rijen, waaruit de markantste uitgekozen is).

De gecorrigeerde kans is 0,0003, dus nog ver beneden de extreme betrouwbaarheidsdrempel 0,0027. Weliswaar kan men nog eens met 12 vermenigvuldigen (geen van de overige maanden toonde een wespentaille), maar dan is de kans nog niet groter dan 0,0036. Bovendien is de laatste weg gevaarlijk. Want op deze manier kan men ieder significant resultaat insignificant maken. Haalt men er nl. vele (aantal = s) gevallen bij, waarin het bewuste resultaat niet optreedt, dan wordt tenslotte de voor het selectieeffect gecorrigeerde kans ($= s$ maal de oorspronkelijke kans) $> 0,0027$.

Het is uiterst onwaarschijnlijk, dat het volgende jaar in de kolom „Min + 2” weer een luchtdruk in Juli van 1013,7 mb zal opleveren. Dit houdt in, dat zelfs aan het voorkomen van zo’n kleine overschrijdingskans $= 0,0003$ geen betekenis gehecht behoeft te worden, ook wanneer het een voor het selectieeffect gecorrigeerde kans betreft.

Deze overwegingen geven aanleiding, de volgende regel in het leven roepen: Betrouwbaarheidsdrempel voor *a priori* hypothesen: 0,05 (gebruikelijke grens). Betrouwbaarheidsdrempel voor *a posteriori* hypothesen (met geëlimineerd selectieeffect): 0,0001 (zie het laatste voorbeeld, waar een overschrijdingskans $= 0,0003$ zonder betekenis was).

Eventueel kan men niemandslanden scheppen ($10^{-1} - 10^{-2}$ resp. $10^{-4} - 10^{-5}$). Valt de gevonden overschrijdingskans aan de hoge kant van het niemandsland (kans $> 10^{-1}$ resp. 10^{-4}), dan heeft de statisticus er geen bezwaar tegen, dat men de hypothese als werkhypothese gebruikt. Valt de kans aan de lage kant (kans $< 10^{-2}$ resp. 10^{-5}), dan mag men de hypothese als verworpen beschouwen en is bij een „tertium non datur” de andere hypothese bewezen.

Op p. 112 hebben we aangetoond, dat het tijd en ruimte besparen zal, wanneer we het begrip „steekproefeffect” zoveel mogelijk gebruiken. De toepassing van het begrip „selectieeffect” zal niet minder besparen. Wij gehoorzamen daarmee ook aan „het gebod van de tijd” (zie p. 146).

Men leze bv. de uitvoerige gedeelten, die KENDALL (II 23.27 en 26.36–26.42), BARTELS (1925, 452 en 1949, 45), WALKER (1925, 338), WALDEN (1950, diens factor c), HEMELRIJK (1951, 14), NASS (1950a, 97, diens criteriumfraude), WÜNSCHMANN (1949, 8), DANIELS (1946), HOTELLING (1937) en nog vele anderen aan deze moeilijkheden besteed hebben. Ter vergelijking stelle men daartegenover de weinige woorden, die HOFMANN (1951a, 40 en 1952, 48), SCHWERTFEGER (1952, 18) en WEGER (1952, 230) — door gebruik te maken van de sinds 1950 gepropageerde term — nodig hebben om zich begrijpelijk uit te drukken.

3.3 HET PERSISTENTIEEFFECT

We have to assume that Nature is trying to ruin the statistician.
TINTNER 1952

Bijna alle statistische wetten zijn gebaseerd op de onderlinge onafhankelijkheid van de te onderzoeken gebeurtenissen. – De meeste geofysische verschijnselen, waarvan de metingen meer dan ongeveer zeven dagen uiteenliggen, kan men beschouwen als verschijnselen, die aan de strenge eisen van de theoretische statistica voldoen. Bij een tijdsafstand van de waarnemingen kleiner dan zeven dagen gaat het meestal niet.

Men kan, zoals PANOFSKY (1949) gedaan heeft, toch de statistische methoden zuiver formeel op meteorologische grootheden toepassen; men krijgt dan echter – zoals hij – zinloze resultaten.

Het gaat echter niet aan, met PANOFSKY te beweren: „As yet, it is impossible to decide the degree of significance of a correlation coefficient obtained in a given sample”. Want Prof. BARTELS heeft reeds 25 jaar geleden nieuwe begrippen geformuleerd en methoden aangegeven, waarmee men deze hindernis enigszins kan ontlopen.

De foutenvoortplantingswet berust op het volgende: De gegeven waarnemingsreeks $x_1, x_2, \dots, x_n$ met standaarddeviatie σ wordt in groepen van elk g waarden verdeeld. Iedere groep heeft een gemiddelde $y_i = \frac{1}{g} \sum_{k=1-g}^0 x_{i+g+k}$.

Deze gemiddelden hebben een standaarddeviatie $\sigma(g)$. Zijn de waarnemingen onafhankelijk van elkaar, dan geldt de foutenvoortplantingswet $\sigma(g) = \sigma/\sqrt{g}$.

Volgens BARTELS (1947, 7 Fiat Review 90) heeft onbedachtzame toepassing van deze wet op geofysische (en andere) voorbeelden, waarbij de veronderstelde onafhankelijkheid der afzonderlijke waarden niet bestaat, reeds tot veel verkeerde conclusies aanleiding gegeven. Want dat leidt er toe, de gemiddelden als nauwkeuriger bepaald te beschouwen dan ze in werkelijkheid zijn.

De geldigheid van de foutenvoortplantingswet kunnen wij demonstreren aan de hand van de December-temperaturen te Zwanenburg-De Bilt (1741–1940). De standaarddeviatie, berekend uit deze 200 maandtemperaturen, is $2,44^\circ\text{C}$. Bepalen wij het gemiddelde van elk der 20 decennia in deze periode, dan krijgen we:

1741–50	2,92°C	1791–00	1,46°C	1841–50	2,64°C	1891–00	2,95°C
51–60	2,86	1801–10	2,23	51–60	2,42	1901–10	2,41
61–70	1,87	11–20	1,18	61–70	2,80	11–20	3,94
71–80	3,21	21–30	3,21	71–80	2,15	21–30	2,47
81–90	0,87	31–40	3,30	81–90	1,71	31–40	2,14

Gemiddelde over 1741–1940: $2,43^\circ\text{C}$.

De 20 decenniumgemiddelden hebben een standaarddeviatie $= 0,72^{\circ}\text{C}$. Volgens de foutenvoortplantingswet zou deze moeten zijn $= s(10) = s: \sqrt{10} = 2,44: \sqrt{10} = 0,77^{\circ}\text{C}$. De overeenstemming kunnen we bevredigend noemen.

Op soortgelijke wijze kunnen we de ongeldigheid van de foutenvoortplantingswet bewijzen aan de hand van afhankelijke gebeurtenissen, bv. van alle etmaalgemiddelden van de temperatuur te De Bilt in December 1950.

Deze 31 temperaturen (hoogste $7,8^{\circ}\text{C}$ op 1.12, laagste $-9,9^{\circ}\text{C}$ op 30.12) hebben een gemiddelde $= -1,2^{\circ}\text{C}$ en een standaarddeviatie $s = 3,7^{\circ}\text{C}$. Verwaarlozen we de jaarlijkse gang en passen we de foutenvoortplantingswet formeel toe, dan is het resultaat: $s(31) = s: \sqrt{31} = 3,7: \sqrt{31} = 0,66^{\circ}\text{C}$.

De December-temperaturen zullen dus om de normaal van December schommelen met een standaarddeviatie van $0,66^{\circ}\text{C}$. Zijn de December-temperaturen volgens GAUSS verdeeld, dan mag men beweren, dat het bijna zeker is, dat de normaal in December in het interval $-1,2^{\circ}\text{C} \pm 3 \times 0,66$ of tussen $-3,2$ en $+0,8^{\circ}\text{C}$ ligt. In werkelijkheid is het gemiddelde over 1741–1940 gelijk aan $2,4^{\circ}\text{C}$.

Er klopt iets niet, wat ook blijkt uit de grootte van de t van STUDENT. De afwijking van de normaal is $-1,2 - 2,4 = -3,6^{\circ}\text{C}$. De berekende standaarddeviatie is $0,66^{\circ}\text{C}$. $t = \frac{3,6}{0,66} = 5,4$. De overschrijdingskans P bij 30 vrijheidsgraden is ongeveer 10^{-7} .

Zou de toegepaste methode juist zijn, dan moeten we op grond van de zeer kleine overschrijdingskans (10^{-7}) ontkennen, dat de December-normaal $2,4^{\circ}\text{C}$ is. Gelukkig weten we, dat de gemaakte fout ingesloten is in de veronderstelling, dat de 31 etmaalgemiddelden onafhankelijk van elkaar waren.

Slaan we een t -tabel op, dan zien we, dat een t -waarde $= 5,4$ voorkomt bij $P = 0,01$ en 3 tot 4 vrijheidsgraden. Aangezien de Decembermaand van 1950 geen bijzonder extremum was, schijnt een $P = 0,01$ zeer aannemelijk. Daaruit blijkt, dat de 31 temperaturen hoogstens 3 tot 4 vrijheidsgraden bezitten, m.a.w. dat deze 31 afhankelijke temperatuurgegevens even betrouwbare uitkomsten leveren als 3 of 4 onafhankelijke temperaturen.

Prof. BARTELS heeft deze gedachtengang gepreciseerd door de invoering van het begrip „effectief aantal $\gamma(g)$ onafhankelijke waarden in groepen van g elementen” (zie 1.12.2). $\gamma(g)$ kan berekend worden uit $\gamma(g) = \sigma^2/\sigma^2(g)$. Door de invoering van $\gamma(g)$ is het mogelijk met de gegeneraliseerde foutenvoortplantingswet te werken:

$$\sigma(g) = \sigma/\sqrt{\gamma(g)}$$

In ons eerste voorbeeld (de 20 decenniumgemiddelden uit de periode 1741–1940) krijgen we voor $\gamma(10)$ de schatting $c(10) = \frac{2,44^2}{0,72^2} = 11,5$.

In het algemeen is $\gamma(g) \leq g$; slechts reeksen, die op een bijzondere wijze geconstrueerd zijn, hebben $\gamma(g) > g$ (bv. de c -reeks van 1.12.3). Hier is $c(10)$ ook > 10 , maar het geringe verschil mogen we wel aan het steekproefeffect toeschrijven.

In ons tweede voorbeeld is $c(31) = \frac{3,7^2}{2,44^2} = 2,3$. De raming van hoogstens 3 tot 4 klopt dus zeer goed. De waarde 2,3 is niet zeer betrouwbaar, omdat $s = 3,7$ berekend is uit de gegevens van slechts één Decembermaand.

In 1.12.3 hebben wij nog een tweede door BARTELS ingevoerd begrip gedefinieerd: het aequivalente herhalingsgetal („die äquivalente Wiederholungszahl“) $\omega(g) = g/\gamma(g)$.

Ter verduidelijking geven we nu een voorbeeld van BARTELS. Uit een rij toevalsgetallen met standaarddeviatie σ vormen we een nieuwe reeks door ieder element driemaal te herhalen. Er ontstaat: $x_1, x_1, x_1, x_1, x_2, x_2, x_2, x_2, x_3, x_3, x_3, x_3, \dots$

In deze nieuwe reeks zijn de gemiddelden over elke $g = 400$ elementen gelijk aan de gemiddelden over elke 100 elementen van de oorspronkelijke reeks. Voor de eerste reeks met onderling onafhankelijke elementen geldt de foutenvoortplantingswet streng. Voor de nieuwe reeks is dus $\sigma(400) = \sigma/\sqrt{100}$, d.w.z. $\gamma(400) = 100$ en $\omega(400) = 4$. Het aequivalente herhalingsgetal is gelijk aan het aantal gelijke elementen.

In toevalsreeksen is $\gamma(g) = g$ en $\omega(g) = 1$. In reeksen met persistentie is $\gamma(g) < g$ en $\omega(g) > 1$.

Zoals in 1.12.3 uiteengezet is, geldt voor een reeks, bestaande uit voortschrijdende gemiddelden over k termen ($b_i = \frac{1}{k} \sum_{t=0}^{k-1} a_{i+t}$):

$$\omega(g) = k - \frac{k^2 - 1}{g} \quad (\text{voor } g > k - 2)$$

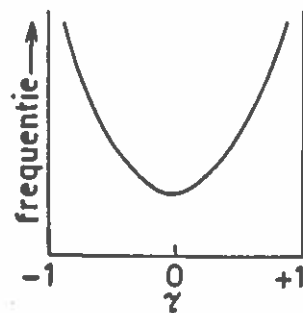
Voorbeeld: Voor $g = 60$ en $k = 30$ is $\omega(60) = 30 - \frac{899}{180} = 25$.

$$\gamma(60) = \frac{60}{\omega(60)} = \frac{60}{25} = 2,4$$

Deze getallen komen o.a. voor bij correlatiecoëfficiënten, die LYSGAARD in zijn standaardwerk „Recent Climatic Fluctuations“ (1949, 31-2) berekend heeft.

De vele hoge door LYSGAARD gevonden correlatiecoëfficiënten zijn voor dergelijke klimatologische werken zeer ongewoon. Wanneer men aanneemt, dat elk van zijn correlatiecoëfficiënten op 31 onafhankelijke paren waarnemingen berust, dan zijn alle correlatiecoëfficiënten significant. Neemt men echter de waarde van $\gamma(60) = 2,4$ als juist aan, dan blijven maar weinig significante correlaties over.

De frequentiefunctie van de steekproefwaarden r voor $\rho = 0$ (= universumwaarde van de correlatiecoëfficiënt) is $f(r, 0) = \text{const.} \times (1-r^2)^{\frac{n-4}{2}}$ (zie ook



1.11.1). De grafische voorstelling van deze functie is dus voor $n < 4$ een U-kromme. Daarmee kunnen we verklaren, waarom LYSGAARD zo vele hoge correlatiecoëfficiënten gevonden heeft.

Hier treedt weer een bezwaar tegen vereffening naar voren. Door de vereffening wordt het aantal onafhankelijke gebeurtenissen verkleind en de toevallige overeenstemming tussen de te onderzoeken krommen vergroot.

In de meeste gevallen is $g \gg k$ en dan kan men $\frac{k^2-1}{3g}$ tegenover k verwaarlozen. In zulke gevallen geldt: $\omega(g) = k$ en $\gamma(g) = g/k$.

Wanneer men dus de voortschrijdende gemiddelden over 30 jaren bepaalt van de langste geofysische reeksen (die ongeveer 250 jaren tellen), dan hebben de nieuwe reeksen een effectief aantal onafhankelijke waarden van niet meer dan 8. Door deze methoden van vereffenen bereikt men, dat waardevolle geofysische tijdreeksen teruggebracht worden tot steekproeven van geringe omvang.

De in 1.12 behandelde theorie van BARTELS geeft bij de reeksen van voortschrijdende gemiddelden bevredigende resultaten. Niet op elk gebied echter schenkt BARTELS' theorie voldoening, want bij de overmaatreeksen van BUYS BALLOT stuiten we nog op moeilijkheden.

Het uitgangspunt zij een x -reeks: $E(x_i) = 0$, $E(x_i^2) = 1$ en $E(x_i x_j) = 0$ met $i \neq j$.

Uit deze leiden we een e -reeks (of overmaatreeks) af, waarbij: $e_i = \sum_{j=1}^i x_j$.

Wat is nu $\gamma(g)$?

$$\sigma^2 = E(e_i^2) = i \quad 1 \leq i \leq g$$

$$\sigma^2(g) = E\left\{\left(\frac{1}{g} \sum_{i=1}^g e_i\right)^2\right\} = \frac{1}{g^2} \cdot \frac{1}{6} g(g+1)(2g+1)$$

$$\gamma(g) = \frac{\sigma^2}{\sigma^2(g)} = \frac{i \cdot 6g}{(g+1)(2g+1)} = \frac{3i}{g} \text{ (voor niet te kleine } g)$$

$$\text{Dus } \frac{3}{g} < \gamma(g) < 3.$$

Dat zou betekenen, dat correlatiecoëfficiënten tussen overmaatreeksen meestal hoge waarden moeten aannemen (frequentiefunctie een U-kromme!).

In deze richting zou dan de verklaring te zoeken zijn van de befaamdste hoge correlatie in de meteorologie. We bedoelen het verband tussen de relatieve zonnevlekgetallen en de hoogte van het wateroppervlak van het Victoriameer ($r = 0,87$ voor het tijdvak 1901–22; BROOKS 1923). Met $\gamma(g) = \gamma(22) = 22$ is deze steekproefwaarde zeer significant. De gegevens van de laatste 30 jaar zijn, zoals BAUR (1948, 111) en STEINHAUSER (1948 en 1949) aangetoond hebben, niet in overeenstemming met de voor de hand liggende hypothese van een directe afhankelijkheid tussen zonnevlekken en waterstand. Toch moet de hoge waarde $r = 0,87$ statistisch verklaard kunnen worden.

De zwakste schakel in bovenstaand betoog is de veronderstelling $\gamma(g) = 22$. Wegens de periodiciteit in de zonnevlekken zijn de relatieve getallen zeker niet onafhankelijk. Het effectieve aantal onafhankelijke waarden in de jaarlijkse zonnevlekgetallen is < 22 . De jaarlijkse gemiddelden van de waterstanden zijn ook niet onafhankelijk, deze reeks kan als een overmaatreeks beschouwd worden.

Immers: $H_n - H_{n-1} = f(N, T, V, A)$, waarbij H_n = hoogte van het wateroppervlak in het jaar n ;

$$\left. \begin{array}{l} N = \text{neerslag} \\ T = \text{toevloeiing} \\ V = \text{verdamping} \\ A = \text{afvloeiing} \end{array} \right\} \text{ in het jaar } n.$$

Omdat de afhankelijkheid tussen de jaarlijkse bedragen van N verwaarloosd mag worden (hetzelfde geldt voor T , V en A), is $E(H_{n+1} - H_n)(H_n - H_{n-1}) = 0$. Dit houdt in, dat H_n een overmaatreeks is.

Volgens een voorstel van HAFSTAD (1940) is het aantal paren onafhankelijke waarden n_r , waaruit de correlatiecoëfficiënt berekend gedacht kan worden, gelijk aan het meetkundig gemiddelde van de effectieve aantallen onafhankelijke waarden van zonnevlekgetallen en waterstanden. Dit gemiddelde is natuurlijk < 22 , maar of het een waarde $= 5$ zal bereiken, is twijfelachtig. Bij 5 onafhankelijke paren waarden nl. heeft $r = 0,88$ bij $p = 0$ een overschrijdingskans $= 0,05$; bij 7 onafhankelijke paren waarden heeft $r = 0,87$ bij $p = 0$ een overschrijdingskans $= 0,01$. Dit toont, hoe klein het meetkundig gemiddelde moet zijn om een aannemelijke verklaring (steekproefeffect) te krijgen voor de hoge steekproefwaarde $r = 0,87$, wanneer men uitgaat van volledige onafhankelijkheid ($p = 0$).

Hierboven (p. 136) hebben we getracht, $\gamma(g)$ theoretisch te bepalen. Het is ons slechts gedeeltelijk gelukt. Daarom hebben we ook de experimentele weg gevolgd. Eenvoudigheidshalve namen we $x_i = +1$ of -1 . Wij construeerden 54 e -reeksen, die elk 25 elementen telden. Daaruit berekenden we 79 correlatiecoëfficiënten. Hiervan hadden 23 een absolute waarde $> 0,50$. —De 54 oorspronkelijke x -reeksen leverden 27 correlatiecoëfficiënten, waarvan de grootste een

absolute waarde = 0,446 had. Dit is in zeer goede overeenstemming met $P(r = 0,445, n = 25; \rho = 0) = 0,02$ —.

De frequentietabel van de 79 correlatiecoëfficiënten van de overmaatreeksen toonde overduidelijk aan, dat hier van een U-kromme geen sprake kan zijn.

Drie van de 79 correlatiecoëfficiënten van de overmaatreeksen waren $> 0,87$ of $< -0,87$ (nl. $-0,97$, $-0,88$ en $+0,94$).

De verklaring van C. E. P. BROOKS voor het verband tussen de zonnevlekken en de waterstand gold voor niet steekhoudend in de ogen van de meeste geofysici. Zij behoeven, tot hun vreugde, nu slechts historische betekenis aan dit geval toe te kennen, daar de waarnemingen van de laatste dertig jaren de samenhang niet bevestigd hebben. De statisticus echter werd daarmee voor de moeilijkheid gesteld, de paradox op te lossen en te verklaren, waarom de eerste twee decennia van de twintigste eeuw zulk een hoge correlatiecoëfficiënt voortgebracht hadden.

Door bovenstaande beschouwingen is de extreem lage, onjuiste overschrijdingskans door middel van benaderingen en schattingen voorlopig herleid tot een overschrijdingskans van $3/79$, een zeer aannemelijke waarde.

Eenvoudiger voorbeelden van de verwaarlozing van het persistentieeffect zijn er ook. We noemen slechts een vroeger werk van FLOHN (1941). BARTELS schrijft daarover (1948, 125–6):

„FLOHN gelooft de realiteit van hoge ‚Oberschwingungen‘ van de jaarlijkse gang in de luchtdruk en daarmee de golfstructuur van de singulariteiten met behulp van het expectancy-criterium van SCHUSTER op de volgende wijze te kunnen bewijzen.”

„De standaarddeviatie van het daggemiddelde van de luchtdruk te Frankfurt am Main in de 83 onderzochte jaren ($N = 30295$ dagen) is $s = 8,4$ mbar. Zich baserend op de veronderstelling, dat de dagelijkse waarden op toevallige wijze verdeeld zijn, berekent hij volgens de formule van SCHUSTER de verwachtingswaarde E van de amplitude van een sinusoïde, die men door harmonische analyse van deze N waarden zou krijgen. Het blijkt, dat $E = 2s/\sqrt{n} = 0,096$ mbar. Inderdaad wordt voor de periode 30,4 dagen ($= 1/12$ jaar) een amplitude van de grootte $A = 0,3$ mb, dus meer dan $3E$ verkregen. De werkhypothese (nl. de toevallige volgorde der daggemiddelden), die aan de formule voor E ten grondslag ligt, is dus blijkbaar onjuist.”

Prof. BARTELS vervolgt: „Tegen dit betoog is niets in te brengen, maar wel tegen de interpretatie van het resultaat. FLOHN gelooft nl. daarmee het bestaan van persistente perioden van $1/12$ jaar bewezen te hebben. Het beschouwde model van toevallig verdeelde dagelijkse waarden faalt echter wegens een andere,

triviale reden: de daggemiddelden volgen niet toevallig op elkaar, de luchtdruk verandert continu van de ene op de andere dag." Of, om in onze terminologie van 1.12.3 te spreken: de etmaalgemiddelden van de luchtdruk vormen niet een *a*-reeks, maar meer een *b*- of *d*-reeks.

3.4 HET PSEUDOCORRELATIEEFFECT

Men vindt soms tussen een tweetal grootheden geen correlatie van betekenis, wel echter tussen grootheden, die uit de oorspronkelijke afgeleid zijn. Men werkt dan verder met de afgeleide grootheden, waarbij het gevaar, onjuiste conclusies te trekken, niet denkbeeldig is.

PEARSON was de eerste, die de biometrici op dit gevaar attent maakte. In 1897 publiceerde hij een artikel onder de titel: „On a form of spurious correlation which may arise when indices are used in the measurement of organs". Uit deze titel blijkt reeds, dat PEARSON doelde op de valse correlatie, die ontstaat bij het correleren van de grootheden $\frac{a}{c}$ en $\frac{b}{c}$, waarbij *a*, *b* en *c* eventueel onderling onafhankelijk kunnen zijn.

In de geofysica wordt weinig met aan indices analoge grootheden gewerkt, nog minder worden zij met elkaar gecorreleerd. Des te meer wordt in de meteorologie gebruik gemaakt van verschillen of sommen van atmosferische grootheden.

Kent elke bioloog het artikel van PEARSON (1897) en zijn fouten door het pseudocorrelatieeffect op biometrisch gebied zeldzaam, in de geofysica wees voor 1950 alleen een artikel van WALKER (1926) op het verraderlijke effect. Dit bleek onvoldoende. Verderop zullen we enkele gevallen behandelen, waarin de pseudocorrelatie verwarring stichtte.

PEARSON ging uit van vier veranderlijke grootheden (series metingen) a_i , b_i , c_i en d_i . De correlatiecoëfficiënt tussen $\frac{a_i}{c_i}$ en $\frac{b_i}{d_i}$ noemde hij $\rho\left(\frac{a}{c}, \frac{b}{d}\right)$ en de correlatiecoëfficiënt tussen a_i en b_i : ρ_{ab} ; enz. Standaarddeviaties en gemiddelden werden met σ en m aangegeven.

Na weglating van de termen van de derde en hogere orden vond PEARSON

$$\rho\left(\frac{a}{c}, \frac{b}{d}\right) = \frac{\rho_{ab}\sigma_a\sigma_b m_c m_d - \rho_{ad}\sigma_a\sigma_d m_b m_c - \rho_{bc}\sigma_b\sigma_c m_a m_d + \rho_{cd}\sigma_c\sigma_d m_a m_b}{\sqrt{(\sigma_a^2 m_c^2 + \sigma_c^2 m_a^2 - 2\rho_{ac}\sigma_a\sigma_c m_a m_c)} \sqrt{(\sigma_b^2 m_d^2 + \sigma_d^2 m_b^2 - 2\rho_{bd}\sigma_b\sigma_d m_b m_d)}}$$

Het meest verrassende resultaat krijgen we, wanneer we $c_i = d_i$ stellen (dus $\rho_{cd} = 1$, $\sigma_c = \sigma_d$, $m_c = m_d$) en $\rho_{ab} = \rho_{ac} = \rho_{bc} = 0$, $\sigma_a = \sigma_b = \sigma_c$ en $m_a = m_b = m_c$.

Dan wordt $\rho\left(\frac{a}{c}, \frac{b}{c}\right) = 1/2$.

WICKSELL (1921) leidde een exacte formule af voor $\rho\left(\frac{a}{c}, \frac{b}{c}\right)$ onder de voorwaarde $\rho_{ab} = \rho_{ac} = \rho_{bc} = 0$.

$$\rho\left(\frac{a}{c}, \frac{b}{c}\right) = \frac{m_a m_b \frac{\sigma_1^2}{c}}{\sqrt{\left(\sigma_a^2 \frac{\sigma_1^2}{c} + \sigma_1^2 m_a^2 + \sigma_a^2 m_1^2\right)} \sqrt{\left(\sigma_b^2 \frac{\sigma_1^2}{c} + \sigma_1^2 m_b^2 + \sigma_b^2 m_1^2\right)}}$$

Vervangen we in deze formule overal $\frac{1}{c}$ door c , dan hebben we de correlatiecoëfficiënt $\rho(ac, bc)$ uitgedrukt in $\sigma_a, \sigma_b, \sigma_c, m_a, m_b$, en m_c . Met $\sigma_a = \sigma_b = \sigma_c$ en $m_a = m_b$ is

$$\rho(ac, bc) = \frac{m_a^2}{\sigma_a^2 + m_a^2 + m_c^2}$$

Dat de correlatiecoëfficiënten in al deze formules afhankelijk zijn van de gemiddelden van de oorspronkelijke grootheden, mag opmerkelijk heten, omdat men gewend is aan de onafhankelijkheid bij rechtstreekse correlaties, zoals die tussen a_i en b_i . De coëfficiënt $\rho(ac, bc)$ is maximaal, wanneer $m_c = 0$ en minimaal ($= 0$), wanneer $m_a m_b = 0$.

Na deze correlaties tussen quotiënten en producten gaan we over op correlaties tussen verschillen en sommen.

Het is zeer gemakkelijk af te leiden, dat er tussen a en $(a-b)$ een hoge correlatie moet bestaan, ook indien – of beter: juist indien – er tussen a en b geen correlatie aanwezig is ($\rho_{ab} = 0$).

$$\begin{aligned} \rho(a, a-b) &= \frac{E\{a_i(a_i-b_i)\}}{E\{a_i^2\}^{1/2} E\{(a_i-b_i)^2\}^{1/2}} = \frac{E(a_i^2) - E(a_i b_i)}{\sigma_a \{\sigma_a^2 + \sigma_b^2 - 2E(a_i b_i)\}^{1/2}} \\ &= \frac{\sigma_a^2 - \rho_{ab} \sigma_a \sigma_b}{\sigma_a (\sigma_a^2 + \sigma_b^2 - 2\rho_{ab} \sigma_a \sigma_b)^{1/2}} = (\text{stel: } \sigma_a = \sigma_b) = \frac{1 - \rho_{ab}}{(2 - 2\rho_{ab})^{1/2}} = \sqrt{\frac{1 - \rho_{ab}}{2}} \end{aligned}$$

$$\text{Evenzo geldt: } \rho(a, a+b) = \sqrt{\frac{1 + \rho_{ab}}{2}}$$

Voor $\rho_{ab} = 0$ is $\rho(a, a-b) = \sqrt{1/2} = 0,71 = \rho(a, a+b)$.

Voor $\rho_{ab} > 0$ is $\rho(a, a-b) < 0,71$ en $\rho(a, a+b) > 0,71$.

Voor $\rho_{ab} = -0,28$ is $\rho(a, a-b)$ zelfs $= 0,80$ en $\rho(a, a+b) = 0,60$.

Deze relaties hebben vele moeilijkheden veroorzaakt. Een van de eerste slachtoffers werd een Amerikaan, die in het Journal of the American Statistical Association de volgende paradox beschreef: —In een reeks toevalsgetallen kan men uit de grootte van het n^{de} getal niets over de grootte van het volgende getal opmaken. Inderdaad geldt in toevalsreeksen $\rho(x_n, x_{n+1}) = 0$. Correleer ik echter x_n met het verschil $x_n - x_{n+1}$, dan krijg ik de zeer hoge correlatiecoëfficiënt 0,71.

x_n is bekend en door de hoge correlatie kan ik voor $x_n - x_{n+1}$ een nauwkeurige schatting maken. Tenslotte kan ik x_{n+1} met een zeer grote nauwkeurigheid bepalen uit de geschatte waarde voor $x_n - x_{n+1}$ en de bekende waarde voor x_n .

Gelukkig kwam in een van de volgende nummers van het JASA de oplossing van de paradox: — Het is verkeerd te beweren, dat men over x_{n+1} niets weet. Men kent zijn verwachtingswaarde (= het gemiddelde van alle getallen in het universum) en zijn standaarddeviatie σ , want verzwegen was de veronderstelling, dat alle toevalsgetallen normaal verdeeld waren. Voorspellen we dus: x_{n+1} = gemiddelde waarde, dan is de fout in 68 % van de gevallen $< \sigma$. Het is ook onjuist te beweren, dat we $x_n - x_{n+1}$ zo nauwkeurig kunnen schatten. Dit verschil heeft een standaarddeviatie groter dan die van x_{n+1} , nl. $= \sigma\sqrt{2}$. Daarmee wordt het voordeel van de hoge correlatiecoëfficiënt $= 1/\sqrt{2}$ weer te niet gedaan.

Ook op een enigszins andere wijze kan de paradox opgelost worden. De beste schatting y voor het verschil $x_n - x_{n+1}$ bepalen we uit de regressievergelijking

$$y = \beta x_n + \alpha. \text{ Hiervoor geldt: } \alpha = 0 \text{ en } \beta = \frac{E\{(x_n - x_{n+1}) x_n\}}{E\{x_n^2\}} = \frac{E(x_n^2)}{E(x_n^2)} = 1.$$

Dus: $y = x_n$. De beste schatting voor x_{n+1} verkrijgen we uit het verschil $x_n - y = x_n - x_n = 0$. De standaarddeviatie van $x_{n+1} = \sigma_{x_n - y} = \sigma_y = \sigma_{x_n} = \sigma$. Hieruit blijkt, dat de kunstgreep van het invoeren van het verschil $x_n - x_{n+1}$ ons geen extra inlichtingen over de grootte van x_{n+1} heeft kunnen verschaffen. —

De meteorologische tegenhanger heeft BAUR gegeven in de Juli-afl levering van de „Grosswetterlagen Mitteleuropas” (1949a, 57). Daarin is te lezen: „De correlatiecoëfficiënt tussen de neerslaghoeveelheid van Juli in Duitsland en de verandering van de regensommen van Juli op Augustus bedraagt in de zomers, die — zoals 1949 — 2 of 3 jaren na een zonnevlekkenmaximum kwamen, $-0,72$ (periode 1801–1948)”.

Aan deze hoge correlatie hebben we helaas niets, want na substitutie van

$$r(\text{Juli, Aug.} - \text{Juli}) = -0,72 \text{ in de formule } r(a, b-a) = -\sqrt{\frac{1-\rho_{ab}}{2}} \text{ krijgen we}$$

$r(\text{Juli, Aug.}) = -0,03$. Bij de jacht op hoge correlaties blijkt de geschoten bok een onbetekenende directe correlatiecoëfficiënt te zijn. $\rho(\text{Juli, Aug.})$ zal ongeveer $+0,20$ zijn. Het verschil met de waarde $-0,03$ voor $r(\text{Juli, Aug.})$ kan door het steekproefeffect verklaard worden; op hoogstens $\frac{2}{11} (1949-1801) = \text{ca } 30$ paren waarden steunt de correlatiecoëfficiënt. (De tabel op p. 961 van HANN-SÜRING, vijfde uitgave, openbaart, dat slechts 17 jaren aan BAUR's berekening ten grondslag lagen).

Een eventuele zonnevlekkeninvloed kan hier niet aangetoond worden, omdat het verschil tussen $-0,03$ en $+0,20$ niet significant is.

Een ietwat afwijkend geval vinden we in HANN-SÜRING (p. 951, vijfde uitgave). Daar wordt besproken de correlatie tussen de temperatuur van November tot Januari te Kristiansund (K_n) verminderd met de temperatuur in dezelfde maanden van het jaar daarvoor (K_{n-1}) enerzijds en het overeenkomstige verschil van de temperatuur in Februari en Maart te Berlijn ($B_n - B_{n-1}$) anderzijds.

We kunnen de volgende berekening opstellen:

$$\rho(K_n - K_{n-1}, B_n - B_{n-1}) = \frac{E\{(K_n - K_{n-1})(B_n - B_{n-1})\}}{\{E(K_n - K_{n-1})^2\}^{1/2} \{E(B_n - B_{n-1})^2\}^{1/2}} =$$

$$= \frac{E(K_n B_n) + E(K_{n-1} B_{n-1}) - E(K_n B_{n-1}) - E(K_{n-1} B_n)}{\{E(K_n^2) + E(K_{n-1}^2) - 2E(K_n K_{n-1})\}^{1/2} \{E(B_n^2) + E(B_{n-1}^2) - 2E(B_n B_{n-1})\}^{1/2}}$$

Natuurlijk is: $E(K_n B_n) = E(K_{n-1} B_{n-1}) = \rho(K_n, B_n) \sigma_K \sigma_B$

$$E(K_n^2) = E(K_{n-1}^2) = \sigma_K^2 \text{ en } E(B_n^2) = E(B_{n-1}^2) = \sigma_B^2$$

Verwaarlozen we bovendien $E(K_n B_{n-1})$, $E(K_{n-1} B_n)$, $E(K_n K_{n-1})$ en $E(B_n B_{n-1})$, dan ontstaat:

$$\rho(K_n - K_{n-1}, B_n - B_{n-1}) = \frac{2 \rho(K_n, B_n) \sigma_K \sigma_B}{(2 \sigma_K^2)^{1/2} (2 \sigma_B^2)^{1/2}} = \rho(K_n, B_n)$$

WALKER (1926!! 80) bleek achteraf precies dezelfde afleiding gegeven te hebben. Besteedt de meteoroloog meer tijd aan het schrijven dan aan het lezen van zijn vakliteratuur?

Het verschil tussen de twee steekproefcorrelatiecoëfficiënten $r(K_n, B_n)$ en $r(K_n - K_{n-1}, B_n - B_{n-1})$ verleidde Prof. BAUR tot de uitspraak (eveneens op p. 951):

„In het bijzonder zij er nog op gewezen, dat de correlaties van de temperatuurverschillen met die van het jaar daarvoor aanzienlijk groter zijn dan tussen de temperaturen zelf. Dit is in overeenstemming met het ook bij andere correlaties vastgestelde feit, dat de relaties met processen (veranderingen) meestal nauwer zijn dan met toestanden, hoewel in andere gevallen meestal geen veranderingen van jaar tot jaar, maar gewezen toestanden met daaropvolgende veranderingen (vergelijk p. 953) of plaats gehad hebbende processen met later volgende toestanden gecorreleerd zijn”.

Het bewijs van WALKER maakt duidelijk, dat deze bewering in geen geval van theoretische zijde gesteund wordt. M.i. kunnen dergelijke combinaties van waarnemingen geen betere uitkomsten geven dan de waarnemingen zelf.

Desondanks vindt BAUR een correlatiecoëfficiënt tussen de veranderingen, die groter is dan de correlatiecoëfficiënt tussen de temperaturen zelf. Deze tegenstelling tussen theorie en praktijk is slechts een schijnbare, want de theorie levert ons: $\rho(K_n - K_{n-1}, B_n - B_{n-1}) = \rho(K_n, B_n)$, d.i. de gelijkheid van de correlatie-

coëfficiënten in het universum, terwijl de practijk ons helaas slechts correlaties tussen steekproeven kan schenken. In BAUR's geval is misschien de bijdrage $K_n B_{n-1}$ of $K_{n-1} B_n \neq 0$. Het steekproefeffect moet het verschil tussen theorie en practijk veroorzaakt hebben.

In de zojuist geciteerde regels verwijst Prof. BAUR naar p. 953 van HANN-SÜRING. Daar vinden we correlatiecoëfficiënten van zonale druk- en temperatuurverschillen in de maand n met a) het luchtdrukverschil Ponta Delgada-IJsland in de maand $n+1$ verminderd met het luchtdrukverschil $PD-Y$ in de maand n en met b) de luchtdruk in de maand $n+1$ op IJsland (= Stykkisholm + Berufjord) verminderd met de overeenkomstige luchtdruk in de maand n .

Nemen we bv. het eerste geval van de eerste regel. Daarin wordt het luchtdrukverschil Rome-Ponta Delgada in Januari ($R_J - PD_J$) gecorreleerd met het luchtdrukverschil Ponta Delgada-IJsland in Februari minus het overeenkomstige verschil in Januari, dus met $PD_F - Y_F - PD_J + Y_J$.

De universumwaarde van de gezochte correlatiecoëfficiënt is:

$$\rho = \frac{E\{(R_J - PD_J)(PD_F - Y_F - PD_J + Y_J)\}}{\{E(R_J - PD_J)^2\}^{1/2} \{E(PD_F - Y_F - PD_J + Y_J)^2\}^{1/2}}$$

Door een eenvoudige vermenigvuldiging kan ρ uitgedrukt worden uitsluitend in correlaties tussen „toestanden” en standaarddeviaties van „toestanden”, zodat er in de formule voor ρ niets overblijft, dat op een „Vorgang” of „Änderung” duidt. De grootste bijdrage in de teller levert $+E(PD_J^2)$, dan volgt $-E(PD_J Y_J)$ (eveneens >0), de andere termen spelen geen rol. Daaruit blijkt, dat de gezochte $\rho > 0$ zal zijn. Het vet drukken van de significante correlatiecoëfficiënten heeft weinig zin, want we weten van te voren, dat alle $\rho > 0$ zijn.

BAUR besluit dit gedeelte met de op zichzelf juiste, maar geen betekenis hebbende conclusie: „Es ist besonders zu beachten, dass . . . hier wiederum die Korrelation mit einem Vorgang, mit der Änderung des Gefälles, sich als strammer erweist als zwischen Zustand und Folgezustand” (p. 954, HANN-SÜRING).

Onder de tabel (p. 953) schrijft Prof. BAUR: „Uit de tabel is te zien, dat een groot (d.w.z. groter dan de normaal) drukverschil Rome-Ponta Delgada of Indianapolis (U.S.A.) – Ponta Delgada in de maand n gevolgd wordt door een stijging van de druk in de maand $n+1$ te Ponta Delgada en dit betekent vergroting van de luchtdrukgradiënt Ponta Delgada-IJsland”.

Dit is ook zonder enige correlatierekening onmiddellijk in te zien. Grote („bovennormale”) drukverschillen Rome (resp. Indianapolis) – Ponta Delgada worden in vele gevallen veroorzaakt door een lage („ondernormale”) druk in Ponta Delgada. In de volgende maand zal de druk te Ponta Delgada dus, ge-

middeld genomen, stijgen. Daaruit volgt weer, dat de luchtdrukgradiënt Ponta Delgada-IJsland vergroot wordt, al is het dan slechts ten opzichte van de waarde in de vorige maand. Deze gradiënt is echter, gemiddeld genomen, klein („ondernormaal”) bij hoge („bovennormale”) $P_R - P_{PD}$.

Het hele betoog komt daarop neer, dat na een extreem afwijkende druk te Ponta Delgada in de maand n in de volgende maand een meer normale druk verwacht mag worden. Uit al deze beschouwingen mag men de conclusie trekken, dat het onmogelijk is extra bijzonderheden te ontdekken, wanneer men de ingewikkelde berekeningen toepast, die BAUR gebruikt heeft.

Dergelijke „resultaten” ontstaan evenzo, wanneer men in plaats van met maandgemiddelden met dekade-, seizoen- of jaargemiddelden werkt. Daarmee wordt ook aan de volgende bewering van Prof. BAUR iedere grond ontnomen:

„De correlaties zijn in Februari, Maart, April, Juni en October het grootst, omdat klaarblijkelijk in deze maanden de tussen oorzaak en gevolg verlopende tijd, de werkingsperiode, juist zo groot is, dat de samenhangen door toepassing van maandgemiddelden achterhaald kunnen worden”.

De in deze paragraaf gekritiseerde fouten zijn des te onbegrijpelijker, omdat Prof. BAUR soortgelijke effecten reeds lang kent en ze zelfs beschreven heeft, o.a. een $a(a+b)$ -geval (1931, 346) en een $a(a+b+c+d)$ -geval (1929, 232 en 238).

Mooie voorbeelden van het pseudocorrelatieeffect geven VON RUDLOFF und TRENKLE in hun historisch voortreffelijke „Beiträge zur langfristigen Witterungsvorhersage” (1950).

Op p. 43-4 publiceerden zij de kolommen x , y en r van onderstaande tabel.

	x	y	r	categorie	p	
1.	P_{48}	$P_{49} - P_{48}$	$+ 0,48$	$a(b-a)$	$- 0,71$	Alle r zijn berekend voor het tijdperk 1887-1935.
2.	P_{29}	$P_{30} - P_{29}$	$- 0,62$	$a(b-a)$	$- 0,71$	
3.	$P_{29} - P_{28}$	P_{29}	$- 0,71$	$a(a-b)$	$+ 0,71$	
4.	P_{28}	$P_{30} - P_{28}$	$- 0,81$	$a(b-a)$	$- 0,71$	
5.	$P_{28} - P_{27}$	$P_{30} - P_{28}$	$- 0,66$	$(a-b)(c-a)$	$- 0,50$	
7.	P_{31}	$P_{33} - P_{31}$	$- 0,57$	$a(b-a)$	$- 0,71$	
8.	$P_{34} - P_{33}$	$P_{33} + P_{35} - 2P_{34}$	$- 0,71$	$(b-a)(a-2b+c)$	$- 0,87$	

Het eerste getal in de r -kolom moet natuurlijk $-0,48$ zijn, het derde $+0,71$. De x - en y -kolommen geven aan, wat met elkaar gecorreleerd is. P_n = de gemiddelde luchtdruk in de n^{de} pentade te Karlsruhe. ($n = 1$ is 1-5 Jan., enz.)

Onze laatste kolom toont de universumwaarden van de correlatiecoëfficiënten, indien alle pentadegemiddelden onafhankelijk van elkaar zouden zijn.

De kleine verschillen tussen de r - en ρ -kolommen laten zien, hoe gering de invloed van het persistentieeffect hier is.

De oorspronkelijke tabel bevat ook een f_r -kolom. Op p. 40 definiëren VON RUDLOFF en TRENKLE: $f_r = \frac{1-r^2}{\sqrt{n}}$. Met behulp van deze grootheid trachten zij de significantie van de verschillende correlatiecoëfficiënten aan te tonen. In 4.1 wordt deze ongelukkige methode besproken.

Op p. 115 werd de correlatie behandeld tussen het drukverschil December–November in Edinburgh ($P_D - P_N$) en de temperatuur van de daaropvolgende winter in Centraal Europa ($T = \frac{1}{3}T_D + \frac{1}{3}T_J + \frac{1}{3}T_F$, alle grootheden beschouwd als afwijking van de normaal). Daar werd beweerd, dat de correlatie ($= -0,36$) veroorzaakt werd door persistentie- en pseudocorrelatieeffect.

Voor de universumcorrelatiecoëfficiënt geldt:

$$\rho(P_D - P_N, T) = \frac{E\{(P_D - P_N)(\frac{1}{3}T_D + \frac{1}{3}T_J + \frac{1}{3}T_F)\}}{\{E(P_D - P_N)^2\}^{1/2}\{E(T^2)\}^{1/2}}$$

Na het uitwerken der producten is het mystieke karakter van de oorspronkelijke correlatiecoëfficiënt verdwenen: er blijven slechts standaarddeviaties over en correlaties tussen maandgemiddelden. De grootste bijdrage in de teller levert $+E(P_D T_D) = \rho(P_D, T_D) \sigma_{P_D} \sigma_{T_D}$. Deze bijdrage is negatief, want hoge (lage) druk in Edinburgh gaat veelal gepaard met een lage (hoge) temperatuur in Centraal Europa in de wintermaanden. In de teller heffen $+E(P_D T_J)$ en $-E(P_N T_D)$ elkaar vrijwel op. De overige termen hebben weinig te betekenen, omdat de tijdsafstand tussen de gecorreleerde grootheden twee of meer maanden bedraagt. In de noemer komen nog termen voor als $E(P_D P_N)$, $E(T_D T_J)$, enz. Hun grootte wordt bepaald door het persistentieeffect.

Door substitutie van voorzichtigeschattingen van de na de uitwerking verkregen termen blijkt, dat de hier besproken correlatiecoëfficiënt inderdaad ca $-0,36$ moet zijn. We mogen dus aan deze correlatie geen prognostische betekenis toekennen.

Van SEILKOPF (1948, 202) is afkomstig de methode van de negatieve „Druckänderungsbilder“. Hij beschrijft zijn methode als volgt:

„Perioden und Rhythmen des Luftdrucks kommen auch in Isallobarenkarten zum Ausdruck, wobei allerdings die 12-bis 48stündigen Druckänderungen infolge der oft grossen Amplituden die längerperiodischen unterdrücken. In mehrtägigen Änderungskarten zwischen mehrtägigen Mittelkarten der Luftdruckverteilung werden die kurzfristigen Druckänderungen ausgeglichen. Zeigen dann Tag für Tag aufeinanderfolgende n -tägige Änderungskarten über dem gleichen

Gebiete im Abstand von jeweils n Tagen abwechselnd positive und negative Luftdruckänderungsgebiete von annähernd gleicher Form, gleichliegenden Zentren und gleichen Höchstwerten in den Zentren, so ist dies ein Zeichen von einer über dem gleichen Gebiet auftretenden Druckperiode van $2n$ Tagen, die für kartenmässige Extrapolationen gebraucht werden kann".

Het is zeer waarschijnlijk, dat Prof. SEILKOPF het pseudocorrelatieeffect over het hoofd heeft gezien en wel de categorie: $(b-a)$ $(c-b)$. Immers, geven we

$$\frac{1}{n} \{ P_{k-1/2n} + P_{k-1/2n+1} + \dots + P_k + \dots + P_{k+1/2n-1} \} \text{ aan door } a, \frac{1}{n} \{ P_{k+1/2n} + \dots + P_{k+n} + \dots + P_{k+3/2n-1} \} \text{ aan door } b \text{ en } \frac{1}{n} \{ P_{k+1/2n} + \dots + P_{k+2n} + \dots + P_{k+3/2n-1} \}$$

aan door c , dan stellen $(b-a)$ en $(c-b)$ de drukken voor op twee achtereenvolgende „ n -tägige Änderungskarten im Abstand von jeweils n Tagen". De correlatiecoëfficiënt tussen de grootheden $(b-a)$ en $(c-b)$ bedraagt $-0,5$. Dit geeft de verklaring van de „abwechselnd positive und negative Luftdruckänderungsgebiete". De correlatie is echter weer niet zo stram, dat de afwisseling steeds moet optreden. Het steekproefeffect zal de tegenstelling nu eens voor de éne waarde van n markant doen zijn, dan weer voor een andere waarde van n . Geheel in overeenstemming met deze theoretische opmerking is de ervaring van Prof. SEILKOPF: „Wenn die entgegengesetzten Änderungskarten nicht mehr genügend übereinstimmen, wird durch Isallobarenkarten von $n+1$, $n+2$ oder $n-1$, $n-2$ Tagen gegebenenfalls eine Änderung der Periodenlänge erkannt." Door onze beschouwingen is opnieuw gebleken, dat verschilkaarten van deze vorm weinig extra informatie leveren.

Het voorwoord van MEYERS Handatlas bevat de leuze: „Schlagworte zu prägen, ist heute das Gebot der Zeit". Wil men voor alle mogelijke categorieën, die in deze paragraaf behandeld zijn, een gemeenschappelijke vaste term gebruiken, dan is het aan te bevelen, steeds van pseudocorrelatie te spreken. Is het nodig in bijzonderheden te treden, dan zou men bv. het volgende kunnen schrijven: „In dit of dat onderzoek heeft het pseudocorrelatieeffect een rol gespeeld. Het betrof hier een $(a-b)$ $(b+c-2a)$ -geval".

In hoofdstuk 4 zullen nog enkele gecompliceerde uitingen van de pseudocorrelatie besproken worden.

3.5 HET KROMMINGSEFFECT

Zij ons doel het bepalen van het dagelijks verloop van de luchtdruk op zonnige Decembertagen in De Bilt.

Uit het tijdvak 1912-1923 zoeken we die dagen, waarop de heliograaf minstens 5,7 uren zonneschijn registreerde. Twaalf dagen voldeden aan de eis.

De klassieke methode om de dagelijkse gang te verkrijgen is deze: vul een matrix van 25 kolommen en 12 rijen met luchtdrukwaarden in. Elke rij bevat de uurlijkse barometerstanden van een bepaalde dag. Kolom i bevat de 12 luchtdrukken, waargenomen op uur i van de 12 verschillende dagen. De 25 kolomgemiddelden $\bar{p}_i$ geven het dagelijks verloop van de luchtdruk; althans bij benadering, want de toevallige schommelingen zullen nog niet geheel weggemiddeld zijn. Storingen, onafhankelijk van de zonneinvloed, zijn er de oorzaak van, dat de gemiddelden van uur 0 en van uur 24 niet aan elkaar gelijk zijn. Dit ongewenste effect verwijderde VON LAMONT (1868) door het gemiddelde van uur i te vermeerderen met $\frac{i-12}{24} (\bar{p}_0 - \bar{p}_{24})$, de zgn. correctie van LAMONT.

De dagelijkse gang van de luchtdruk op heldere dagen te De Bilt, weergegeven in fig. 17a, is op deze wijze berekend, alleen zijn ter besparing van rekenwerk slechts die uren genomen, die een veelvoud van 3 zijn. Hoewel de matrix dus zowel een geringe breedte als hoogte bezit, suggereert fig. 17a toch enigszins, hoe de uiteindelijke vorm van de dagelijkse gang zal zijn.

Uitgebreidere onderzoeken hebben deze suggestie bevestigd. Steeds werd om middernacht een diep minimum en tussen 10 en 12 uur een steil maximum gevonden, benevens een secundair maximum omstreeks 18 uur.

Analoge berekeningen voor sombere dagen geven een volkomen ander beeld. Nu ontwikkelen zich een nachtelijk maximum en een minimum in het begin van de middag. Niets ligt meer voor de hand dan de markante tegenstelling toe te schrijven aan de invloed van de zonnestraling. Deze verklaring bevredigde de klimatologen meer dan een halve eeuw lang.

Hoe verhelderend de verklaring ook werkt, men moet steeds blijven zoeken naar dissonante resultaten; dat zijn resultaten, die in strijd zijn met de verklarende hypothese (Spartaanse methode: VAN DER BIJL 1951a). Slaagt men daarin niet, dan wordt de verklaring meer en meer zeker. Slaagt men er wel in, dan moet naar een betere hypothese gezocht worden. Het is onbegrijpelijk, dat niemand voor 1920 naar dissonante uitkomsten gezocht heeft; zelfs HANN niet, die zijn hele leven aan het probleem van de luchtdrukschommelingen gewerkt heeft.

„Onbegrijpelijk” is inderdaad het woord, want op vrij eenvoudige wijze kan de onvoldoendheid van bovengenoemde verklaring aangetoond worden. Nemen we nl. niet de luchtdrukwaarden van 0–24 uur, maar van 6–6, van 12–12 of van 18–18 uur, dan is het resultaat telkens in het midden een maximum en aan begin en eind een duidelijk minimum (fig. 17 b, c, d).

Tegenover dit verschijnsel staat de zonnetheorie machteloos. Zij kan alleen een middernachtelijk minimum en een maximum omstreeks 12 uur verklaren.

Als eerste verkreeg BARTELS zulke paradoxale uitkomsten. In 1923 kon hij de schijnbaarheid van de paradox aantonen. De oplossing is deze: zonnige dagen komen meestal voor bij hogedruksituaties. Een voorbijtrekkend hogedrukgebied tekent op de barograafstrook een convexe kromme op. Op welke wijze we nu ook een dergelijke kromme in even lange stukjes knippen, steeds zal het gemiddelde van de stukjes eveneens convex zijn, zoals gemakkelijk en geometrisch en algebraïsch is in te zien. De convexiteit is bestand tegen lineaire processen. Beschouwen we de barogrammen van de Decembermaanden 1912-1923 als een continue kromme (met 11 discontinuïteiten), dan is onze berekening van het dagelijkse luchtdrukverloop op zonnige dagen als volgt geschied:

Uit de convexe gedeelten van de gehele kromme zijn 12 barogramstukken gelicht. Elk stuk, ter lengte van 24 uur, is gekoppeld aan een zonnige dag. De 12 luchtdrukwaarden van uur i ($i = k + 0, k + 3, \dots, k + 21$) werden gemiddeld, wat de gemiddelde gang van de barometerstand op zonnige dagen leverde en wel van middernacht tot middernacht voor $k = 0$ (fig. 17a). Voor $k = 6, 12$ en 18 schuift het beginuur steeds 6 uur op. Hogedruksituaties zijn op geen enkele wijze gebonden aan bepaalde uren of gedeelten van de dag. Dit maakt plausibel, waarom het eindresultaat in sterke mate onafhankelijk is van begin- en einduur. De ware dagelijkse gang verdrinkt in de veel grotere golf, die door het passeren van het hogedrukgebied veroorzaakt wordt (vgl. fig. 17 e en e' met fig. 17 f en f').

Dit is dan de verklaring, waarvan BARTELS later uit misplaatste bescheidenheid beweerd heeft, dat hij *bijna triviaal* is. Daarentegen mogen wij de oplossing van de paradox zeker wel *geniaal* noemen.

In fig. 17 is achtereenvolgens weergegeven de gemiddelde gang van de luchtdruk te De Bilt tijdens zonnige dagen voor de etmalen 0-24 uur (a), 6-6 (b), 12-12 (c) en 18-18 uur (d); e is ontstaan door middeling van de zich daarboven bevindende gedeelten van de krommen a, b, c en d . In deze vier schuift het krommingseffect steeds 6 uur op, de ware dagelijkse gang is in alle vier natuurlijk dezelfde. Door de uitgevoerde optelling wordt het krommingseffect zo goed als geëlimineerd. Trekken we nu de dagelijkse gang e af van a , dan houden we niets anders over dan de convexe kromme f , die zijn ontstaan alleen te danken heeft aan het voorbijtrekken van hogedrukgebieden en dus niets met het dagelijkse luchtdrukverloop te maken heeft.

Naast deze methode, die niet in de geofysische literatuur vermeld staat, is er de methode van BARTELS, die de storing direct bepaalt. BARTELS adviseert de vier krommen a, b, c en d zo te verschuiven, tot de beginpunten onder elkaar gekomen zijn. De maxima zijn dan ook ongeveer bij dezelfde verticaal gelegen. Het bepalen van de gemiddelden van p_t (van a), p_{t+6} (van b), p_{t+12} (van c) en

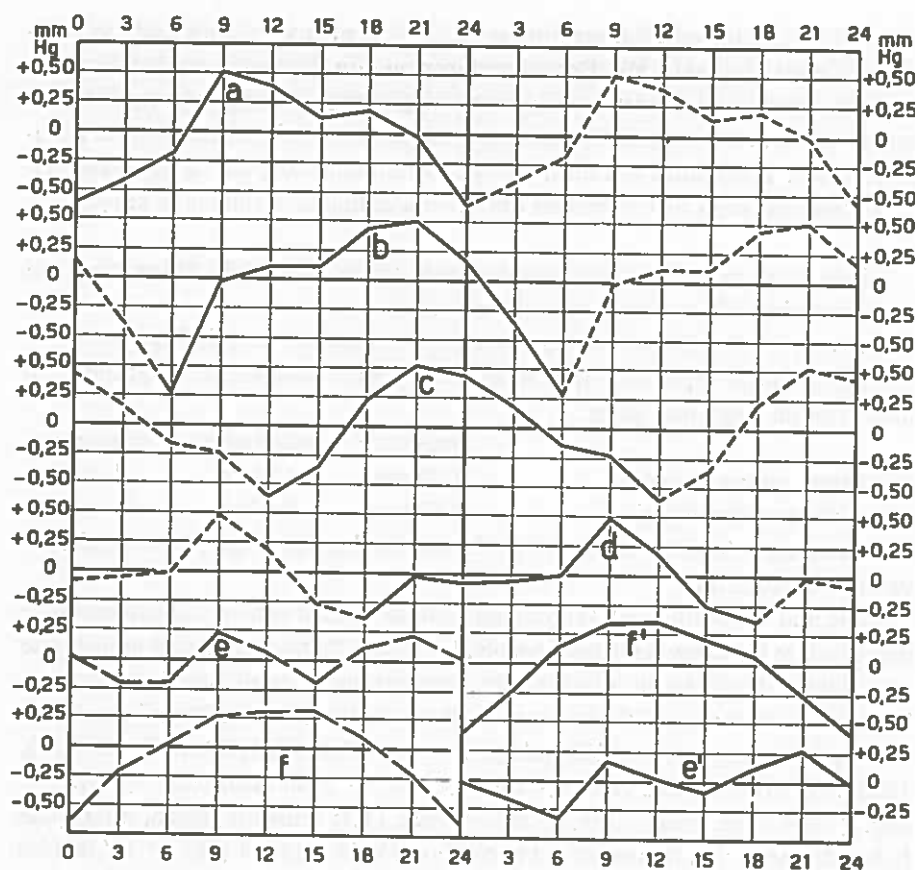


fig. 17

p_{t+18} (van d) geeft als resultaat f' . Deze kromme is met opzet geplaatst in het verlengde van de strook, die de getrokken gedeelten van a , b , c en d bevat, ter verkrijging van een visuele contrôle op het middelingsproces.

Trekken we f' van a af, dan houden we de dagelijkse gang e' over. De overeenkomst tussen f en f' is, ook in getallenwaarden, uitstekend. Bovendien zijn ze bijna identiek met de convexe curve, die BARTELS (1938, 279) vond bij 41 heldere dagen in Potsdam (December, 1893–1916, gemiddelde bewolking $\leq 0,2$ en maximale bewolking $\leq 0,8$).

Minder fraai is de overeenkomst tussen e en e' . Kennelijk is het aantal van 12 dagen te gering om de ware dagelijkse gang voldoende nauwkeurig te kunnen bepalen. Hiermee in overeenstemming is de bewering van BARTELS (1938, 277), dat waarnemingen van niet minder dan 1100 Decemberdagen nodig zijn om het

aandeel van de toevalsschommelingen tot 20% van de ware amplitude te reduceren. Omdat BARTELS alle Decemberdagen in zijn berekeningen betreft (dus ook die met sterke luchtdrukdalingen en -stijgingen), geldt zijn betoog niet voor ons geval, want zonnige dagen gaan gepaard met een rustig verloop van de barometerstand. Toch blijkt wel uit het grote aantal van 1100, dat de gegevens van 12 zonnige etmalen niet toereiken om tot een definitief resultaat te komen.

Uit de lange rij verhandelingen van onderzoekers, die slachtoffer werden van het krommingseffect, zullen we er twee kiezen.

BODMAN (1910, 32) vond op een Zweedse expeditiereis naar de Zuidpool voor het station Snow Hill (64,4°S, 57,0°W) de volgende gemiddelde amplituden en fasen van de dagelijkse gang:

	amplitude	maximum	minimum
12 heldere voorjaarsdagen	1,79 mm	11 uur	1 uur
12 sombere winterdagen	1,61 mm	24 uur	15 uur

Daarentegen had het jaargemiddelde van de dagelijkse gang een amplitude van slechts 0,22 mm!

ROBB and TANNAHILL (1935) zochten naar de invloed van de maanstand op de luchtdruk te Glasgow. Om aperiodieke, toevallige fluctuaties zoveel mogelijk te vermijden gebruikten zij alleen dagen, waarop het verschil tussen hoogste en laagste barometerstand kleiner was dan 0,1 inch Hg (3,3 mb). Hoe zij hun waarnemingsmateriaal ook in drieën splitsten — eerst 1868–1882 (1a), 1883–1897 (1b), 1898–1912 (1c); daarna Nov. + Dec. + Jan. + Feb. (2a), Mar. + Apr. + Sep. + Oct. (2b), Mei + Jun. + Jul. + Aug. (2c); tenslotte dagen, waarop de halve diameter van de maan $\leq 14',99$ (3a), $15',00 - 15',99$ (3b), $\geq 16',00$ (3c) was — steeds ontstond een gladde convexe kromme met een diep minimum in begin- en einduur en een duidelijk maximum bij de bovenste culminatie. De negen krommen gaven in totaal slechts vijf uiterst kleine secundaire maxima te zien. Onbekend met BARTELS' ontdekking schreven zij deze gemiddelde gang toe aan de invloed van de maan.

De kritiek, die CHAPMAN (1936) leverde, onthulde een tragische geschiedenis.

In 1918 had CHAPMAN het lunaire atmosferische getij te Greenwich bepaald. Hoewel ook hij slechts dagen gebruikte met een spreidingsbreedte in de luchtdruk $< 0,1$ inch Hg, ontsnapte hij aan de desastreuze gevolgen van het krommingseffect door een gelukkige transpositie. Deze kunstgreep, door de auteur zelf later (1934, 25) *fortunate, though not designed* genoemd, vergt te veel ruimte om hier beschreven te worden.

Precies dezelfde methode gebruikend voor Glasgow (1868–1912) vond ROBB in 1926 een minimale invloed van de maan. De fourieranalyse gaf voor de half-

daagse component een amplitude van 0,005 mb. CHAPMAN moedigde hem aan, deze uitkomst te publiceren, maar waarschijnlijk teleurgesteld door de kleinheid van het resultaat besloot ROBB de waarnemingen volgens de klassieke methode te gaan bewerken, dus door de waarnemingen van overeenkomstige uren te middelen. Met de hulp van TANNAHILL werden de bovengenoemde, meer markante uitkomsten bereikt. Deze schonken blijkbaar wel bevrediging, zodat overgegaan werd tot de publicatie (ROBB and TANNAHILL 1935), die tot de reactie van CHAPMAN (1936) leidde.

BARTELS (CHAPMAN and BARTELS 1940, 548) signaleert het krommingseffect ook in andere gevallen: HUNTINGTON (1923, 249) voor het tijdvak 1749–1913, CLAYTON (1923, 229) voor het tijdvak 1856–1912 en BAUER and DUVAL (1925) voor 1913–1922 constateerden onafhankelijk van elkaar een voorkeur van hoge zonnevlekgetallen voor de maanden April–September. Zou deze jaarlijkse periode reëel zijn, dan impliceert dat een invloed van de aarde op de zonnevlekken. BARTELS ontkent de realiteit door te beweren, dat de gevonden jaarlijkse gang door het krommingseffect veroorzaakt werd. Dat deze bewering niet meer dan een postulaat is, blijkt, wanneer we de jaarlijkse gang berekenen voor de 12 maanden Juli–Juni en wel voor het tijdvak 1913–1922. Nu ligt het minimum in het midden, het maximum blijft bij Juni en Juli.

Toepassing van één der verdelingsvrije methoden (slot 1.10.2), nl. die der *m* rangschikkingen, geeft een overschrijdingskans, die $\gg 5\%$ is. Dit betekent, dat we het veelvuldig voorkomen van hoge zonnevlekgetallen tijdens onze zomer een toevalsresultaat mogen noemen. In een afzonderlijke publicatie zal uitvoeriger op dit onderzoek ingegaan worden.

Een verdere aanwijzing, dat het toeval in de drie bovengenoemde artikelen een rol speelde, blijkt uit de grootte van de gevonden jaarlijkse gangen. HUNTINGTON verkreeg minimale afwijkingen; het tijdvak, dat hij gebruikte, is het langste (1749–1913). Voor BAUER and DUVAL geldt het omgekeerde en CLAYTON treft steeds de middenweg; ergo alles in strikte overeenstemming met de foutenvoortplantingswet.

Als laatste argument tegen BARTELS' opvatting – de convexiteit van de zonnevlekkencurve oorzaak van de gewaande terrestrische invloed op de zonnevlekken – moge gelden, dat de tweede differenties van de maandelijks zonnevlekgetallen voor nauwelijks meer dan 50% negatief zijn.¹⁾ De kromming gaat dus geheel schuil onder andersoortige afwijkingen, bijgevolg kan de convexiteit niet de veronderstelde hoofdoorzaak zijn.

¹⁾ $y = f(x)$ is convex, indien $\frac{d^2 y}{dx^2} < 0$.

Een nevenvorm, die aan het krommingseffect doet denken, vinden we bij KOPPE (1951, 88). Zijn Abb. 1 vertoont o.a. het gemiddelde van stukken, gesneden uit de grafiek van de dagelijkse zonnevlekgetallen. Als middelpunt van de stukken nam Prof. KOPPE steeds een magnetisch bijzonder sterk gestoorde dag. In de jaren 1946-1949 vond hij 52 dagen (zgn. sleuteldagen), waarop het magnetisch karaktergetal C in Wingst = 2 was en die minstens drie dagen uit elkaar lagen. Nu werd een matrix ingevuld. De ruggegraat, de middelste kolom, 52 elementen lang, bevatte de zonnevlekgetallen van de sleuteldagen. Elke afzonderlijke rij werd gecompleteerd met de gegevens van de 8 aan de sleuteldag voorafgaande en de 8 daaropvolgende dagen.

Na sommering van de 17 kolommen en deling door 52 ontstond het gemiddelde verloop van de zonnevlekgetallen in de omgeving van magnetisch sterk gestoorde dagen. KOPPE's figuur toonde een duidelijke en regelmatige daling van het zonnevlekgetal.

Dit resultaat staat in scherp contrast tot een welbekend magnetisch feit, dat in de titel van paragraaf 11.15 van CHAPMAN and BARTELS (1940) weerspiegeld wordt: „The lack of general correlation between the sunspot numbers and the magnetic activity on individual days”.

De verklaring van de tegenstelling vinden we, wanneer we onze aandacht richten naar de periode, die Prof. KOPPE gebruikte: 1946-1949. Het maximum van de zonnevlekgetallen ligt in het begin van het tijdvak, zodat de rijen van de matrix overwegend genomen zijn uit de dalende tak van de zonnevlekkenkromme. Deze gang, die in de afzonderlijke rijen verduisterd wordt door toevallige invloeden, moet in het gemiddelde van de 52 rijen sterk tot uiting komen.

Strikt genomen mogen we hier niet spreken van een krommingseffect – het zou eerder een dalingseffect kunnen heten – maar de verschijnselen zijn zo verwant, dat de reeds lang ingeburgerde naam als verzamelbegrip kan blijven fungeren.

Rijzen echter bezwaren, dan zou de suggestie „bartelseffect” voor en krommingseffect en dalingseffect en eventueel later te ontdekken analoge effecten niet ongegrond zijn.

HOOFDSTUK 4

SLOTBESCHOUWINGEN OVER
GEOFYSISCH-STATISTISCHE RESULTATEN4.1 DE σ_r -FOUT

Op p. 144 werden correlatiecoëfficiënten besproken, waarvan VON RUDLOFF und TRENKLE de significantie trachtten aan te tonen met behulp van f_r . Deze grootheid noemen zij de gemiddelde of toevallige fout. Hun definitie is $f_r = \frac{1-r^2}{\sqrt{n}}$.

Niet tevreden met deze incorrecte definiëring en ongelukkige benaming maken ze nog een fout: f_r wordt niet op de juiste wijze gebruikt, want zij eisen, dat voor een bruikbare correlatie r minstens het zesvoudige van f_r moet zijn.

De volgende reeks van bezwaren kan geopperd worden:

punt 1. Het moet zijn $f_r = \frac{1-\rho^2}{\sqrt{n}}$ in plaats van $\frac{1-r^2}{\sqrt{n}}$. In de formule voor de uni-

versumstandaarddeviatie mogen geen steekproefgrootheden gebruikt worden.

punt 2. Men moet rekening houden met het benaderingskarakter van deze formule. Reeds in 1917 schreef SOPER c.s.: — It is well known that for the „probable error” to have meaning the distribution must approach the Gaussian distribution. It is clear that these conditions are by no means fulfilled for samples of 25 and 50, whatever be the value of ρ . There is nearer approach in the low values of ρ in samples of 100, but there is considerable deviation for $\rho = 0.5$, and upwards. —

Exacte overschrijdingskansen voor verschillende waarden van r bij enige waarden van ρ (meestal $\rho = 0$) zijn in de meeste Angelsaksische statistische leerboeken opgenomen. In de Duitse literatuur zijn alleen de werken van BAUR (1948, 142) en KOLLER (1943) van deze onmisbare gegevens voorzien.

punt 3. Het voor toepassing van de exacte methode benodigde werk zinkt in het niet, vergeleken bij het rekenwerk, dat de andere methode eist.

punt 4. Het is beter, f_r met σ_r aan te duiden en dit symbool de internationale naam standaarddeviatie te geven.

punt 5. De foutieve toepassing. Het quotiënt r/f_r (met $f_r = \frac{1-r^2}{\sqrt{n}}$), dat minstens

6 diende te bedragen, is een maat voor de waarschijnlijkheid, dat een negatieve steekproefcorrelatiecoëfficiënt optreedt bij een universumwaarde $= |r|$. Het is ook een maat voor de kans, dat een positieve steekproefcorrelatiecoëfficiënt optreedt bij een universumwaarde $= -|r|$.

Deze waarschijnlijkheden zijn echter totaal onbelangrijk. Gevraagd wordt altijd naar de significantie van een steekproefcorrelatie. Het antwoord wordt gegeven door de waarschijnlijkheid van het optreden van een steekproefcorrelatiecoëfficiënt $> |r|$, wanneer de universumwaarde $= 0$ is (de overschrijdskans van r).

Is men er zeker van, dat de frequentieverdeling van de steekproefcorrelatiecoëfficiënten voldoende bij de gaussverdeling aansluit, dan heeft het zin de standaarddeviatie σ_r te bepalen. Voor σ_r geldt in dit geval: $\sigma_r = \frac{1}{\sqrt{n}}$.

De overschrijdskans van de grootheid $r/\sigma_r = r \sqrt{n}$ kan afgelezen worden in fig. 5 op p. 21.

Mrs. NEWNHAM (1928) zegt van deze „important source of error in the use of the correlation coefficient”, dat „meteorological literature abounds in examples of a quite indefensible process whereby the existence of a genuine connexion between two quantities is made to appear far more probable than is actually the case”. Inderdaad, σ_r wordt ten onrechte verkleind en de overschrijdskans ten onrechte vergroot, wanneer men σ_r gelijkstelt aan $\frac{1-r^2}{\sqrt{n}}$, in plaats van aan $\frac{1}{\sqrt{n}}$.

Als voorbeeld geeft Mrs. NEWNHAM: $r = 0,7$ en $n = 16$. Daaruit volgt: $\sigma_r = \frac{1-r^2}{\sqrt{n}} = 0,13$ met $r/\sigma_r = 5,4$ en een overschrijdskans van 0,000.000.02. De betere methode levert: $\sigma_r = \frac{1}{\sqrt{n}} = 0,25$ met $r/\sigma_r = 2,8$ en een overschrijdskans van 0,00256.

FISHER (1928a) reageert hierop onmiddellijk. Hij stelt zich wel achter Mrs. NEWNHAM's eerste beschouwing (de fout in *punt 5* van deze paragraaf beschreven), maar wijst dan op het ongeoorloofde om de frequentieverdeling van r voor grote waarden van r/σ_r aan een gaussverdeling gelijk te stellen. De exacte methode (1.11.1) levert een overschrijdskans van 0,00127, wanneer $r = 0,7$, $n = 16$ en $\rho = 0$. De in *punt 2* van deze paragraaf genoemde fout veroorzaakt voor deze getallenwaarden dus een afwijking van meer dan 100%.

FISHER besluit met de woorden: — the formula for the standard error of the correlation coefficient may soon be classed among the things „which students have to know, but only a fool would use”. —

Om niet onvriendelijk te zijn, moeten we wel aannemen, dat het begrip „soon” bij FISHER meer dan 24 jaar omvat, want een rij van onderzoekers blijft manipuleren met de standaarddeviatie van de correlatiecoëfficiënt.

Naast de in het begin van deze paragraaf genoemde meteorologen beperken we ons tot recente overtreders als BAMPA e MELIS (1949, 46) en NEUWIRTH (1951, 71). Vòòr deze jaren bezigde welhaast iedereen de foutieve formule, in het bijzonder in continentaal Europese publicaties.

Een groot gedeelte van de schuld draagt het meteorologische handboek van HANN-SÜRING. Nog in de vijfde uitgave van 1938 is op p. 93 de methode te vinden, waarover FISHER zich zo denigrerend uitliet.

Onbekendheid met de Engelse literatuur kan SÜRING niet als verontschuldiging aanvoeren, want twee jaar voor Mrs. NEWNHAM wees SCHMIDT (1926, 332) reeds op de in *punt 5* besproken fout. De geringe invloed van SCHMIDT's artikel kan verklaard worden uit de aanwezigheid van enkele zeer storende zelffouten.

Het is daarom beter, hier niet van de Ad. SCHMIDT-fout (VAN DER BIJL 1951, 211) te spreken, maar eenvoudig van de σ_r -fout.

Een merkwaardige combinatie van selectieeffect en σ_r -fout treffen we aan bij MEISSNER (1943). Uit dit voorbeeld leren we, dat kleine steekproeven vaak onverwacht hoge correlatiecoëfficiënten opleveren.

Decennium	Aantal	In zijn tabel 1 publiceert MEISSNER de aantallen
1821—30	6,3	ijsdagen in December te Praag per decennium. Cor-
31—40	8,7	releren we de decenniumgemiddelden van 1821—30
41—50	9,0	tot 1871—80 met de tijd, dus met de reeks 1, 2, 3, 4,
51—60	10,6	5 en 6, dan krijgen we $r = 0,99 \pm 0,06$ volgens
61—70	9,5	MEISSNER. De toevoeging $\pm 0,06$ betekent, dat de
71—80	11,4	standaarddeviatie, berekend naar een formule, die
81—90	9,2	iets anders is dan de in deze paragraaf gebruikte,
91—00	7,9	0,06 oplevert.
1901—10	7,0	De hoge correlatie is ten dele ontstaan door het
11—20	4,7	selectieeffect (3.2). De selectiefactor is echter niet

uitzonderlijk groot.

Zou de zojuist bepaalde correlatiecoëfficiënt inderdaad volgens GAUSS verdeeld zijn met een standaarddeviatie = 0,06, dan is het verband zeer significant, ook als we de selectie in rekening brengen.

Gelijke beschouwingen kunnen we houden voor het tweede gedeelte van de tabel. De correlatiecoëfficiënt tussen de aantallen ijsdagen in de decennia 1871—80 tot 1911—20 enerzijds en de getallen 1—5 anderzijds is $-0,99 \pm 0,06$. Terecht noemt MEISSNER dit „hohe und doch irreführende Korrelationsfaktoren”, maar hij geeft niet de juiste overschrijdingskansen aan.

Vermijdt men de σ_r -fout, dan ziet men uit een tabel, zoals tabel VI van FISHER and YATES (1949), dat $r = 0,992$ (berekend uit vijf) en $r = 0,974$ (berekend

uit zes paren waarnemingen) een overschrijdingskans van 0,001 hebben. De door MEISSNER gevonden correlaties hebben dus kleine waarschijnlijkheden, die echter wel door het selectieeffect veroorzaakt zullen zijn.

4.2 EEN BIJZONDER INGEWIKKELD GEVAL VAN PSEUDOCORRELATIE

De tabellen XXXV–LXII van VISSER (1946) bevatten correlaties tussen een meteorologische grootheid x van het station y in één van de maanden Jan.–Dec. en de neerslag in Nederland of de temperatuur te De Bilt z maanden later. Op grond van fysische overwegingen zijn x en y zorgvuldig uitgekozen, terwijl z meestal 1–12 bedraagt, maar soms tot 24 oploopt.

Deze tabellen vertonen in grote meerderheid een bonte afwisseling van positieve en negatieve waarden. Slechts twee tabellen maken daarop een uitzondering, nl. tabel XLVIII en LXII.

Het gemeenschappelijke van dit tweetal is, dat ze betrekking hebben op correlaties van de perioden van $2\frac{1}{4}$ en $5\frac{1}{4}$ jaar. Eerst zijn deze perioden, maand voor maand, gecorreleerd met de neerslag in Nederland (tabel XLVIII), later met de temperatuur in De Bilt (tabel LXII), eveneens voor elke maand.

Alle twaalf correlatiecoëfficiënten van de $P\ 2\frac{1}{4}$ met de neerslag in Nederland zijn positief, terwijl de drie overige rijen ieder slechts één negatieve correlatiecoëfficiënt bezitten ($P\ 2\frac{1}{4}$ = periode van $2\frac{1}{4}$ jaar).

Naast deze overeenkomsten zijn er ook verschillen. Voor niet minder dan tien van de twaalf maanden was de correlatie zo stram, dat de periode van $2\frac{1}{4}$ jaar in de regressievergelijkingen ¹⁾ voor de neerslag gebruikt kon worden; daarentegen komen $P\ 2\frac{1}{4}$ en $P\ 5\frac{1}{4}$ in de overige 3×12 gevallen helemaal niet voor: de periode van $5\frac{1}{4}$ jaar in geen enkele regressievergelijking voor temperatuur of neerslag en de periode van $2\frac{1}{4}$ jaar in geen enkele temperatuurvergelijking.

Bepalen we van de correlatiecoëfficiënten de gemiddelden over de twaalf maanden, dan krijgen we het volgende staatje:

	neerslag Nederland (tabel XLVIII)	temperatuur De Bilt (tabel LXII)
$P\ 2\frac{1}{4}$	0,282	0,112
$P\ 5\frac{1}{4}$	0,182	0,165

Wanneer we bedenken, dat deze gemiddelden standaarddeviaties van ca 0,04 hebben (zie later), dan zijn er markante verschillen te noteren.

¹⁾ Het correlatieonderzoek was bij VISSER (1946) slechts middel. Het doel was regressievergelijkingen (1.11.3) op te stellen, waarin links de temperatuur te De Bilt of de neerslag in Nederland in een bepaalde maand voorkwam en rechts 5 à 7 grootheden x van stations y . Om een vroegtijdige voorspelling in het bezit te hebben, werd z zo groot mogelijk gekozen.

Ook hier moet, zoals steeds, de vraag „toeval of realiteit?” gesteld worden. De hypothese, dat alle 48 correlaties door het toeval veroorzaakt werden, schijnt – oppervlakkig beschouwd – zonder meer verworpen te kunnen worden.

Want het lijkt onmogelijk, dat de toevalshypothese kwesties als de volgende kan ophelderen:

Waarom komt de periode van $2\frac{1}{4}$ jaar in de neerslag zoveel beter tot zijn recht dan in de temperatuur?

Hoe is het te verklaren, dat er 45 positieve correlaties voor de dag kwamen tegenover slechts 3 negatieve?

Waarom is in de neerslag de $P\ 2\frac{1}{4}$ de belangrijkste en in de temperatuur de $P\ 5\frac{1}{4}$?

Niet alleen konden deze kwalitatieve vragen bevredigend beantwoord worden, maar bovendien werd ook nog een kwantitatieve overeenstemming bereikt. We hebben dit te danken aan de gemakkelijkerheid, waarmee ingewikkelde correlatiecoëfficiënten zich laten herleiden tot een functie van eenvoudige correlatiecoëfficiënten (vergelijk VAN DER BIJL 1951c).

We beginnen ons onderzoek met de correlatie tussen de $P\ 2\frac{1}{4}$ en de neerslag in Nederland.

De periode van $2\frac{1}{4}$ jaar wordt voor elke maand weergegeven door een serie van negen getallen, die zich om de negen jaren herhalen. Deze periode is berekend uit waarnemingen van het tijdvak 1849–1938 en wel op de volgende wijze:

Eerste getal voor Jan. = gemiddelde van de neerslagafwijkingen in de
Januarimaanden 1849, 1858, 1921 en 1930

Tweede getal voor Jan. = gemiddelde van de neerslagafwijkingen in de
Januarimaanden 1850, 1859, 1922 en 1931

Dit gaat zo door tot:

Negende getal voor Jan. = gemiddelde van de neerslagafwijkingen in de
Januarimaanden 1857, 1866, 1929 en 1938

Voor de elf andere maanden werden de „periodegetallen” (wellicht een betere naam dan „periode van $2\frac{1}{4}$ jaar”) op analoge wijze berekend.

Correleren we nu de neerslag in de maand Januari met die „periodegetallen” voor het tijdvak 1901–1939, dan krijgen we deze producten in de teller ($J\ 1bcd$ N = afwijking van de neerslaghoeveelheid in Januari van het jaar $1bcd$ ten opzichte van de normaal):

$$\begin{aligned}
 J 1901 N \times \frac{1}{10} (& J 1856 N + J 1865 N + \dots + J 1901 N + \dots + J 1928 N + J 1937 N) \\
 J 1902 N \times \frac{1}{10} (& J 1857 N + J 1866 N + \dots + J 1902 N + \dots + J 1929 N + J 1938 N) \\
 J 1903 N \times \frac{1}{10} (& J 1849 N + J 1858 N + J 1867 N + \dots + J 1903 N + \dots + J 1930 N)
 \end{aligned}$$

Dit eindigt met:

$$\begin{aligned}
 J 1938 N \times \frac{1}{10} (& J 1857 N + J 1866 N + \dots + J 1902 N + \dots + J 1929 N + J 1938 N) \\
 J 1939 N \times \frac{1}{10} (& J 1849 N + J 1858 N + J 1867 N + \dots + J 1903 N + \dots + J 1930 N)
 \end{aligned}$$

We zien hier een uiting van het pseudocorrelatieeffect.

De term vóór het vermenigvuldigteken komt ook steeds voor in het gedeelte daarna (de laatste regel uitgezonderd). Nemen we aan, dat de tien gegevens, die zich in elke regel rechts bevinden, onderling onafhankelijk zijn en passen we vervolgens de theorie van het pseudocorrelatieeffect toe (3.4), dan vinden we als de (voorlopige) universumwaarde ρ' van de in tabel XLVIII berekende correlatiecoëfficiënten $\rho' = 1/\sqrt{10} = 0,316$.

Deze waarde dient verminderd te worden met $1/39 \rho'$, omdat het laatste product geen gemeenschappelijke termen bevatte. De juiste waarde wordt: $\rho = {}^{38}/_{39} \rho' = 0,308$.

De twaalf berekende correlatiecoëfficiënten schommelen om een gemiddelde van 0,282.

Elke correlatiecoëfficiënt heeft een standaarddeviatie $= \frac{1 - (0,308)^2}{\sqrt{39}} = 0,145$.

Uit de twaalf gevonden waarden (tabel XLVIII) werd als standaarddeviatie berekend $s = 0,158$. Het verschil ($= 0,013$) is kleiner dan de universumstandaarddeviatie σ_s van de steekproefstandaarddeviaties, want maken we gebruik van de in 1.6.2 besproken benaderingsmethode, dan is $\sigma_s = \frac{\sigma}{\sqrt{2n-2}} = \frac{0,145}{\sqrt{22}} = 0,031$.

Het gemiddelde dezer twaalf correlatiecoëfficiënten heeft volgens de foutenvoortplantingswet een universumstandaarddeviatie $= \frac{0,145}{\sqrt{12}} = 0,042$.

Het verschil tussen ρ en het berekende gemiddelde ($= 0,308 - 0,282 = 0,026$) is kleiner dan de standaarddeviatie van het gemiddelde ($= 0,042$).

Het is onnodig om overschrijdingskansen te bepalen.

Zou het verband tussen de neerslag in Nederland en de periode van $5\frac{1}{4}$ jaar op dezelfde wijze onderzocht zijn, dan zouden rechts niet $10 (= \frac{90}{4 \times 2\frac{1}{4}})$, maar $4 \text{ à } 5 (= \frac{90}{4 \times 5\frac{1}{4}})$ elementen staan.

De universumwaarde van de correlatiecoëfficiënt zou dan $\frac{1}{\sqrt{4}} \hat{=} \frac{1}{\sqrt{5}} = 0,45$ à 0,50 zijn. Deze waarde wijkt significant af van het gemiddelde 0,182.

Bij nadere beschouwing bleek inderdaad een andere methode gevolgd te zijn.

Eerst werd elke maandhoeveelheid vervangen door het derde gedeelte van de totaalhoeveelheid, gevallen in die maand, de voorafgaande en de volgende. Dan werden alle hoeveelheden uit het tijdvak 1849-1938 opgeteld, die $5\frac{1}{4}$ jaar uit elkaar lagen.

Het eerste periodegetal voor Januari is derhalve het gemiddelde van:

$$\begin{aligned} & \frac{1}{3} (D 1848 N + J 1849 N + F 1849 N) + \\ & + \frac{1}{3} (M 1854 N + A 1854 N + M 1854 N) + \\ & + \frac{1}{3} (J 1859 N + J 1859 N + A 1859 N) + \\ & + \frac{1}{3} (S 1864 N + O 1864 N + N 1864 N) + \\ & + \frac{1}{3} (D 1869 N + J 1870 N + F 1870 N) + \\ & \text{enz. enz. tot} \\ & + \frac{1}{3} (S 1927 N + O 1927 N + N 1927 N) + \\ & + \frac{1}{3} (D 1932 N + J 1933 N + F 1933 N) + \\ & + \frac{1}{3} (M 1938 N + A 1938 N + M 1938 N) + \end{aligned}$$

Het tweede periodegetal voor Januari wordt verkregen door alle jaartallen met 1 te vermeerderen (de laatste term valt dan weg, omdat het tijdvak tot aan 1939 loopt). Voor het k^{de} periodegetal worden de jaartallen $(k-1)$ groter.

De periodegetallen voor Februari ontstaan door alle maanden door hun opvolger te vervangen. De berekeningen voor de overige tien maanden worden op analoge wijze uitgevoerd.

Uit een en ander kan afgeleid worden, dat het p^{de} periodegetal van de q^{de} maand gelijk is aan het $(p+5)^{\text{de}}$ periodegetal van de $(q+3)^{\text{de}}$ maand.

Correleren we nu de neerslag in Nederland (n) voor het tijdvak 1901-1939 met de periodegetallen (p) van $P 5\frac{1}{4}$, dan heeft de factor p van het product in de teller gemiddeld $3 \times \frac{90}{5\frac{1}{4}} = \frac{360}{7}$ termen. De factor n bestaat steeds uit één term, maar deze bevindt zich jaar op jaar ook onder de ruim 50 termen van factor p (uitgezonderd het product voor 1939).

De theorie van het pseudocorrelatieeffect leert, dat de (voorlopige) universumwaarde ρ' in dit geval zal zijn:

$$\rho' = \frac{38}{39} \times \frac{1}{\sqrt{\frac{360}{7}}} = 0,136$$

We hebben een fout gemaakt door het persistentieeffect over het hoofd te zien. Immers, als we de producten in de teller van de correlatiecoëfficiënt uitwerken, blijft naast $(J 1bcd N)^2$ staan $(J 1bcd N) \times (D 1bcd-1 N)$ en $(J 1bcd N)$

$\times (F \text{ } 1bcd \text{ } N)$. (N = neerslagafwijking; D , J en F zijn de elkaar opvolgende maanden; $1bcd$ geeft het jaartal aan). De verwachtingswaarde van deze producten is $\neq 0$. In de noemer spelen zij geen rol van betekenis, in de teller wel.

De autocorrelatiecoëfficiënten van de eerste orde (1.12.1) zijn ongeveer 0,1. We moeten ρ' een bijslag geven van $2 \times 10\% = 0,027$. De juiste universumwaarde ρ voor de correlatiecoëfficiënt van $P \text{ } 5\frac{1}{4}$ met de neerslag in Nederland wordt $0,136 + 0,027 = 0,163$.

Gevonden is als gemiddelde van de twaalf correlatiecoëfficiënten 0,182 en als steekproefstandaarddeviatie van dit gemiddelde 0,040. De universumstandaarddeviatie van het gemiddelde is $\frac{1 - (0,163)^2}{\sqrt{39 \times 12}} = 0,045$.

Alle verschillen zijn ook nu weer niet significant.

De periode van $5\frac{1}{4}$ jaar in de temperatuur te De Bilt werd op dezelfde wijze als in de neerslag van Nederland bepaald. De hierboven gegeven universumwaarden gelden dus ook voor de berekeningen met de correlatiecoëfficiënten uit tabel LXII van VISSER (1946).

De twaalf correlaties geven een gemiddelde van 0,165, de steekproefstandaarddeviatie van dat gemiddelde is 0,046 (universumwaarden 0,163, resp. 0,045). De zeer nauwe overeenstemming behoeft niet tot verwerping te leiden, daar de invloed van het persistentieeffect globaal met een geschatte waarde van $2 \times 10\%$ is aangegeven.

De bepaling van de periode van $2\frac{1}{4}$ jaar in de temperatuur geschiedde door de gemiddelden van die maandelijkse afwijkingen te berekenen, die 27 maanden van elkaar verwijderd lagen. Echter, zonder eerst de maandelijkse afwijking „glad te strijken”, d.w.z. te vervangen door het gemiddelde van drie maanden met de oorspronkelijke in het centrum.

De voorgaande uitvoerige statistische beschouwingen kunnen thans herhaald worden. Gemakkelijk is in te zien, dat de nieuwe universumwaarde gelijk is

$$\text{aan } \frac{38}{39} \times \sqrt{\frac{1}{90 \cdot 2\frac{1}{4}}} = \frac{38}{39} \times \frac{1}{\sqrt{40}} = 0,154.$$

De universumstandaarddeviatie van het gemiddelde der twaalf correlatiecoëfficiënten is $\frac{1 - (0,154)^2}{\sqrt{39 \times 12}} = 0,045$.

De steekproef, de gegevens uit tabel LXII, leverde ons de waarden 0,112, resp. 0,042.

In het begin van deze paragraaf hebben we de regressievergelijkingen reeds genoemd. We richten nu onze aandacht op de regressiecoëfficiënten van de $P \text{ } 2\frac{1}{4}$, optredend in de vergelijkingen voor de neerslag in Nederland (VISSER 1946, 117).

Het is geoorloofd, op deze coëfficiënten de theorie van de enkelvoudige regressie (1.11.2) toe te passen. We weten dus, dat de universumwaarde β van de regressiecoëfficiënten bepaald wordt door

$$\beta = \frac{\sum_{i=1}^N (x_i - \mu_1) (y_i - \mu_2)}{\sum_{i=1}^N (x_i - \mu_1)^2}$$

Een eenvoudige schrijfwijze voor deze formule is:

$$\beta = \frac{E(x_0 x_1)}{E(x_1^2)}$$

x_0 is de neerslagafwijking in Nederland, bv. in Januari en x_1 heeft termen als $\frac{1}{10} (J 1856 N + J 1865 N + \dots + J 1937 N)$.

Stel nu $E(x_0^2) = \sigma^2$, dan is $E(x_1^2) = \frac{1}{100} (\sigma^2 + \dots + \sigma^2) = \frac{1}{10} \sigma^2$ en $E(x_0 x_1) = \frac{38}{39} \times \frac{1}{10} \sigma^2$.

Hieruit volgt: $\beta = \frac{38}{39} = 0,974$.

De universumstandaarddeviatie σ_b is $= \frac{\sigma_0 \sqrt{1-\rho^2}}{\sigma_1 \sqrt{n}}$ (zie 1.11.2).

In deze formule moeten we voor ρ substitueren de vroeger reeds berekende waarde

$$\rho = \frac{38}{39} \times \frac{1}{\sqrt{10}} = 0,308.$$

Verder geldt: $n = 39$ $\sigma_0 = \sigma$ $\sigma_1 = \frac{\sigma}{\sqrt{10}}$

Tenslotte is het resultaat: $\sigma_b = \sqrt{10} \times \sqrt{\frac{1-(0,308)^2}{39}} = 0,482$. De schattingen s_b voor σ_b hebben een universumstandaarddeviatie $= \frac{0,482}{\sqrt{22}} = 0,103$.

De universumstandaarddeviatie van het gemiddelde van de twaalf steekproefregressiecoëfficiënten is $\frac{0,482}{\sqrt{12}} = 0,139$.

Hiermede zijn alle benodigde theoretische waarden bepaald. De steekproefwaarden b van de regressiecoëfficiënten ontleen we aan tabel LXV, deel A (VISSE 1946, 117). Ze zijn in volgorde van grootte geplaatst: October 1,65, Juni 1,27, Januari 1,24 (gemiddelde van 1,22 en 1,26), Juli 1,19, Februari 1,14, September 1,07, Augustus 0,873, Maart 0,835, November 0,809 (gemiddelde uit 0,790 en 0,828) en als kleinste Mei 0,795.

Er zijn zes waarden groter dan β en vier kleiner.

Een kwestie, die gedeeltelijk met dit onderzoek samenhangt, is de vraag naar de realiteit van de perioden van $2\frac{1}{4}$ en $5\frac{1}{4}$ jaar. De statistische zijde van dit probleem kan het best met de variantieanalyse (1.10) aangepakt worden.

Uit ervaring weten we, dat zelfs deze er niet in zal slagen, de significantie van de genoemde perioden aan te tonen, daar enerzijds de perioden te rudimentair ontwikkeld zijn en anderzijds het aantal gegevens te gering is. Zodoende blijven alleen theoretische argumenten voor het bestaan van de perioden pleiten. Voor deze argumenten, waarin de geografische verspreiding een grote rol speelt, verwijzen wij naar de uitvoerige beschouwingen, die VISSER (1946, 26-9) aan de periode van $2\frac{1}{4}$ jaar wijdt.

Dat de vraag naar de realiteit der perioden inderdaad samenhangt met het hier behandelde, blijkt uit de veronderstelling, die we steeds moesten gebruiken om toepassing van de theorie van de pseudocorrelatie mogelijk te maken. We hebben nl. aangenomen – bv. bij de $P\ 2\frac{1}{4}$ van de neerslag in Nederland – dat de tien termen $J\ 1856\ N$, $J\ 1865\ N$, $J\ 1874\ N$, . . . en $J\ 1937\ N$ als onderling onafhankelijk beschouwd mochten worden. Deze onafhankelijkheid impliceert de afwezigheid van een periode van $2\frac{1}{4}$ of 9 jaren.

Ook hypothesen, die een weinig markante, een zwakke periode postuleren, zullen het gevonden gedrag van de correlaties en regressies kunnen verklaren. Maar het bestaan van een duidelijke periode van $2\frac{1}{4}$, $5\frac{1}{4}$ of 9 jaren is onverenigbaar met de becijferde gegevens.

Met dit onderzoek hebben we enkele conclusies van een stevige theoretische basis voorzien. Deze conclusies, die VISSER (1946) empirisch samenstelde, zijn: „ . . . dat in de temperatuur de periode van 27 maanden veel minder tot ontwikkeling komt dan in de neerslag” (p. 33).

„Voor de temperatuur heeft de periode van 9 jaar niet of nauwelijks betekenis” (p. 34).

„Voor de weersverwachtingen op lange termijn moet de periode van $5\frac{1}{4}$ jaar zowel voor neerslag als voor temperatuur verworpen worden” (p. 44).

4.3 DE GRAFIEKEN-METHODE IN HET SIGNIFICANTIE-TIJDPERK

Wij zullen voorloopig maar een enkel voorbeeld geven van meteorologische lijnen waaruit duidelijk kan blijken hoe men alleen door de aanwending van grafische lijnen en door hare onderlinge vergelijking in staat wordt gesteld om de betrekkingen, die tusschen Kosmische verschijnselen bestaan, goed in 't oog te doen vallen.

MAREY 1881

Deze langademige ontboezeming verschaft ons een duidelijk inzicht in de wijze van wetenschapsbeoefening tijdens de 19^{de} eeuw. Blijkbaar vond de ontdekking, dat grafische voorstellingen essentieel zijn bij het opsporen van nog onbekende relaties, geen algemene ingang. Want MAREY (1881, 62) voegt tenminste op de volgende bladzijde aan enkele opmerkingen over het door SABINE (1851) vastgestelde verband tussen de zonnevlekken en de magnetische declinatie

toe: – Hoogst waarschijnlijk zullen in de sterrekunde en in de meteorologie nog zeer belangrijke ontdekkingen op dit gebied worden gedaan; wij bevestigen hier alleen dat het nieuwe veld, dat hiermede ter bearbeiding voor de wetenschap is opgesteld, hoofdzakelijk is aangewezen door de grafische methode, die alleen in staat is een helder licht te werpen op dergelijke zonderlinge betrekkingen. –

Het laatste gedeelte van de afgelopen eeuw kunnen we derhalve het grafieken-tijdperk noemen. De allengs groter wordende invloed van mathematisch-statistische methoden stempelt onze eeuw tot het significantie-tijdperk.

In 4.3.1 en 4.3.2 behandelen we werken van twee onderzoekers, die qua methode nog in de vorige eeuw leven.

4.3.1 DE INVLOED VAN DE MAAN OP DE TEMPERATUUR

Notre satellite est souvent mal vu au pied des girouettes.

VITTRANT 1942

Op p. 122 hebben we aangetoond, dat zelfs van een zo suggestieve grafische voorstelling als de tweede kromme van fig. 15 niet bewezen kon worden, dat de afwijkingen door een reële oorzaak ontstaan waren. Uit deze van HOFMANN afkomstige figuur kunnen we leren, aan welke twee voorwaarden een grafische voorstelling moet voldoen, wil hij „in 't oog vallend” genoemd kunnen worden. De twee voorwaarden zijn: 1) de eenheid langs de verticale as moet zo groot mogelijk gekozen worden en 2) de abscissenas moet de ordinatenas niet in het punt 0 snijden, maar zo dicht mogelijk bij de kromme zelf.

De geofysicus, die wel het best deze eisen ingewilligd heeft, is VITTRANT (1942). Zijn fig. 14, die hier als onze fig. 18 op dezelfde schaal is weergegeven, kwam op de volgende wijze tot stand: Uit de gegevens van de 69 jaren 1873–1941 werden de zgn. „maximumdagen” in Zi Ka Wei (China) berekend. Een dag is een „maximumdag”, wanneer het temperatuurgemiddelde T_i van die dag groter is dan dat van de voorafgaande of volgende dag, dus wanneer geldt:

$$T_{i-1} < T_i > T_{i+1}$$

De 69 jaren bevatten 4 041 maximumdagen (d.i. gemiddeld 1 per 6,24 dagen). Nu werd van alle maximumdagen bepaald, welke positie zij innamen in de synodische maan-cyclus van 29,53 dagen. Door de cyclus beurtelings 30 en 29 dagen te laten tellen en daarbij nog een geringe verfijning toe te passen, kon het werken met gebroken getallen vermeden worden. Op deze wijze was het mogelijk, van elke dag in de maan-cyclus te berekenen, hoeveel maximumdagen er op gevallen waren. Het resultaat is in tabel 22 weergegeven.

Op de 30^{ste} dag van de maancyclus kwamen niet meer dan 86 maximumdagen voor, maar deze dag (met $i = 30$) was in slechts 53 van elke 100 cycli aanwezig. Derhalve is 86 met 100/53 vermenigvuldigd, uitkomst 162.

Tabel 22

i	m	i	m	i	m	i	m	i	m	i	m
1	120	6	114	11	151	16	151	21	144	26	154
2	144	7	142	12	150	17	143	22	125	27	140
3	140	8	137	13	127	18	120	23	127	28	130
4	145	9	135	14	110	19	127	24	149	29	145
5	140	10	132	15	161	20	130	25	122	30	162 (86)
	689		660		699		671		667		731

i: dag in de maancyclus

m: aantal maximumdagen op dag i (periode 1873-1941)

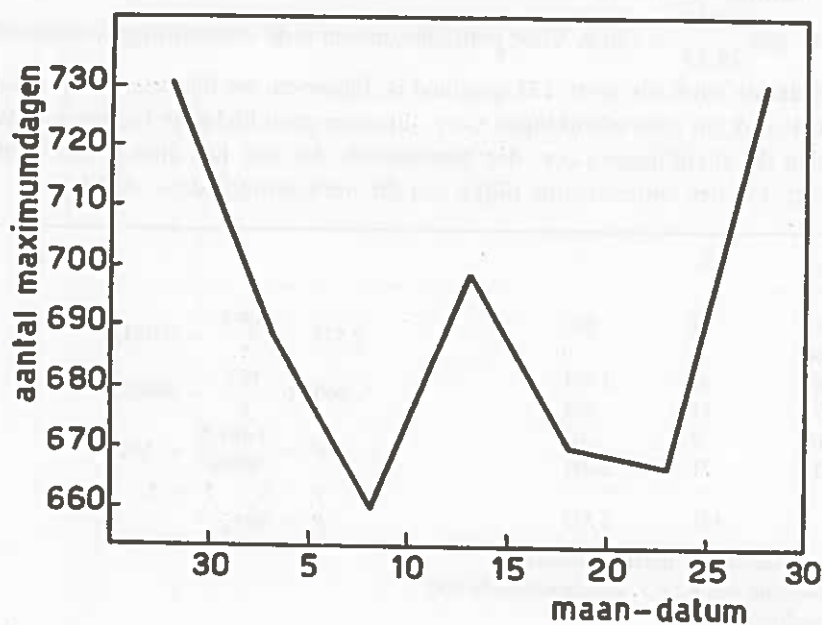


fig. 18

De grafische voorstelling van VITTRANT toont de pentadesommen. Omdat zo uitstekend aan de twee in het begin genoemde voorwaarden voldaan is, schijnt de kromme een overtuigende mate van realiteit te bezitten. In ieder geval werd VITTRANT er zelf door overtuigd. Dit was niet zo moeilijk, daar VITTRANT reeds een vooropgezette mening (vgl. p. 110) over dit vraagstuk had, getuige de fraaie verdediging van zijn opvattingen, waarmee VITTRANT zijn artikel begint: – Comme les campagnards de tous les pays du monde et comme le maréchal

BUGEAUD, le maréchal FOCH croyait à l'influence de la lune sur le temps, ce qui faisait sourire les météorologistes officiels de l'armée; car notre satellite est souvent mal vu au pied des girouettes et il y a mauvaise presse. Pourtant puisque la lune remue périodiquement la mer et que son action y est plus du double de celle du soleil, pourquoi serait-elle incapable de balancer l'atmosphère? –

Ondanks dit hartstochtelijke betoog zal de overgrote meerderheid der geofysische onderzoekers het bestaan van een merkbare maaninvloed op de luchttemperatuur blijven ontkennen. Dat deze algemene opvatting niet in tegenspraak is met VITTRANT's afbeelding, kan met behulp van de G^2 -toets aangetoond worden.

Is de luchttemperatuur te Zi Ka Wei onafhankelijk van de maanfase, dan is de verwachtingswaarde van het aantal maximumdagen voor elke maan-datum gelijk en wel $\frac{4\ 041}{29,53} = 136,8$. Voor pentadesommen is de verwachtingswaarde 684.

Op dezelfde wijze als op p. 123 geschied is, behoeven we bij toepassing van de G^2 -toets ook nu geen afwijkingen t.o.v. dit ware gemiddelde te berekenen. We bepalen de afwijkingen t.o.v. dat gemiddelde, dat ons het minste rekenwerk bezorgt. Uit het onderstaande blijkt, dat dit werkgemiddelde = 660 is.

a	b	c	
689	29	841	
660	0	0	$7\ 573 - \frac{157^2}{6} = 3\ 464,8$
699	39	1\ 521	$660 + \frac{157}{6} = 686,2$
671	11	121	$G^2 = \frac{3\ 464,8}{686,2} = 5,05$
667	7	49	$\nu = 6 - 1 = 5$
731	71	5\ 041	$P = 40\%$
	157	7\ 573	

a: pentadesom van maximumdagen
b: afwijking van a t.o.v. werkgemiddelde 660
c: kwadraten van a

Het comprimeren van gegevens brengt altijd verlies aan inlichtingen met zich mee. Het zou mogelijk kunnen zijn, dat de G^2 -toets, toegepast op de frequenties van de dertig afzonderlijke dagen van de maancyclus, wel een significante maaninvloed onthult.

Nemen we bij de gegevens van tabel 22 als voorlopig gemiddelde 120 aan, dan is de algebraïsche som der afwijkingen $d_i = \sum d_i = 517$ en de som der kwadraten $\sum d_i^2 = 14\ 049$.

$$G^2 = \frac{14\,049 - \frac{517^2}{30}}{120 + \frac{517}{30}} = \frac{5\,139,4}{137,2} = 37,5 \quad v = 29 \quad P = 15\%$$

Ook deze uitgebreidere gegevens verlenen dus geen steun aan de opvattingen van VITTRANT.

De volgende artikelen van VITTRANT (1943, 1944 en 1945) bevatten veel meer suggestieve grafieken. Zij alle hebben betrekking op de invloed van de maan op geofysische grootheden. Ondanks de bewering van VITTRANT (1945, 291), dat: – *La météorologie n'est pas encore une science, „cognitio per causas”. Il faut la créer. Tout nous porte à croire que la plus intense, parmi les causes que l'on désire connaître, est l'action lunaire* –, zal het toch niet moeilijk zijn, de bereikte resultaten met behulp van statistische methoden als insignificant te toetsen.

4.3.2 DE INVLOED VAN DE MAAN OP DE NEERSLAG

Na een slap aftreksel van de vurige pleidooien aangehoord te hebben van een advocaat, die het bestaan van een maaninvloed op de temperatuur verdedigde, gaan we nu luisteren naar MINEUR (1951). Van zijn hand verscheen het resultaat van een „*Recherche d'une influence possible de la Lune sur les précipitations atmosphériques*”.

Voor we verder gaan, moet ons eerst een opmerking van het hart: de twee voorbeelden van deze paragraaf zijn in zoverre ongelukkig, dat zij de indruk wekken, dat voornamelijk Fransen de invloed van de maan trachten te bewijzen. Dit is niet waar. Elk land herbergt lotgenoten van VITTRANT en MINEUR; de meesten zullen wel in Duitsland te vinden zijn.

MINEUR bewerkte de etmaalsommen van de neerslag van de vier stations Clermont-Ferrand, Parijs (Parc Saint-Maur), Nantes en Algiers voor een tijdvak van ongeveer 75 jaren.

Om de eventuele invloed van de maanfase op de neerslaghoeveelheid te onderzoeken moeten de omstreeks 75×365 waarnemingen in een matrix van dertig kolommen gerangschikt worden. De dertig etmaalsommen van een bepaalde synodische maand worden achter elkaar in dezelfde rij genoteerd. De regensommen, waargenomen op alle dagen, die dezelfde maan-datum hebben, bevinden zich in dezelfde kolom. Ca 75 jaren omvatten ca 900 synodische maanden, zodat elke kolom 900 getallen telt. De dertig kolomgemiddelden geven een beeld van het gemiddelde neerslagverloop in de beschouwde 900 synodische maanden. Slechts indien de toevalsschommelingen voor het grootste gedeelte door het middelingsproces verdwenen zijn, kan uit de verkregen maandelijks gang de maaninvloed op de neerslag gelezen worden.

De meest typische resultaten, die MINEUR vond, werden in diagrammen gereproduceerd (onze fig. 19). Ook zij voldoen alle aan de in het begin van deze paragraaf vermelde voorwaarden. Des te groter moet onze voorzichtigheid zijn bij het interpreteren van de grafische voorstellingen en het trekken van conclusies. Voor Parijs (Parc Saint-Maur) en Nantes werden de gemiddelde etmaalhoeveelheid $\bar{h}$, de waarschijnlijkheid p van een droge dag en de gemiddelde neerslaghoeveelheid $\bar{h}'$ op regendagen tegen de maanfasen uitgezet. De kruisjes, die deze waarden aangeven, vormen een tamelijk onregelmatig beeld.

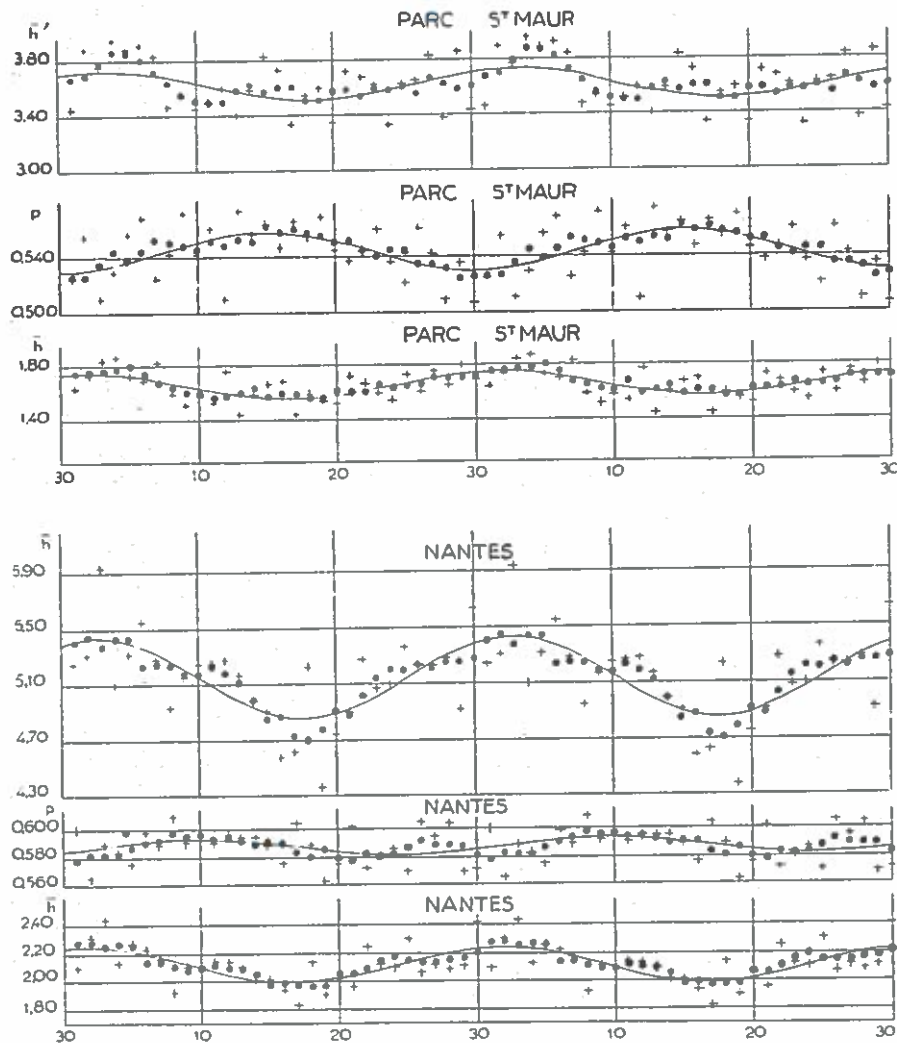


fig. 19

Door een lineaire functie van vijf opeenvolgende waarden te gebruiken slaagt MINEUR erin de onregelmatigheid van het beeld te verminderen. Dit moeten we echter niet als een verdienste van MINEUR zien, maar eerder als een gevolg van de foutenvoortplantingswet. De gladgestreken waarden zijn door punten aangegeven. Tenslotte werd de best aangepaste sinusoïde bepaald.

De diagrammen (fig. 19) geven aanleiding tot de volgende statistische beschouwingen, die alle min of meer fnuikend zijn voor de hypothese van een aanwijsbare lunaire invloed op de neerslag:

1. De berekende gemiddelden $\bar{h}$, $\bar{h}'$ en p wijken herhaaldelijk verder van de sinusoïde af dan de amplitude van de sinusoïde groot is.
2. De standaarddeviatie s van de etmaalsommen, in De Bilt afgetapt, is 4,2 mm (berekend uit de 18 262 waarnemingen van de periode 1901–1950). Er zijn goede gronden om aan te nemen, dat het geringe onderlinge verschil tussen de jaarlijkse neerslaghoeveelheden in De Bilt, Parijs en Nantes ons veroorlooft, de standaarddeviaties van de etmaalsommen van de drie stations aan elkaar gelijk te stellen, althans voor dit onderzoek. De jaargemiddelden zijn voor De Bilt 729 mm (1849–1920), voor Parc Saint-Maur 576 mm (1874–1923) en voor Nantes 786 mm (1881–1920).

Als de neerslag onafhankelijk is van de maanfase, dan zal h een standaarddeviatie s_n hebben $= s/\sqrt{n}$ volgens de foutenvoortplantingswet. (n geeft het aantal waarnemingen aan, waaruit $\bar{h}$ bepaald is). Slechts indien s_n significant groter is dan $s/\sqrt{n}$, staat de realiteit van de maaninvloed vast.

Zonderen we in de h -diagrammen de vijf hoogste en de vijf laagste kruisjes af, dan houden we twintig van de dertig waarden van $\bar{h}$ over (=67%). Dit percentage valt zo goed als samen met het percentage (nl. 68%), dat betrekking heeft op het aantal waarnemingen, gelegen in het interval $< M - s_n, M + s_n >$. M is het algemeen gemiddelde van h .

Deze eenvoudige, zij het ruwe schattingsmethode voor de standaarddeviatie geeft $s_n = 0,15$ mm voor Parijs en $s_n = 0,14$ mm voor Nantes. Nemen we $n = 900$, dan is $s/\sqrt{n} = 0,14$ mm.

3. Een analoge methode kunnen we toepassen voor de spreidingsbreedte. Het verschil tussen de grootste en kleinste etmaalsom in een periode van dertig (eigenlijk 29,53) dagen zij R .

De gemiddelde waarde $\bar{R}$ bepalen we als het gemiddelde van de gemiddelde grootste dagsommen in de twaalf maanden Jan.–Dec. Dit mogen we doen, *a*) omdat het aantal dagen in een kalendermaand niet veel van 29,53 (de lengte van een synodische maand) afwijkt en *b*) omdat we kunnen aannemen, dat, evenals dit te De Bilt tot nu toe steeds het geval is, ook op de Franse stations

iedere maand minstens één droge dag bevat. Daar de kleinste etmaalsom in een maand dus altijd $= 0$ is, kunnen we de spreidingsbreedte gelijk stellen aan de grootste waarde.

De gegevens van De Bilt, die Dr C. LEVERT ter beschikking stelde, gaven als resultaat voor de periode 1897-1946: $\bar{R} = 15,5$ mm.

Bij afwezigheid van een maaninvloed moet de spreidingsbreedte R_n van $\bar{h}$ gelijk zijn aan $\bar{R}/\sqrt{n}$. Uit fig. 19 is af te lezen, dat $R_n = 0,45$ mm voor Parijs en $R_n = 0,6$ mm voor Nantes. Deze waarden zijn niet significant verschillend van $\bar{R}/\sqrt{900} = 0,52$ mm.

4. In een reeks van toevalsgetallen a_i is het aantal toppen a_n (gedefinieerd door: $a_{n-1} < a_n > a_{n+1}$) gelijk aan een derde van het totale aantal elementen in de reeks (BOTTEMA 1941). Elk van de zes reeksen uit fig. 19 telt dertig termen en resp. 10, 11, 10, 8, 12 en 10 toppen.

5. Als de lunaire invloed inderdaad reëel zou zijn, dan moet het maximum van de kromme voor de waarschijnlijkheid van een droge dag samenvallen met het minimum voor $\bar{h}$. Dit heeft niet plaats bij de curven, die op Nantes betrekking hebben.

Elk van deze vijf argumenten draagt er weinig toe bij om de aanvaardbaarheid van MINEUR's hypothese te vergroten. Integendeel, het is uit het voorgaande duidelijk, dat de resultaten van MINEUR's onderzoek geheel in overeenstemming zijn met de hypothese, die de onafhankelijkheid tussen maanfase en neerslag in Frankrijk postuleert.

MINEUR besluit zijn artikel met de conclusies:

— Il me paraît résulter de cette étude quelques résultats certains:

3a Clermont-Ferrand ne révèle pas d'influence lunaire sur la pluie.

3b Au Parc Saint-Maur (Paris) et à Nantes $\bar{h}$ et $\bar{h}'$ sont maxima le 3^e jour de la Lune et minima le 18^e.

3c En Alger, au contraire, $\bar{h}$ et $\bar{h}'$ sont minima vers le 4^e et le 24^e jour de la Lune et maxima vers le 15^e et le 19^e. —

Deze drie gevolgtrekkingen pleiten eerder *tegen* het bestaan van een algemene maaninvloed dan er *voor*.

Wellicht ten overvloede moge er de aandacht op gevestigd worden, dat de grafische methode als zodanig niet fout is. Hij kan echter aanleiding geven tot het trekken van voorbarige, zo niet onjuiste conclusies. De grafische methode kan voortreffelijk zijn als begin van een wetenschappelijk onderzoek, maar als eindfase van een dergelijk onderzoek is hij ongeschikt. Daartoe moeten significantie-methoden te hulp geroepen worden.

4.4 MAANDELIJKE LUCHTDRIKVERANDERINGEN, IN HET BIJZONDER DIE, VOORAFGAANDE AAN STRENGE WINTERS

Oft expectation fails, and most oft there
Where most it promises.

SHAKESPEARE (All's well that ends well)

Het in 3.1 (p. 114) beschreven duel tussen de Duitse meteorologen RODEWALD en BAUR werd veroorzaakt door een artikel van eerstgenoemde (1948). Daarin vermeldde hij, dat de vijf strenge winters van de laatste 20 jaar voorafgegaan werden door sterke luchtdrukdalingen en -stijgingen boven Groot Brittannië en omliggende gebieden. De dalingen kwamen tot uiting in het zgn. „anomaliever-schil” tussen October en November ($\Delta P_{\text{Nov.}} - \Delta P_{\text{Oct.}} < < 0$), de drukstijgingen in het verschil tussen November en December ($\Delta P_{\text{Dec.}} - \Delta P_{\text{Nov.}} > > 0$).

In 3.4 (p. 145) werd duidelijk gemaakt, waaraan we het verband tussen de wintertemperatuur in Centraal Europa en de luchtdrukstijging van November op December te danken hebben. Nu gaan we ons bezig houden met het vraagstuk, waarom een negatief verschil $\Delta P_{\text{Nov.}} - \Delta P_{\text{Oct.}}$ bij voorkeur gevolgd wordt door een positief verschil $\Delta P_{\text{Dec.}} - \Delta P_{\text{Nov.}}$. Aan de hand van het in 3.4 behandelde is het niet moeilijk in te zien, dat ook in dit onderzoek de pseudocorrelatie een belangrijke rol speelt. Het betreft hier een (b-a) (c-b)-geval met een correlatie-coëfficiënt $= -0,50$. Na een hoge (b-a) $= \Delta P_{\text{Nov.}} - \Delta P_{\text{Oct.}}$ volgt dus vrij vaak een lage (c-b) $= \Delta P_{\text{Dec.}} - \Delta P_{\text{Nov.}}$. Dit is in volledige overeenstemming met wat waargenomen is. Een analoge verklaring heeft ook RODEWALD (1950a) gegeven.

Toch is het voorgaande niet geheel correct, daar stilzwijgend onderlinge onafhankelijkheid tussen de luchtdrukgemiddelden der drie maanden verondersteld is. Geven we met O, N, D aan de afwijkingen (ten opzichte van de normaal) van de luchtdrukgemiddelden in de maanden October, November en December, dan geldt:

$$\begin{aligned} \rho(N-O, D-N) &= \frac{E\{(N-O)(D-N)\}}{E\{(N-O)^2\}^{1/2} E\{(D-N)^2\}^{1/2}} = \frac{E(ND) - E(N^2) - E(OD) + E(ON)}{E(N^2 + O^2 - 2NO)^{1/2} E(D^2 + N^2 - 2DN)^{1/2}} = \\ &= \frac{\rho_1 - 1 - \rho_2 + \rho_1}{(2-2\rho_1)^{1/2} (2-2\rho_1)^{1/2}} = \frac{-1 - \rho_2 + 2\rho_1}{2-2\rho_1} = -\frac{1}{2} + \frac{\rho_1 - \rho_2}{2-2\rho_1} \\ &\quad \rho_1 = \rho(O, N) = \rho(N, D) \quad \rho_2 = \rho(O, D) \end{aligned}$$

Het aan elkaar gelijkstellen van $\rho(O, N)$ en $\rho(N, D)$ veroorzaakt geen grote fout, daar beide niet groter zijn dan ca 0,20. Verder hebben we in bovenstaande afleiding gepostuleerd, dat $\sigma_O = \sigma_N = \sigma_D$. Hiervoor vinden we steun bij RODEWALD (1950a, 363), wat uit het volgende moge blijken.

Zijn figuur 3 bevat de jaarlijkse gang van de gemiddelde absolute „anomalie-verschillen” (Uppsala 1860–1920). Deze zijn als volgt berekend: voor elke maand

wordt het luchtdrukverschil met de vorige maand bepaald als afwijking ten opzichte van zijn normale waarde. Dit verschil wordt absoluut genomen en van de absolute verschillen uit de 60 jaren, die betrekking hebben op dezelfde twee maanden, wordt het gemiddelde berekend. Door dit niet alleen te doen voor bv. $|\Delta P_{\text{Feb.}} - \Delta P_{\text{Jan.}}|$, maar ook voor $|\Delta P_{\text{Mar.}} - \Delta P_{\text{Feb.}}|$ enz. tot $|\Delta P_{\text{Jan.}} - \Delta P_{\text{Dec.}}|$ verkrijgen we tenslotte de jaarlijkse gang van de gemiddelde absolute anomalieverschillen.

We mogen aannemen, dat de anomalieverschillen gaussisch verdeeld zijn. Dit veroorlooft ons, de jaarlijkse gang van de standaarddeviatie van de anomalieverschillen gelijk te stellen aan die van de gemiddelde absolute anomalieverschillen met die restrictie, dat alle waarden in de verhouding $\sqrt{\pi}/2 = 1,24$ groter worden.

Daar verder geldt, dat bv. $\sigma_{N-O}^2 = \sigma_N^2 + \sigma_O^2 - 2 \rho_1 \sigma_N \sigma_O$, geeft RODEWALD's fig. 3 ook het jaarlijkse verloop van de standaarddeviatie van de maandelijksse luchtdrukgemiddelden. Voor de laatste maanden van het jaar maakt RODEWALD's kromme een zeer vlakke indruk, zodat we terecht van de gelijkstelling van σ_O , σ_N en σ_D gebruik mochten maken.

Nemen we $\rho_1 = 0,20$ en $\rho_2 = 0,05$, dan is $\rho(N-O, D-N) = -0,40$. Voor $\rho_1 = 0,10$ en $\rho_2 = 0,05$ wordt $\rho(N-O, D-N) = -0,47$. Betrekken we andere maanden in ons onderzoek, dan zullen ook dergelijke correlatiewaarden optreden, bv. voor $\rho(D-N, J-D)$, $\rho(J-D, F-J)$, enz.

Zouden we alle anomalieverschillen correleren met de daaropvolgende anomalieverschillen, dan mag een waarde van $-0,40$ à $-0,47$ verwacht worden.

RODEWALD (1950a, 362) laat deze correlatie tot uiting komen in zijn fig. 1. Daaronder publiceert hij zijn fig. 2, waarin de toevalscijfers 0-9 de plaats van de maandelijksse luchtdrukgemiddelden te Uppsala innemen.

Inderdaad toont fig. 2 een correlatiecoëfficiënt van omstreeks $-0,50$ geheel in overeenstemming met de theorie, daar bij deze toevalscijfers $\rho_1 = \rho_2 = 0$ is.

RODEWALD's eerste figuur wekt echter de indruk, dat de correlatie daar niet $-0,50$ is, maar meer afzakt naar de waarde $-0,40$ toe. Voor deze opvatting pleit, dat een steilere regressierechte beter bij de waargenomen grootheden aansluit. Verder verklaart een minder stramme correlatie onmiddellijk, waarom de gemiddelde absolute afwijking van de regressierechte in fig. 1 groter is dan in fig. 2. Deze verklaring is ongedwongener dan wat RODEWALD veronderstelt: „man kann darin vielleicht einen Rest Erhaltungstendenz der Druckanomalien erblicken”.

Tot dusverre heeft men nog niet expliciet aangegeven, waarom de markantste anomalieverschillen bij voorkeur in het Oceaangebied in de omgeving van Groot Brittannië en IJsland aangetroffen worden. Toch is dit vrij eenvoudig in te zien.

Uit de bijzonder waardevolle kaarten van KLEIN (1951) zijn standaarddeviaties af te lezen van interdiurne drukveranderingen en van etmaalgemiddelden van de luchtdruk voor het grootste gedeelte van het Noordelijk halfrond en dat niet alleen voor de druk gereduceerd op zeeniveau, maar ook voor de druk op de standaardhoogten 3 km (700 mb), 6 km (500 mb) en 9–10 km (300 mb). Dikwijls zijn ook de waarden afzonderlijk voor zomer en winter aangegeven.

KLEIN's publicatie, die dus de tweede momenten van diverse klimatologische grootheden in kaart brengt, vormt daarmee een belangrijke aanvulling op de zgn. klimatologische kaarten, want deze geven in feite niets anders weer dan de geografische verspreiding van de gemiddelden (de eerste momenten) der klimatologische grootheden.

De nieuwe kaarten tonen, dat de grootste standaarddeviaties zich zomer en winter juist in het gebied tussen Schotland en IJsland bevinden. Weliswaar geldt dit voor de dagelijkse waarden, maar KLEIN heeft ook kaarten getekend, waaruit blijkt, dat de autocorrelatiecoëfficiënten (één dag verschil) over het Noordelijk deel van de Atlantische Oceaan zo goed als constant zijn. Dit veroorlooft ons om het gedrag van de spreiding van de maandgemiddelden gelijk te stellen aan dat van de etmaalgemiddelden.

De opmerking van RODEWALD (1950a, 363), dat na een anomalieverschil $\Delta P_{\text{Aug.}} - \Delta P_{\text{Juli}} = 3$ mm de waarschijnlijkste waarde van $\Delta P_{\text{Sep.}} - \Delta P_{\text{Aug.}}$ niet $-1,5$ mm, maar ca 4 mm is, wordt door een blik op RODEWALD's eerste figuur reeds gelogenstraft. Andere bezwaren zijn vermeld in de Anhang 2 van VAN DER BIJL (1951).

Meer dan een volle bladzijde besteedt RODEWALD daarna aan de beschrijving van meteorologische curiosa – ze zijn het althans in zijn ogen – maar na enige bestudering blijkt, dat alle tot de normale uitingen van de atmosfeer behoren. Zo bv.: in de reeks maandelijks drukafwijkingen te Dorpat (1866–1915) hadden 141 maanden een absolute afwijking groter dan 4 mm. Van de daaropvolgende maanden hadden 134 een kleinere en 5 een grotere afwijking van de normaal. Bij de overige twee bleef de afwijking precies even groot.

RODEWALD ziet geen kans gebruik te maken van deze en dergelijke klimatologische bijzonderheden voor prognostische doeleinden. Hij gooit het nu over een andere boeg, de zuiver statistische bewerking wordt verlaten en een „synoptische” gevolgd:

„Die Positiv- und Negativgebiete in Anomalie-Änderungskarten lassen sich oft von einer Karte zur nächstfolgenden identifizieren, und es lassen sich daraus Bahnen der Zentren bzw. der Ausläufer der Änderungsgebiete zeichnen. Nicht nur die (sich verlagernden) Druckanomalien bzw. Anomalie-Änderungen zeigen eine gewisse Erhaltungsneigung, sondern auch ihre Bahnen. Deshalb lässt sich

mit Anomalie-Änderungskarten formal ähnlich wie mit Karten der 24-stündigen Druckänderung arbeiten". Er is echter een groot verschil: „Bei der Prognose der 24-stündigen Druckänderungen benutzt man die Höhenströmung, die Höhenisobaren-Divergenz, die Luftmassenverteilung und sonstige physikalische Anhaltspunkte. Bei der Prognose der Druckänderung von einem Monat zum nächsten hat man heutigentags noch kaum physikalische Anhaltspunkte; man kann im wesentlichen nur aus der Vorgeschichte heraus extrapolieren. Diese Vorgeschichte liefert zunächst in den ‚Bahnen‘ – wie bei den 24-stündigen Isallobaren – einen Anhaltspunkt für die künftige Verlagerung".

In de twee voorbeelden, die RODEWALD geeft, laten zich wel banen aanbrengen, die in het algemeen oostwaarts gericht zijn, maar het kost even weinig moeite om westwaarts gerichte banen te construeren. Elke kaart toont een afwisselende opvolging van positieve en negatieve gebieden; de daaropvolgende kaart, die dus betrekking heeft op het verschil tussen de twee volgende maanden, biedt ongeveer hetzelfde beeld, alleen hebben de centra hun teken gewisseld. En nu kan men wel zeggen: dit gebied is naar het Oosten getrokken, maar er is evenveel recht, het tegenovergestelde te beweren. Een beroep op de westcirculatie is tevergeefs, daar de centra van de maandelijks luchtdrukafwijkingen van kaart tot kaart nauwelijks te volgen zijn. In de verplaatsing van de anomalieverschillen zal nog minder regelmaat gevonden kunnen worden.

RODEWALD heeft nog een ervaringsregel gevonden: de „Tendenz der Verstärkung polwärts rückender negativer Anomalie-Änderungen". Als we de kaarten van KLEIN (1951) nog eens raadplegen, dan zien we, dat de spreiding van de etmaalgemiddelden van de luchtdruk in het algemeen met de geografische breedte toeneemt. Zo zal het ook wel gesteld zijn met de standaarddeviatie van de maandgemiddelden vanwege de constantheid van de autocorrelatiecoëfficiënten en dit verklaart dan meteen de geciteerde wetmatigheid.

Al te zeker voelt RODEWALD zich niet, getuige dit „zinnetje", dat de grondigheid van onze oosterburen illustreert: „Es wäre notwendig, die Anomaliekarten und Änderungskarten einiger Jahrzehnte für die gesamte Nordhemisphäre zu studieren, um mittlere, häufigste, typische Bahnen und Verlagerungsstrecken, etwaige Vorzugsgebiete für Stationärwerden, Häufungsstellen, Verstärkungs- und Abschwächungsstellen, mögliche Typen der Entwicklung, der Neubildung und des Abbruchs von Anomalien und Anomalie-Änderungen, Beziehungen zu mittleren Höhenwinden und dergleichen festzustellen".

Een discussie van zes prognostische anomalieverschilkaarten beëindigt RODEWALD's artikel. In het begin van een maand werd de dan bekend geworden anomalieverschilkaart geëxtrapoléerd, wat leidde tot een verwachting van de op te

treden anomalieverschillen. Voor de eerste zes maanden van 1949 werden de voorspelde en de waargenomen anomalieverschillen met elkaar gecorreleerd. Het resultaat was: 0,76, 0,80, 0,48, 0,64, 0,93 en 0,78, dus gemiddeld 0,73, alle berekend over ca 80 punten van het gehele gebied van de Noord Atlantische Oceaan en Europa.

Vijf van de zes waarden liggen ver boven de zgn. toevalscoëfficiënt van $+0,50$. Dit laatste getal berust op het volgende: „Moet ik het anomalieverschil van maand N op maand D voorspellen, dan krijg ik al een correlatiecoëfficiënt van $+0,50$ tussen verwachte en opgetreden waarden, wanneer ik de te verwachten waarde $D-N$ gelijk stel aan de helft van de waarde $O-N$ (het negatieve van $N-O$).” Deze bewering, volgens RODEWALD afkomstig van Prof. BARTELS, ziet over het hoofd, dat de factor $\frac{1}{2}$ door een willekeurige vervangen kan worden. Ook al voorspellen we anomalieverschillen van duizenden mb, als de evenredigheidsfactor maar constant gehouden wordt, is de correlatiecoëfficiënt steeds $= +0,50$. Hiermee wordt ons nu een onverwacht resultaat in de schoot geworpen: *de correlatierekening dient verre gehouden te worden van beschouwingen over de kwaliteit van prognostische kaarten.*

De vergelijking van de zes hoge correlaties in 1949 met de toevalscoëfficiënt van $+0,50$ gaat echter mank. De regel van BARTELS-RODEWALD over de laatste coëfficiënt heeft betrekking op gaussisch verdeelde grootheden. De toevalscoëfficiënt is berekend uit waarden voor hetzelfde station, maar voor achtereenvolgende tijdstippen. We kunnen hem de ‘locale’ correlatiecoëfficiënt noemen. Daartegenover staat de ‘geografische’ coëfficiënt. Deze wordt berekend uit de gegevens van diverse punten, maar op hetzelfde tijdstip. De anomalieverschillen zijn echter wel wat betreft de tijd gaussisch verdeeld, maar niet geografisch. Eén blik op een kaart is voldoende om te constateren, dat de gebieden met geringe anomalieverschillen kleiner zijn dan de gaussische verdeling zou voorschrijven. De centra met grote afwijkingen beslaan juist een aanzienlijk gedeelte van de weerkaart.

Deze overwegingen maken het plausibel, dat de geografische coëfficiënt groter zal zijn dan de locale.

4.5 PROEVE VAN REGIONALE NEERSLAGVERWACHTINGEN OP LANGE TERMIJN

Ook dit weet Mnemosyne:

Dat herinn'ren en voorspellen

Naar hun wezen één zijn . . .

VESTDIJK (Mnemosyne in de bergen)

In zijn laatste paragraaf (p. 129) behandelt VISSER (1946) de vraag, of de temperatuur te De Bilt wel representatief is te achten voor de temperatuur in andere

plaatsen en of een neerslagverwachting voor Nederland voor alle delen van het land geldig is.

Correlatieberekeningen leerden, dat het eerste gedeelte van de vraag, betreffende de temperatuur te De Bilt, positief beantwoord kon worden (WOUDENBERG 1946, 15-6). Het jaarlijkse neerslaggemiddelde van Nederland is echter slechts representatief voor een grote middenmoot, maar daarbuiten niet meer. De neerslagverwachting kan alleen gelden voor het midden des lands, in grote trekken het gebied, begrensd door een lijn over Leeuwarden, Leeghwater, Oudensch, Nijmegen, Dalmsholte, Leeuwarden, zo luidde de conclusie.

Mede op grond van dit oordeel werden vijf stations uitgezocht, die aan de volgende drie eisen moesten voldoen.

1. De stations dienden een voldoende lange waarnemingsreeks te bezitten.
2. Ze moesten zo goed mogelijk buiten het genoemde middengebied vallen.
3. Uit landbouwkundig oogpunt moesten ze het meest daarvoor in aanmerking komen.

De vijf stations Leeuwarden, Alkmaar, Leeghwater, Groot Ammers en Heino werden uitgekozen. Het onderzoek werd begonnen met het opstellen van regressievergelijkingen voor de maand Augustus. De grootte van de correlatiecoëfficiënten, die reeds bij het onderzoek naar de neerslag in Nederland berekend waren, werd gebruikt als selectiemiddel ten behoeve van het nieuwe onderzoek.

Het bleek, dat het in het algemeen weinig uitmaakte, of men nu de neerslagcijfers van Nederland, dan wel die van bv. Leeuwarden met een meteorologische grootheid correleerde. Het grootste verschil gaf de correlatie met de temperatuur te De Bilt in September. De correlatiecoëfficiënt van de temperatuur in De Bilt in September met het neerslaggemiddelde van Nederland in Augustus, elf maanden later, was 0,22 en die met de neerslag in Leeuwarden in Augustus, ook elf maanden later, was 0,04, beide berekend voor het basistijdvak 1901-1939.

Bij het correlatieonderzoek van de vier andere stations kwamen geen grotere verschillen voor. Het vermoeden, dat de regressievergelijkingen voor de vijf stations daarom wel weinig van die voor de neerslag van Nederland zullen afwijken, werd niet in alle opzichten bevestigd, zoals uit het onderstaande blijkt.

Regressievergelijkingen voor de Augustus-neerslag in

$$\begin{aligned} \text{Nederland: } x_0 = & 6,67 x_9 \text{ V} + 3,00 x_9 \text{ VIII}_{-1} + 4,96 x_2 \text{ III}_{-1} + 0,873 x_{11} - \\ & - 3,86 x_6 \text{ IV} - 8,59 x_2 \text{ IX}_{-1} \quad R = 0,774 \quad m = 6 \end{aligned}$$

$$\begin{aligned} \text{Leeuwarden: } x_0 = & 7,04 x_9 \text{ V} - 2,92 x_3 \text{ III} - 3,33 x_7 \text{ I} - 8,78 x_3 \text{ VIII}_{-1} + \\ & + 2,67 x_4 \text{ IV} + 0,596 x_{10} \text{ VIII}_{-1} - 8,39 x_2 \text{ IX}_{-1} \quad R = 0,870 \quad m = 7 \end{aligned}$$

$$\text{Alkmaar: } x_0 = 7,37 x_7 \text{ VII}_{-1} - 0,691 x_{10} \text{ VII}_{-1} + 4,89 x_9 \text{ XI} - 8,58 x_3 \text{ V} + \\ + 3,07 x_4 \text{ IV} - 0,127 x_1 \text{ VIII}_{-1} - 0,062 x_1 \text{ III} \quad R = 0,827 \quad m = 7$$

$$\text{Leeghwater: } x_0 = 10,83 x_7 \text{ VII}_{-1} - 2,65 x_5 \text{ V} - 0,226 x_1 \text{ VIII}_{-1} - 0,694 x_{10} \text{ VII}_{-1} - \\ - 8,18 x_2 \text{ IX}_{-1} \quad R = 0,762 \quad m = 5$$

$$\text{Groot Ammers: } x_0 = 6,55 x_9 \text{ V} + 10,08 x_6 \text{ VI} - 5,12 x_3 \text{ V} + 3,59 x_2 \text{ III}_{-1} - \\ - 8,80 x_3 \text{ VII}_{-1} \quad R = 0,708 \quad m = 5$$

$$\text{Heino: } x_0 = -3,17 x_5 \text{ VI} + 6,29 x_7 \text{ VII}_{-1} + 6,66 x_2 \text{ XI} - 2,16 x_9 \text{ III} - 1,88 x_5 \text{ V} \\ R = 0,759 \quad m = 5$$

De notatie, identiek aan die van VISSER (1946), is:

- x_0 = uit de regressievergelijking berekende neerslag in Augustus
- x_1 = neerslag te Paramaribo
- x_2 = temperatuur te Ivigtut
- x_3 = druk te Ponta Delgada
- x_4 = druk te Reykjavik
- x_5 = drukverschil Ponta Delgada—Reykjavik
- x_6 = temperatuur te De Bilt
- x_7 = druk te Berlijn
- x_9 = drukverschil Berlijn—Helsinki
- x_{10} = zonnevlekkenvariatie
- x_{11} = periode van $2\frac{1}{4}$ jaar

Elke x stelt de afwijking voor ten opzichte van het gemiddelde over de jaren 1901–39.

Aan iedere term van de vergelijkingen is toegevoegd de maand in Romeinse cijfers. De index -1 geeft aan, dat het tijdsverschil meer dan elf maanden bedraagt.

De beschouwingen van 4.2 geven de verklaring, waarom x_{11} , de periode van $2\frac{1}{4}$ jaar, alleen in de vergelijking van Nederland voorkomt. De periodegetallen berusten op de regencijfers van Nederland en geven daarom wel een goede correlatie met de neerslaghoeveelheden van Nederland, maar een minder goede met die van de vijf regionale stations.

Alleen voor het station Leeuwarden werden de gegevens voor de jaren 1901–39 in de regressievergelijking gesubstitueerd. Wanneer men de uitkomsten naast de waargenomen neerslaghoeveelheden stelt, blijkt de overeenkomst heel goed te zijn. Slechts zes maal werd het teken onjuist voorspeld in de 39 jaren van het

basistijdvak. Deze mislukkingen vonden meestal plaats bij geringe afwijkingen van de normaal. De extremen werden alle goed berekend en dit is juist wat men als het eigenlijke doel van een verwachting op lange termijn mag beschouwen.

Geheel onverwacht zijn deze resultaten niet, vooral als we nader ingaan op de slotconclusies van 1.11.4. De regressievergelijkingen zijn zodanig opgebouwd uit de klimatologische gegevens, dat de berekende hoeveelheden optimaal aangepast zijn aan de waargenomen hoeveelheden. Verder geldt de formule $R^2 = \Sigma x_B^2 / \Sigma x_W^2$ (x_B = berekende, x_W = waargenomen neerslaghoeveelheid). De in alle vergelijkingen hoge waarde van R maakt het plausibel, dat x_B in het algemeen slechts weinig van x_W zal verschillen. Speciaal geldt dit voor de extremen: bij een grote $|x_W|$ behoort een grote $|x_B|$, anders zou R de hoge waarde niet kunnen bereiken. Voorts volgt uit de formule voor R^2 , dat $|x_B| < |x_W|$, immers $R < 1$ en zowel x_B als x_W volgen bij benadering een gaussverdeling, zodat uit $\Sigma x_B^2 < \Sigma x_W^2$ volgt $\Sigma |x_B| < \Sigma |x_W|$. Inderdaad bleek $|x_B| = 26,5$ mm en $|x_W| = 30,9$ mm te zijn.

Hun quotiënt zou een waarde = 0,858 geven voor R . In overeenstemming daarmee resulteerde uit de correlatieberekeningen $R = 0,870$ en uit het quotiënt van de tweede momenten van x_B en x_W de waarde $R = 0,876$.

De ware toetsing van het regressieonderzoek ligt in het vergelijken van berekening en waarneming buiten het basistijdvak, het tijdvak dus, waarvan de gegevens gediend hebben om de regressievergelijkingen op te stellen. Deze vrij simpele waarheid, o.a. door VISSER (1946, 119) naar voren gebracht, wordt vaak over het hoofd gezien. Als voorbeeld kunnen we KHA (1945) citeren.

Deze leidde regressievergelijkingen af voor de moessonregenval in drie kustgebieden van Burma. Zijn basistijdvak was 1891–1930. De multiële correlatiecoëfficiënten zijn resp. 0,80, 0,80 en 0,87. In drie figuren geeft KHA de berekende en waargenomen neerslaghoeveelheden voor de veertig moessons van het basistijdvak. De overeenkomst laat natuurlijk weinig te wensen over, maar geen enkele waarneming uit de jaren 1930–45 wordt gebruikt om de regressie onafhankelijk te toetsen.

Tabel 23 levert de gegevens, waarmee de toetsing der regressievergelijkingen kan plaats vinden.

We zien al dadelijk, dat bij vier van de vijf stations het teken van de afwijking in de meerderheid juist was voorspeld (Lwd 10 goed, 3 fout; A 7 goed, 6 fout; Lwt 7 goed, 6 fout; GA 9 goed, 4 fout, waartegenover slechts Heino met 4 goed, 9 fout).

Het totale aantal juiste voorspellingen (althans wat het teken betreft) is 37, het aantal foutieve is 28.

Tabel 23

	Leeuwarden		Alkmaar		Leeghwater		Groot Ammers		Heino	
	x_B	x_W	x_B	x_W	x_B	x_W	x_B	x_W	x_B	x_W
1940	-26	15	22	-16	-4	-18	-5	-30	-12	17
1941	16	86	3	32	-39	81	11	106	-3	90
1942	35	6	-1	-14	23	-23	30	-1	17	-9
1943	30	24	-5	27	-21	35	7	28	-0,2	28
1944	44	-53	-10	-43	-8	-37	-8	-69	1	-59
1945	16	10	-1	32	5	18	52	-14	-45	94
1946	-65	-12	8	12	21	63	-18	-3	2	7
1947	-43	-26	-60	-26	-19	-27	9	-36	29	-39
1948	38	22	5	10	19	53	14	-13	2	39
1949	-21	-16	20	-20	24	-40	-17	-47	6	-20
1950	-3	31	39	-75	-7	35	40	45	8	40
1951	30	27	-2	32	5	36	22	62	38	7
1952	31	40	50	6	56	-10	25	5	30	-15
Σ pos. afw.	240	261	108	226	153	321	210	246	133	322
Σ neg. afw.	-158	-107	-118	-119	-98	-155	-48	-213	-60	-142
$\frac{1}{13} \Sigma x_B $	30,6		17,4		19,3		19,8		14,8	
$\frac{1}{13} \Sigma x_W $		28,3		26,5		36,6		35,3		35,7

Op dezelfde wijze als in het slot van 2.3.2 bepalen we de significantie van de verhouding 37 : 28. Daar $p = q = \frac{1}{2}$ is, is de relatieve afwijking $T = \frac{37-28}{\sqrt{37+28}} = \frac{9}{\sqrt{65}} = 1,12$. Deze heeft een dubbelzijdige overschrijdingskans $P = 26\%$. We zijn dus nog niet gerechtigd, altijd betere dan toevalsresultaten te verwachten.

Dat de maandverwachtingen inderdaad niet zo fraai zijn, blijkt, als we de grootte van de verschillen $x_B - x_W$ in ogenschouw nemen. Het gemiddelde der absolute verschillen is voor Leeuwarden in de 39 basisjaren 19,1 mm. Het gemiddelde der absolute verschillen $x_B - x_W$ in de dertien toetsjaren is voor de vijf stations achtereenvolgens 31,5, 34,9, 43,5, 37,0 en 47,6 mm. Het gemiddelde over de vijf stations is 38,9 mm.

Een goede vergelijkingsmaatstaf voor dit getal vinden we in de uitkomsten van een verwachting, die met opzet geheel onbetrouwbaar gemaakt is. Dit bereiken we, indien we de verwachting voor het jaar n (x_B^n) geldig verklaren voor een ander jaar. We deden dit voor het volgende en het voorafgaande jaar. Door een cyclische verwisseling toe te passen wisten we moeilijkheden aan begin en eind van het basistijdvak te vermijden. Het bleek, dat

$$\frac{1}{39} \Sigma |x_B^n - x_W^{n-1}| = 39,9 \text{ mm en } \frac{1}{39} \Sigma |x_B^n - x_W^{n+1}| = 38,8 \text{ mm was.}$$

Een waarde van 39 mm mogen we dus als limiet stellen bij de neerslagverwachtingen voor de maand Augustus. Elke prognose, die even grote of nog grotere afwijkingen $|x_B - x_W|$ oplevert, is onbruikbaar.

Nu kunnen we zelfs aantonen, dat een „Lazy Man”, die steeds de normale regenhoeveelheid voorspelt, betere resultaten weet te verkrijgen. Luidt de verwachting immers stevast $x_B = 0$, dan is $|x_B - x_W| = |x_W|$. Voor de jaren 1940—52 zijn de gemiddelden van $|x_W|$ in tabel 23 opgenomen. Voor elk station is $\frac{1}{13} \sum |x_W| < \frac{1}{13} \sum |x_B - x_W|$. Voor het algemeen gemiddelde geldt dit natuurlijk ook, de getallen zijn 32,5, resp. 38,9 mm.

We moeten onze eisen aan een betrouwbare maandverwachting derhalve hoger stellen. De gemiddelde fout $|x_B - x_W|$ dient beneden de klimatologische grens van 32 mm te liggen.

Dat de verwachting $x_B = 0$ de beste resultaten geeft, moge op het eerste gezicht verbazen — bij nadere beschouwing blijkt er zelfs een bewijs mogelijk te zijn.

Gaan we uit van gaussisch verdeelde weersfactoren, dan zal bij een prognose $x_B = 0$ het gemiddelde der absolute fouten $|x_B - x_W|$ gelijk zijn aan $0,7979 \sigma_W$ (σ_W = standaarddeviatie der waargenomen neerslaghoeveelheden).

Deze gemiddelde fout zal groter zijn, wanneer we regressievergelijkingen gebruiken en de fout neemt met de grootte van de multiële correlatiecoëfficiënt toe! De paradox wordt eerst bewezen en daarna verklaard.

De verwachte waarden x_B zijn eveneens gaussisch verdeeld, daar ze opgebouwd zijn uit 5 à 7 gaussisch verdeelde meteorologische grootheden.

De standaarddeviatie σ_B van x_B hangt met σ_W en de multiële correlatiecoëfficiënt R samen volgens $R = \sigma_B / \sigma_W$.

De te verwachten absolute fout voor een zekere berekende waarde x_B is

$$\int_{-\infty}^{\infty} \frac{1}{\sigma_W \sqrt{2\pi}} |x_B - x_W| e^{-\frac{x_W^2}{2\sigma_W^2}} dx_W =$$

$$= \frac{1}{\sigma_W} \sqrt{\frac{2}{\pi}} \left(\sigma_W^2 e^{-\frac{x_B^2}{2\sigma_W^2}} + x_B \int_0^{x_B} e^{-\frac{x_W^2}{2\sigma_W^2}} dx_W \right)$$

Tenslotte laten we nog x_B variëren en vinden dan voor de te verwachten gemiddelde afwijking $|x_B - x_W|$ de uitdrukking:

$$E(|x_B - x_W|) = \int_{-\infty}^{\infty} \frac{1}{\sigma_B \sqrt{2\pi}} e^{-\frac{x_B^2}{2\sigma_B^2}} dx_B \frac{1}{\sigma_W} \sqrt{\frac{2}{\pi}} \left(\sigma_W^2 e^{-\frac{x_B^2}{2\sigma_W^2}} + x_B \int_0^{x_B} e^{-\frac{x_W^2}{2\sigma_W^2}} dx_W \right) =$$

$$= \sigma_W \sqrt{\frac{2R^2 + 2}{\pi}} = 0,7979 \sigma_W \sqrt{R^2 + 1}$$

Voor $R = 0$ is $\sigma_B = 0$ of $x_B \equiv 0$ en $E(|x_B - x_W|) = E(|x_W|) = 0,7979 \sigma_W$.

Dit bevestigt het resultaat, verkregen voor de gemiddelde absolute fout bij het voortdurend voorspellen van de normale waarde.

Voor $R = 1$ is $\sigma_B = \sigma_W$. Dan wordt het gemiddelde absolute verschil $|x_B - x_W| = \sqrt{2}$ maal het gemiddelde van $|x_B|$ of $|x_W|$. Dit is eveneens in overeenstemming met de theorie.

Het eindresultaat geeft duidelijk aan, dat de te verwachten discrepantie tussen waargenomen en uit de regressievergelijking berekende waarde groter is, naarmate de multiële correlatiecoëfficiënt dichter bij 1 ligt.

Deze schijnbare tegenstelling is reeds bij het begin van onze afleiding naar binnen gesmokkeld. We hebben nl. de afwijking $|x_B - x_W|$ vermenigvuldigd met de frequentieverdeling van de waargenomen waarden x_W . Dit impliceert een volkomen onafhankelijkheid tussen waargenomen en berekende waarden, dus o.a. een multiële correlatiecoëfficiënt = 0. R geeft in deze berekening alleen maar de grootte van het quotiënt van de twee standaarddeviaties aan.

Er dient terdege rekening gehouden te worden met de dubbele betekenis van R . Voorts moet onderscheid gemaakt worden tussen de resultaten, verkregen in het basistijdvak en die daarbuiten.

We hebben de regressievergelijking en zijn multiële correlatiecoëfficiënt R berekend op grond van de waarnemingen uit het basistijdvak. Deze gevonden steekproefwaarde van R kan dicht bij zijn universumwaarde liggen. In dat geval

blijft de formule $R = \frac{\sum x_B x_W}{N \sigma_B \sigma_W}$ ook buiten de basisperiode geldig, want R en $\frac{1}{N} \sum x_B x_W$ blijven constant.

De universumwaarde van R kan echter ook = 0 zijn. Weliswaar behoudt de

formule $R = \frac{\sum x_B x_W}{N \sigma_B \sigma_W}$ buiten de basisperiode zijn eigenlijke geldigheid, maar

de waarden van R en $\frac{1}{N} \sum x_B x_W$ vertonen op de grenzen van het basistijdvak

discontinuïteiten. Bij een tengevolge van steekproef- en selectieeffect ontstane hoge steekproefwaarde van R in het basistijdvak is er grote overeenkomst tussen x_B en x_W . Aan de eis $R = \frac{\sum x_B x_W}{N \sigma_B \sigma_W}$ wordt dan ook voldaan. Buiten het basistijdvak is er helemaal geen overeenstemming tussen x_B en x_W ($\sum x_B x_W = 0$) en liggen de steekproefwaarden van R vlak bij 0. Derhalve geldt ook nu weer

$$R = \frac{\sum x_B x_W}{N \sigma_B \sigma_W}$$

De formule $R = \frac{\sigma_B}{\sigma_W}$ gedraagt zich anders dan de zojuist behandelde. Voor het eerste geval, waarin de hoge steekproefwaarde voor R in het basistijdvak weinig afwijkt van de universumwaarde, gaat de formule steeds op.

In het tweede geval (hoge steekproefwaarde voor R in het basistijdvak, universumwaarde echter = 0) is de formule in de basisperiode juist. Daarbuiten is $R = 0$, maar σ_B/σ_W houdt zijn hoge waarde, omdat σ_B bepaald wordt door de regressiecoëfficiënten en deze blijven dezelfde.

De dertien toetsjaren leveren ons de schattingen $\sum |x_B| / \sum |x_W|$ voor σ_B/σ_W en wel Leeuwarden 1,082 (R uit basistijdvak = 0,870), Alkmaar 0,655 ($R = 0,827$), Leeghwater 0,527 ($R = 0,762$), Groot Ammers 0,562 ($R = 0,708$) en Heino 0,416 ($R = 0,759$). De z -toets leert, dat geen enkele schatting significant afwijkt van de bijbehorende waarde R .

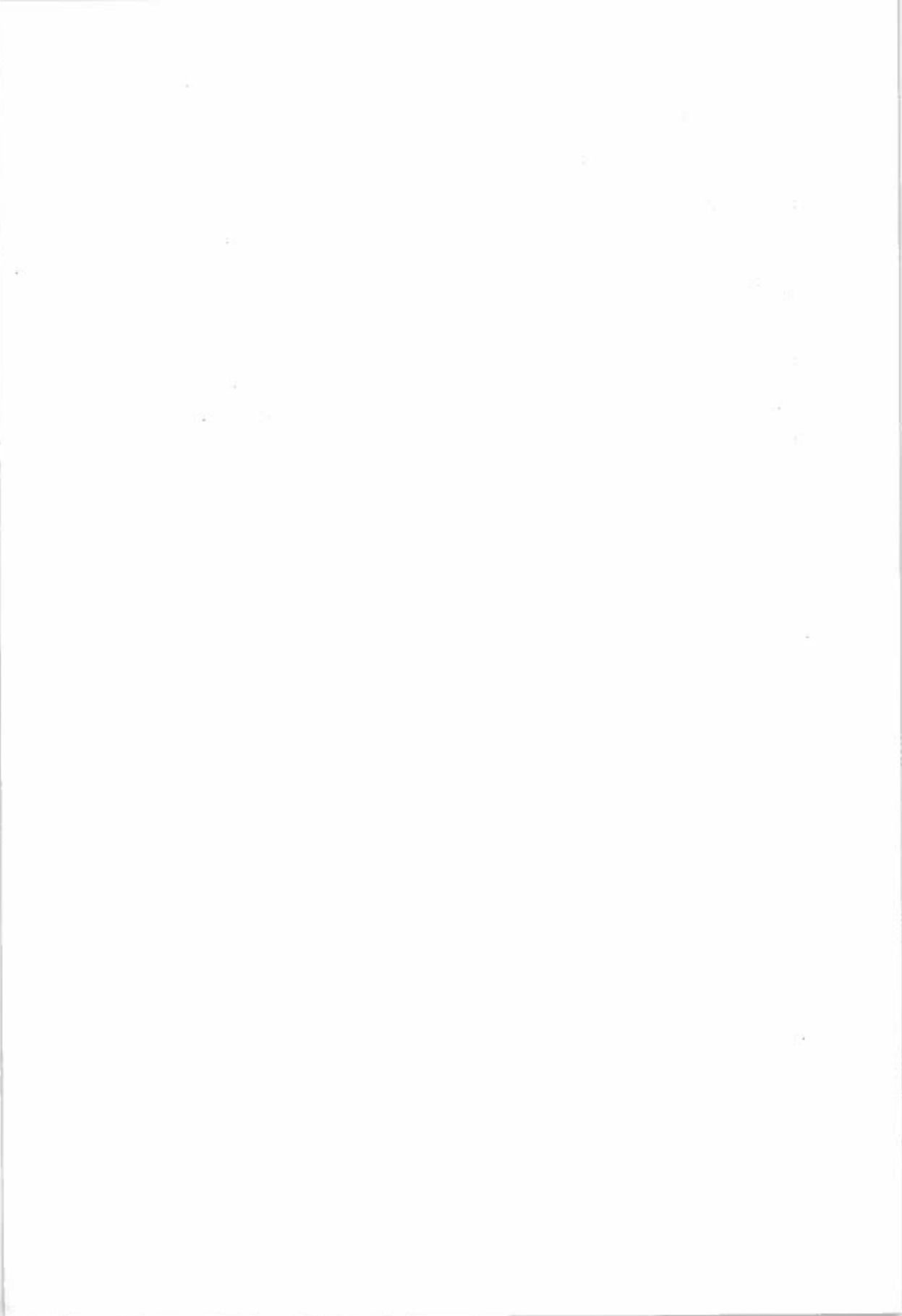
De voor gaussische verdelingen afgeleide gemiddelde absolute afwijkingen in x_W en $x_B - x_W$ zijn nog niet vergeleken met de experimentele resultaten voor de toetsperiode 1940-52.

Uit de gegevens van de basisjaren bepalen we als schattingen voor σ_W : Leeuwarden 36,2 mm, Alkmaar 39,1 mm, Leeghwater 44,4 mm, Groot Ammers 39,1 mm en Heino 34,5 mm. Nederland met 32,6 mm (VISSE 1946, 117) blijft beneden deze cijfers, wat begrijpelijk is. Vermenigvuldiging met de factor 0,7979 moet een schatting voor $E(|x_W|)$ geven. We vinden achtereenvolgens 28,9 mm (L 28,3), 31,2 mm (A 26,5), 35,4 mm (L 36,6), 31,2 mm (GA 35,3) en 27,5 mm (H 35,7). Tussen haakjes zijn geplaatst de aan tabel 23 ontleende waarden van $\frac{1}{13} \sum |x_W|$.

Een verdere vermenigvuldiging met de van station tot station verschillende factor $\sqrt{R^2 + 1}$ levert ons schattingen voor $E(|x_B - x_W|)$ en wel 38,3 mm (L 31,5), 41,0 mm (A 34,9), 39,8 mm (L 43,5), 38,7 mm (GA 37,0) en 39,7 mm

(H 47,6). De tussen haakjes geplaatste waarden zijn de reeds eerder vermelde experimentele resultaten $\frac{1}{13} \sum |x_B - x_W|$. De overeenkomst, die ook hier weer bevredigend is, verleent een aanzienlijke steun aan de hypothese, die aan de paradox ten grondslag lag, nl. de onafhankelijkheid tussen waargenomen en berekende regenhoeveelheden.

Nemen we enerzijds in ogenschouw, dat een regionale neerslagverwachting op lange termijn langs de lijnen van de regressiemethode hiermede onmogelijk is gebleken en anderzijds Mnemosyne's ervaring, dat „herinn'ren en voorspellen naar hun wezen één zijn”, dan kunnen we uit de resultaten van dit onderzoek de min of meer gedurfde conclusie trekken, dat het herinneringsvermogen van de atmosfeer maar zeer infantiel en rudimentair ontwikkeld is.



LITERATUUR

Het getal voor de naam van een tijdschrift geeft de jaargang aan, de getallen daarna: begin- en eindpagina van het artikel.

Afkortingen: N = nummer

- AdM Annalen der Meteorologie
 AHMM Annalen der Hydrographie und maritimen Klimatologie
 AMGB Archiv für Meteorologie, Geophysik und Bioklimatologie
 BAMS Bulletin of the American Meteorological Society
 BDW Berichte des deutschen Wetterdienstes in der U.S. Zone
 JASA Journal of the American Statistical Association
 JRSS Journal of the Royal Statistical Society
 MR Meteorologische Rundschau
 MZ Meteorologische Zeitschrift
 QJ Quarterly Journal of the Royal Meteorological Society
 RMA Rivista di Meteorologia Aeronautica
 TAGU Transactions of the American Geophysical Union

- A. C. AITKEN 1947 Statistical Mathematics
 5th ed. Oliver and Boyd London
 M. BAMPAC Sulla correlazione fra nuvolosità e tendenza barometrica
 L. MELIS 1949 9 N2 RMA 46-7
 J. BARTELS 1922 Neue Methoden zur Berechnung und Darstellung der täglichen
 Luftdruckschwankung bei starken unperiodischen Störungen
 Diss. Göttingen, gedruckt in BARTELS 1923
 J. BARTELS 1923 Zur Berechnung der täglichen Luftdruckschwankung
 51 AHMM 153-60
 J. BARTELS 1925 Besprechung von WAGNER 1925
 42 MZ 451-3
 J. BARTELS 1930 Geophysikalischer Nachweis von Veränderungen der Sonnen-
 strahlung
 9 Ergebnisse exakter Naturw. 38-79
 J. BARTELS 1938 Sonnen- und mondentägige Luftdruckschwankungen
 HANN-SÜRING (5de ed.) 276-306
 J. BARTELS 1943 Gesetz und Zufall in der Geophysik
 31 Naturwissenschaften 421-35
 J. BARTELS 1947 Mathematische Methoden der Geophysik
 7 Fiat Review 89-99

- J. BARTELS 1948 Anschauliches über den statistischen Hintergrund der sogenannten Singularitäten im Jahresgang der Witterung
1 AdM 106-27
- J. BARTELS 1948a Zufallszahlen für statistische Versuche
1 AdM 209-16
- J. BARTELS 1949 Null kann vier bedeuten
2 AdM 44-8
- L. A. BAUER and
C. R. DUVAL 1925 Studies concerning the relation between the activity of the sun and of the earth's magnetism
30 Terrestrial Magnetism 191-213
- F. BAUR 1929 Zur Theorie der linearen Mehrfachkorrelation
9 Zeitschrift für angewandte Math. und Mechanik 231-41
- F. BAUR 1931 Über das Abhängigkeitsgesetz stochastisch verbundener Veränderlichen
48 MZ 346-9
- F. BAUR 1948 Einführung in die Grosswetterkunde
Dieterich'sche Verlagsbuchhandlung Wiesbaden
- F. BAUR 1949 Zurückführung des Grosswetters auf solare Erscheinungen
A1 AMGB 358-74
- F. BAUR 1949a Die Grosswetterlagen Mitteleuropas
Juli 1949
- F. BAUR 1950 Über falsche und richtige Statistik in der Meteorologie
3 AdM 74-83
- F. BAUR 1950a Die Erscheinungen des Grosswetters
8. Teil, HANN-SÜRING, Lehrbuch der Meteorologie,
5. Auflage, S. 903-76
- W. J. B. BEVERIDGE 1950 The art of scientific investigation
William Heinemann London
- M. BIDER 1948 Die ungewöhnliche Wärmeperiode 1947 statistisch untersucht
4 N4 Experientia 1-7
- R. BILANCINI 1948 Sulla distribuzione delle temperature intorno alla normale
8 NI RMA 13-29
- W. BLEEKER 1936 De gemiddelde hoogtewind boven De Bilt volgens loodsballonwaarnemingen (1920-1931)
Med. en Verh. 38
- W. BLEEKER en
H. C. BUVOET 1949 De klimatologische betekenis van de begrippen luchtsoort en luchtmassa
56 Tijdschrift Kon. Ned. Aardrijkskundig Genootschap 162-87
- G. BODMANN 1910 Schwedische Südpolar Expedition 1901-1903
Meteorologie Band 2 4. Lieferung

- S. T. BOK 1946 De gedachtengang van de statistica
Stenfert Croese Leiden
- O. BOTTEMA 1941 Toevalsperioden
39 Hemel en Dampkring 303-7
- A. BRAVAIS 1846 Analyse mathématique sur les probabilités des erreurs de
situation d'un point
9 Mémoires présentés par divers savants à l'Académie Royale
des Sciences de l'Institut de France 255-332
- C. E. P. BROOKS 1923 Variations in the levels of the Central African Lakes Victoria
and Albert
N20 Geophysical Memoirs Met. Office London
- C. E. P. BROOKS and
N. CARRUTHERS 1944 Statistical methods in climatology
MOM 467 Met. Office London
- K. A. BROWNLEE 1947 Industrial experimentation
2nd ed. H.M.S.O.
- C. H. D. BUYS BALLOT 1882 Beredeneerd register op de werken van het K.N.M.I. tot 1882
Kemink & Zn Utrecht
- W. v. D. BIJL 1948 Changes in the climate of London
3 Weather 369
- W. v. D. BIJL 1949 Hoe wordt de komende winter?
47 Hemel en Dampkring 215-9
- W. v. D. BIJL 1950 Bemerkungen zu meteorologisch-statistischen Problemen
3 AdM 163-6
- W. v. D. BIJL 1951 Fünf Fehlerquellen in wissenschaftlicher statistischer Forschung
4 AdM 183-212
- W. v. D. BIJL 1951a Diskussionsbemerkung zu KOPPE 1951
4 AdM 94
- W. v. D. BIJL 1951b Methodical considerations on the statistical operations con-
cerning climatic variations
B2 AMGB 468-85
- W. v. D. BIJL 1951c Steekproeven, die meer opleveren dan steekproefwaarden
3 Mededelingen Medisch-Biologische Sectie, V.v.S. 21-3
- N. CARRUTHERS 1945 The optimum period for a British rainfall normal
71 QJ 144-50
- S. CHAPMAN AND
M. AUSTIN 1934 The lunar atmospheric tide at Buenos Ayres, 1891-1910
60 QJ 23-8
- S. CHAPMAN 1936 The lunar atmospheric tide at Glasgow
56 Proc. Royal Soc. Edinburgh 1-5
- S. CHAPMAN and
J. BARTELS 1940 Geomagnetism
Oxford Univ. Press

- H. H. CLAYTON 1923 Hoofdstuk 11: "The sunspot period as a result of planetary influences" in HUNTINGTON 1923
- W. G. COCHRAN 1947 Some consequences when the assumptions for the analysis of variance are not satisfied
3 Biometrics 22-38
- W. G. COCHRAN and G. M. COX 1950 Experimental Designs
John Wiley and Sons New York
- V. CONRAD 1942 Fundamentals of physical climatology
Harvard Univ. Press
- V. CONRAD 1944 Methods in Climatology
1st ed. Harvard Univ. Press
- V. CONRAD 1950 Zie CONRAD and POLLAK 1950
- V. CONRAD and L. W. POLLAK 1950 Methods in Climatology
2nd ed. Harvard Univ. Press
- A. F. CROSSLEY 1948 Changes in the climate of London
3 Weather 283
- L. B. C. CUNNINGHAM and W. R. B. HYND 1946 Random processes in problems of air warfare
8 Supplement JRSS 62-85
- M. CURRY 1946 Bioklimatik
American Bioclimatic Research Institute Riederau (Dld)
- H. E. DANIELS 1946 Discussion of CUNNINGHAM and HYND 1946
8 Supplement JRSS 87
- F. N. DAVID 1938 Tables of the correlation coefficient
Cambridge Univ. Press
- R. H. DAW 1948 Changes in the climate of London
3 Weather 370-1
- W. DERHAM 1714-6 An account of the Rain which fell every Year at Upminster in Essex, the last Eighteen Years, with Remarks upon that of the Year 1714. Also a Comparison of what has been observed of that kind at Paris, by M. de la Hire
29 Phil. Trans. London 130-2
- R. DESCARTES 1637 Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences
- E. L. DODD 1941 The problem of assigning a length to the cycle to be found in a simple moving average and in a double moving average of chance data
9 Econometrica 25-37
- F. Y. EDGEWORTH 1892 Correlated averages
5th series, 34 Phil. Mag. 190-204

- T. A. ELKINS 1952 The effect of random errors in gravity data on second derivative values
17 *Geophysics* 70-88
- R. A. FISHER 1915 Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population
10 *Biometrika* 507-21
- R. A. FISHER 1921 On the 'probable error' of a coefficient of correlation deduced from a small sample
1 N4 *Metron* 3-32
- R. A. FISHER 1928 The general sampling distribution of the multiple correlation coefficient
A121 *Proc. Royal Soc. London* 654-73
- R. A. FISHER 1928a Correlation coefficients in meteorology
121 *Nature* 712
- R. A. FISHER 1948 Statistical methods for research workers
10th ed. Oliver and Boyd London
- R. A. FISHER and F. YATES 1949 Statistical Tables for biological, agricultural and medical research
3rd ed. Oliver and Boyd London
- H. FLOHN 1941 Über Begriff und Wesen der Singularitäten der Witterung
58 *MZ* 229-33
- H. FLOHN 1948 Diskussionsbemerkung zu J. BARTELS
1 *AdM* 252-3
- F. GALTON 1888 Co-relations and their measurement, chiefly from anthropometric data
45 *Proc. Royal Soc. London* 135-45
- W. U. GARSTKA 1949 Interpretation of snow surveys
30 *TAGU* 412-30
- W. U. GARSTKA 1950 Discussion of GARSTKA 1949
31 *TAGU* 937-9
- E. L. GRANT 1940 The probability viewpoint in hydrology
21 *TAGU* 7-13
- L. R. HAFSTAD 1940 On the BARTELS technique for time-series analysis, and its relation to the analysis of variance
35 *JASA* 347-61
- H. C. HAMAKER 1948 Foutentheorie en wiskundige statistiek
2 *Statistica* 177-205
- H. C. HAMAKER 1948a Toevals-cijfers
2 *Statistica* 97-106 + 129-37
- G. HAMMING 1952 Interactie-problemen
64 *Landbouwkundig Tijdschrift* 162-73

- J. VON HANN und
R. SÜRING 1938 Lehrbuch der Meteorologie
5. Auflage Willibald Keller Leipzig
- J. HEMELRIJK 1951 KENDALL's rangcorrelatiecoëfficiënt
Rapport S59 Math. Centrum Amsterdam
- A. HOFMANN 1947 Atmosphärische Zirkulation über Mitteleuropa und Sonnen-
flecken
1 MR 161-2
- A. HOFMANN 1951 Zur Methodik der langfristigen Witterungsvorhersage
4 AdM 160-3
- A. HOFMANN 1951a Die Grosswetterlagen Mitteleuropas Mai 1951
- A. HOFMANN 1952 Die Grosswetterlagen Mitteleuropas Juni 1952
- H. HOTELLING 1937 Recensie van R. A. FISHER: "Statistical methods for research
workers" (6th ed.)
32 JASA 218-9
- E. HUNTINGTON 1923 Earth and Sun. An hypothesis of weather and sunspots
Yale Univ. Press New Haven
- W. HUSSLEIN 1949 Gibt es eine barische Vorbereitung mitteleuropäischer Winter?
2 AdM 36-44
- M. O. JOHNSON 1946 Correlation of cycles in weather, solar activity, geomagnetic
values and planetary configurations
Phillips and van Orden Cy San Francisco
- M. G. KENDALL 1947 The advanced theory of statistics I (3rd ed.) and II (1946)
Charles Griffin and Co London
- M. G. KENDALL and
B. BABINGTON SMITH 1939 Tables of Random Sampling Numbers
Tracts for Computers N24
Cambridge Univ. Press
- M. M. KHA 1945 Forecasting the coastal rainfall of Burma
71 QJ 114-25
- W. H. KLEIN 1951 A hemispheric study of daily pressure variability at sea level
and aloft
8 Journal of Meteorology 332-46
- E. KLEINSCHMIDT 1932 Zur Anwendung der Statistik in der Meteorologie
49 Wetter 353-8
- S. KOLLER 1943 Graphische Tafeln zur Beurteilung statistischer Zahlen
2. Aufl. Th. Steinkopff Dresden u. Leipzig
- H. KOPPE 1951 Sonnenaktivität, Grosswetter und wetterbezogene Reaktionen
4 AdM 87-94 Diskussionsbemerkungen dazu
4 AdM 94-6

- W. KÖPPEN und
R. GEIGER 1936 Handbuch der Klimatologie
Bd. 1B (V. CONRAD) S. 92-8
- A. LABRIJN 1945 Het klimaat van Nederland gedurende de laatste twee en een
halve eeuw
Med. en Verh. 49
- J. VON LAMONT 1868 Monatliche und jährliche Resultate der an der königlichen
Sternwarte bei München von 1857 bis 1866 angestellten mete-
orologischen Beobachtungen
VI. Supplementband zu den Annalen der Münchener Stern-
warte S. VII—XI
- H. LEVENE 1952 On the power function of tests of randomness based on runs
up and down
23 Annals Math. Statistics 34-56
- C. LEVERT 1948 Een statistisch probleem in de klimatologie
2 Statistica 107-19
- L. LYGGAARD 1949 Recent climatic fluctuations
5 Folia Geograph. Dan. Copenhagen
- P. C. MAHALANOBIS 1938 Biography of Prof. R. A. FISHER
4 Sankhya 265-72
- E. J. MAREY 1881 De grafische methode. Vertaald door F. M. Jaeger
Henri J. Stenberg 's-Hage
- O. MEISSNER 1943 Hohe und doch irreführende Korrelationsfaktoren
60 MZ 359-60
- H. MINEUR 1951 Recherche d'une influence possible de la Lune sur les précipi-
tations atmosphériques
233 Comptes Rendus (Paris) 1426-8
- M. J. MORONEY 1951 Facts from Figures
Pelican book A 236
Penguin Books Harmondsworth
- C. A. G. NASS 1948 De grondslagen der regressieanalyse met een toepassing op een
doseringsletaliteitsproef
2 Statistica 74-94
- C. A. G. NASS 1950 Recensie van COCHRAN and COX
4 Statistica 229-32
- C. A. G. NASS 1950a Commentaar op „Een eenvoudig schema voor organoleptisch
onderzoek” door Ir H. de Mirande
2 Mededelingen Medisch-Biologische Sectie, V.v. S. 95-7
- R. NEUWIRTH 1951 Die Anwendung der Labilitätskarten
Jahresbericht 1950 Bad. Landeswetterdienst 66-78

- E. V. NEWNHAM 1928 Correlation coefficients in meteorology
121 Nature 421
- H. A. PANOFKY 1949 Significance of meteorological correlation coefficients
30 BAMS 326-7
- K. PEARSON 1897 Mathematical contributions to the theory of evolutions. – On a
form of spurious correlation which may arise when indices are
used in the measurement of organs
60 Proc. Royal Soc. London 489-98
- K. PEARSON and
A. LEE 1897 On the distribution of frequency (variation and correlation) of
the barometric height at diverse stations
A 190 Phil. Trans. London 423-69
- W. F. PETERSEN 1947 Man, Weather and Sun
Ch. C. Thomas Springfield
- L. PONCELET 1946 Sur l'étude du climat en Belgique
Kon. Belgisch Met. Inst. Mengelingen 25
- R. A. ROBB and
T. R. TANNAHILL 1935 The lunar atmospheric pressure inequalities at Glasgow
55 Proc. Royal Soc. Edinburgh 91-6
- M. RODEWALD 1948 Die barische Vorbereitung strenger und milder mitteleuropä-
ischer Winter
1 AdM 99-106
- M. RODEWALD 1950 Zur Frage der Winter-Vorbereitung durch die Zirkulations-
entwicklung vom Oktober bis Dezember
3 AdM 225-40
- M. RODEWALD 1950a Die Verwendung monatlicher Druckanomalien – Änderungs-
karten zu langfristigen Zirkulationsprognosen
3 AdM 361-7
- H. VON RUDLOFF und
H. TRENKLE 1950 Beiträge zur langfristigen Witterungsvorhersage
Abh. Bad. Landeswetterdienst 1-76
- E. SABINE 1851 On periodical laws discoverable in the mean effects of the
larger magnetic disturbances
141 Phil. Trans. London 123-39
- R. SCHERHAG 1948 Die Wintertemperaturen in Berlin und die Sonnenflecken-
relativzahlen
A1 AMGB 233-46
- A. SCHMIDT 1926 Zur Kritik des Korrelationsfaktors
43 MZ 329-34
- M. SCHNEIDER 1950 Laubentfaltung der Rosskastanien, Sonnenflecken und März-
temperatur
N12 BDW 104-9

- M. SCHWARZBACH 1950 Das Klima der Vorzeit. Eine Einführung in die Paläoklimatologie
Ferdinand Enke Verlag Stuttgart
- W. SCHWERDTFEGGER 1952 Über die Entstehung mehrtägiger Rhythmen im Wetterablauf
N38 BDW 16-20
- H. SEILKOPF 1948 Langfrist-Vorhersagen. Bevinde zich in: R. MÜGGE, Meteorologie und Physik der Atmosphäre
19 Fiat Review 196-204
- N. SHAW 1928 Manual of Meteorology
Cambridge Univ. Press
- E. SLUTSKY 1937 The summation of random causes as the source of cyclic processes
5 Econometrica 105-46
- H. E. SOPER c.s. 1917 On the distribution of the correlation coefficient in small samples
11 Biometrika 328-413
- F. STEINHAUSER 1938 Die Meteorologie des Sonnblicks. I. Teil. Beiträge zur Hochgebirgsmeteorologie nach Ergebnissen 50-jähriger Beobachtungen des Sonnblickobservatoriums, 3106 m
Springer Wien
- F. STEINHAUSER 1948 Die Sonnenflecken und die Seespiegelschwankungen des Viktoriasees
1 Wetter und Leben 260-2
- F. STEINHAUSER 1949 Ergänzung zur Mitteilung über die Sonnenflecken und die Seespiegelschwankungen des Viktoriasees
1 Wetter und Leben 357-9
- „STUDENT“ 1908 Probable error of a correlation coefficient
6 Biometrika 302-10
- C.W. THORNTON 1949 Report of the committee on climatology, 1947-1948
30 TAGU 439-43
- G. TINTNER 1952 Abraham WALD's contributions to Econometrics
23 Annals Math. Statistics 21-8
- L. H. C. TIPPETT 1927 Random sampling numbers
Tracts for Computers N15
Cambridge Univ. Press
- S. I. TOMKEIEFF 1947 Recensie van VISTELIUS 1947
160 Nature 846-7
- V. VARANGOT 1948 Over correlatierekening en de herleiding van klimatologische reeksen
2 Statistica 250-2

- S. W. VISSER 1946 Weersverwachtingen op lange termijn in Nederland
Med. en Verh. 51
- S. W. VISSER 1949 Het klimaat van Nederland
Wetensch. suppl. 2 van Hemel en Dampkring 3-15
- S. W. VISSER en
W. VAN DER BIJL 1952 Een experiment met toevalsreeksen
Verschijnt in Statistica
- A. B. VISTELIUS 1947 On the correlation of mesorhythms in the lower Permian
deposits of Trans-Kama Tataria, and their stratigraphic
significance
55 Comptes Rendus (Doklady) Acad. Sci. URSS 237-40
- M. VITTRANT 1942 Périodes lunaires en météorologie
Série 3. 3 Bulletin de l'Université l'Aurore 909-30
- M. VITTRANT 1943 Périodes lunaires des typhons et dépressions
Série 3. 4 Bulletin de l'Université l'Aurore 729-60
- M. VITTRANT 1944 Cadre lunaire et cyclones
Série 3. 5 Bulletin de l'Université l'Aurore 643-62
- M. VITTRANT 1945 Météorologie luni-solaire
Série 3. 6 Bulletin de l'Université l'Aurore 289-306
- A. WAGNER 1925 Eine bemerkenswerte 16jährige Klimaschwankung, mit einem
Anhang: Mögliche Perioden verschiedener Wellenlänge
133 Sitzungber. d. Akademie d. Wiss. in Wien, Math.-naturw.
Klasse, Abt. IIa, S. 171-224
- H. WALDEN C.S. 1950 Zur Vorhersage von Hochdruckgebieten über Nordwesteuropa
3 AdM 267-83
- G. T. WALKER 1925 On periodicity
51 QJ 337-45
- G. T. WALKER C.S. 1926 On correlation coefficients, their calculation and use
52 QJ 73-84
- A. WEGENER 1910 Über die Ableitung von Mittelwerten aus Drachenaufstiegen
ungleicher Höhe
3 Beiträge zur Physik der freien Atmosphäre 13-9
- N. WEGER 1952 Weinernten und Sonnenflecken
N38 BDW 229-37
- L. L. WEISS 1944 Preliminary report on duration of stormy periods at selected
localities and intervals between periods
N3 Research paper USWB, Division of special scientific ser-
vices
- S. D. WICKSELL 1921 An exact formula for spurious correlation
1 N4 Metron 33-40

- W. F. WILLCOX 1935 Definitions of statistics
3 Revue de l'Institut Int. de Statistique 388-99
- J. P. M. WOUDEBERG 1946 Het verband tussen het weer en de opbrengst van winterarwe
in Nederland
Med. en Verh. 50
- R. WÜNSCHMANN 1949 Abschätzung von Zufallsschwankungen bei Verhältniszahlen
4 Statistische Praxis 7-8

SUMMARY

- 0 Some general aspects of climatology and statistics are discussed in the introductory chapter.
- 1 The needed statistical formulae are stated in chapter 1. They are well known (see *e.g.* KENDALL 1947), except for a section (1.12) on persistence.
- 2 Various sets of geophysical data are analysed statistically, especially the long series of climatological observations at Zwanenburg, begun in 1735 (see LABRIJN 1945). Some remarkable evidence concerning the BRÜCKNER cycle was revealed.
- 3 The essential parts of a paper in German (VAN DER BUL 1951) are given in this chapter. The pitfalls in handling geophysical records statistically are discussed under five headings (sampling, selection, persistence, pseudo-correlation and curvature effect).
- 3.1 Sampling errors arise if properties of a sample are assumed valid for the whole universe without a thorough investigation of their statistical significance.
- 3.2 In general, statistical methods must be applied to random samples. The selection effect comes in when two (or more) special samples, *e.g.* those with the largest or smallest means, are considered without allowing for the non-randomness of the samples. The selection effect can be removed by using the so-called selection factor. This factor takes into account the number of samples from which the special samples are chosen.
- 3.3 We speak of persistence when successive observations are not independent of each other. A given number of dependent observations is bound to give less weighty information than the same number of independent ones.
- 3.4 Two variables which are both functions of a third variable are generally related. Simplest example: the correlation coefficient between a_i and $a_i + b_i$ is 0.71 if a_i and b_i are uncorrelated and have equal variances. The expression 'pseudo-correlation' (suggested by Dr M. E. Wise) is preferred to 'spurious correlation' (introduced by PEARSON).
- 3.5 $f(x)$ be convex (concave) for $a < x < b$. This interval is divided into n equal parts. Now $\sum_{k=0}^{n-1} f(x + k \frac{b-a}{n})$ with $a < x < a + \frac{b-a}{n}$ will also be convex (concave). This curvature effect discovered by BARTELS in 1922 deprived many publications of their value, especially those on subjects like the moon's influence on air pressure.
- 4.1-4.4 These sections of the concluding chapter present critical examinations of various statistical investigations of geophysical problems (VON RUDLOFF und TRENKLE 1950, MEISSNER 1943, VISSER 1946, VITTRANT 1942, MINEUR 1951, RODEWALD 1950a).
- 4.5 The last section contains a multiple regression analysis of Dutch monthly records. This attempt to derive long range forecasts of regional rainfall in August proved to be unsuccessful.

Most results obtained in the chapters 2 and 4 will also be published in international journals.

