

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 106

De beïnvloeding van de fosphaatstofwisseling in de
rattenlever door vasten en grote doses vitamine D₂

door

J. de Boer

	<u>Inhoud</u>	blz
1.	Inleiding	1
2.	Statistische analyse van de waarnemingen	1
3.	Conclusies	5

1. Inleiding.

Het doel van het onderhavige onderzoek is: na te gaan of de aanmaak- en aanvoersnelheden van intracellulair anorganisch fosphaat en phospholipoiden in de rattenlever beïnvloed worden door vasten en grote doses vitamine D₂. Een directe meting of berekening van de aanvoer- resp. aanmaaksnelheden of ook maar van de na een bepaalde tijd aangevoerde of gemaakte hoeveelheid was niet mogelijk, daar de afvoer niet gemeten kan worden. De woorden "aanvoer" en "aanmaak" worden daarom in dit verslag gebruikt voor "aanvoer verminderd met afvoer of omzetting" en "aanmaak verminderd met afvoer of omzetting", waarbij "afbraak" in "omzetting" begrepen is.

Er zijn 3 groepen ratten bij het onderzoek betrokken:

N-groep, d.w.z. normale dieren

V-groep, d.w.z. dieren die 24 uur gevast hebben

D-groep, d.w.z. dieren die 24 uur tevoren 2 mg vitamine D₂ gekregen hebben.

Alle ratten krijgen op een bepaald tijdstip voor de dood, die door decapitatie plaats vindt, een even grote injectie van radioactieve phosphor (P₃₂), in de vorm van anorganisch fosphaat in de buikholte. Het fosphaat komt dan via het bloed achtereenvolgens in de extracellulaire ruimte en de intracellulaire ruimte van de lever, om daarna omgezet te worden in (phospho)lipoid. Na de dood wordt de hoeveelheid anorganisch fosphaat in het bloedserum en in de intracellulaire ruimte van de lever, en de hoeveelheid phospholipoid in de lever bepaald. De waarnemingen die gedaan worden voor I (= Intracellulair fosphaat) zijn onderling onafhankelijk, evenals de waarnemingen voor L (= Lipoid).

Nagegaan wordt of de toename van de genoemde hoeveelheden in de tijd, dus de aanmaak- en aanvoersnelheden, voor de groepen ratten al of niet systematisch verschillen.

2. Statistische analyse van de waarnemingen.

Van de lever wordt een zuur extract gemaakt, waaruit het anorganisch fosphaat verwijderd wordt. Met een Geiger-teller wordt het aantal aanslagen per minuut bepaald dat door een deel van het extract veroorzaakt wordt, zowel vóór als na de genoemde verwijdering. Deling van het verschil van deze aantallen door het aantal aanslagen, veroorzaakt door een standaardhoeveelheid phosphor (gecorrigeerd op een nuleffect), geeft een waarneming van de hoeveelheid radioactief anorganisch fosphaat, in de lever aanwezig. Eigenlijk moeten we zeggen: hoeveelheid op een constante factor na, maar omdat deze factor niet essentiëel is voor de vergelijking die we op het oog hebben, spreken we hier en in het vervolg kortweg van hoeveelheid. Het anorganisch fosphaat in de lever bevindt zich in de intracel-

lulaire en de extracellulaire ruimte. De concentratie van het fosfaat in de extracellulaire ruimte wordt gelijk verondersteld aan die in het bloedserum, waarmee een vrije uitwisseling plaats vindt. Uit een radiometing als boven vinden we de hoeveelheid fosfaat in het serum en dus ook in de extracellulaire ruimte, als we het volume daarvan kennen. Daaruit volgt weer de hoeveelheid fosfaat in de intracellulaire ruimte, door aftrekking.

Om de hoeveelheid lipoid te bepalen, voeren we weer een deling uit van het aantal aanslagen, veroorzaakt door resp. een deel van het lipoid en de standaard.

De verschaftte gegevens geven dan bij rangschikking naar de groepen en de tijd t, verlopen tussen injectie en dood (5, 10 of 20 minuten), de volgende resultaten:

Tabel I
Hoeveelheden radioactief intracellulair fosfaat en lipoid ¹⁾

t	I (Intracellulair fosfaat)			L (Lipoid)		
	5	10	20	5	10	20
N	173	303	361	9	39	61
	217	320	393	24	43	62
	424	401	599	30	51	107
					71	162
V	276	389	574	18	39	69
	296	465	699	22	46	115
	329	630	829	27	48	116
					111	157
D	142	198	349	18	19	64
	179	293	431	33	37	67
	234	312	457	45	62	85
				61	120	168

De waarden in deze tabel zijn in fig. 1 en fig. 2 getekend, en de gemiddelden van een groep op de verschillende tijdstippen door een gebroken lijn verbonden. Dit zijn tevens de regressielijnen (zie onder c)).

We bestuderen nu de aanvoersnelheden van het intracellulair fosfaat en de aanmaaksnelheden van het lipoid voor de 3 groepen:

¹⁾ Zie pag. 1.

- a) Over de gehele tijd
- b) In de eerste 5 minuten
- c) In de tweede 5 minuten en de tweede 10 minuten.

a) Over de gehele tijd (zie fig. 1 en 2).

We vergelijken de drie groepen, twee aan twee bij I (voor L analoog) met de toets van Wilcoxon (zie memorandum S 47 (M 7))²⁾ op $t = 5$, $t = 10$ en $t = 20$. Dit levert ons 3 toetsingsgrootheden op: U_5 , U_{10} en U_{20} . Zijn de gemiddelden van deze 3 grootheden onder de nulhypothese resp. μ_5 , μ_{10} en μ_{20} , en de spreidingen resp. σ_5 , σ_{10} en σ_{20} , dan is

$$U = \frac{U_5 + U_{10} + U_{20} - (\mu_5 + \mu_{10} + \mu_{20})}{\sqrt{\sigma_5^2 + \sigma_{10}^2 + \sigma_{20}^2}}$$

bij benadering normaal verdeeld met gemiddelde 0 en spreiding 1, onder de hypothese dat er geen systematisch verschil tussen de hoeveelheden I (resp. L) van de 2 betrokken groepen op $t = 5$, $t = 10$ en $t = 20$.

We toetsen deze hypothese nu door te toetsen of U bovengenoemde normale verdeling bezit, en vinden de volgende overschrijdingskansen:

Tabel II

Tweezijdige overschrijdingskansen van de gecombineerde toetsen van Wilcoxon

	I	L
N - V	0,06	0,58
N - D	0,17	0,38
V - D	$2 \cdot 10^{-3}$	0,93

Er is, over de gehele tijd, dus alleen een duidelijk aantoonbaar systematisch verschil tussen de V- en de D- groep bij I, en wel zijn de hoeveelheden I van de V-groep groter dan die der D-groep. Verder is er een zwakke aanwijzing, dat, eveneens bij I, de V-groep hoger ligt dan de N-groep. Overigens zijn er geen redenen, om de getoetste hypothese te verwerpen.

b) Eerste 5 minuten.

Omdat de hoeveelheden op $t = 0$ alle nul zijn, vergelijken we op

2) De memoranda, waarnaar verwezen wordt, bevinden zich aan het eind van dit rapport. Ter algemene oriëntatie is memorandum S 47 (M 6) over de gang van zaken bij het toetsen van een hypothese, toegevoegd.

t = 5 weer met de toets van Wilcoxon. Deze passen we toe op dezelfde combinaties als onder a).

Als resultaat vinden we:

Tabel III

Tweezijdige overschrijdingskansen van de toets van Wilcoxon bij t = 5

	I	L
N - V	0,51	0,83
N - D	0,51	0,60
V - D	0,05	0,60

Alleen voor I bij V - D vinden we dus een aanwijzing voor een verschil; de aanvoersnelheid van I in de eerste 5 minuten is volgens deze aanwijzing bij de V-groep systematisch hoger dan bij de D-groep.

c) Tweede 5 en tweede 10 minuten.

We vergelijken weer de drie groepen twee aan twee, bij I en bij L, voor t = 5, 10 en voor t = 10, 20. Met een toets voor de gelijkheid van twee regressiecoëfficiënten (zie memorandum S 73 (M 37), § 3) toetsen we of er systematisch verschil is tussen de betreffende snelheden.

In fig. 1 en fig. 2 zijn de regressielijnen getekend die bij deze toetsing een rol spelen. We toetsen nu eerst de gevallen waarbij de hellingen der regressielijnen het meest uiteenlopen. Dit zijn de gevallen

Bij I: N - V 5 - 10, N - V 10 - 20 en V - D 5 - 10

Bij L: V - D 5 - 10, V - D 10 - 20.

Vinden we hierbij geen kleine overschrijdingskansen, dan behoeven we de overige toetsingen niet uit te voeren, omdat die dan met hun minder uiteenlopende hellingen zeker ook grote overschrijdingskansen zullen geven.

Tabel IV

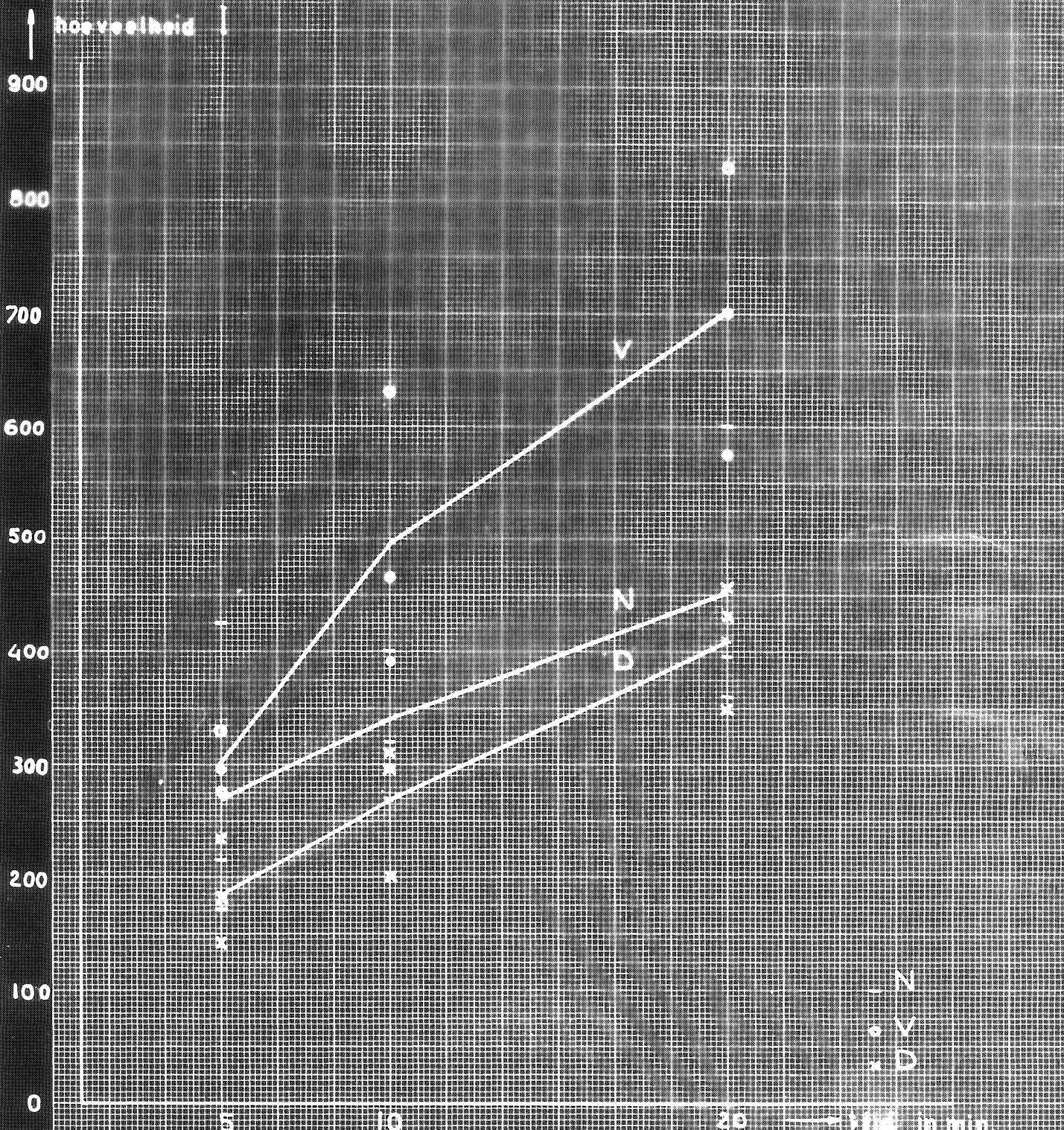
Overschrijdingskansen van de regressiecoëfficiëntentoets

		I	L
N - V	5 - 10	0,32	
N - V	10 - 20	0,33	
V - D	5 - 10	0,50	0,10
V - D	10 - 20		0,31

We vinden dus geen systematische verschillen. De aanvaardbaarheid van de aan deze toets ten grondslag liggende normaliteitsonderstelling kan wegens het, statistisch gezien, kleine waarnemingsmateriaal niet getoetst worden. Een parameter vrije toets, waarbij deze veronderstelling niet wordt gemaakt, is voor dit probleem echter nog niet ontwikkeld.

3. Conclusies.

De hoeveelheid intracellulair fosfaat is bij de V-groep gedurende de gehele tijd systematisch hoger dan bij de D-groep. Verder zijn er zwakke aanwijzingen, dat deze hoeveelheid voor de V-groep ook hoger is dan voor de N-groep en dat de aanvoersnelheid van het intracellulaire fosfaat in de eerste 5 minuten voor de V-groep systematisch groter is dan voor de D-groep. Statistisch kunnen eventuele verdere verschillen niet aangetoond worden. Hierdoor wordt echter de conclusie, dat dergelijke **verschillen niet** bestaan, niet gerechtvaardigd, vooral niet, daar de omvang van het waarnemingsmateriaal zeer klein is. Een gevolg daarvan is nl., dat men in de regel slechts **grote** systematische verschillen zal ontdekken, terwijl voor het vinden van **kleinere** verschillen meer waarnemingen vereist zijn.



Hoeveelheid in racolubitionsmiddel in de N.V. en Dimes op 5, 10 en 20 minuten na de injectie

De getrokken lijnen verbinden de gemiddelden

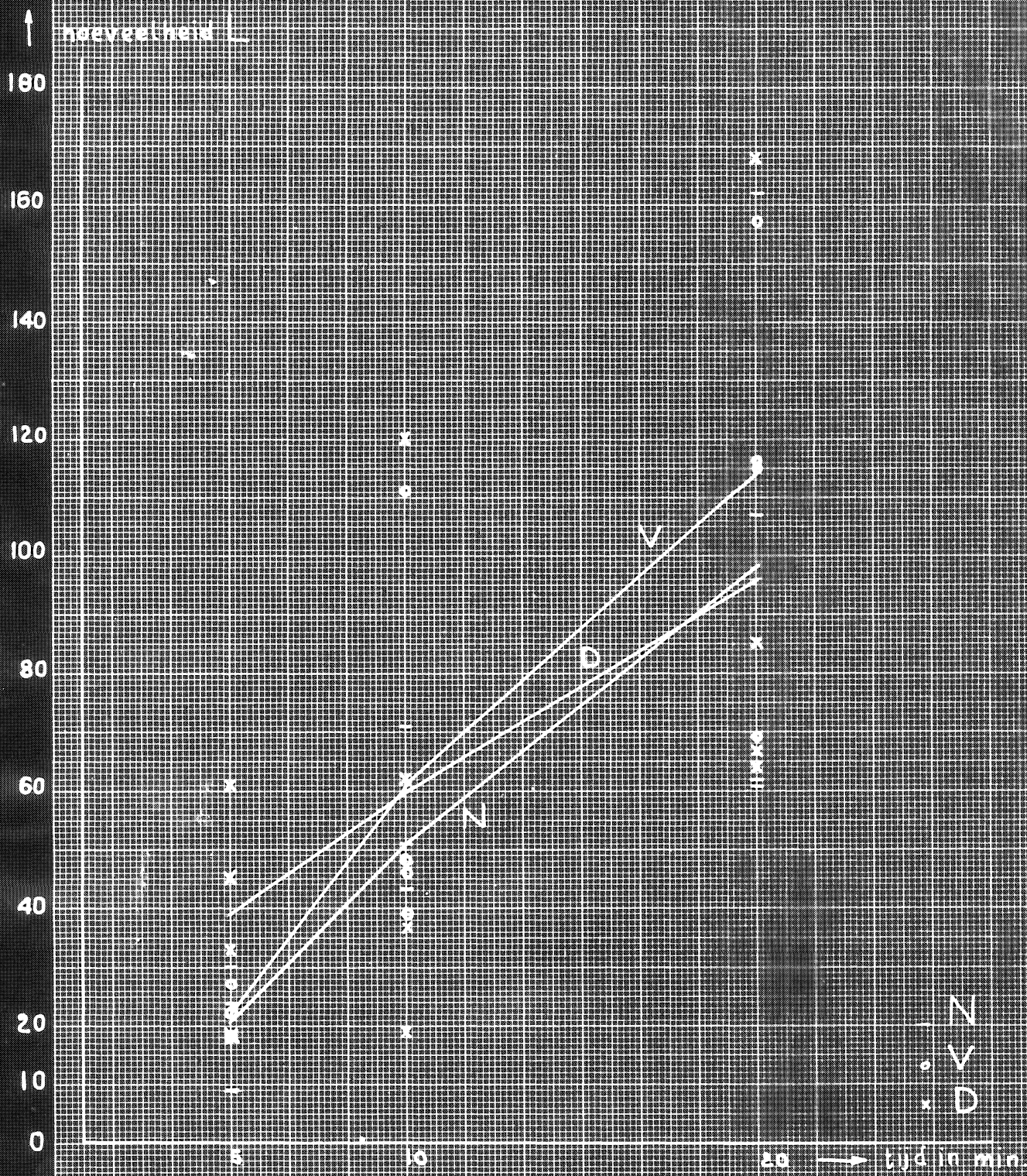


FIG. 2

Hoeveelheid Lipoid in de N-, V- en D groep op 5, 10 en 20 minuten na de injectie

De getrokken lijnen verbinden de gemiddelde en

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 grootheden²⁾, 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 zone 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 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 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 toetsingsgrootte nog juist in de (bij α behorende) kritieke zone 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.

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 toetsingsgrootte \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 $\frac{1}{2}$ 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 toetsingsgrootte \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 toetsingsgrootte \underline{U} onder de hypothese H_0 voor grote waarden van n en m (beide ≥ 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, \dots, 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

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

en

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

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 verworpt indien de gevonden waarde U van \underline{U} te sterk van μ afwijkt, d.w.z. als

$$\frac{|U - \mu|}{\sigma} > \xi_{\alpha} \quad 2)$$

waarin α de onbetrouwbaarheidsdrempel is en ξ_{α} volgt uit

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

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

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

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

en kan ook in een tabel van de normale verdeling worden gevonden.

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

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

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

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

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

Indien n en m kleiner zijn dan 10, zijn tabellen beschikbaar voor het berekenen van de overschrijdingskans k voor de uit de steekproef bepaalde waarde U van \underline{U} (zie [2] en [4]).

Dergelijke tabellen bestaan echter niet voor het geval, dat gelijke waarnemingen optreden.

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

Litteratuur:

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

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - 0.

Statistische Afdeling
Rapport S 73(M 37)
door
Mevr. G. Klerk-Grobbe.

Toetsing van de gelijkheid van een aantal gemiddelden of een aantal regressiecoëfficiënten als toepassing van de grondstelling der variantieanalyse¹⁾.

We zullen uitgaan van de bij de variantieanalyse gebruikelijke onderstelling, dat de oorspronkelijke waarnemingen alle onderling onafhankelijk normaal²⁾ verdeeld zijn met dezelfde (onbekende) spreiding σ .

1. De grondstelling der variantieanalyse.

De grondstelling der variantieanalyse luidt als volgt:

Zij gegeven waarnemingen w_1, w_2, \dots, w_n van de stochastische variabelen $\underline{w}_1, \underline{w}_2, \dots, \underline{w}_n$ ³⁾, die onderling onafhankelijk normaal verdeeld zijn met dezelfde (onbekende) spreiding σ , en waarvan de mathematische verwachtingen (theoretische gemiddelden), aangegeven door

$$\begin{aligned} E\underline{w}_1 &= \mu_1, \\ E\underline{w}_2 &= \mu_2, \\ &\vdots \\ E\underline{w}_n &= \mu_n, \end{aligned}$$

voldoen aan k onderling onafhankelijke lineaire relaties:

$$(1.1) \quad \begin{aligned} L_1(\mu_1, \dots, \mu_n) &= 0 \\ &\vdots \\ L_k(\mu_1, \dots, \mu_n) &= 0. \end{aligned}$$

-
- 1) Dit memorandum dient slechts ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.
2) De stochastische grootte \underline{x} ³⁾ bezit een normale verdeling met gemiddelde μ en spreiding σ , indien voor iedere a geldt:

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

- 3) De onderstropping geeft aan, dat de grootte stochastisch is, d.w.z. een waarschijnlijkheidsverdeling bezit.

Verder bestaat de te toetsen hypothese uit nog ℓ onderling en van (1.1) onafhankelijke lineaire relaties tussen μ_1, \dots, μ_n :

$$(1.2) \quad \begin{aligned} L_{k+1}(\mu_1, \dots, \mu_n) &= 0 \\ \vdots \\ L_{k+\ell}(\mu_1, \dots, \mu_n) &= 0 \end{aligned}$$

Als toetsingsgrootheid voor de toetsing der hypothese nemen we:

$$F = \frac{k}{\ell} \frac{Q_r - Q_a}{Q_a},$$

waarin Q_a de waarde is van $\sum_{i=1}^n (w_i - \mu_i)^2$, wanneer we hierin op grond van de waarnemingen w_1, \dots, w_n de kleinste kwadraten schattingen voor μ_1, \dots, μ_n , welke voldoen aan (1.1) substitueren, terwijl Q_r de waarde van $\sum_{i=1}^n (w_i - \mu_i)^2$ is, wanneer we de kleinste kwadraatschattingen voor μ_1, \dots, μ_n , welke aan (1.1) én (1.2) voldoen, invullen.

Beschouwen we niet alleen de gevonden waarden w_1, \dots, w_n , maar de verzameling van alle waarden, die w_1, \dots, w_n kunnen aannemen, dan bezitten ook de grootheden Q_a en Q_r en dus F een waarschijnlijkheidsverdeling. Is de te toetsen hypothese vervuld dan is F verdeeld als de F van Snedecor met ℓ en k vrijheidsgraden. Is de hypothese niet vervuld, dan zullen de grotere waarden van F een grotere waarschijnlijkheid bezitten. Als kritieke zône bij de toetsing gebruiken we daarom de eenzijdige zône:

$$F \geq F_0,$$

waarin F_0 , behorende bij een bepaalde onbetrouwbaarheidsdrempel ε , opgezocht kan worden in tabellen van de F -verdeling met ℓ en k vrijheidsgraden (zie b.v. [3] of [4]). De beschreven toets is een λ -toets in de zin van Neyman (zie [2], hoofdstuk 5).

Een bewijs van bovenstaande stelling is o.a. te vinden bij Mann [1] hoofdstuk 4.

2. Toetsing van de gelijkheid van een aantal gemiddelden.

Bij deze toetsing zullen we voor de grootheden w_1, \dots, w_n een nieuwe notatie invoeren, nl.:

$$\begin{aligned} &\underline{y}_{11}, \underline{y}_{12}, \dots, \underline{y}_{1n_1}; \\ &\underline{y}_{21}, \underline{y}_{22}, \dots, \underline{y}_{2n_2}; \\ &\vdots \\ &\underline{y}_{s1}, \underline{y}_{s2}, \dots, \underline{y}_{sn_s}. \end{aligned}$$

Van deze grootheden is gegeven, dat hun verwachtingen voldoen

aan de volgende $\sum_{i=1}^s (n_i - 1)$ onafhankelijke lineaire relaties:

$$(2.1) \quad \begin{aligned} EY_{1j} &= \mu_1; & j &= 1, \dots, n_1 \\ EY_{2j} &= \mu_2; & j &= 1, \dots, n_2 \\ &\vdots & & \\ EY_{sj} &= \mu_s; & j &= 1, \dots, n_s. \end{aligned}$$

De te toetsen hypothese legt aan de verwachtingen nog $s-1$ van (2.1) onafhankelijke lineaire restricties op:

$$(2.2) \quad \mu_1 = \mu_2 = \dots = \mu_s.$$

De grootheden Q_a en Q_r worden nu:

$$Q_a = \sum_{i=1}^s \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2,$$

waarin

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad (i = 1, \dots, s)$$

de kleinste kwadratenschatting voor μ_i ($i = 1, \dots, s$) is, welke aan (2.1) voldoet, en

$$Q_r = \sum_{i=1}^s \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2,$$

waarin

$$\bar{y} = \frac{\sum_{i=1}^s \sum_{j=1}^{n_i} y_{ij}}{\sum_i n_i}$$

de kleinste kwadratenschatting voor μ_i ($i = 1, \dots, s$) is, welke aan (2.1) en (2.2) voldoet. Hieruit vinden we:

$$Q_r - Q_a = \sum_{i=1}^s n_i \bar{y}_i^2 - \bar{y}^2 \sum n_i.$$

De toetsingsgrootte is dus:

$$\underline{F} = \frac{\sum_i (n_i - 1)}{s - 1} \cdot \frac{\sum_{i=1}^s n_i \bar{y}_i^2 - \bar{y}^2 \sum n_i}{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2}$$

onder de te toetsen hypothese verdeeld als de F van Snedecor met $s-1$ en $\sum_i (n_i - 1)$ vrijheidsgraden.

Opmerking: Is $s = 2$, vergelijken we dus de gemiddelden van twee steekproeven, dan is de bovengebruikte toetsingsgrootte \underline{F} het kwadraat van de toetsingsgrootte \underline{t} , welke we bij toetsing van de gelijkheid van twee gemiddelden met behulp van de toets van Student gebruiken. Bovenstaande toetsing is dus op te vatten als een generalisatie voor meer steekproeven

van de toets van Student voor twee steekproeven. Dezelfde opmerking kunnen we maken bij de volgende toepassing.

3. Toetsing van de gelijkheid van een aantal regressiecoëfficiënten.

Voor de grootheden w_1, \dots, w_n gebruiken we ook hier de notatie:

$$\begin{aligned} & y_{11}, \dots, y_{1n_1} \\ & y_{21}, \dots, y_{2n_2} \\ & \cdot \\ & \cdot \\ & y_{s1}, \dots, y_{sn_s} \end{aligned}$$

Van de verwachtingen van deze grootheden is nu gegeven, dat ze voldoen aan de relaties:

$$(3.1) \quad \begin{aligned} E y_{1j} &= \beta_1 \xi_{1j} + \alpha_1 & (j = 1, \dots, n_1) \\ &\cdot \\ &\cdot \\ E y_{sj} &= \beta_s \xi_{sj} + \alpha_s & (j = 1, \dots, n_s) \end{aligned}$$

waarin de β 's en α 's onbekende parameters zijn, maar de waarden ξ_{ij} ($j = 1, \dots, n_i; i = 1, \dots, s$) exact bekend zijn. Deze betrekkingen vormen $\sum_{i=1}^s (n_i - 2)$ onafhankelijke relaties tussen de grootheden $E y_{ij}$ ($j = 1, \dots, n_i; i = 1, \dots, s$). De te toetsen hypothese zegt:

$$(3.2) \quad \beta_1 = \beta_2 = \dots = \beta_s (= \beta),$$

wat neerkomt op nog $(s-1)$, van (2.1) onafhankelijke, relaties tussen de grootheden $E y_{ij}$.

De kleinste kwadratenschattingen voor β_i en α_i , alleen gelet op (3.1), zijn

$$\text{voor } \beta_i: \quad \underline{b}_i = \frac{\sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi}_i) y_{ij}}{\sum_j (\xi_{ij} - \bar{\xi}_i)^2} \quad (i = 1, \dots, s),$$

$$\text{voor } \alpha_i: \quad \underline{a}_i = \bar{y}_i - \underline{b}_i \bar{\xi}_i \quad (i = 1, \dots, s),$$

waarin

$$\bar{y}_i = \frac{1}{n_i} \sum_j y_{ij} \quad \text{en} \quad \bar{\xi}_i = \frac{1}{n_i} \sum_j \xi_{ij}.$$

Gelet op (3.1) en (3.2) gaan deze schattingen over in:

$$\text{voor } \beta_i: \quad \underline{b} = \frac{\sum_{i=1}^s \sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi}_i) y_{ij}}{\sum_{i=1}^s \sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi}_i)^2},$$

$$\text{voor } \alpha_i: \quad \underline{a}_i = \bar{y}_i - \underline{b} \bar{\xi}_i \quad (i = 1, \dots, s).$$

Ter vereenvoudiging zullen we in de volgende formules

$$K_i \stackrel{\text{def}}{=} \sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi}_i)^2$$

substitueren. Voor de grootheden Q_a en Q_r volgt nu:

$$\begin{aligned} Q_a &= \sum_i \sum_j (y_{ij} - b_i \xi_{ij} - a_i)^2 = \\ &= \sum_i \sum_j (y_{ij} - \bar{y}_i - b_i (\xi_{ij} - \bar{\xi}_i))^2 = \\ &= \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 - \sum_i b_i^2 K_i \end{aligned}$$

en

$$Q_r = \sum_i \sum_j (y_{ij} - \bar{y}_i - b(\xi_{ij} - \bar{\xi}_i))^2$$

en dus

$$Q_r - Q_a = \sum_i b_i^2 K_i - b^2 \sum_i K_i$$

zodat de toetsingsgrootte overgaat in

$$F = \frac{\sum_i (n_i - 2) \sum_i b_i^2 K_i - b^2 \sum_i K_i}{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2 - \sum_i b_i^2 K_i}$$

welke dus onder de te toetsen hypothese verdeeld is als de F van Snedecor met $(s-1)$ en $\sum_i (n_i - 2)$ vrijheidsgraden.

Litteratuur

- [1] H.B. Mann, Analysis and design of experiments analysis of variance and analysis of variance designs. Dover Publications, Inc., New York 1949.
- [2] J. Neyman, First course in probability and statistics. Henry Holt and Co., New York 1950. Hoofdstuk 5.
- [3] R.A. Fisher and F. Yates, Statistical Tables for Biological agricultural and medical research, 3^d Ed., Oliver & Boyd, London 1949 Table V.
- [4] A.M. Mood, Introduction to the theory of Statistics, Mc Graw-Hill, New York-Toronto-London, 1950, hoofdstuk 13 (regressietheorie), pp. 426-427 (tabel F-verdeling).