MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

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

Rapport S 169

Voedingsonderzoek bij zwangere vrouwen

III Vergelijking van de gehalten van stoffen in het bloed, resp. in de urine in de verschillende perioden der zwangerschap.

door

Constance van Eeden

1955

1. Inleiding

Bij een onderzoek naar de voedings- en gezondheidstoestand van zwangere vrouwen werden ons van 498 gezonde zwangere vrouwen o.a. gegevens verstrekt over het gehalte in het bloed en de urine van de volgende stoffen:

1. caroteen, in honderdsten γ per ml. serum,
2. vitamine A, in tienden I.E. per ml. serum,
3. vitamine E, in tienden γ per ml. serum,
4. vitamine B_1, A.P.P., in γ per 100 ml. serum,
5. vitamine B_1, in tientallen γ per 24 uur urine,
6. vitamine B_2, in tientallen γ per 24 uur urine,
7. vitamine C, in γ per ml. serum,
8. cholesterol, in tientallen mg. % serum,
9. cholesterolester, in tientallen mg.% serum,
10. alk.phosphatase, in tienden Bessey m.M. units in serum,
11. haemoglobine, in tienden hgl.g.% in bloed,
12. s.g. serum,
13. stikstof, in grammen in 24 uur urine,
14. verhouding tussen albumen en globuline in serum x 10.

Voor ieder der vrouwen werden deze gehalten meestal driemaal bepaald en wel in de 3e, de 6e en de 9e maand van de zwangerschap.[1] De waarnemingsperiode strekt zich, voor de eerste periode der zwangerschap, uit van Maart 1950 tot en met Juli 1951, voor de tweede periode van Juni 1950 tot en met October 1951 en voor de derde periode van Augustus 1950 tot en met December 1951.

Naar aanleiding van deze gegevens hebben wij de volgende vragen beantwoord:

1. Is er een verschil tussen 1950 en 1951?
2. Is er een verschil tussen de drie perioden der zwangerschap?

Beiden vragen voor ieder der bovengenoemde gegevens apart.

2. Resultaten

2.1. Onderzoek naar een verschil tussen 1950 en 1951

Dit onderzoek is uitgevoerd voor ieder der maanden en ieder der zwangerschapsperioden apart, voor zoverre dezelfde maanden

---

1) Deze maanden zullen wij in hetvolgende aanduiden als resp. de eerste, de tweede en de derde periode der zwangerschap.

in de beide jaren voor ieder der 3 perioden binnen de waarnemingstermijn lagen. Voor ieder der in de inleiding genoemde gegevens zijn voor de eerste periode de waarnemingen uit Maart 1950 vergeleken met die uit Maart 1951, April 1950 met April 1951 enz. tot en met Juli 1950 met Juli 1951; analoog zijn voor de tweede resp. de derde periode de maanden Juni 1950 tot en met October 1950 (resp. Augustus 1950 tot en met December 1950) vergeleken met de overeenkomstige maanden van 1951.

Voor de vergelijking van twee maanden hebben wij de toets van WILCOXON toegepast (zie memorandum S 47(M 7)). Wij krijgen dan voor iedere periode der zwangerschap en voor ieder der maanden:

1. een waarde voor de toetsingsgrootheid $\underline{U}$,
2. een waarde voor $\mathcal{E}(\underline{U} \mid H_0)$,
3. een waarde voor $\sigma^2(\underline{U} \mid H_0)$,

waarbij de hypothese H_0 inhoudt, dat er voor die periode en die maand geen verschil is tussen 1950 en 1951. Wij geven deze waarden aan met resp. $\underline{U}_i$, μ_i en σ_i^2 .

Deze resultaten hebben wij, voor iedere periode apart, op twee manieren gecombineerd:

A. Wij berekenen de grootheid:

$$\underline{u} = \sum_i \frac{\underline{U}_i}{n_i m_i} .$$

Het gemiddelde μ en de variantie σ^2 van $\underline{u}$ onder de hypothese H_0', inhoudende, dat er, voor de beschouwde periode der zwangerschap, voor <u>ieder</u> der maanden geen verschil is tussen 1950 en 1951, worden

$$\mu = \sum_i \frac{\mu_i}{n_i m_i} ,$$

$$\sigma^2 = \sum_i \frac{\sigma_i^2}{n_i^2 m_i^2} .$$

De grootheid $\underline{u}$ is, onder de hypothese H_0' , bij benadering normaal verdeeld met gemiddelde μ en spreiding σ en de kritieke zône bestaat uit grote waarden van $|\underline{u}|$. Deze toets leidt speciaal dan tot verwerping van H_0' als er voor tenminste één maand een verschil is tussen 1950 en 1951 en de aanwezige verschillen overwegend hetzelfde teken hebben.

----

2) Door de grootheden $\underline{U}_i$ te delen door $n_i m_i$ bereikt men dat de toets voor de hypothese H_0' bruikbaar is voor een klasse van alternatieve hypothesen die onafhankelijk is van de verhoudingen der steekproefgrootten.

B. Wij berekenen de grootheid

$$\chi^2 = \sum_i \frac{(u_i - \mu_i)^2}{\sigma_i^2}.$$

Deze grootheid bezit, als H_0' juist is, bij benadering een χ^2-verdeling met een aantal vrijheidsgraden dat gelijk is aan het aantal grootheden u_i dat men combineert. De kritieke zône bestaat uit grote waarden van χ^2 en deze toets leidt ook dan tot verwerping van H_0' als de verschillen tussen 1950 en 1951 voor de verschillende waarden niet alle hetzelfde teken hebben.

De reden, dat wij beide toetsen toepassen, is dat als de verschillen tussen 1950 en 1951 overwegend hetzelfde teken hebben toets A een groter onderscheidingsvermogen bezit dan toets B, terwijl toets B een groter onderscheidingsvermogen bezit dan toets A als de verschillen tussen 1950 en 1951 niet overwegend hetzelfde teken hebben.
Wij krijgen dus voor ieder der in de inleiding genoemde gegevens zes overschrijdingskansen nl. één voor ieder der perioden en voor ieder der toetsen A en B. Deze overschrijdingskansen staan vermeld in tabel I (zie blz. 4).

Uit tabel I zien wij dat er voor practisch alle gegevens een verschil is tussen de jaren 1950 en 1951. Ter verduidelijking van de resultaten uit tabel I zijn voor enige der gegevens in de figuren 1 t/m 5 (zie blz. 9 t/m 13) de gemiddelden voor de verschillende maanden grafisch uitgezet.
Of de gevonden verschillen tussen 1950 en 1951 verschillen zijn tussen de werkelijke gehalten of geheel of gedeeltelijk te wijten zijn aan veranderingen in de bepalingstechniek is met deze gegevens niet na te gaan. Dit zou wel mogelijk geweest zijn, als men gedurende het onderzoek regelmatig contrôlebepalingen had verricht van een bekend praeparaat of iets dergelijks; men had dan kunnen onderzoeken of de waarnemingen en de controlebepalingen een overeenstemmend verloop vertoonden.
Een gevolg hiervan is dat wij niet kunnen onderzoeken of er seizoensschommelingen in de gehalten optreden: wij kunnen nl. wel nagaan of er verschillen zijn tussen de maanden van één jaar, maar deze verschillen kunnen eveneens het gevolg zijn van veranderingen in de bepalingstechniek. Dientengevolge zullen wij de rest van het onderzoek uitvoeren voor ieder der maanden apart, aannemende dat de eventuele veranderingen in het gehalte en in de bepalingstechniek in de loop van één maand te verwaarlozen zijn.

Tabel I [3)4)]

Overschrijdingskansen gevonden bij de vergelijking van 1950 en 1951.

Methode	A			B		
Periode	1	2	3	1	2	3
Caroteen	0,54+	0,84+	0,32-	0,44	0,95	0,006
Vitamine A	$< 10^{-4}$+	0,77+	0,05-	0,0003	0,035	0,023
Vitamine B_1, A.P.P.	0,21+	0,05+	0,04+	0,33	0,009	0,0005
Vitamine B_1	0,71+	0,15-	0,012-	0,42	0,70	0,0045
Vitamine C	0,001-	0,70-	0,54+	0,01	0,76	0,25
Cholesterol	0,19-	0,05-	0,17-	0,07	0,10	0,18
Cholesterolester	0,007-	0,0004-	0,007-	0,001	0,003	0,004
Alk.Phosph.	$< 10^{-8}$+	5.10^{-6}+	5.10^{-5}+	$\ll 10^{-4}$	$\ll 10^{-4}$	$\ll 10^{-4}$
Haemoglobine	$< 10^{-8}$+	$< 10^{-4}$+	0,21-	$< 10^{-5}$	7.10^{-5}	0,28
s.g.serum	0,53-	0,90-	0,36+	5.10^{-4}	0,035	0,83
stikstof	0,77+	0,28+	0,23+	0,37	0,08	0,68
Verhouding a/g	3.10^{-8}+	4.10^{-5}+	0,74+	$< 10^{-5}$	0,0002	0,002

----

3) Het teken + bij een overschrijdingskans betekent, dat de waarnemingen in het tweede jaar hoger zijn dan die in het eerste jaar.

4) Daar in het tweede jaar vrijwel geen vitamine E en in het het eerste jaar vrijwel geen vitamine B_2-bepalingen verricht zijn, konden wij voor deze gegevens de twee jaren niet vergelijken.

2.2. <u>Onderzoek naar een verschil tussen de perioden der zwangerschap</u>.

Hier is eerst onderzocht, voor ieder der gegevens, of het gehalte stijgt of daalt gedurende de zwangerschap. Hiertoe hebben wij, voor ieder der maanden apart, TERPSTRA's toets tegen verloop voor groepen waarnemingen toegepast (zie bijlage S 168 (M 61)). Wij krijgen dan voor iedere maand

1. een waarde voor de toetsingsgrootheid $\underline{W}$,
2. een waarde voor $\sigma^2(\underline{W} \mid H_0)$,

waarbij de hypothese H_0 inhoudt, dat er voor de betreffende maand geen verschil is tussen de drie perioden der zwangerschap. Deze waarden geven wij resp. aan met $\underline{W}_i$ en σ_i^2 .
Stel nu

$$\underline{V} = \sum_i \underline{W}_i ,$$

dan is de grootheid $\underline{V}$ onder de hypothese H_0' , inhoudende dat er voor iedere maand geen verschil is tussen de drie perioden, bij benadering normaal verdeeld met gemiddeld 0 en variantie $\sigma^2 = \sum_i \sigma_i^2$. De kritieke zône bestaat uit grote positieve resp. grote negatieve waarden van V en de toets leidt speciaal dan tot verwerping van H_0' als er voor minstens één maand een verloop is gedurende de zwangerschap en als dit verloop, voor die maanden waarin het aanwezig is, overwegend een stijging (resp. overwegend een daling) is.
De resultaten van deze toets staan vermeld in tabel II (zie blz. 6).

Uit tabel II zien wij:

a. er is geen reden om aan te nemen, dat de gehalten vitamine A, vitamine B_2 en vitamine C stijgen of dalen gedurende de zwangerschap,
b. voor de overige gegevens vinden wij een duidelijk stijging of daling.

Voor die gegevens, waarvoor deze toets geen aanwijzingen voor een verloop gaf hebben wij nog een toets van algemenere aard toegepast, nl. TERPSTRA's toets voor het probleem van m rangschikkingen (zie voor de formulering van dit probleem memorandum S 47 (M 14); de toets van TERPSTRA stelt ons in staat deze methode ook toe te passen als de aantallen waarnemingen in de vakjes van het schema van m rijen verschillend zijn.

Tabel II [5]

Overschrijdingskansen gevonden bij het onderzoek naar een stijging of daling van de gehalten gedurende de zwangerschap.

Gegeven	Overschrijdingskans
Caroteen	$< 10^{-8}$ -
Vitamine A	0,19-
Vitamine E	0,036+
Vitamine B_1 (A.P.P.[1])	0,04-
Vitamine B_1	4.10^{-5} -
Vitamine B_2	0,81+
Vitamine C	0,50+
Cholesterol	$\ll 10^{-8}$ +
Cholesterolester	$\ll 10^{-8}$ +
Alk.phosphatase	$\lll 10^{-8}$ +
Haemoglobine	$< 10^{-8}$ -
S.G. serum	4.10^{-9} -
Stikstof	3.10^{-6} -
Verhouding a/g	$< 10^{-8}$ -

De door hem voorgestelde methode is geen directe generalisatie van de in het genoemde memorandum beschreven toets, maar van de rangcorrelatiemethode van M.G. KENDALL). Als het j^e vakje van de i^e rij van het schema n_{ij} waarnemingen van een stochastische grootheid x_{ij} bevat $(j=1,2,\ldots,k; i=1,2,\ldots,m)$, dan luidt de hypothese H_0 dat voor iedere rij, dus voor iedere i , de grootheden $x_{i1}, x_{i2}, \ldots, x_{ik}$ dezelfde waarschijnlijkheidsverdeling bezitten; voor de kolommen behoeft dit niet te gelden, d.w.z. de grootheden x_{ij} en $x_{i'j}$ $(i \neq i')$ behoeven niet dezelfde verdeling te bezitten

----

5) Het teken + bij een overschrijdingskans betekent, dat de waarnemingen een stijgend verloop vertonen.

Als $\underline{u}_{jj'}^{(i)}$ de toetsingsgrootheid van de toets van WILCOXON is voor de vergelijking van de waarnemingen van $\underline{x}_{ij}$ die van $\underline{x}_{ij'}$ uit de i^e rij, dan is de toetsingsgrootheid

$$\underline{S} \stackrel{\text{def}}{=} \sum_{i<i'}\sum \sum_{j<j'}\sum \frac{(2\underline{u}_{jj'}^{(i)} - n_{ij}n_{ij'})(2\underline{u}_{jj'}^{(i')} - n_{i'j}n_{i'j'})}{n_{ij}\, n_{ij'}\, n_{i'j}\, n_{i'j'}} .$$

De kritieke zône bestaat uit grote waarden van $\underline{S}$ en de toets leidt speciaal dan tot verwerping van H_0 als er een overeenstemming tussen de m rijen is; dus niet alleen als er een stijging of een daling is gedurende de zwangerschap, maar ook als de waarden in de tweede periode hoger of lager zijn dan die in de eerste en de derde. In ons geval heeft het schema drie kolommen (één voor iedere periode); de aantallen n_{ij} zijn de aantallen waarnemingen die in de i^e maand en de j^e periode verricht zijn. De resultaten van deze toets staan vermeld in tabel III.

Tabel III

Overschrijdingskansen gevonden bij het onderzoek naar een verschil tussen de perioden.

Gegeven	Overschrijdingskans
Vitamine A	$>0{,}25$
Vitamine B_2	$>0{,}50$
Vitamine C	$>0{,}50$

Hieruit zien wij dat er ook op grond van deze toets geen reden is om aan te nemen dat er een verschil is tussen de perioden voor vitamine A, vitamine B_2 en vitamine C.

Als gevolg van de in de vorige paragraaf genoemde eventuele veranderingen in de bepalingstechniek heeft het geen zin schattingen te geven van het absolute gehalten der verschillende stoffen. Wel kunnen wij, aannemende dat een eventuele wijziging der waarnemingstechniek alleen betrekking heeft op het niveau der uitkomsten, een schatting geven van de verschillen tussen de drie perioden. Hiertoe berekenen wij voor iedere maand het het gemiddelde van de waarnemingen uit de eerste van die uit de tweede en van die uit de derde periode. Deze gemiddelden geven wij resp. aan met $\bar{\underline{x}}_i$, $\bar{\underline{y}}_i$ en $\bar{\underline{z}}_i$, waarbij de index i de maand aangeeft. Dan zijn

$$v_1 = \frac{1}{k} \sum_{i=1}^{k} (\bar{y}_i - \bar{x}_i),$$

en

$$v_2 = \frac{1}{k} \sum_{i=1}^{k} (\bar{z}_i - \bar{y}_i),$$

waarin k het aantal maanden is, schattingen van het verschil tussen de tweede en de eerste periode resp. het verschil tussen de derde en de tweede periode. Deze schattingen staan vermeld in tabel IV.

Tabel IV

Schattingen van de verschillen tussen de tweede en de eerste, resp. de derde en de tweede periode.

	verschil tussen	
	2e en 1e	3e en 2e
	periode	
Caroteen	12,48	7,98
Vitamine A	-0,17	+0,03
Vitamine E	3,24	7,19
Vitamine B_1 A.P.P.	-0,68	-0,31
Vitamine B_1	-0,59	-3,43
Vitamine B_2	4,14	-2,03
Vitamine C	0,07	-0,08
Cholesterol	4,46	2,35
Cholesterolester	1,95	1,49
Alk.phosphatase	5,93	13,76
Haemoglobine	-6,76	-5,74
S.G. serum	-0,0006	-0,00006
Stikstof	-0,80	-1,36
Verhouding a/g	-0,88	-1,53

Deze verschillen zijn uitgedrukt in de in de inleiding genoemde eenheden.

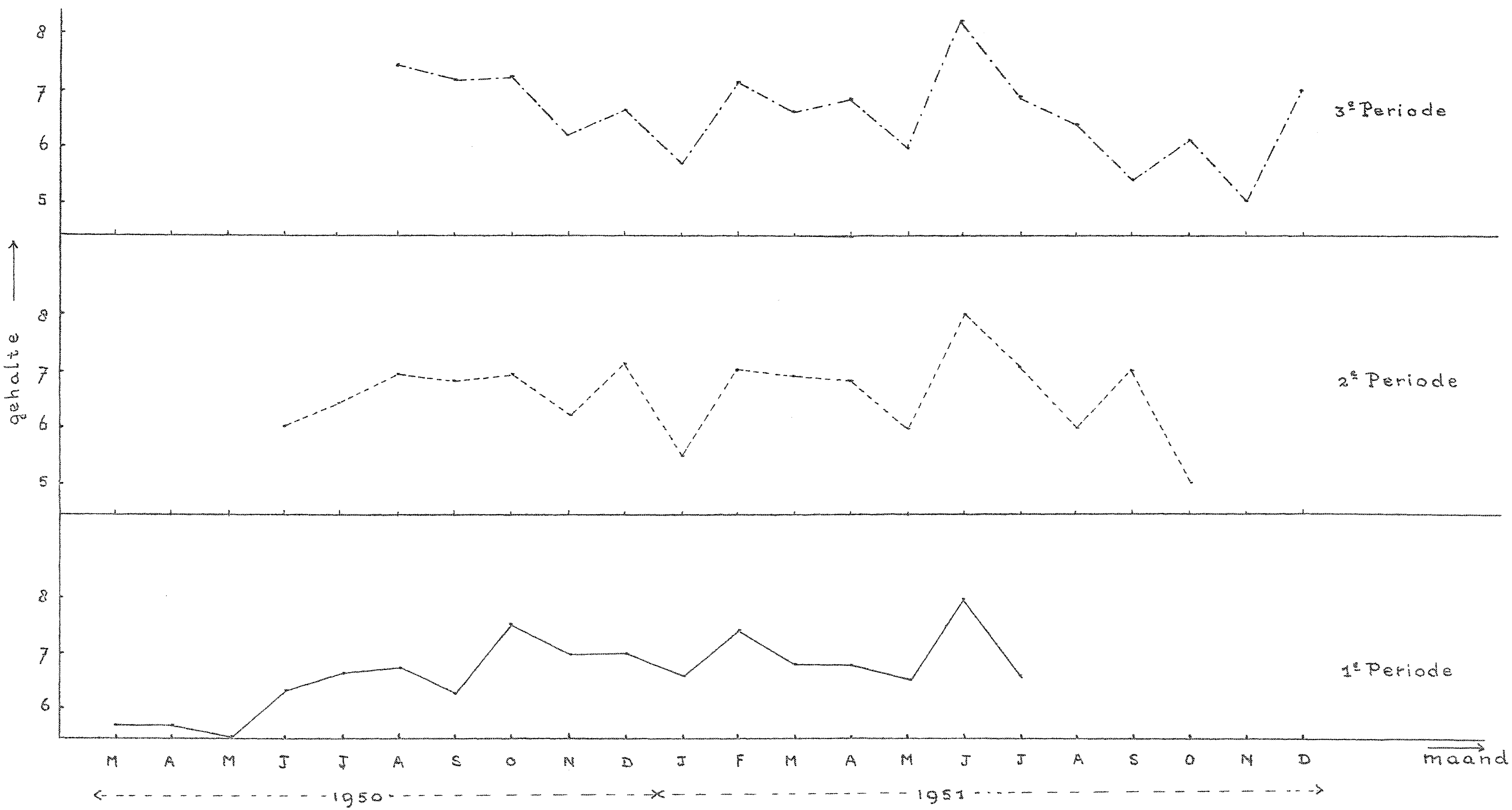

Fig. 1: Gemiddelde gehalte vitamine A voor de verschillende maanden en perioden.

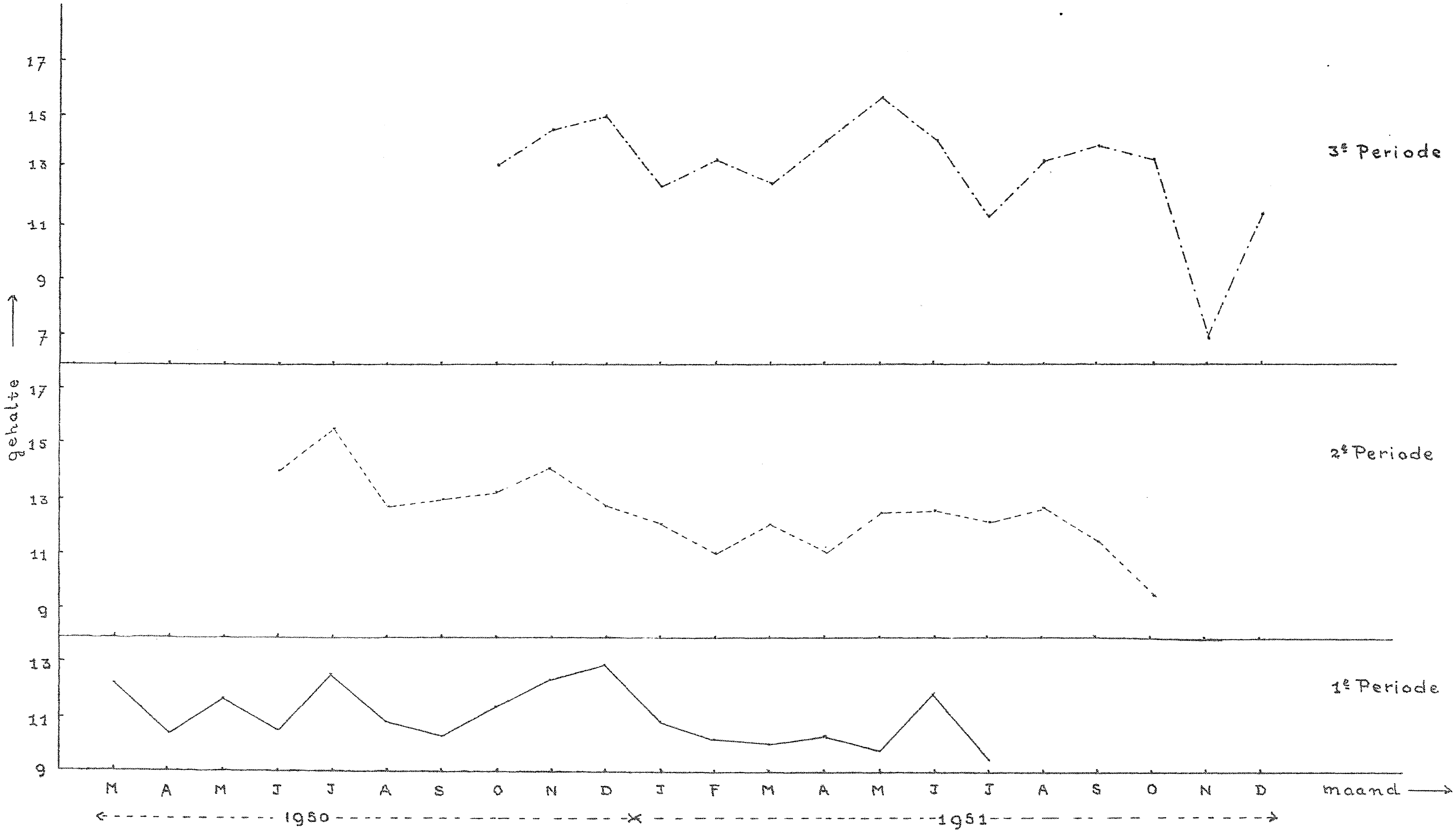

Fig. 2: Gemiddelde gehalte cholesterolester voor de verschillende maanden en perioden.

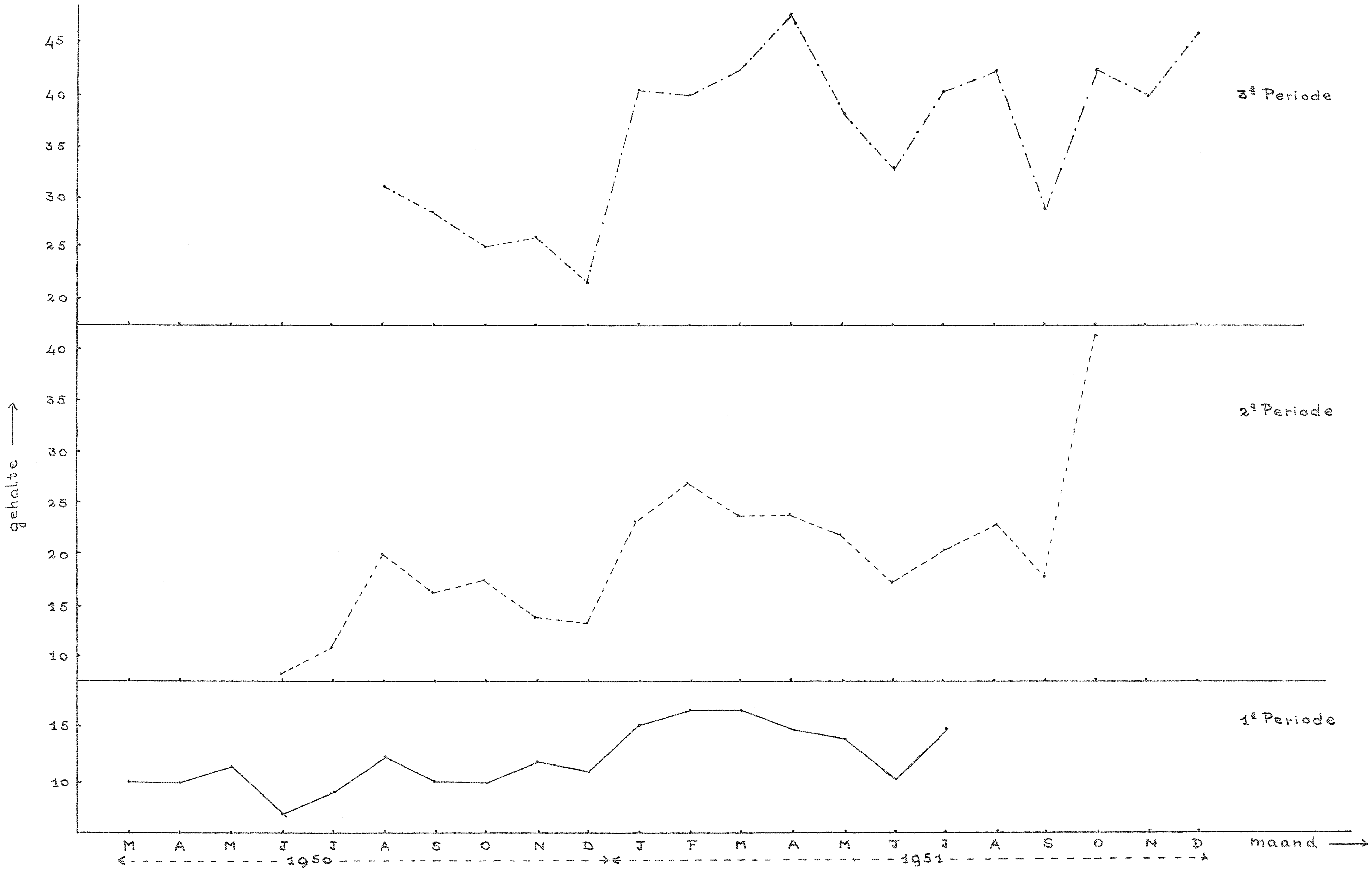

Fig. 3: Gemiddelde gehalte alk. phosphatase voor de verschillende maanden en perioden

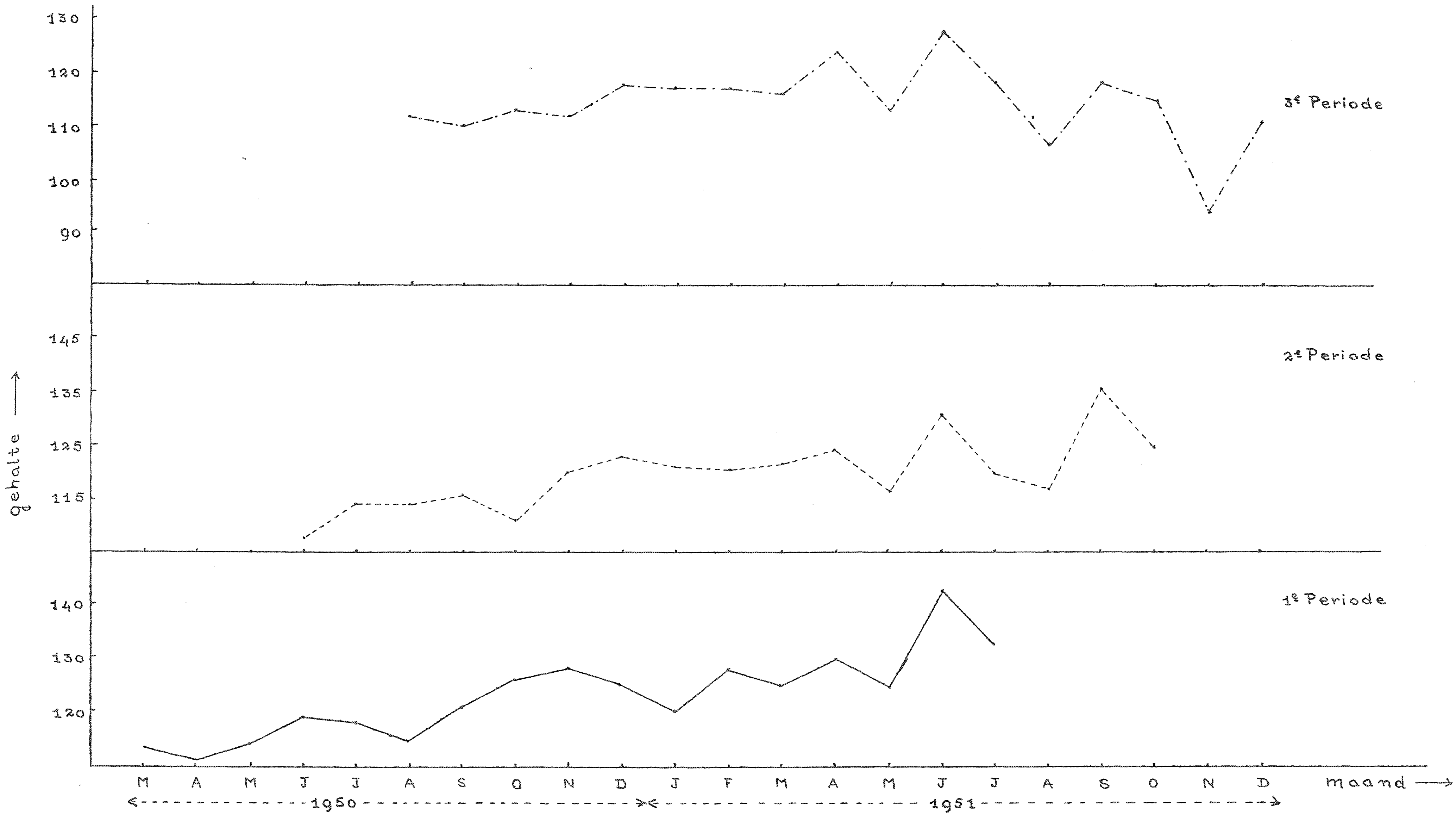

Fig. 4: Gemiddelde gehalte haemoglobine voor de verschillende maanden en perioden.

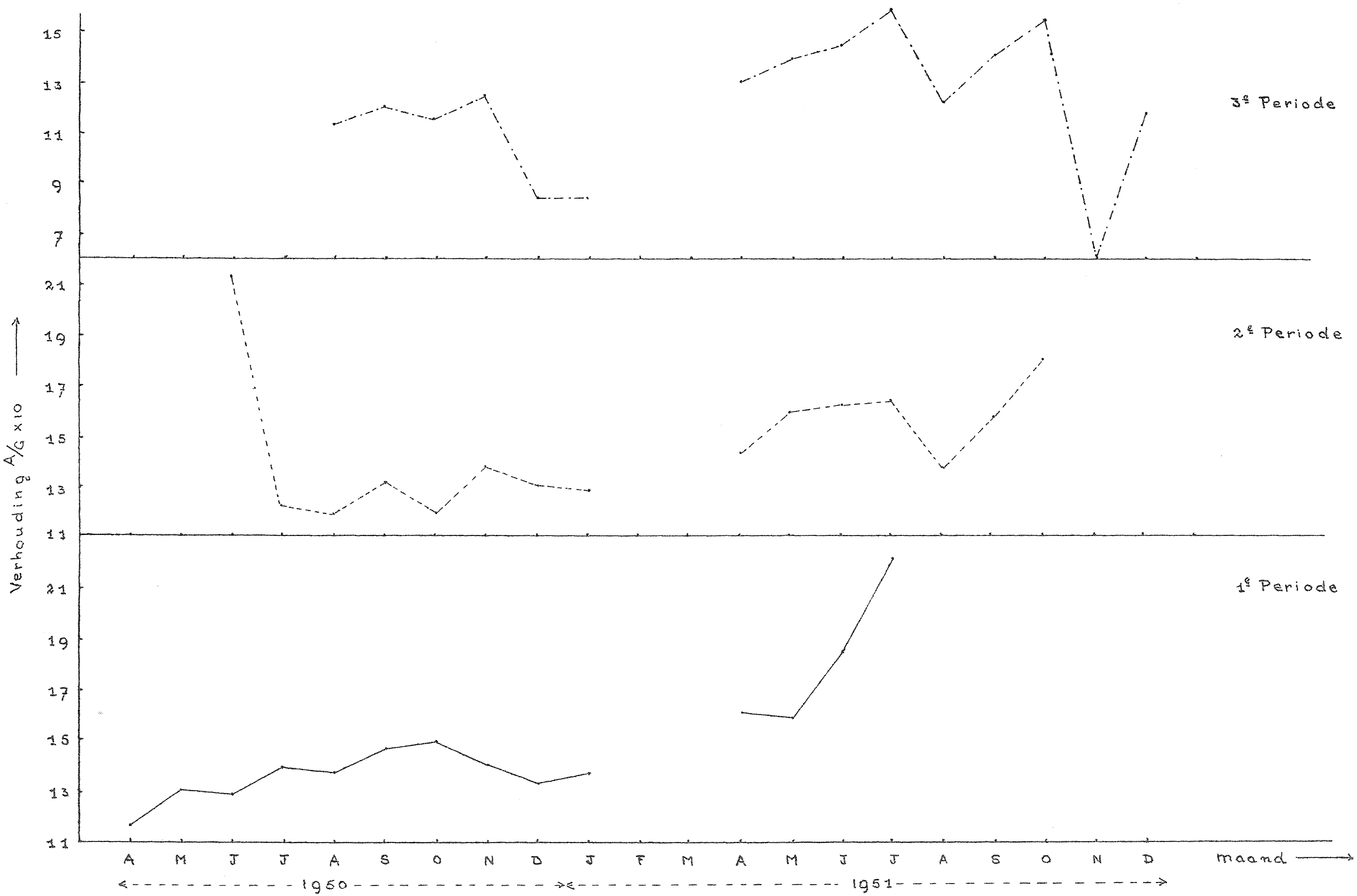

Fig. 5: Gemiddelde verhouding tussen albumen en globuline voor de verschillende maanden en perioden.

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.
S47(M6)

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

De toetsing van een hypothese $\mathcal{H}_0$ berust steeds op een aantal waarnemingen $x_1, x_2, \ldots, x_n$ van één of meer stochastische grootheden[2], of op enige groepen van waarnemingen (bv. twee steekproeven).

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

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat $\mathcal{H}_0$ juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van $\underline{u}$, en wel op zodanige wijze, dat de kans, dat $\underline{u}$ een in Z gelegen waarde aanneemt, onder de hypothese $\mathcal{H}_0$, gelijk is aan een gegeven getal α, zodat Z dus van α afhankelijk is[3]. Z heet de kritieke zône van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu $\mathcal{H}_0$ op grond van de waarnemingen $x_1, x_2, \ldots, x_n$, indien de bij deze waarnemingen behorende waarde van $\underline{u}$ in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien $\mathcal{H}_0$ juist is, gelijk aan α. Derhalve is α de kans op ten onrechte verwerping van de juiste hypothese, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met $\alpha = 0{,}05$ resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, waarin de hypothese die men toetst juist is, deze toch verwerpen.

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

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

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

De toetsingstheorie biedt in het algemeen geen mogelijkheid om tot aanvaarding van een hypothese te komen. Indien een bepaalde hypothese H_0 niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

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

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

In plaats van de onbetrouwbaarheidsdrempel α wordt vaak bij de uitslag van een toetsing de overschrijdingskans k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van H_0 zou zijn overgegaan; anders gezegd: de kleinste α, waarvoor de gevonden waarde der toetsingsgrootheid nog juist in de (bij α behorende) kritieke zône Z ligt. Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leq \alpha$ is.

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijke men bv. de tweede hieronder gegeven littèratuurplaats. Wij moeten hier volstaan met de opmerking, dat éénzijdige toetsing veelal eerder tot verwerping van H_0 leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

Litteratuur:

J. Neyman, First course in probability and statistics, New York, 1950, Chapter 5.

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

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

Maart, 1952.

De toets van Wilcoxon.[1])

Deze methode dient tot het toetsen van de hypothese H_0, inhoudende, dat twee steekproeven $x_1,\dots,x_n$ en $y_1,\dots,y_m$ afkomstig zijn uit één collectie (ook wel populatie of universum genaamd).

Voor het toetsen van de hypothese H_0 wordt gebruik gemaakt van een toetsingsgrootheid $\underline{U}$ [2]), die als volgt uit de waarnemingen berekend wordt. Onderstellen we, dat de waarnemingen $x_1,\dots,x_n$ en $y_1,\dots,y_m$ naar opklimmende grootte gerangschikt zijn, dan bepalen we eerst het aantal waarnemingen uit de tweede steekproef, dat kleiner is dan de kleinste waarneming x_1 uit de eerste steekproef (bij gelijkheid tellen wij $\frac{1}{2}$ in plaats van 1). Noem dit aantal V_1. Vervolgens wordt het aantal waarnemingen uit de tweede steekproef bepaald, dat kleiner is dan de op één na kleinste waarneming x_2 uit de eerste steekproef (bij gelijkheid wordt weer een $\frac{1}{2}$ in plaats van 1 geteld). Dit aantal noemen we V_2. Evenzo worden met betrekking tot $x_3,x_4,\dots,x_n$ de aantallen $V_3,V_4,\dots,V_n$ bepaald. De waarde U van de toetsingsgrootheid $\underline{U}$ wordt voor de twee steekproeven dan gegeven door

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

Wanneer onder de waarnemingen niet te veel gelijken voorkomen, kan bewezen worden, dat de toetsingsgrootheid $\underline{U}$ onder de hypothese H_0 voor grote waarden van n en m (beide $\geq$ 10) bij benadering een normale verdeling bezit. De waarnemingen $x_1,\dots,x_n$ en $y_1,\dots,y_m$ tezamen genomen vallen uiteen in een aantal groepen van gelijke waarnemingen. Noem het aantal van deze groepen k, dan is k minstens 1 (als alle waarnemingen gelijk zijn) en hoogstens m+n (als alle waarnemingen verschillend zijn).

---

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

[2]) Stochastische grootheden worden door onderstreping aangeduid.

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

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

en

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

De grootheid $\mu(\underline{U})$ is dus onafhankelijk van de waarden vast. Indien de hypothese H_0 niet vervuld is, zal de grootheid $\underline{U}$ grote of kleine waarden bezitten, al naar gelang $\underline{y}$ systematisch kleiner of groter is dan $\underline{x}$.

De (tweezijdige) toets bestaat nu daarin, dat men H_0 verwerpt indien de gevonden waarde U van $\underline{U}$ te sterk van μ afwijkt, d.w.z. als

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

waarin α de onbetrouwbaarheidsdrempel is en ξ_α volgt uit

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

en in een tabel van de normale verdeling kan worden opgezocht.

De (tweezijdige) overschrijdingskans k , behorende bij T, is gedefiniëerd als

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

en kan ook in eenttabel van de normale verdeling worden gevonden.

Bij eenzijdige toetsing wordt α door 2α vervangen, resp. k gehalveerd.

Een bijzonder geval van het bovenstaande is, dat onder de waarnemingen voor $\underline{x}$ en $\underline{y}$ in 't geheel geen gelijken voorkomen. In dat geval kan de uitdrukking voor de variantie herleid worden tot

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

---

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

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

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

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

<u>Litteratuur</u>:

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

MATHEMATISCH CENTRUM
2de Boerhaavestr.49
A M S T E R D A M O.
Statistische Afdeling
S47 (M14)

Methode der m rangschikkingen [1])

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

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

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

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

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

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

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

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

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

In ons voorbeeld is $W = \frac{12 \times 64}{16 \times 210} = 0{,}229$.
W varieert dus tussen 0 en 1.

De verdeling van $\underline{S}$ onder de hypothese H_o is exact berekend voor een aantal waarden van n en m [1] , terwijl voor grote m en n benaderingen bekend zijn.

De meeste gebruikelijke benaderingen zijn de volgende.

1°. De χ^2-benadering:

$\chi_r^2 = m(n-1)\underline{W} = \frac{12\ \underline{S}}{mn(n+1)}$ heeft voor $m \to \infty$ een χ^2-verdeling met n-1 vrijheidsgraden ([1] pg. 84 [2] pg. 36-37).

2°. De z-benadering:

$\underline{V} = (m-1)\frac{\underline{W}}{1-\underline{W}}$ is bij benadering verdeeld als $\underline{F} = e^{2\underline{z}}$
($\underline{F}$ is de $\underline{F}$ van Snedecor, $\underline{z}$ de $\underline{z}$ van Fisher) met

$\nu_1 = n-1-\frac{2}{m}$

$\nu_2 = (m-1)\nu_1$ vrijheidsgraden ([1] pg. 84 [2] pg.33-3[illegible]

Met behulp van de verdelingen van $\underline{S}$ of $\underline{W}$ onder de hypothese H_o, kan deze hypothese getoetst worden, waarbij H_o verworpen wordt als $\underline{W}$ waarden dichtbij 1 (resp. $\underline{S}$ dichtbij $\frac{1}{12} m^2(n^3-n)$) aanneemt, de kritieke zône is dus van de vorm $W \geq W_o$ (resp. $S \geq S_o$).

Het kan voorkomen dat de waarnemers geen onderscheid ontdekken in de mate waarin verschillende elementen het kenmerk bezitten. Ze geven deze elementen dan gelijke rangnummers.

Veronderstel, dat door een waarnemer geen onderscheid wordt gemaakt tussen de elementen, die de rangnummers 3 t/m 6 moeten dragen. Dan wordt als rangnummer van iedere van deze elementen het gemiddelde van de rangnummers $\frac{1}{4}(3 + 4 + 5 + 6) = 4\frac{1}{2}$ gebruikt.

Daar het maximum van $\underline{S}$ nu verandert, moeten wij een correctie op de formule voor $\underline{W}$ toepassen. Deze vindt men in [1] (pg.82) en [2] (pg. 28-30). Eveneens veranderen dan de formules voor de χ^2-benadering ([1] pg. 86, [2] pg. 37) en voor de z-benadering ([1] pg. 86 [2] pg. 34), doch deze correcties zijn van weinig betekenis, tenzij het aantal gelijken groot is.

Literatuur: [1] M.G.Kendall, Rank correlation methods, London 1948, Hoofdstuk 6, pag. 80.
Tabel van de verdelingsfunctie van S voor:

n = 3   m = 2 t/m 10
n = 4   m = 2 t/m 6
n = 5   m = 3
op pag. 146-149.

Tabel van de waarden van S, waarvan de overschrijdingskansen onder de hypothese H_0 gelijk zijn aan 0,05 of 0,01, berekend met behulp van de z-benadering voor:

n = 3   m = 8,9,10,12,14,15,16,18,20
n = 4   m = 4,5,6,8,10,15,20
n = 5 t/m 7   m = 3,4,5,6,8,10,15,20
op pag. 150.

[2] Ph.van Elteren, Methode der m rangschikkingen, Cursus "Parametervrije Methoden", Hoofdstuk II, Rapport S 59, Mathematisch Centrum (1951), Blz. 18-45.

MATHEMATISCH CENTRUM,
2e Boerhaavestraat 49,
Amsterdam - O.
Statistische Afdeling
Rapport S 168 (M 61)

Een parametervrije toets tegen verloop voor groepen waarnemingen.[1]

Wij beschouwen k onderling onafhankelijke stochastische grootheden $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$. Van de grootheid $\underline{x}_i$ zijn n_i onderling onafhankelijke waarnemingen $x_{i,1}, \ldots, x_{i,n_i}$ gegeven $(i=1,\ldots,k)$.

De hypothese H_0 die wij willen toetsen luidt, dat $\underline{x}_1, \ldots, \underline{x}_k$ dezelfde waarschijnlijkheidsverdeling bezitten, terwijl de alternatieve hypothesen inhouden, dat de stochastische grootheden $\underline{x}_1, \ldots, \underline{x}_k$ een stijgend of dalend verloop vertonen, gedefinieerd door:

$$(1) \qquad \sum_{i<j}\sum \left\{ P[\underline{x}_i < \underline{x}_j] - P[\underline{x}_i > \underline{x}_j] \right\} \neq 0.$$

De toetsingsgrootheid $\underline{W}$ wordt als volgt gedefiniëerd. Stel voor $i < j$ is $\underline{U}_{i,j}$ het aantal paren waarnemingen $(x_{i,\lambda}, x_{j,\mu})$ met $x_{i,\lambda} < x_{j,\mu}$, vermeerderd met de helft van het aantal paren $(x_{i,\lambda}, x_{j,\mu})$ met $x_{i,\lambda} = x_{j,\mu}$ $(\lambda \leq n_i, \mu \leq n_j)$.[2] Stel verder

$$(2) \qquad \underline{W}_{i,j} \stackrel{def}{=} 2\left\{ \underline{U}_{i,j} - \mathcal{E}(\underline{U}_{i,j} \mid H_0) \right\} = 2\underline{U}_{i,j} - n_i n_j ,$$

dan is

$$(3) \qquad \underline{W} \stackrel{def}{=} \sum_{i<j}\sum \frac{\underline{W}_{i,j}}{n_i n_j} .$$ [3]

Als onder de $N = \sum_i n_i$ waarnemingen h groepen gelijke waarnemingen optreden en als $\underline{t}_\ell$ het aantal waarnemingen in de ℓ^e groep is $(\ell = 1, 2, \ldots, h)$ dan geldt

$$(4) \qquad \mathcal{E}[\underline{W} \mid t_1, t_2, \ldots, t_\ell ; H_0] = 0 ,$$

---

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

2) $\underline{U}_{i,j}$ is dus de toetsingsgrootheid van de toets van WILCOXON, toegepast op de waarnemingen van $\underline{x}_i$ en $\underline{x}_j$ en $\underline{U}_{j,i} = n_i n_j - \underline{U}_{i,j}$.

3) De hier beschreven toets is een wijziging van een door T.J. TERPSTRA [4] ontwikkelde toets voor dit probleem. Hij gebruikt de toetsingsgrootheid $\underline{T} = \sum_{i<j}\sum \underline{W}_{i,j}$. Door de toetsingsgrootheid $\underline{W}$ te gebruiken bereikt men, dat de toets bruikbaar is voor een klasse van alternatieve hypothesen, die niet afhangt van de verhoudingen der steekproefgrootten (zie ook [5]).

$$(5)\left\{\begin{array}{l}\sigma^2[\underline{W} \mid t_1, t_2, \ldots, t_h; H_0] = \\ = \dfrac{\{N^3 - \Sigma t_\ell^3 - 3(N^2 - \Sigma t_\ell^2)\} \sum \frac{(k+1-2i)^2}{n_i} - \{2(N^3 - \Sigma t_\ell^3) - 3N(N^2 - \Sigma t_\ell^2)\} \sum\sum_{i<j} \frac{1}{n_i n_j}}{3N(N-1)(N-2)}\end{array}\right. .$$

Treden er geen gelijke waarnemingen op, dan is $t_\ell = 1$ voor iedere ℓ en:

$$(6) \qquad \sigma^2[\underline{W} \mid H_0] = \frac{1}{3}\left\{ \sum_i \frac{(k+1-2i)^2}{n_i} + \sum\sum_{i<j} \frac{1}{n_i n_j} \right\}.$$

Als H_0 juist is, is de grootheid $\underline{W}$ voor grote waarden van N bij benadering normaal verdeeld met gemiddelde en variantie volgens (4) en (5).
Het is duidelijk, dat $\underline{W}$ in het algemeen grote positieve waarden zal aannemen als er een stijgend en grote negatieve als er een dalend verloop is; de tweezijdige kritieke zône bestaat dus uit grote waarden van $|W|$.

Opmerkingen.

1. Als $k = 2$, dus als wij twee stochastische grootheden $\underline{x}_1$ en $\underline{x}_2$ hebben met resp. n_1 en n_2 waarnemingen, dan geldt:

$$(7) \qquad \underline{W} = \frac{2\underline{U} - n_1 n_2}{n_1 n_2},$$

waarin $\underline{U}$ de toetsingsgrootheid van de toets van WILCOXON is, toegepast op de steekproeven van $\underline{x}_1$ en $\underline{x}_2$. Uit (4) en (5) volgt:

$$\mathcal{E}[\underline{W} \mid t_1, t_2, \ldots, t_h; H_0] = 0,$$

$$\sigma^2[\underline{W} \mid t_1, t_2, \ldots, t_h; H_0] = \frac{N^3 - \Sigma t_\ell^3}{3 n_1 n_2 N(N-1)}$$.(zie memorandum S 47 (M 7))

Als er in dit geval geen gelijke waarnemingen optreden, kan men de exacte overschrijdingskans opzoeken in tabellen van de exacte verdeling van $\underline{U}$ (zie b.v. [1]).

2. Als $n_i = 1$ voor iedere i dan is

$$\underline{W} = \sum\sum_{i<j} \operatorname{sgn}(\underline{x}_j - \underline{x}_i),$$

waarin

$$\operatorname{sgn} z \overset{\text{def}}{=} \begin{cases} 1 & \text{als } z > 0 \\ 0 & \text{als } z = 0 \\ -1 & \text{als } z < 0. \end{cases}$$

Uit (4) en (5) volgt:

$$\mathcal{E}[\underline{W} \mid t_1, t_2, \ldots, t_h; H_0] = 0,$$

$$\sigma^2[\underline{W} \mid t_1, t_2, \ldots, t_h; H_0] = \frac{2(N^3 - \Sigma t_\ell^3) + 3(N^2 - \Sigma t_\ell^2)}{18}$$

De toets is in dit geval identiek met de rangcorrelatie methode van KEDALL voor het geval, dat één der rijen bestaat uit de getallen $1, 2, \ldots, k$.(zie memorandum S 47 (M 13)). Treden er in dit geval geen gelijke waarnemingen op en is $k \leq 40$ dan kan men de exacte overschrijdingskans opzoeken in de tabellen van de exacte verdeling van de toetsingsgrootheid $\underline{S}$ van KENDALL (zie [2]).
Is $k \leq 10$ en $t_\ell \leq 3$ voor iedere ℓ dan kan men gebruik maken van de tabellen van de exacte verdeling van $\underline{S}$ van SILLITTO [3]

Literatuur.

[1] Van der Vaart, H.R., Gebruiksaanwijzing voor de toets van Wilcoxon, Rapport S 32 (M 4) van de Statistische Afdeling van het Mathematisch Centrum.

[2] Kaarsmaker, L. en A. van Wijngaarden, Tables for use in rankcorrelation, Rapport R 73 van de Rekenafdeling van het Mathematisch Centrum.

[3] Sillitto, G.P., The distribution of Kendall's coefficient of rankcorrelation in rankings containing ties, Biometrika 34 (1947), 36-40.

[4] Terpstra, T.J., The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking, Proc.Kon.Ned.Ak. 55 (1952).

[5] Van Eeden, C., A test for the equality of probabilities against a class of specified alternatives, including trend, Rapport S 157(VP 3) van de Statistische Afdeling van het Mathematisch Centrum.
