

Statistische Afdeling
Rapport S 71,
door J. van Klinken
en Dr J. Hemelrijk Jr.

De invloed van de voeding op het cholesterol-
gehalte van het bloed.

1. Inleiding

Uit de medische literatuur blijkt, dat het cholesterol, dat de mens via zijn voeding opneemt, uitsluitend van dierlijke oorsprong is. De sterolen in plantaardige voedingsstoffen, die in structuur weinig van cholesterol verschillen, worden in het darmkanaal van de mens niet geresorbeerd.

Het onderzoek had tot doel een antwoord te vinden op de volgende vragen.

- a) Is er een verband tussen het cholesterolgehalte van de voeding en het cholesterolgehalte van het bloedserum.
- b) Zijn er individuele eigenaardigheden wat betreft die samenhang.
- c) Is er een verschil in dit opzicht tussen de beide geslachten.
- d) Zijn er nog andere oorzaken, dan de voeding, die het cholesterolgehalte van het bloedserum bepalen.

De opstelling van de proef was als volgt.

Uit 100 kandidaten werden 60 proefpersonen gekozen, die geen klinische afwijkingen vertoonden; 44 van deze proefpersonen hadden een leeftijd tussen 20 en 30 jaar, 13 tussen 30 en 40 en 3 tussen 40 en 50 jaar. Het aantal vrouwen was gelijk aan dat van de mannen. De gehele proef strekte zich uit over 36 weken, verdeeld in 3 perioden van 12 weken. Iedere proefpersoon gebruikte gedurende een bepaalde periode een bepaald dieet. Er waren drie diëten:

- 1e. Een volstrekt vegetarisch dieet (V-dieet).
- 2e. Een normaal dieet, bevattend aan dierlijke producten: 50 gr. vlees, 30 gr. kaas, 1/2 liter melk en 2 eieren per week (L-dieet).
- 3e. Een dieet met 200 gr. vlees, 50 gr. kaas en 2 eieren per dag, dus zeer rijk aan dierlijke voedingsmiddelen.

Uit voedselanalyses blijkt, dat het V-dieet onge-

veer 3, het L-dieet 300 en het H-dieet 900 m.g. cholesterol (per dag) bevat.

Bij de aanvang van de proef werd de groep van 60 proefpersonen in 3 groepen van 20 verdeeld. Elk der groepen omvatte 10 mannen en 10 vrouwen. Groep I kwam op het V-dieet, groep II op het L-dieet en groep III op het H-dieet. Aan het einde van de eerste periode werd elke groep gesplitst in 2 groepen van 10, 5 mannen en 5 vrouwen, die naar de beide andere diëten gingen. Er ontstonden zo 6 groepen n.l. Ia, Ib, IIa, IIb, IIIa, IIIb. Steeds bevonden zich 2 groepen op eenzelfde dieet. De verdeling van de 6 groepen over de 3 diëten tijdens elk der 3 perioden is in tabel I weergegeven. (De tabellen vindt men aan het einde van het rapport.)

Eens per week werd het cholesterolgehalte van het bloedserum bepaald volgens de methode van Saiffer en Kammerer.

De in dit rapport toegepaste statistische methoden zijn kort beschreven in een aantal aan het einde van het rapport toegevoegde memoranda.

2. Opmerkingen over de waarnemingstechniek.

2.1 Een opvallende overeenstemming in het verloop van de gemiddelden van het cholesterolgehalte van het serum voor mannen resp. vrouwen uit eenzelfde groep was aanleiding om te onderzoeken of dit overeenkomstig gedrag in de tijd aan het toeval geweten kan worden of niet.

Uit het waarnemingsmateriaal blijkt dat de aanpassing aan een nieuw dieet in de eerste 6 weken van een periode plaatsvindt. We betrokken daarom in het onderzoek van deze overeenstemming alleen de cholesterolwaarden van de laatste 6 weken van de verschillende perioden. Een eventueel te vinden overeenstemmend gedrag in de tijd kan dan niet geweten worden aan een stijging of daling ten gevolge van een dieetovergang. De gebruikte toetsingsmethode, de methode der m rangschikkingen, berust op rangordeningsmethoden, ontworpen door M.G.KENDALL. Zie voor een korte beschrijving hiervan het aan dit rapport toegevoegde memorandum S 47 (M14) In de eerste plaats werden de cholesterolwaarden van de 10 personen uit eenzelfde groep voor de laatste 6 weken van een bepaalde periode onderling vergeleken. De verschillende groepen en perioden zijn gedeeltelijk gekozen naar aanleiding van opvallende uitspringende waarden van enkele gemiddelden, gedeeltelijk juist wegens het

ontbreken daarvan. In tabel II zijn de resultaten weergegeven. Zoals de tabel doet zien, zijn alle resultaten significant. Op dezelfde wijze werd onderzocht of de groepsgemiddelden van mannen en vrouwen samen ook een overeenstemmend gedrag in de tijd vertoonden.

Hierbij werden weer alleen de tweede helften van de perioden in het onderzoek betrokken. De resultaten zijn in tabel III weergegeven. Hoewel de eerste beide overschrijdingskansen niet significant zijn, zijn ze alle drie wel aan de lage kant. De drie toetsingen kunnen tot een enkele toetsing worden gecombineerd. De bij deze toets gebruikte toetsingsgrootheid $\frac{12 S}{mn(n+1)}$ bezit bij benadering een χ^2 -verdeling met $n-1$ vrijheidsgraden (n : aantal kolommen, m : aantal rijen; zie memorandum M 14). De drie toetsingsgrootheden zijn in ons geval onderling onafhankelijk. De som van de drie toetsingsgrootheden is nu verdeeld volgens χ^2 met als aantal vrijheidsgraden de som van de drie aantallen vrijheidsgraden.

We vinden een significante overschrijdingskans: $< 0,01$. Ook de groepsgemiddelden vertonen dus een overeenkomstig gedrag in de tijd. We kunnen uit dit alles concluderen, dat de waarnemingsuitkomsten van betrekkelijk grote groepen proefpersonen, zelfs van verschillende diëten, gelijktijdig sterke stijgingen of dalingen vertonen. Om dit verschijnsel nader te kunnen onderzoeken, was het noodzakelijk eerst een onderzoek in te stellen naar de waarnemingstechniek.

2.2 De waarnemingstechniek.

De cholesterolbepalingen vonden als volgt plaats. Er wordt een standaardoplossing gemaakt, die in samenstelling zoveel mogelijk overeenkomst met bloedserum dat 250 mg. % cholesterol bevat vertoont. (Mg. % betekent: mg. per 100 cc.) Met deze standaardoplossing worden telkens 8 proefsera vergeleken. Het cholesterolgehalte van elk der 8 bloedsera wordt dan bepaald volgens de formule:

$$\text{chol. gehalte} = \frac{B_1}{S} \times 250 \text{ mg. \%} \quad i = 1 \dots 8$$

Hierin is S de colorimeteraflezing van de standaardoplossing en B_1 die van het i -de proefserum. Voordat de colorimeteraflezingen kunnen plaatsvinden, ondergaan de 8 proefsera en de standaardoplossing verschillende bewerkingen, zoals baden met een bepaalde temperatuur etc. Er zijn bewerkingen die op alle 8 proefsera en de stan-

daardoplossing gelijktijdig uitgevoerd worden en bewerkingen, die alleen op de standaardoplossing toegepast worden.

De mogelijkheid bestaat, dat door deze laatste groep bewerkingen S een sterk afwijkende waarde verkrijgt. Alle 8 cholesterolwaarden die met behulp van deze waarde van S bepaald worden, zullen dan te laag of te hoog zijn. Verder zijn, zoals direct uit het waarnemingsmateriaal blijkt, de series van 8 op één standaard betrokken sera niet regelmatig over de 6 groepen van proefpersonen verdeeld. Zeer vaak valt de helft of meer der sera van één serie binnen eenzelfde groep. Het heeft dus zin om te onderzoeken of het gevonden overeenstemmend gedrag in de tijd het gevolg is van onderlinge onafhankelijkheden tussen de uitkomst van de bepalingen uit eenzelfde serie.

2.3 Oriënterend onderzoek naar afhankelijkheden.

We bepaalden voor elk der proefpersonen de rangcorrelatiecoëfficiënt \bar{r} tussen de cholesterolwaarden van de tweede helften van elk der drie perioden en de daarbij behorende colorimeteraflezingen S van het standaardserum. Voor de berekening van \bar{r} verwijzen wij naar het bijgevoegde memorandum M 13. \bar{r} kan waarden aannemen tussen -1 en +1. Waarden van \bar{r} in de buurt van -1 betekenen dat lage cholesterolwaarden corresponderen met hoge waarden van S en omgekeerd. Indien de uitkomst voor het cholesterolgehalte niet gecorreleerd was met de waarde S van de bijbehorende standaard, zou ieder van de gevonden correlatiecoëfficiënten \bar{r} beschouwd kunnen worden als een waarneming van een stochastische grootheid, die symmetrisch verdeeld is ten opzichte van nul. Deze stochastische grootheden zijn echter niet onafhankelijk, daar de sera van verschillende personen in groepen van 8 gemeenschappelijk behandeld zijn. De resultaten zijn in tabel IV weergegeven. Zoals de tabel doet zien, worden er veel meer negatieve dan positieve waarden voor \bar{r} gevonden. Verder blijkt uit de tabel, dat het aantal negatieve correlatiecoëfficiënten \bar{r} naar het einde van de proef sterk toeneemt. In periode III wijken zelfs vele der negatieve waarden van \bar{r} significant van nul af (toetsing volgens memorandum S 47 (M 13)). We kunnen dus wel besluiten tot een sterke afhankelijkheid tussen uitkomst en standaard. Dit geldt dan speciaal voor de perioden II en III. Hoge cholesterolwaarden corresponderen vaak met lage waarden van S en omgekeerd. De variaties in de cholesterolwaarden, die het gevolg zijn van

de variaties in S , veroorzaakt door de bewerkingen die uitsluitend S ondergaat, zijn dus stellig niet te verwaarlozen. Het vermoeden rijst zelfs dat het opvallend overeenstemmende gedrag der uitkomsten binnen de groepen een gevolg is van deze afhankelijkheid. Om dit te toetsen met behulp van een toets die nauwer bij het gestelde probleem aansluit dan de vorige, werd de volgende weg ingeslagen.

2.4 Definitieve toetsing der afhankelijkheid.

We beschouwen tweetallen opeenvolgende bepalingen van het cholesterolgehalte van twee willekeurige proefpersonen, waarbij:

- a) beide proefpersonen in beide bepalingen op eenzelfde standaard betrokken zijn.
- b) ze in geen van beide bepalingen op eenzelfde standaard betrokken zijn.

Deze tweetallen opeenvolgende waarnemingen werden uit de oorspronkelijke aantekeningen van de waarnemingsresultaten bijeengezocht volgens een vast systeem, waarbij uitsluitend gelet werd op het voldoen aan de eisen a en b. Ook werd er voor gezorgd dat de verschillende viertallen onderling onafhankelijk zijn. Wegens de grote uitgebreidheid van het materiaal zijn niet alle paren die aan a en b voldoen in het onderzoek betrokken.

Geven wij de cholesterolwaarden van de twee proefpersonen aan met W_1 en W_2 , de daarmee corresponderende standaardwaarden met S_1 en S_2 , terwijl we verder de getallen behorende bij de tweede bepaling voorzien van een accent, dan kunnen wij het volgende schema opstellen.

proefpersoon	1e bepaling	2e bepaling
1	(S_1, W_1)	(S'_1, W'_1)
2	(S_2, W_2)	(S'_2, W'_2)
a	$S_1 = S_2$	$S'_1 = S'_2$
b	$S_1 \neq S_2$	$S'_1 \neq S'_2$

Wij onderscheiden nu twee gevallen.

1e. overeenkomstig gedrag (beide stijgen: $W_1 < W'_1$; $W_2 < W'_2$ of dalen: $W_1 > W'_1$; $W_2 > W'_2$);

2e. tegengesteld gedrag ($W_1 < W'_1$; $W_2 > W'_2$ of omgekeerd).

Men kan dit dus ook als volgt aangeven:

Overeenstemming: $(W_1 - W'_1)(W_2 - W'_2) > 0$;

Tegengesteld gedrag: " " < 0 .

De verschillende gevonden aantallen zijn in tabel V weergegeven. Toetsen we met de tekentoets voor de eerste regel van tabel V, dus voor het geval van gelijke waarden van S, de hypothese dat de kans op gelijk gedrag gelijk is aan die op tegengesteld gedrag, dan vinden wij een overschrijdingskans $< 0,01$. (Zie memorandum M 22.) Voeren wij hetzelfde uit voor de tweede rij, dus verschillende standaarden, dan vinden wij in het geheel geen significant resultaat. Verder kunnen wij beide toetsingen nog combineren. Is p de kans op gelijk gedrag bij gelijke standaardwaarden en p' in het geval van verschillende standaarden, dan blijkt de hypothese $p = p'$ niet gehandhaafd te kunnen worden. We vinden een overschrijdingskans 0,005. (Zie voor het toetsen van een 2 x 2-tabel het bijgevoegde memorandum M 23.)

2.5 Conclusies.

Uit dit alles kunnen de volgende conclusies worden getrokken.

1e. Het overeenstemmend gedrag in de tijd wordt geheel of ten dele veroorzaakt door het betrekken van steeds 8 waarnemingen op eenzelfde standaardwaarde S, waarbij de 8 waarnemingen uit eenzelfde serie, niet gelijkmatig over de 6 groepen verdeeld zijn.

Opm. Dat niet gelijkmatig verdeeld zijn van de 8-taalen waarnemingen over de 6 groepen is een gevolg van het feit, dat de proefpersonen met de nummers 1 tot en met 20 voornamelijk in de groepen Ia en Ib voorkomen, 20 tot en met 40 in de groepen IIa en IIb en de nummers 40 tot en met 60 in de groepen IIIa en IIIb, terwijl verder de bepalingen in de regel naar opklimmend rangnummer zijn verricht.

2e. Ook de grote wekelijkse variaties kunnen zeer goed geheel of grotendeels voortvloeien uit de gevolgde bepalingstechniek. Deze heeft in ieder geval dergelijke schommelingen ten gevolge.

2.6 Onderzoek der duplobepalingen.

2.6.1 Bevestiging van de afhankelijkheid tussen uitkomst en standaard.

Om een idee te krijgen van de reproduceerbaarheid van de cholesterolwaarden waren tijdens de proef een groot aantal duplobepalingen verricht, waarbij beide

waarden van W op verschillende waarden van S betrokken zijn. De duplo's zullen een bevestiging moeten geven van de zojuist gevonden resultaten. In de eerste plaats zal nu ook de afhankelijkheid van de standaard S voor de dag moeten komen. Duplo-waarnemingen geven wij schematisch aan met:

$$\begin{array}{cc} W_1 & W_2 \\ S_1 & S_2 \end{array}$$

We verdelen de duplo's in twee klassen A en B. In klasse A komen de duplo's met de eigenschap: $(S_1 - S_2)(W_1 - W_2) > 0$ en in klasse B die waarvoor geldt: $(S_1 - S_2)(W_1 - W_2) < 0$. In tabel VI zijn de verschillende aantallen weergegeven. Toetsen wij met behulp van de tekentoets (zie memorandum S 53 (M 23)) de hypothese, dat de kans dat een duplo tot A behoort even groot is als de kans tot B te behoren, dan vinden wij een overschrijdingskans $< 0,01$. We hebben dus weer een bevestiging van wat we reeds vonden: er is een sterke negatieve correlatie tussen de cholesterolwaarden en de daarbij behorende waarden van S .

2.6.2 Schatting van de nauwkeurigheid.

Uit de duplo's kunnen wij als volgt een indruk van de nauwkeurigheid der bepaling verkrijgen. Onderstellen wij dat de fout in de bepaling een normale waarschijnlijkheidsverdeling met constante spreiding σ bezit dan is het gemiddelde van de waarden $W_1 - W_2$ van alle duplo's gedeeld door 1,128 een zuivere schatting van deze . De factor 1,128 is afkomstig uit de relatie

$$\sum w = 1,128 \sigma,$$

waarin $w = W_1 - W_2$ de range van een steekproef van 2 waarnemingen uit een normale verdeling voorstelt en \sum het symbool voor de mathematische verwachting is. In het algemeen is de onderstelling van normaliteit, die hierbij wordt gebruikt, nogal gewaagd. Bij meetfouten zijn de afwijkingen van normaliteit vaak niet zeer groot. In het onderhavige geval bestaat er tegen de schatting van een onnauwkeurigheid voor de hele proeftijd nog een ander bezwaar. Indien σ niet constant is tijdens de proefperiode, geeft de verkregen schatting een soort gemiddelde indruk van de onnauwkeurigheid. De verkregen uitkomst geeft dan niet meer dan een ruwe indruk van de onnauwkeurigheid. Voor het gemiddelde van $W_1 - W_2$ vonden wij 15,4 mg. $\%$. Als schatting voor hebben wij dan: $\hat{\sigma} = \frac{15,4}{1,128} \approx 15$. Noemen wij de waarde die wij door meting gevonden hebben $W_{\text{gem.}}$, dan bevat het

interval ($W_{\text{gem.}-30}$, $W_{\text{gem.}+30}$) met een kans van ongeveer 95% de werkelijke waarde van W . De lengte van het 95% betrouwbaarheidsinterval is dus ongeveer 60 mg. %. Hieruit zien wij dat nog betrekkelijk grote toevallige variaties verwacht moeten worden.

3. Beantwoording van de vragen a, b, c en d uit de inleiding.

3.1 Invloed van de voeding.

Om te onderzoeken of de voeding invloed heeft op het cholesterolgehalte van het bloedserum, werden voor elk van de proefpersonen de cholesterolwaarden van steeds twee op elkaar volgende diëten met elkaar vergeleken. Wij beperkten ons tot het vergelijken van de tweede helften van de perioden. Voor elk der proefpersonen werd het verschil bepaald van de som van de cholesterolwaarden van de laatste 6 weken van een bepaalde periode met de som van de cholesterolwaarden van de laatste 6 weken van de daarop volgende periode¹⁾. Men mag aannemen dat door de sommen van de cholesterolwaarden te nemen, de invloed van de standaardwaarden S voldoende geëlimineerd is, om vertrouwen te kunnen hebben in eventueel te vinden conclusies. Ook nemen wij aan dat er nu geen afhankelijkheden meer bestaan tussen de verschillen van deze sommen. In totaal zijn er 6 dieetovergangen n.l.:

$L \rightarrow H$	$L \rightarrow V$	$H \rightarrow V$
$H \rightarrow L$	$V \rightarrow L$	$V \rightarrow H$

Elke dieetovergang wordt door 20 proefpersonen, 10 mannen en 10 vrouwen, gemaakt. Wij krijgen dus voor elke dieetovergang 20 verschillen. Heeft een bepaalde dieetovergang geen invloed op het cholesterolgehalte van het bloedserum, dan zullen de verschillen, die behoren bij deze dieetovergang, een symmetrische verdeling ten opzichte van nul bezitten. De te toetsen hypothese is dus: de verschillen hebben een symmetrische verdeling ten opzichte van nul. Dit werd uitgevoerd met de symmetrietoets (zie memorandum S 47 M 10). De resultaten voor de verschillende dieetovergangen zijn in tabel VII weergegeven. Zoals uit de tabel blijkt, zijn alle overschrijdingskansen behalve die, welke behoort

1) Het werken met sommen komt op hetzelfde neer als het werken met gemiddelden, daar iedere som uit een gelijk aantal (n.l. 6) termen bestaat.

bij de overgang $L \rightarrow V$, zeer klein, dus de effecten zeer duidelijk. De overschrijdingskans behorend bij $L \rightarrow V$ is, hoewel niet significant, toch ook klein ($\approx 0,05$). De hypothese, dat de verschillen symmetrisch verdeeld zijn ten opzichte van nul, is niet te handhaven. We kunnen dus besluiten tot een duidelijke invloed van de voeding op het cholesterolgehalte van het bloedserum. Bij overgang op een dieet, dat rijker is aan voedingsstoffen van dierlijke oorsprong stijgt in het algemeen het cholesterolgehalte. De overgang $L \rightarrow V$ verdient een nadere beschouwing. We vonden hier een enigszins afwijkend resultaat. Dit resultaat vindt waarschijnlijk zijn verklaring in de griepepidemie en de daardoor ontstane voedingsstoornissen, afwijkingen van het dieet etc., die een groot aantal mensen uit de "vegetarische" groep van de tweede periode omvatte. Zie hiertoe ook § 3.4.

3.2 Het verschillende reageren van de proefpersonen.

Dit gedeelte van het onderzoek heeft tot doel, na te gaan of er voor de verschillende proefpersonen een verband is tussen de grootte der reactie op de dieetovergangen en het cholesterolgehalte van hun bloedserum. Als maat voor de reactie namen wij voor ieder der proefpersonen de spreiding in de drie sommen van de cholesterolwaarden over de laatste 6 weken van elk der drie perioden. Deze spreiding gingen wij nu vergelijken met het gemiddelde van die drie sommen. De spreiding kan men beschouwen als een maat voor de grootte der reactie en het gemiddelde als een maat voor het cholesterolgehalte van het bloed der betrokken proefpersonen. Het is mogelijk dat de spreiding beïnvloed wordt door de volgorde van de diëten. We kunnen daarom alleen maar mensen uit eenzelfde groep op rechtstreekse wijze onderling vergelijken. De 10 spreidingen uit elk der 6 groepen gaven we naar opklimmende grootte de rangnummers 1 tot en met 10, en plaatsten ze verder in volgorde van opklimmende gemiddelde van de drie sommen. Wij krijgen zo een schema van 6 rijen van 10 rangnummers; zie tabel VIII. Is er geen verband tussen spreiding en gemiddelde, dan zal elke volgorde (permutatie) van de 10 rangnummers uit een bepaalde rij even waarschijnlijk zijn. We te toetsen hypothese is dus: elke permutatie van de 10 rangnummers van elk der 6 rijen is even waarschijnlijk. De toetsing werd uitgevoerd volgens de methode van de m rangschikkingen (zie memorandum S 47 M 14). Wij vonden een sig-

nificante overschrijdingskans: 0,03. We mogen dus wel besluiten tot een verband tussen reactie en gehalte. Uit het getallenschema (zie tabel VIII) blijkt verder, dat de eerste kolommen over het algemeen de lage rangnummers bevatten, de overige rangnummers lijken meer naar willekeur over de andere 8 kolommen verdeeld. Nu is het bekend dat deze toets een zeer groot onderscheidingsvermogen heeft voor de alternatieve hypothese, dat in enkele kolommen alle hoge of lage rangnummers voorkomen, terwijl de rest van de rangnummers gelijkmatig over de resterende kolommen verdeeld is. Het is dus mogelijk dat het significante resultaat uitsluitend te danken is aan het feit, dat in de eerste beide kolommen zo goed als alle lage rangnummers voorkomen. Om dit te controleren werd uit elke groep de twee proefpersonen met het laagste gemiddelde cholesterolgehalte weggelaten. Daarna werd het bovenbeschreven procédé nogmaals uitgevoerd. Het resultaat was nu een overschrijdingskans 0,75 (zie tabel VIII,2). Het zijn dus inderdaad de eerste kolommen die de significantie hebben veroorzaakt. Wij kunnen nu de volgende conclusie trekken:

De proefpersonen met de laagste gemiddelde cholesterolwaarden reageren in het algemeen minder sterk op de dieetovergangen dan de overige proefpersonen. Bij deze laatsten is geen samenhang meer tussen het cholesterolgehalte van het bloedserum en de reactie op de dieetovergangen te constateren.

3.3 Verschillende gedrag tussen de beide geslachten.

Dit gedeelte van het onderzoek had tot doel na te gaan of er een verschil is tussen mannen en vrouwen wat betreft het cholesterolgehalte van hun bloedserum en hun reactie op de verschillende dieetovergangen. Wij kunnen in verband met de verschillende volgorden van diëten slechts weer mannen en vrouwen uit eenzelfde groep onderling vergelijken. Als waarde voor het cholesterolgehalte werd evenals in de vorige paragraaf gemiddelde van de 3 bovengenoemde sommen genomen en als maat voor de reactie de spreiding van die sommen,

Beperken wij ons eerst tot de gemiddelden. We hebben dan 6 tweetallen groepen van ieder 5 getallen. Elk tweetal bestaat uit de 5 gemiddelden van de vrouwen en de 5 gemiddelden van de mannen uit eenzelfde groep. Ter verduidelijking nog onderstaand schema (ieder streepje stelt een gemiddelde voor).

	mannen	vrouwen
Ia	- - - - -	- - - - -
Ib	- - - - -	- - - - -
IIa	- - - - -	- - - - -
IIb	- - - - -	- - - - -
IIIa	- - - - -	- - - - -
IIIb	- - - - -	- - - - -

Is er geen verschil tussen de beide geslachten wat betreft het cholesterolgehalte, dan kunnen de twee 5-tallen waarden van een bepaalde rij (groep) beschouwd worden als twee steekproeven uiteenzelfde collectie. De te toetsen hypothese is nu dus, dat de twee groepen van 5 waarden uit elke rij (groep) twee steekproeven zijn uit eenzelfde verdeling. Om deze hypothese te kunnen toetsen, werd voor ieder tweetal de toets van Wilcoxon (zie S 47 (M 7)) toegepast, terwijl deze toetsingen vervolgens werden gecombineerd. De toetsingsgrootte u van Wilcoxon bezit onder de te toetsen hypothese H_0 een symmetrische verdeling. Zijn beide steekproeven groter dan 10, dan is de normale verdeling al een goede benadering. Voor het gemiddelde en de spreiding van deze normale verdeling geldt dan:

$$\mu = \frac{m \cdot n}{2}, \quad \sigma = \sqrt{\frac{mn(m+n+1)}{12}}$$

(m, n grootte van de steekproeven). In ons geval is $m=n=5$. Voor elke rij (groep) kunnen wij de waarde van de toetsingsgrootte u berekenen. Wij hebben 6 onafhankelijke grootte u . De som hiervan ($\sum u$) zal nu bij benadering een normale verdeling bezitten met

$$\mu = 6 \cdot \frac{m \cdot n}{2} \quad \text{en} \quad \sigma = \sqrt{\frac{6 \cdot mn(m+n+1)}{12}}$$

dus voor $m=n=5$; $\mu = 75$; $\sigma = 11,7$.

We toetsen nu de hypothese, dat $\sum u$ uit een normale verdeling: $N(75; 11,7)$ komt. Het resultaat was niet significant. Wij vonden een overschrijdingskans 0,4.

Het zojuist beschreven procédé kunnen wij ook uitvoeren voor de spreidingen. Hier vonden wij een overschrijdingskans 0,8. Ook wat betreft het reageren op de verschillende dieetovergangen is er dus geen verschil te constateren.

We kunnen samenvattend zeggen: er viel geen verschil in gedrag tussen de mannelijke en vrouwelijke proefpersonen te constateren.

3.4 Andere oorzaken dan de voeding die het cholesterolgehalte van het bloedserum bepalen.

Naast de voeding kunnen er nog andere oorzaken zijn die het cholesterolgehalte van het bloedserum beïnvloeden. In de eerste plaats werd hiernij gedacht aan ziekte-toestanden. Tijdens de proefperiode hebben een betrekkelijk groot aantal proefpersonen een griepaanval doorgemaakt. De data van de ziekte en de nummers van de zieke proefpersonen zijn in onderstaand tabelletje opgegeven.

20 - 21 November	42, 45, 48, 50, 52, 53, 58
5 - 8 Februari	27, 28, 34, 43, 47, 48, 49, 52, 54

Een groot aantal van de zieke proefpersonen behoorde tijdens de ziekte tot de vegetarische groep.

Vermoed werd, dat tijdens de ziekte een daling in het cholesterolgehalte zou uittreden, gevolgd door een aanmerkelijke stijging tijdens het herstel. Het is mogelijk, dat tijdens de ziekte en een korte tijd daarna onregelmatigheden in de voeding hebben plaatsgevonden. Stijgingen na de ziekte zouden dan het gevolg kunnen zijn van meer en vleesrijker voedsel. Een daling tijdens de ziekte in de vegetarische periode is waarschijnlijk niet te wijten aan dieetveranderingen of voedselstoornissen, daar tijdens dit dieet praktisch geen cholesterol wordt opgenomen (zie Inleiding).

Wij beperkten ons daarom tot een onderzoek naar de dalingen tijdens de ziekte. Verder is het noodzakelijk in verband met de invloed van de standaardwaarden S , dat alleen de dalingen van die gezonde en zieke mensen onderling worden vergeleken, waarbij de bepalingen op eenzelfde standaardwaarde S zijn betrokken. Voor deze laatste conditie werd het bruikbare getallenmateriaal erg gering en konden wij ons ook niet uitsluitend beperken tot de proefpersonen van het vegetarische dieet. Met behulp van de boven beschreven combinatie van een aantal toetsingen volgens Wilcoxon (met een kleine wijziging tengevolge van de ongelijke uitgebreidheden der steekproeven) is er toch nog getracht een uitspraak te verkrijgen. Wij hebben weer een rij van paren onderling vergelijkbare groepen (zie tabel IX). De te toetsen hypothese is: de dalingen van de zieke en gezonde mensen uit elke rij kunnen beschouwd worden als twee steekproeven uit eenzelfde collectie. Wij vonden een (tweezijdige) overschrijdingskans 0,064. Deze overschrijdingskans is

wel aan de lage kant, maar kan toch niet significant genoemd worden. Hoewel men hieruit niet kan besluiten tot een daling van het cholesterolgehalte ten gevolge van de ziekte, zou men deze uitkomst kunnen beschouwen als een aanwijzing om bij verder onderzoek dit vraagstuk zorgvuldig in het oog te houden.

3.5 Opmerking.

Of de volgorde van de diëten ook van invloed is op het gedrag van het cholesterolgehalte kon niet worden onderzocht, daar de proefpersonen zeer waarschijnlijk niet willekeurig over de 6 groepen verdeeld waren. Dit blijkt als volgt.

Getoetst werd de hypothese, dat de cholesterolwaarden van groep Ia, b en die van groep IIIa, b twee steekproeven zijn uit eenzelfde verdeling. Hiervoor werden juist de groepen I en III genomen, omdat bij eventuele fysiologisch bepaalde voorkeur, deze voorkeur zich eerder zal uiten in de extreme richtingen V en H dan in de richting van het neutrale dieet L. De toetsing werd uitgevoerd met de reeds eerder genoemde methode van Wilcoxon, zowel voor de waarde bij binnenkomst als voor de gemiddelden van de drie bovengenoemde sommen. De resultaten waren in beide gevallen significant. Voor de binnenkomstwaarden werd een overschrijdingskans 0,007 en voor de gemiddelden een overschrijdingskans 0,02 gevonden. De proefpersonen die op het V-dieet kwamen (groep I) hadden systematisch hoger cholesterolgehalte dan zij die het H-dieet kregen.

4. Conclusies.

1. De bepalingstechniek, waarbij een aantal bepalingen op één standaard-waarde worden betrokken, heeft ten gevolge dat er een duidelijke afhankelijkheid tussen de waarnemingen ontstaat. Deze afhankelijkheid maakt de statistische analyse moeilijker en werkt het ontstaan van schijneffecten in de hand. Aan de conclusies kan slechts dan waarde worden gehecht, indien zij verkregen zijn volgens een methode, waarbij met deze afhankelijkheid rekening is gehouden.
2. Uit de verrichte duplobepalingen volgt, dat de spreiding van onafhankelijke duplobepalingen ongeveer gelijk is aan
3. De voeding heeft een duidelijke invloed op het cholesterolgehalte van het bloedserum. Dit gehalte neemt

neemt toe in de richting V L H.

4. Personen met een laag cholesterolgehalte reageren minder sterk op dieetverandering dan personen met een middelmatig of hoog cholesterolgehalte.

5. Er is geen verschil gevonden tussen de cholesterolgehalten of de veranderingen daarvan bij mannen en bij vrouwen,

6. Hoewel niet kon worden aangetoond, dat griep het cholesterolgehalte in het bloedserum doet dalen, was er toch een zwakke aanwijzing in deze richting.

7. De verdeling der personen over de diverse groepen bij het begin van de proef was niet onafhankelijk van hun cholesterolgehalte.

Opmerkingen: De onder 1 en 7 genoemde afhankelijkheden bemoeilijken het statistisch onderzoek. Het valt aan te bevelen te trachten bij een volgend onderzoek deze afhankelijkheden te vermijden.

Tabel I

Verdeling van de groepen over de drie diëten voor elk der perioden

Groep	1e periode	2e periode	3e periode
Ia	V	L	H
Ib	V	H	L
IIa	L	V	H
IIb	L	H	V
IIIa	H	L	V
IIIb	H	V	L

Tabel II

Overeenstemmend gedrag in de tijd tussen de proefperso-
nen uit eenzelfde groep

Groep en dieet	Ia,H	Ib,L	IIa,V	IIb,L	IIIa,H	IIIb,V
Overschrij- dingskans	0,02	0,02	0,01	0,02	0,05	0,05

Tabel III

Overeenstemmend gedrag in de tijd tussen de groepgemid-
delden.

periode	1	2	3
overschrij- dingskans	< 0,1	< 0,1	< 0,01

Afhankelijkheden tussen de cholesterolwaarden en de daarmee corresponderende waarden van S.

Periode I 28/11 - 5/1			Periode II 13/2 - 30/3			Periode III 15/5 - 23/6											
Groep	nr.pr.p.	Groep nr.pr.p.	Groep nr.pr.p.	Groep nr.pr.p.	Groep nr.pr.p.	Groep nr.pr.p.	Groep nr.pr.p.										
Ia	1	+ 0,40	I Ib	25	+ 0,20	Ia	1	- 0,43	I Ib	25	- 0,24	Ia	1	- 0,62	I Ib	25	- 0,43
	3	- 0,20		29	- 0,33		3	- 0,14		29	+ 0,40		3	- 0,61		29	- 0,71
	5	- 0,20		31	0,00		5	0,00		31	- 0,39		5	- 0,67		31	- 0,33
	7	+ 0,60		35	+ 0,50		7	- 0,33		35	- 0,43		7	- 0,47		35	- 0,20
	9	+ 0,60		39	- 0,33		9	- 0,22		39	+ 0,34		9	- 0,31		39	- 0,47
	2	+ 0,40		24	+ 0,33		2	- 0,36		24	- 0,17		2	- 0,54		24	- 0,47
	4	- 0,33		26	- 0,20		4	- 0,60		26	+ 0,06		4	+ 0,20		26	- 0,43
	8	- 0,20		30	- 0,40		8	- 0,13		30	- 0,17		8	- 0,38		30	- 0,73
	10	+ 0,47		38	- 0,33		10	+ 0,03		38	+ 0,66		10	- 0,57		38	- 0,47
	12	- 0,14		40	- 0,17		12	- 0,03		40	- 0,15		12	- 0,51		40	- 0,50
Ib	11	- 0,07	IIIa	41	- 1,00	Ib	11	- 0,29	IIIa	41	+ 0,30	Ib	11	- 0,52	IIIa	41	- 0,25
	13	- 0,21		51	- 0,07		13	- 0,10		51	- 0,39		13	- 1,00		51	- 0,14
	15	+ 0,33		53	- 0,40		15	- 0,05		53	- 0,22		15	- 0,17		53	- 0,87
	17	- 0,80		55	- 0,20		17	- 0,11		55	- 0,35		17	- 0,33		55	- 0,55
	19	+ 0,33		59	0,00		19	+ 0,24		59	- 0,50		19	- 0,44		59	- 0,14
	6	+ 0,40		42	- 0,33		6	+ 0,20		42	- 0,31		6	- 0,57		42	- 0,43
	14	- 0,13		44	+ 0,73		14	+ 0,07		44	- 0,53		14	- 0,78		44	- 0,43
	16	- 0,33		46	+ 0,13		16	- 0,43		46	- 0,64		16	- 0,40		46	- 0,04
	18	- 0,40		50	- 0,52		18	+ 0,18		50	- 0,42		18	- 0,39		50	- 0,42
	20	+ 0,33		60	- 0,47		20	- 0,50		60	- 0,47		20	- 0,33		60	- 0,20
IIa	21	+ 0,07	IIIb	48	- 0,51	IIa	21	+ 0,22	IIIb	48	- 0,46	IIa	21	+ 0,20	IIIb	48	- 0,24
	23	- 0,40		52	0,00		23	- 0,07		52	- 0,32		23	- 0,52		52	- 0,51
	27	- 0,20		54	0,07		27	0,00		54	- 0,38		27	- 1,00		54	- 0,67
	33	- 0,30		56	- 0,33		33	- 0,07		56	- 0,24		33	- 0,39		56	- 0,60
	37	- 0,33		58	- 0,07		37	- 0,02		58	- 0,44		37	- 0,05		58	- 0,47
	22	+ 0,27		43	- 0,47		22	- 0,04		43	- 0,55		22	- 0,52		43	- 0,67
	28	- 0,20		45	- 0,20		28	- 0,13		45	- 0,10		28	- 0,50		45	- 0,71
	32	- 1,00		47	- 0,05		32	- 0,17		47	- 0,24		32	- 0,22		47	+ 0,07
	34	+ 1,00		49	- 0,40		34	- 0,20		49	- 0,43		34	- 0,20		49	- 0,25
	36	+ 0,20		57	- 0,14		36	+ 0,13		57	- 0,39		36	+ 0,33		57	- 0,73
	$\tau > 0$		$\tau = 0$		$\tau < 0$		$\tau > 0$		$\tau = 0$		$\tau < 0$		$\tau > 0$		$\tau = 0$		$\tau < 0$
	19		3		38		12		0		48		4		0		56

Tabel V

Aantallen paren proefpersonen met overeenkomstig gedrag, resp. tegengesteld gedrag in de gevallen a en b.

	$(W_1 - W'_1)(W_2 - W'_2) > 0$	$(W_1 - W'_1)(W_2 - W'_2) < 0$
a: $S_1 = S_2$	63	20
$S'_1 = S'_2$		
b: $S_1 \neq S_2$	44	36
$S'_1 \neq S'_2$		

Tabel VI

Duplobepalingen

De aantallen in de klassen A en B (2.5.1)

A	B
$(S_1 - S_2)(W_1 - W_2) > 0$	$(S_1 - S_2)(W_1 - W_2) < 0$
162	64

Tabel VII

Onderzoek van de dieetovergangen met de symmetrietoets T_2

nr. pr.p.	L	H	nr. pr.p.	H	L	nr. pr.p.	L	V	nr. pr.p.	V	L	nr. pr.p.	H	V	nr. pr.p.	V	H
1	+	240	11	-	97	41	-	239	1	+	199	43	-	206	21	+	207
3	+	174	13	-	21	51	+	157	3	+	303	45	-	167	23	+	591
5	+	108	15	-	78	53	-	15	5	+	268	47	-	181	27	+	144
7	+	180	17	+	174	55	-	55	7	+	174	49	-	322	33	+	588
9	+	191	19	-	111	59	-	30	9	+	379	57	-	235	37	+	584
2	+	222	6	-	138	42	-	60	2	+	257	48	-	153	22	+	369
4	+	126	14	-	119	44	+	30	4	+	105	52	-	297	28	+	595
8	+	291	16	+	20	46	-	99	8	+	439	54	-	25	32	+	381
10	+	185	18	-	50	50	-	4	10	+	134	56	-	43	34	+	310
12	+	364	20	-	21	60	+	113	12	+	348	58	-	147	36	+	237
25	+	75	41	-	276	21	-	11	43	+	236	25	-	62	11	+	217
29	+	39	51	-	224	23	-	167	45	+	329	29	-	175	13	+	423
31	-	24	53	-	65	27	+	123	47	+	326	31	+	20	15	+	485
35	+	145	55	-	122	33	-	134	49	+	87	35	-	164	17	+	93
39	+	40	59	-	145	37	-	109	57	+	277	39	-	138	19	+	387
24	+	92	42	-	47	22	-	259	40	+	201	24	-	181	6	+	495
26	+	373	44	-	170	28	-	112	52	+	336	26	-	407	14	+	472
30	+	138	46	-	96	32	-	166	54	+	108	30	-	15	16	+	174
38	+	207	50	-	72	34	-	34	56	+	170	38	-	246	18	+	253
40	+	149	60	-	28	36	+	5	58	+	3	40	-	195	20	+	136
Overschrijdingskansen																	
0,00002			0,00393			0,05303			0,00000			0,00002			0,00000		

+ betekent: het cholesterolgehalte is bij de dieetovergang gestegen.

Tabel VIII

Verband tussen reactie en gehalte
(De tabellen bevatten de rangnummers der spreidingen bij opklimmende gemiddelden.)

1. Ia	3	2	6	5	1	8	7	9	10	4
Ib	1	3	10	8	9	6	5	4	2	7
IIa	1	4	8	10	7	5	3	9	2	6
IIb	2	9	7	4	6	1	10	8	3	5
IIIa	2	1	3	7	8	6	9	5	4	10
IIIb	1	2	5	9	7	3	6	8	10	4
totaal	10	21	39	43	38	29	40	43	31	36
	$v = 9 \quad \chi^2 = 18,4 \quad \text{o.k.} = 0,03$									

2. Ia	4	3	1	6	5	7	8	2
Ib	8	6	7	4	3	2	1	5
IIa	6	8	5	3	2	7	1	4
IIb	6	3	5	1	8	7	2	4
IIIa	1	5	6	4	7	3	2	8
IIIb	3	7	5	1	4	6	8	2
totaal	28	32	29	19	29	32	22	25
	$v = 7 \quad \chi^2 = 4,2 \quad \text{o.k.} = 0,75$							

Tabel IX

Vergelijking van zieke en gezonde proefpersonen.

nr.pr.	27	28	34	22	21	23	
daling	23	35	40	38	- 18	- 22	
nr.pr.	49	54	45				
daling	25	4	- 2				
nr.pr.	53	48	59	57			
daling	5	77	- 24	14			
nr.pr.	48	60	46				
daling	100	31	22				
nr.pr.	50	56					
daling	42	19					
nr.pr.	45	40	38	41	36		
daling	10	25	- 15	16	23		
nr.pr.	42	33	43	35	34	32	31
daling	20	28	- 3	- 30	- 22	- 2	22

Algemene gang van zaken bij het toetsen van een ¹⁾
hypothese.

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

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

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

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van u , en wel op zodanige wijze, dat de kans, dat u een in Z gelegen waarde aanneemt, onder de hypothese H_0 , gelijk is aan een gegeven getal α , zodat Z dus van α afhankelijk is. 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 verworpt nu H_0 op grond van de waarnemingen x_1, x_2, \dots, x_n , indien de bij deze waarnemingen behorende waarde van u in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien H_0 juist is, gelijk aan α . Derhalve is α de kans op ten onrechte verwerping van de juiste hypothese, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met $\alpha = 0,05$ 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 zoverre 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 toetsingsgrootte 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 vergelijk men bv. de tweede hieronder gegeven literatuurplaats. 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.

Tekentoets¹⁾

Deze toets dient voor het toetsen van de hypothese H_0 , dat een aantal grootheden z_1, \dots, z_n alle nul tot mediaan hebben, d.w.z. dat

$$P [z_i > 0] = P [z_i < 0] \quad i = 1, \dots, n,$$

is. De toets geldt zonder enige verdere beperking dan de eis, dat de grootheden z_i onderling onafhankelijk verdeeld zijn; zij behoeven niet dezelfde waarschijnlijkheidsverdeling te bezitten.

De toets berust op één waarneming van ieder der grootheden z_i , dus op n waarnemingen z_1, \dots, z_n . De waarnemingen, die de waarde 0 bezitten, laten wij buiten beschouwing²⁾. Als toetsingsgrootte gebruiken wij nu n_1 , het aantal positieve waarnemingen. Zijn er m waarnemingen $\neq 0$, dan bezit n_1 een binomiale verdeling, onderstellende, dat H_0 juist is:

$$P [n_1 = n_1 | H_0] = \binom{m}{n_1} 2^{-m}.$$

Als kritieke zône worden de grote en kleine waarden van n_1 genomen. De kritieke zône is, voor onbetrouwbaarheidsdrempels 0,01; 0,05; 0,10 en 0,25 en $m = 1$ tot 100 getabelleerd door

W.J. Dixon and A.M. Mood, The statistical sign test, Jrn. Am. Stat. Ass. 41 (1946) p. 556-566.

Voor een groter aantal waarnemingen gebruikt men als benadering van de binomiale verdeling de aangepaste normale verdeling.

Opmerking: De toets wordt vaak gebruikt, indien men een aantal grootheden twee maal heeft waargenomen, voor en na een bepaalde gebeurtenis, om na te gaan of deze gebeurtenis invloed op de grootheden heeft uitgeoefend. Noemen wij de waarnemingen vóór het optreden der gebeurtenis x_i ($i=1, \dots, n$) en erna y_i , dan hebben de grootheden $x_i - y_i$ alle 0 als mediaan, indien x_i dezelfde verdeling bezit als y_i (dus als de gebeurtenis geen invloed heeft gehad). De toets wordt nu toegepast op $z_i = x_i - y_i$ ($i=1, \dots, n$).

-
- 1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.
 - 2) In tegenstelling tot de gewoonte deze waarnemingen voor de helft bij de positieve en voor de helft bij de negatieve te tellen; de door ons gebruikte methode geeft de toets een groter onderscheidingsvermogen.

De toets van Wilcoxon. 1)

Deze methode dient tot het toetsen van de hypothese H_0 , inhoudend dat twee steekproeven x_1, \dots, x_n en y_1, \dots, y_m afkomstig zijn uit één collectie (ook populatie of universum genaamd). Zij is strict genomen, toepasbaar onder de voorwaarde, dat er geen enkel paar waarden (x_i, y_j) is met $x_i = y_j$. Verdere voorwaarden zijn voor de toepassing niet nodig, terwijl de zojuist genoemde, indien er niet teveel dergelijke paren zijn, de toepassing van de toets weinig hindert.

De toetsingsgrootheid U is het aantal paren (x_i, y_j) waarvoor $x_i > y_j$ is (het aantal "inversies"). Daar er $n \cdot m$ dergelijke paren zijn, kan U alle gehele waarden van 0 tot en met $n \cdot m$ aannemen. Is U groot, dan liggen er veel waarden x_i verder naar rechts dan waarden y_j , is U klein, dan juist weinig.

De kritieke zône K neemt nu daarom de kleine en de grote waarden van U en wel van beide zoveel, dat de gekozen onbetrouwbaarheidsdrempel α niet overschreden wordt.

Voor éénzijdige toetsing, te onderscheiden in linker- en rechter-éénzijdige toetsing, gebruikt men kritieke zônes K_1 , resp. K_2 , die geheel bestaan uit kleine, resp. grote waarden van U .

Verwerping van H_0 ten gevolge van het vinden van een grote (resp. kleine) waarde van U wijst erop, dat x_1, \dots, x_n en y_1, \dots, y_m steekproeven uit verschillende collecties zijn, waarbij de op de x -collectie aangenomen waarden systematisch groter (resp. kleiner) dan de op de y -collectie aangenomen waarden zijn.

Litteratuur:

- F. Wilcoxon, Individual comparisons by ranking methods, Biometrics 1 (1945), p.80-83.
- H.B. Mann and D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann.Math.Stat. 18 (1947), p.50-60. Bevat tabellen voor n en $m \leq 8$.
- H.R. van der Vaart, Some remarks on the power function of Wilcoxon's test for the problem of two samples, Proceedings van de Kon.Ned.Ak.v.Wet., 53 (1950), p.494-520.
- H.R. van der Vaart, Gebruiksaanwijzing voor de toets van Wilcoxon, met tabellen voor n en $m \leq 10$, Rapport S32 (M₄)(1950).
- D. van Dantzig, Kadercursus Mathematische Statistiek, Math. Centrum, Amsterdam(1947-50), hoofdstuk 6, 3.

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

S 47 (M 13)

RANGCORRELATIE ¹⁾

De door M.G. Kendall ontwikkelde methode der rangcorrelatie is toepasbaar op de volgende situatie:

De stochastische grootheden x en y bezitten een simultane verdeling. Over deze verdeling zelf behoeft niets ondersteld te worden.

(x_i, y_i) ($i = 1, \dots, n$), zijn onafhankelijk waarnemingsparen van deze stochastische grootheden.

De x_i worden gerangschikt naar opklimmende grootte en vervolgens vervangen door hun rangnummers $1, 2 \dots n$. De y_i verkrijgen op die manier een rangschikking, die niet overeen hoeft te komen met die volgens opklimmende grootte en worden nu vervangen door de rangnummers, die bij deze rangschikking naar opklimmende grootte behoren.

Voorbeeld:

x_i : 0,28 0,42 1,14 1,15 1,21 1,76 1,96 2,53 3,12 3,48
 y_i : 36,4 38,1 28,2 34,1 29,7 25,3 31,3 26,8 32,7 40,2

rangnummers der x_i : 1 2 3 4 5 6 7 8 9 10

" " y_i : 8 9 3 7 4 1 5 2 6 10

De rangcorrelatiecoëfficiënt r dient als maat voor de graad van overeenkomst tussen deze twee rijen rangnummers

Deze coëfficiënt wordt zo gedefinieerd dat bij volledige overeenstemming tussen de twee rijen $r = +1$ en bij precies tegenovergestelde rangschikking $r = -1$ is.

Voor iedere andere rangschikking ligt r tussen de grenzen $+1$ en -1 .

We definiëren daarvoor eerst de grootte P als volgt:

P is de som van het aantal paren rangnummers der (y_i, y_j) dat in de normale volgorde staat, dus waarbij het linker rangnummer het laagste is.

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

We berekenen P als volgt: (zie v.b.)

We beginnen met het eerste rangnummer van de rij der y_i : dit is 8 en staat voor de hogere rangnummers 9 en 10. Het geeft dus een bijdrage 2 aan P .

Het tweede rangnummer van de rij, 9, staat voor 10 en geeft dus een bijdrage 1 aan P .

Het derde van de rij, 3, staat voor 7, 4, 5, 6, 10 en geeft dus een bijdrage 5 aan P , enz.

$$P = 2 + 1 + 5 + 1 + 3 + 4 + 2 + 2 + 1 = 21.$$

P wordt maximaal als de y_i dezelfde rangorde bezitten als de x_i ; dan is $P = \frac{1}{2} n(n-1)$.

$$\text{We definiëren nu } z = \frac{2P}{\frac{1}{2}n(n-1)} - 1.$$

$$\text{In dit voorbeeld vinden we: } z = \frac{2 \times 21}{\frac{1}{2} \cdot 10 \cdot 9} - 1 = -0.07.$$

Voor het geval dat x en y onafhankelijk zijn is de verzameling van z door M.G. Kendall berekend. De hypothese, dat x en y onafhankelijk zijn, kan dus getoetst worden.

Voor $n > 10$ wordt de verdeling door de normale verdeling zeer goed benaderd. Is de hypothese van onafhankelijkheid niet vervuld, dan is de waarschijnlijkheid van grote waarde van $|z|$ groter, dan wanneer dit wel het geval is. De kritieke zone is daarom van de vorm $|z| \geq z_0$, en bij éénzijdige toetsing van de vorm $z \geq z_0$ (rechtszijdige toetsing) of $z \leq z_0'$ (linkszijdige toetsing). De bij dit systeem van kritieke zones behorende overschrijdingskansen vinden we voor $n \leq 10$ getabelleerd in [I] (tabel 1). (Hierbij wordt in plaats van z de grootheid $S = \frac{1}{2} n(n-1)z$ gebruikt).

Het kan gebeuren, dat onder de x_i of y_i gelijke waarnemingen voorkomen. We geven deze gelijke waarnemingen dan ook gelijke rangnummers en wel op volgende wijze: Veronderstel, dat we aan het 2^e t/m 7^e rangnummer geen rangverschil kunnen toekennen dan kiezen we als rangnummer voor ieder van deze waarnemingen het gemiddelde van deze rangnummers 2, 3, ..., 7, dus

$$\frac{1}{5} (2 + 3 + 4 + 5 + 6 + 7) = 4\frac{1}{2}$$

Daar het maximum van P nu kleiner wordt, moeten we een correctie op de formule voor z toepassen, die in [I], pag. 25 beschreven staat. Ook voor dit geval is de verdeling van z bekend.

Litteratuur: [I] M.G. Kendall, Rank Correlation Methods, London 1948, hoofdstuk 1.
Tabel verdelingsfunctie van S voor $n = 4$ t/m 10, pag. 141.

Symmetrietoets¹⁾.

Hypothese H_0 : de waarnemingen z_1, \dots, z_n , zijn afkomstig van n onafhankelijk verdeelde stochastische grootheden, die alle symmetrisch ten opzichte van 0 verdeeld zijn²⁾. Van deze toets bestaan meerdere versies T_1, \dots, T_2 . We bespreken eerst T_1 en T_2 .

Toetsingsgrootheden. Deze worden als volgt uit z_1, \dots, z_n afgeleid:

1e. de waarnemingen, die gelijk aan 0 zijn worden weggelaten. Stel er blijven over: z_1, \dots, z_{n_1} .

2e. Hieruit worden de positieve waarnemingen gezocht. Stel dit zijn x_1, \dots, x_{n_1} , dus n_1 in aantal.

3e. De overblijvende negatieve waarnemingen worden van teken veranderd, zodat zij ook positief worden. Stel dit zijn dan y_1, \dots, y_{n_2} .

4e. De grootheden x_1, \dots, x_{n_1} en y_1, \dots, y_{n_2} worden, door elkaar, in afdalende grootte-volgorde opgeschreven. Stel dit geeft: w_1, \dots, w_n . (Komen er gelijken voor, dan worden deze in willekeurige volgorde geplaatst.)

5e. De groep waarden w_1, \dots, w_n wordt verdeeld in twee groepen w_1, w_2, \dots, w_r en w_{r+1}, \dots, w_n , waarbij $w_r \neq w_{r+1}$ is en r zo dicht mogelijk bij de waarde $\frac{1}{2}n$ genomen wordt. Is n even, dan wordt $r = \frac{1}{2}n$, indien althans $w_{\frac{1}{2}n} \neq w_{\frac{1}{2}n+1}$ is. Zijn er twee mogelijk keuzen voor r , beide op gelijke afstand van $\frac{1}{2}n$, dan nemen wij $r > \frac{1}{2}n$. Is b.v. n oneven en $w_{\frac{1}{2}n - \frac{1}{2}} \neq w_{\frac{1}{2}n + \frac{1}{2}}$, dan nemen wij $r = \frac{1}{2}n + \frac{1}{2}$. Wij geven de waarden w_1, \dots, w_r aan als groep A (die dus r elementen bevat) en de overigen als groep B. Alle elementen van A zijn dus groter dan ieder element van B³⁾.

-
- 1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.
 - 2) Zetten wij hier a in plaats van 0, dan geldt H_0 voor $x_1 - a, \dots, x_{n_1} - a$.
 - 3) In de oorspronkelijke publicaties over deze toets (zie de literatuurverwijzingen aan het einde van dit memorandum) is een enigzins minder algemene definitie van r gegeven. Alle stellingen blijven echter gelden, indien de hier gegeven definitie gebruikt wordt.

6e. Het aantal waarden van x_1, \dots, x_n die in A voorkomen noemen wij u.

De toetsingsgrootheden zijn n_1 en u , r is een hulpgroothed.

V.B. 22 waarden z_i : 7, 4/6, 3/3, 6/3, 5/3, 4/2, 9/2, 5/1, 1/0/0/-1, 3/-2, 5/-3, 2/-4, 6/-4, 5/-4, 6/-4, 5/-5, 3/-7, 0/-7, 9/-8, 0/-8, 7.

∴ $r = 11, \quad n_1 = 8$
 $u = 2$

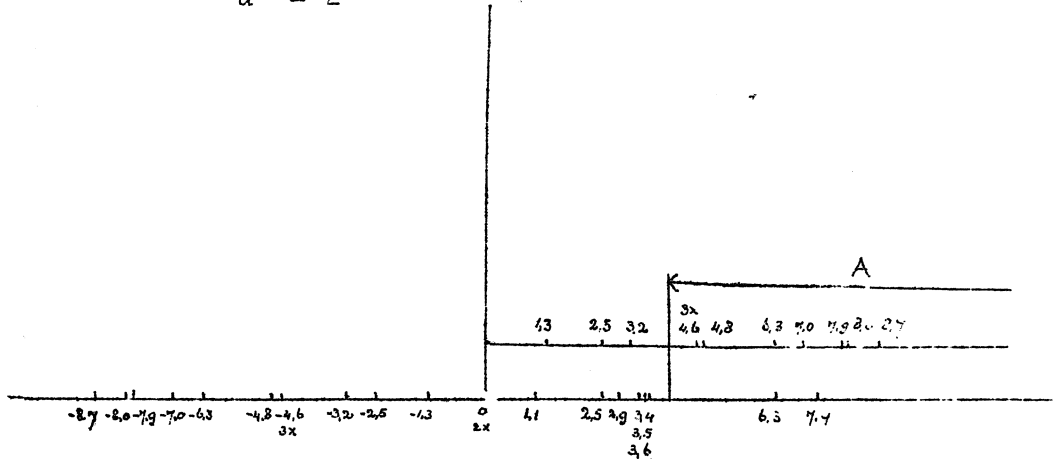


fig. 1

Kritieke zônes. Waarden van n_1 , die dicht bij 0 of dicht bij n liggen, zullen, als H_0 juist is weinig, maar als H_0 onjuist is vaker, voorkomen. Grote resp. kleine waarden van u zullen eveneens, als H_0 juist is weinig voorkomen. Hierop berust de keuze van de bij T_1 en T_2 behorende kritieke zônes Z_1 resp. Z_2 . Z_1 bevat grote en kleine waarden van n_1 en grote en kleine waarden van u, terwijl Z_2 bij grote waarden van n_1 in hoofdzaak grote waarden van u en bij kleine waarden van n_1 in hoofdzaak kleine waarden van u bevat. T_1 leidt bij voldoende grote n vrijwel steeds tot verwerping als de hypothese niet is vervuld. T_2 leidt echter alleen tot verwerping van H_0 als er veel positieve (resp. negatieve) waarden zijn, die verder van 0 verwijderd liggen dan de negatieve (resp. positieve). In sommige gevallen is het juist van belang om deze laatste afwijkingen van H_0 te ontdekken. In dat geval gebruikt men T_2 liever dan T_1 . In fig. 2 is een schematisch voorbeeld gegeven van een serie waarnemingen, waarbij het aantal positieve groter is dan het aantal negatieve, terwijl deze positieve dichter bij 0 liggen dan de negatieve, zodat T_2 niet tot verwerping leidt.

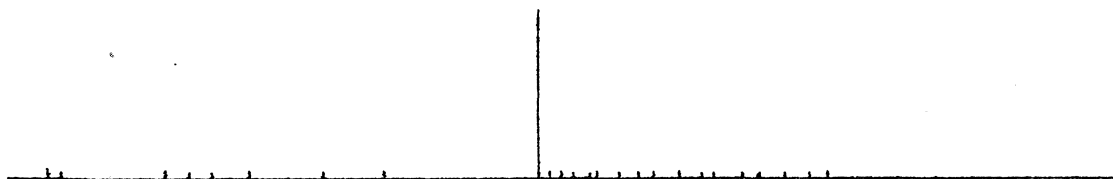


fig. 2

Van T_1 en T_2 bestaan ook éézijdige versies, waarvan de beschrijving te ver zou voeren.

T'_1 en T'_2 .

Toetsingsgrootheden.

1e, 2e en 3e als boven. Eerste toetsingsgrootheid: n_1 .
4e: op x_1, \dots, x_{n_1} en y_1, \dots, y_{n_2} wordt de toets van Wilcoxon toegepast (vgl. S 47 (M 8)). De toetsingsgrootheden zijn n_1 en de U van deze toets van Wilcoxon.

Kritieke zônes. Overwegingen analoog aan die voor T_1 en T_2 (met U in plaats van u) leiden tot analoge kritieke zônes Z'_1 en Z'_2 , behorend bij T'_1 en T'_2 .

Opmerkingen: T_1 en T_2 zijn bijzonder geschikt voor een niet te groot aantal waarnemingen. Zij gelden ook voor niet continue verdelingen. T'_1 en T'_2 zijn alleen geschikt, als er geen of weinig paren (x_i, y_j) met $x_i = y_j$ zijn. Voor grote aantallen zijn T'_1 en T'_2 geschikter dan T_1 en T_2 . Er is ook een versie voor grote aantallen (T''_1 en T''_2), die geheel analoog is met T'_1 en T'_2 met dien verstande, dat u in plaats van U wordt gebruikt (vgl. b.v. [2], blz. 77, § 6.4.5).

Litteratuur:

- [1] J. Hemelrijk, A family of parameterfree tests for symmetry with respect to a given point, I, II. Proceedings van de Kon.Ned.Ak.v.Wet. 53 (1950), p.945-955. Indagationes Mathematicae 12 (1950), p. 340-350.
[2] - " - , Symmetrietoetsen, Diss., Den Haag 1950, Excelsior.

Methode der m rangschikkingen ¹⁾

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

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

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

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

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

1 -3 -4 5 -2 3

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

De som der kwadraten van deze afwijkingen noemen we S

- 2 -

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 W onder de hypothese \mathcal{H}_0 is exact berekend voor een aantal waarden van n en m [1] terwijl voor grote n en m benaderingen bekend zijn. [1] Hiermee kan \mathcal{H}_0 dus getoetst worden, waarbij \mathcal{H}_0 verworpen wordt, als W waarden dichtbij 1 aanneemt. De kritieke zone is dus van de vorm $W \geq W_0$

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

Veronderstel, dat door een waarnemer geen onderscheid wordt gemaakt tussen de elementen, die de rangnummers 3 t/m 6 moeten dragen. Dan wordt als rangnummer van iedere van deze elementen het gemiddelde van de rangnummers

$$\frac{1}{4} (3 + 4 + 5 + 6) = 4\frac{1}{2}$$

gebruikt.

Daar het maximum van S nu verandert, moeten we een correctie op de formule van W toepassen; deze vindt men in het boek van Kendall behandeld.

Litteratuur:

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

tabel verdelingsfunctie van W voor

$$n = 3 \quad m = 2 \text{ t/m } 10$$

$$n = 4 \quad m = 2 \text{ t/m } 6$$

$$n = 5 \quad m = 3$$

op p. 146-149

Toetsing van de hypothese $p_1 = p_2$ met behulp
 van een 2 x 2-tabel¹⁾.

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

$$H_0: p_1 = p_2.$$

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

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

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

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

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

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

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

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

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

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

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

Litteratuur:

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

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