

Statistische Afdeling
Rapport S 60
door Mej. G. Grobben
en
Dr J. Hemelrijk Jr

Statistisch verslag omtrent een ijkmethode
voor leverextracten⁵

1. Inleiding.

Om te onderzoeken of door proefnemingen op konijnen een activiteitsbepaling van leverextracten uitvoerbaar is, werd bij 18 konijnen in de loop van 5 jaar de verandering van de "kleurindex" onderzocht bij injectie met verschillende leverpreparaten na een voorafgaande "vergiftiging" met lood (reeks (8) tot en met (17)).

Ter vergelijking werden eveneens de veranderingen van de kleurindex onderzocht bij injectie met aminozuren (reeks (5)) of iobalt (reeks (6)) of te sterk verdunde preparaten van met lood vergiftigde konijnen (reeks (7)). Als blanco bepalingen werden opgegeven: de normale kleurindexschommelingen bij onbehandelde dieren (reeks (1)), de kleurindexverandering bij injectie van niet vergiftigde konijnen met een leverpreparaat (reeks (2)), de kleurindexverandering na injectie van vergiftigde dieren met fysiologisch zout (reeks (3)) en de kleurindexschommelingen tijdens het toedienen van het lood (reeks (4)).

De kleurindex werd waargenomen om de 6 uur, gedurende een periode van ongeveer 72 uur na de inspuiting (indien er een inspuiting plaats vond; anders gewoon gedurende 72 uur).

Als maat voor de verandering van de kleurindex werd de maximale kleurindexdaling genomen, uitgedrukt in procenten van de hoogste waarde voor de daling.

De uitkomsten der verschillende reeksen werden door de onderzoeker, om een voorlopige indruk te krijgen, vergeleken met behulp van de toets van Student. De uitkomsten van deze toets waren aan het materiaal toegevoegd.

Zoals uit de bijgevoegde histogrammen (zie bijlage I) blijkt is de verdeling der waarnemingen bij de meeste behandelingswijzen verre van normaal; alleen de waarnemingen der normale kleurindexschommelingen (reeks (1)) vormen een uitzondering.

Bij 4 reeksen is op de waarnemingen een toets voor normaliteit volgens R.C.Geary en E.S.Pearson (zie S 47 (M 6) en S 47 (M 16)) toegepast; de resultaten zijn te vinden in tabel I.

Tabel I.
Normaliteitsstoetsing.

reeks	α	$\sqrt{b_1}$	ov.kans voor $\sqrt{b_1}$	ov.kans voor
①	0.763	0.25	$\gg 0.05$	> 0.1
③ _W - Mrt niet	0.799	0.895	< 0.05	> 0.1
①⑦ _Z	0.88	1.03	< 0.05	> 0.1
①⑦	0.785	0.944	0.01	> 0.1

In de eerste kolom betekent b.v. "③_W - Mrt niet": die waarnemingen uit de 3e reeks, die in de wintermaanden gedaan zijn, waarbij we die in Maart weglaten. Als wintermaande werden bij dit onderzoek genomen: Januari, Februari, Maart, October, November en December.

De overige verdelingen wijken, blijkens hun histogrammen nog sterker van de normale af. Door vergelijking van de gevonden waarden van α en $\sqrt{b_1}$ met de door Geary en Pearson opgegeven waarden blijkt:

- 1) De waarden van α wijken niet sterk af van die voor normale verdelingen, waarbij α in de buurt van 0.8 behoort te liggen (d.w.z. het maximum van de verdeling is niet stomper of scherper dan bij een steekproef uit een normale verdeling verwacht zou kunnen worden).
- 2) De waarde van $\sqrt{b_1}$ bij reeks ① wijkt niet te veel af van die bij een normale verdeling ($\sqrt{b_1}=0$).
- 3) Bij de proeven ③_W, ①⑦_Z en ①⑦ is de te grote waarden van $\sqrt{b_1}$ aanleiding tot het verwerpen van een normale verdeling en tot de conclusie dat de verdelingen scheef zijn.

Op grond van deze resultaten is bij de verdere onderzoeken de toets van Wilcoxon (zie S 47 (M 7)) gebruikt, waarbij geen normaliteit behoeft te worden ondersteld.

2. Vergelijking van de reacties op een preparaat-injectie met de blanco-bepalingen.

De resultaten van dit onderzoek, welke dus vergeleken kunnen worden met de uitkomsten, vermeld bij het materiaal, zijn samengevat in tabel II.

Tabel II.
Toetsing van de werking der preparaten.

vergelijking reeks	tussen reeks	$\underline{u}$	$\underline{m}$	kleinste waarn.	overschr. kans	opgeg. ov.kans
① (28)	② (11)	47.5	154	①	0.0009	« 0.01
① (28)	③ (60)	674.5	840	①	0.14	0.03
① (28)	④ (33)	509.5	462	④	0.50	0.28
③ (60)	④ (33)	1229	990	④	0.056	« 0.01
③ (60)	⑤ (11)			③	0.00001	-
③ (60)	⑥ (2)			③	0.048	-
③ (60)	⑦ (7)	208.5	210	③	0.9	-
③ (60)	⑧ (8)	21.5	240	③	0.00003	« 0.01
③ (60)	⑨ (6)	233	180	⑨	0.25	0.45
③ (60)	⑩ (4)	45	120	③	0.044	-
③ (60)	⑪ (21)	118.5	630	③	0.00001	« 0.01
③ (60)	⑫ (6)	191	180	⑫	0.84	-
③ (60)	⑬ (6)	54	180	③	0.0060	-
③ (60)	⑮ (5)	17.5	150	③	0.0014	-
③ (60)	⑯ (3)	13	90	③	0.01	-
③ (60)	⑰ (33)	236.5	990	③	0.00001	« 0.01

In tabel II en de volgende betekenen de tussen haakjes geplaatste getallen de aantallen waarnemingen in de desbetreffende waarnemingsreeksen. De eerste en tweede kolom geven de reeksen die vergeleken worden, in kolom drie is $\underline{u}$ het aantal inversies (= aantal malen dat een waarneming van de eerste reeks groter is dan één van de tweede), kolom vier geeft de verwachtingswaarde van $\underline{u}$ (deze is gelijk aan $\frac{mn}{2}$, als m en n de aantallen der waarnemingen in de beide waarnemingsseries zijn), kolom vijf vermeldt de reeks waarin de waarnemingen het laagst waren en kolom zes geeft de twee-zijdige overschrijdingskansen bij de gevonden waarden voor $\underline{u}$.

Vergelijking met de bij het materiaal opgegeven overschrijdingskansen, die met behulp van de toets van Student berekend zijn en in de laatste kolom opgenomen werden,

geeft een goede overeenstemming te zien, behalve bij de vergelijking van (3) en (4). Dit laatste kan een gevolg zijn van de sterke afwijking van normaliteit, die (3) en (4) vertonen (zie de histogrammen van (3) en (4) in bijlage I). De toets van Student moet in dit geval als onbetrouwbaar worden beschouwd.

3. Onderzoek van het verstrekte materiaal.

Bij voorafgaande onderzoek is er nog geen rekening mee gehouden, dat de proeven onder vele verschillende omstandigheden zijn verricht, waardoor verschillen zouden kunnen optreden, die niet het gevolg zijn van het verschil in de preparaten-toediening.

Teneinde nu de onderlinge vergelijkbaarheid der verschillende reeksen waarnemingen na te gaan, is in principe een splitsing van de waarnemingen van iedere reeks naar de volgende kenmerken gewenst:

- 1) het jaar, waarin de proeven zijn verricht;
- 2) de jaargetijden, waarin de proeven vielen;
- 3) proeven in Maart en April en in overige maanden (Maart en April vormen de bronstijd);
- 4) zomer en winter;
- 5) het geslacht der konijnen;
- 6) de konijnen zelf.

De onder 1) en 2) genoemde splitsingen konden niet worden uitgevoerd, daar er dan in iedere reeks te weinig waarnemingen van verschillende jaren resp. jaargetijden overbleven.

Splitsing 3.

Bij verschillende reeksen (waar de aantallen waarnemingen dit toelieten) werd de invloed van Maart en April onderzocht door de waarnemingen in Maart en April te vergelijken met de overige waarnemingen. De resultaten zijn verwerkt in tabel III.

Tabel III
Invloed van Maart en April.

reeks	vergelijking tussen maanden		<u>u</u>	<u>E_u</u>	kleinere waarn.	overschr. kans
①	Mrt+Apr(6)	rest(22)	62.5	66	Mrt+Apr	0.8
③	Mrt+Apr(20)	rest(40)	294.5	400	Mrt+Apr	0.037
③ _w ♀	Mrt (6)	rest(13)	36.5	39	Mrt	0.78
③ _z ♀	Apr (5)	rest(8)	10.5	20	Apr	0.18
③ _w ♂	Mrt (5)	rest(13)	12.5	32.5	Mrt	0.052
③ _z ♂	Apr (4)	rest(6)	15.5	12	rest	0.54
④	Mrt+Apr(8)	rest(25)	162.5	100	rest	0.012
⑧	Mrt+Apr(3)	rest(5)	12	7.5	rest	0.25
⑪	Mrt+Apr(6)	rest(15)	48.5	45	rest	0.8

In tabel III komen 2 significante verschillen voor tussen de maanden Maart en April en de overige maanden n.l. bij reeks ③ en reeks ④. Daar reeks ③ de voor- naamste blancobepaling is, is bij de volgende onderzoe- kingen steeds nagegaan welk effect het weglaten der waarnemingen in Maart en April heeft op de resultaten.

Splitting 4.

De resultaten van de toetsing tussen zomer- en win- terwaarnemingen van de reeksen, waarbij dit mogelijk was, worden gegeven in tabel IV.

Tabel IV.
Vergelijking van zomer- en winterwaarnemingen.

reeks	vergelijking tussen		<u>u</u>	<u>E_u</u>	kleinere waarn.	overschr. kans
①	zomer(19)	winter(9)	78.5	85.5	zomer	0.76
③	zomer(23)	winter(37)	279	425.5	zomer	0.026
③ ♀	zomer(13)	winter(19)	77	123.5	zomer	0.08
③ ♂	zomer(10)	winter(18)	62	90	zomer	0.19
④	zomer(20)	winter(13)	81.5	130	zomer	0.09
⑦	zomer(3)	winter(4)	7.5	6	winter	0.07
⑪	zomer(9)	winter(12)	66	54	winter	0.42
⑪	zomer(23)	winter(10)	168.5	115	winter	0.038
⑪ ♀	zomer(10)	winter(5)	27.5	25	winter	0.75
⑪ ♂	zomer(13)	winter(5)	56	32.5	winter	0.022

vervolg Tabel IV.

De waarnemingen in Maart en April weggelaten:

reeks	vergelijking tussen		<u>u</u>	<u>Eu</u>	Kleinere waarn.	overschr. kans
①	zomer(17)	winter(3)	21.5	25.5	zomer	0.68
③	zomer(14)	winter(26)	114.5	182	zomer	0.057
③ ♀	zomer(8)	winter(13)	42	52	zomer	0.48
③ ♂	zomer(6)	winter(13)	16.5	39	zomer	0.055
④	zomer(20)	winter(5)	50	50	•	1.
⑪	zomer(7)	winter(8)	39	28	winter	0.23

Het verschil tussen zomer- en winterwaarnemingen is bij de proeven ③ en ⑪ (bij splitsing volgens de geslachten alleen bij mannelijke dieren in ⑪) significant. Hoewel door weglaten voor de waarnemingen in Maart en April deze significantie bij ③ verdwijnt (bij ⑪ kwamen geen waarnemingen in Maart en April voor, dus blijft de significantie bestaan) is toch voor alle zekerheid bij het volgende gelet op ^{een} eventueel verschil tussen zomer- en winterwaarnemingen.

Splitsing 5.

De uitkomsten van de vergelijking der waarnemingsuitkomsten bij mannelijke en vrouwelijke dieren zijn verzameld in tabel V, (weer voor zover mogelijk in verband met de aantallen waarnemingen).

Tabel V.

Onderzoek naar de invloed van het geslacht.

reeks	vergelijking tussen		<u>u</u>	<u>Eu</u>	kleinere waarn.	overschr. kans
① z	♂ (11)	♀ (8)	52	44	♀	0.54
③	♂ (28)	♀ (31)	455.5	434	♀	0.64
③ z	♂ (10)	♀ (13)	71.5	65	♀	0.70
③ w	♂ (18)	♀ (19)	168	171	♂	0.9
⑪	♂ (18)	♀ (15)	129.5	135	♂	0.9
⑪ z	♂ (13)	♀ (10)	77	65	♀	0.48
⑪ w	♂ (5)	♀ (5)	7.5	12.5	♂	0.76
⑪	♂ (7)	♀ (14)	61.5	49	♀	0.36
⑪ z	♂ (2)	♀ (7)	10	7	♀	0.072
⑪ w	♂ (5)	♀ (7)	22.5	17.5	♀	0.5

Vervolg Tabel V.

De waarnemingen in Maart en April weggelaten:

reeks	vergelijking tussen		$\underline{u}$	$\underline{E_u}$	kleinere waarn.	oversch. kans
③	♂ (19)	♀ (21)	206	199.5	♀	0.9
③ z	♂ (6)	♀ (8)	18	24	♂	0.49
③ w	♂ (13)	♀ (13)	93	84.5	♀	0.68
④ z	♂ (9)	♀ (11)	51	49.5	♀	0.9
⑪	♂ (5)	♀ (10)	30	25	♀	0.59
⑪	♂ (18)	♀ (15)	142.5	135	♀	0.84
⑪ z	♂ (13)	♀ (10)	77	65	♀	0.48
⑪ w	♂ (5)	♀ (5)	5.5	12.5	♂	0.20

Uit deze tabel blijkt dat er op grond van dit materiaal geen reden is onderscheid te maken tussen mannelijke en vrouwelijke dieren, daar er geen kleine overschrijdingskansen voorkomen.

Splitsing 6.

Om te onderzoeken of er een verschil bestaat tussen de konijnen werden de konijnen verdeeld in konijnen met een grote en een kleine kleurindexdaling volgens kolom ① en het verschil tussen deze beide groepen onderzocht bij verschillende reeksen.

Ook werd op dezelfde wijze gesplitst volgens grote en kleine daling in reeks ③ en deze verdeling getoetst bij andere reeksen.

Tenslotte werden, daar er bij splitsing volgens ① een significant verschil tussen beide groepen bij ④ gevonden werd en bovendien vele waarnemingen uit ① en 4 in dezelfde periode verricht waren n.l. Juli en Augustus 1947, de konijnen, alleen voor zover het de waarnemingen in deze maanden betrof, nog eens gesplitst in hoge en lage waarnemingen en deze verdeling werd getoetst in reeks ①.

De resultaten zijn opgenomen in tabel VI.

Tabel VI.

Onderzoek naar individuele verschillen tussen de konijnen.

Splitsing in hoog en laag volgens ①.						
proef	vergelijking tussen		<u>u</u>	<u>Eu</u>	lagere waarn.	oversch. kans
③	laag (27)	hoog (25)	271.5	337.5	lage	0.40
④	laag (16)	hoog (15)	69	120	lage	0.05
⑤	laag	hoog	16	15	hoog	0.92
⑪	laag	hoog	63.5	55	hoog	0.57
⑰	laag	hoog	66	55	hoog	0.46

Splitsing in hoog en laag volgens ③

Maart en April waarnemingen weggelaten

① _Z	laag (9)	hoog (10)	54	45	hoog	0.49
④ _Z	laag (9)	hoog (8)	43.5	32	hoog	0.51
⑪	laag (6)	hoog (6)	12	18	laag	0.39
⑰	laag (16)	hoog (15)	122	120	hoog	1.
⑰ _Z	laag (11)	hoog (10)	47.5	55	laag	0.60
⑰ _W	laag (5)	hoog (5)	14.5	12.5	hoog	0.69

Splitsing in hoog en laag volgens ④

alleen voor de waarnemingen in Juli en Augustus 1947

①	laag (9)	hoog (10)	38.5	45	laag	0.60
---	----------	-----------	------	----	------	------

Uit de grote overschrijdingskansen van deze tabel volgt dat er geen reden is om systematische verschillen van enige betekenis tussen de bij de verschillende konijnen verkregen uitkomsten aan te nemen.

4. Definitief onderzoek naar verschillen tussen de kolommen.

De reeksen onderling zijn nog eens vergeleken, maar nu met de onderscheiding tussen zomer en winter en voor zover mogelijk nog met weglating van de waarnemingen in Maart en April.

De resultaten staan in tabel VII. In de significanties komt door de weglating van Maart en April geen verandering, uitgezonderd bij de vergelijking tussen ③_W en ⑰_W.

Tabel VII.

Vergelijking der preparaten en blanco bepalingen.

vergelijking tussen		<u>u</u>	<u>Eu</u>	kl. waarn.	overschm. kans	Mrt+Apr. weggelaten aant. waarn. ov. kans	
① _Z (19)	② _Z (4)	1.5	38	①	0.0038	-	-
① _W (9)	② _W (7)	15	31.5	①	0.09	-	-
① _W (9)	② _W (4)	0	18	①	0.0028	-	-
① _Z (19)	③ _Z (23)	215	218.5	①	0.9	(19)-(14)	0.60
① _W (9)	③ _W (37)	120	166.5	①	0.20	-	-
③ _Z (23)	④ _Z (20)	254.5	230	④	0.60	(14)-(20)	0.26
③ _W (37)	④ _W (13)	286	240.5	④	0.26	(26)-(5)	0.01
③ _W (37)	⑤ _W (11)	32	203.5	③	0.00006	(26)-(7)	0.0044
③ _W (37)	⑥ _W (2)	10	37	③	0.102	-	-
③ _W (37)	⑦ _W (4)	69.5	74	③	0.8	-	-
③ _Z (23)	⑦ _Z (3)	21.5	34.5	③	0.32	-	-
③ _Z (23)	⑧ _Z (3)	0	34.5	③	0.007	-	-
③ _W (37)	⑧ _W (5)	7	92.5	③	0.0009	-	-
③ _W (37)	⑧ _Z (3)	14.5	55.5	③	0.036	-	-
③ _Z (23)	⑨ _Z (4)	45.5	46	③	1.	-	-
③ _W (37)	⑨ _W (2)	59	37	⑨	0.19	-	-
③ _W (37)	⑩ _W (4)	37.5	74	③	0.114	-	-
③ _Z (23)	⑪ _Z (9)	5	103.5	③	0.00004	(14)-(7)	0.0026
③ _W (37)	⑪ _W (12)	70.5	222	③	0.00048	(26)-(8)	0.014
③ _Z (23)	⑫ _Z (6)	58.5	69	③	0.58	-	-
③ _Z (23)	⑬ _Z (3)	7	34.5	③	0.024	-	-
③ _W (37)	⑬ _W (3)	15.5	55.5	③	0.044	-	-
③ _W (37)	⑮ _W (5)	9.5	92.5	③	0.0002	(26)-(5)	0.012
③ _W (37)	⑯ _W (3)	13	15.5	③	0.024	-	-
③ _Z (23)	⑰ _Z (23)	14	264.5	③	0.00001	(14)-(23)	0.00001
③ _W (37)	⑰ _W (10)	94	185	③	0.018	(26)-(10)	0.060
③ _W (37)	⑰ _Z (23)	116	425.5	③	0.00001	(26)-(23)	0.00014
⑤ _W	⑪ _W			⑪		(7)-(8)	0.152
⑤ _W	⑮ _W			⑮		(7)-(5)	0.48
⑤ _W	⑰ _W			⑰		(7)-(10)	0.012
⑤ _W (11)	⑰ _Z (23)	123.5	126.5	⑤	0.9	(7)-(23)	0.69
⑪ _W	⑮ _W			⑪		(8)-(5)	0.044
⑪ _Z	⑰ _Z			⑰		(7)-(23)	0.64
⑪ _W	⑰ _W			⑰		(8)-(10)	0.18

Wanneer Tabel IV met Tabel II en de bij het materiaal opgegeven overschrijdingskansen vergeleken wordt, blijkt dat op drie na alle significante verschillen blijven bestaan. Deze 3 uitzonderingen zijn de verschillen tussen ①_W en ②_W, ③_W en ⑥_W, ③_W en ⑩_W, welke niet

meer significant zijn. Bovendien is bij vergelijking van ⑪_w met ⑮_w zonder de Maart- en April-waarnemingen een significant verschil gevonden (wanneer als grens voor significantie en overschrijdingskans 0.05 genomen wordt); dit verschil werd door de onderzoeker met de toets van Student niet gevonden. In het algemeen zal hier het met de toets van Wilcoxon verkregen resultaat betrouwbaarder zijn dat het met de toets van Student verkregene, wegens de afwijkingen van de normaliteit. Anderzijds is het mogelijk dat door de splitsing in kleinere groepen bepaalde werkelijk bestaande verschillen niet meer worden gevonden. Dit kan gecontroleerd worden door het uitvoeren van een groter experiment met proeven, die onder gelijke omstandigheden genomen worden.

5. Conclusies.

- 1). De gebruikte maat (maximale kleurindexdaling) is niet normaal verdeeld. Men gebruike daarom een niet op normaliteit gebaseerde toetsingsmethode.
- 2). Er is een aanwijzing dat de reactie 's winters van die des zomers verschilt. De reactie is in de maanden Maart en April onregelmatig in vergelijking met de rest van het jaar.
- 3). De kleurindexdaling is bij inspuiting met aminozuren en met de preparaten A (reeks ⑧), D (reeks ⑪), G (reeks ⑬), K (reeks ⑮), L (reeks ⑯), M (reeks ⑰) duidelijk groter dan bij inspuiting met fysiologisch zout (vgl. tabel VII). Wij vonden echter geen verschil in reactie met de fysiologisch-zout-reeks (⑱³) bij de sterk verdunde preparaten en bij de preparaten B (reeks ⑨), C (reeks ⑩) en F (reeks ⑫).

Verder bleek er een significant verschil te zijn tussen de kleurindexdalingen bij onbehandelde dieren en de niet met lood vergiftigde dieren, die een leverinjectie kregen (reeksen ① en ②). Vergiftiging met lood en injectie van fysiologisch zout deed de dalingen echter niet significant stijgen (reeksen ① en ③).

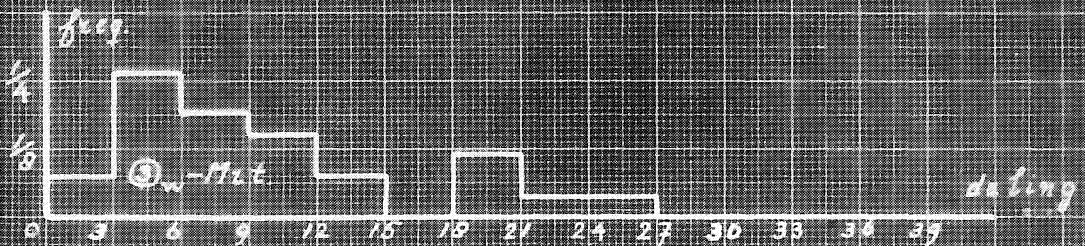
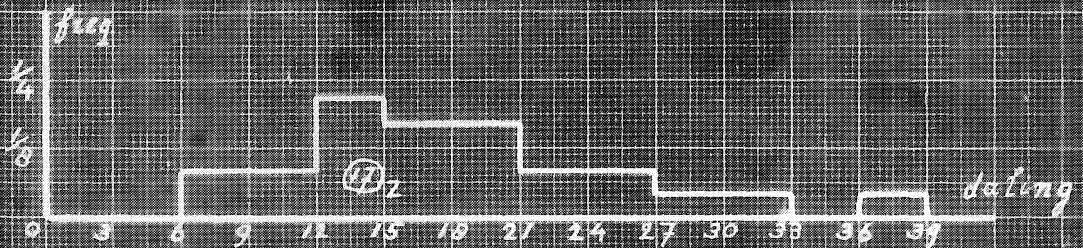
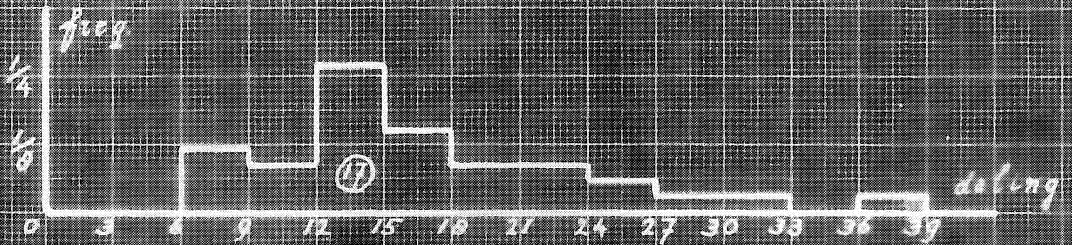
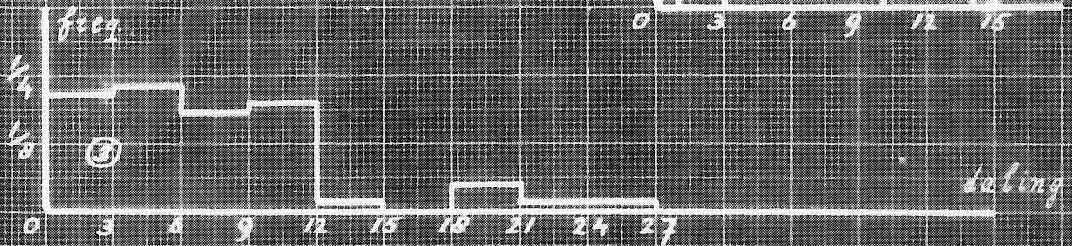
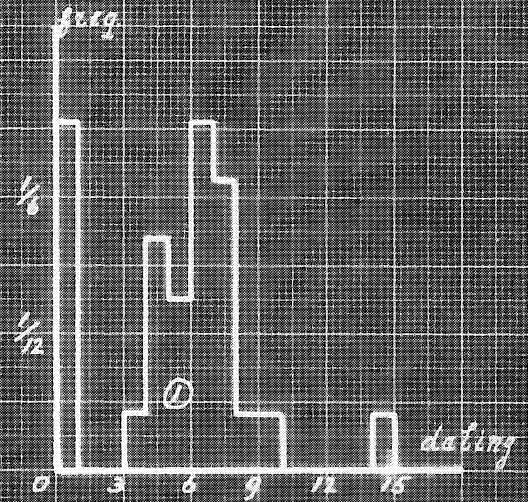
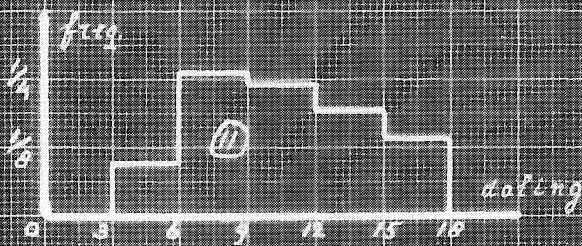
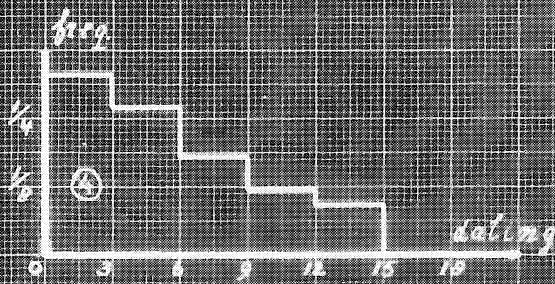
Wij zien dus, dat de aminozuren een dergelijke reactie veroorzaken als de leverpreparaten en dat ook zonder loodvergiftiging reacties optreedt tengevolge van leverpreparaten.

6. Opmerking.

Bij verdere onderzoeking van deze ijkmethode verdient het aanbeveling voor de statistische verwerking van de resultaten de waarnemingen regelmatig over de konijnen te verdelen en rekening te houden met een mogelijke seizoeninvloed.

Een deel van de berekeningen zijn uitgevoerd door de rekenafdeling van het Mathematisch Centrum.

Histogrammen der maximale procentuele
kleurindexdalingen.



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 zone van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwierpt 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 verworping 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 vergelijke men bv. de tweede hieronder gegeven literatuurplaats. Wij moeten hier volstaan met de opmerking, dat éénzijdige toetsing veelal eerder tot verworping 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.

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 ^{ook} 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 λ 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 λ_1 , resp. λ_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.

Normaliteitstoetsen van Geary en Pearson¹⁾

Om de hypothese H_0 te toetsen, dat een reeks waarnemingen $x_1, x_2, \dots, x_n$ beschouwd kan worden als een steekproef uit een normale verdeling, d.i. een verdeling met verdelingsdichtheid

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

zijn door Geary en Pearson de verdelingen, onder hypothese H_0 , berekend van enkele statistische grootheden, welke betrekking hebben op de vorm van de kromme $f(x)$.

De eerste grootheid is de volgende.

$$(1) \quad \underline{a} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{\sqrt{n \sum_{i=1}^n (x_i - \bar{x})^2}} \quad \text{waarin} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{is.}$$

Deze grootheid $\underline{a}$ is een maat voor de "kurtosis" ("platheid" of "slankheid" der kromme $f(x)$). In figuur 1 zijn 3 verdelingsdichtheden $f(x)$, $f_1(x)$ en $f_2(x)$ getekend waarvan $f(x)$ een normale is, terwijl $f_1(x)$ te slank is (en te dik in de staarten), en $f_2(x)$ te plat (en te dun in de staarten). Pearson noemt $f_1(x)$ "leptocurtic" en $f_2(x)$ "platycurtic".

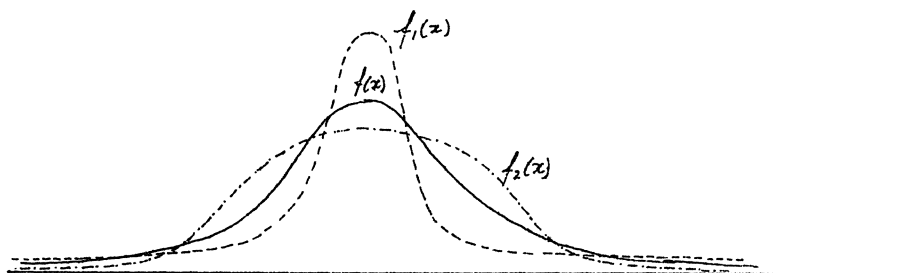


fig. 1

De kritieke zône voor $\underline{a}$ bestaat uit grote en kleine waarden van $\underline{a}$.

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

Als tweede toetsingsgrootheid wordt gebruikt:

$$(2) \quad \sqrt{b_1} = \sqrt{n} \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{\left\{ \sum_{i=1}^n (x_i - \bar{x})^2 \right\}^{\frac{3}{2}}}$$

Deze grootheid is een maat voor de "scheefheid" (Engels: "skewness") van de verdeling. In figuur 2 zijn 3 verdelingsdichtheden getekend, waarvan $f(x)$ symmetrisch is, terwijl $f_1(x)$ "positief scheef" (omdat de "dikke staart" rechts ligt) en $f_2(x)$ "negatief scheef" is.

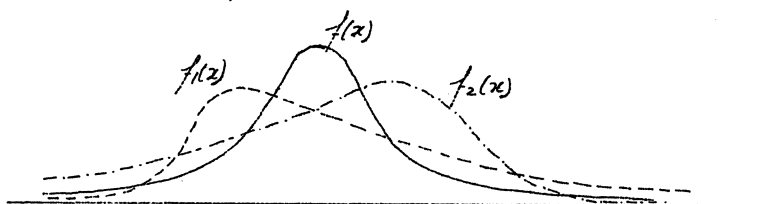


fig. 2

De kritieke zône voor de grootheid $\sqrt{b_1}$, bestaat uit grote en kleine waarden van $\sqrt{b_1}$, (daarbij wordt gewoonlijk ook voor negatieve waarden de notatie $\sqrt{b_1}$ gebruikt).

De derde toetsingsgrootheid

$$(3) \quad b_2 = \frac{n \sum_{i=1}^n (x_i - \bar{x})^4}{\left\{ \sum_{i=1}^n (x_i - \bar{x})^2 \right\}^2}$$

is weer een maat voor de "kurtosis" der kromme en wordt alleen voor zeer grote steekproeven berekend. Voor kleinere steekproeven is de maat a voldoende.

De kritieke zône bestaat uit grote en kleine waarden van b_2 .

Nomogrammen der 3 grootheden:

- 1e. Voor de grootheid a zijn onder de hypothese H_0 nomogrammen en tabellen berekend voor de onbetrouwbaarheidsdrempels 0,01; 0,05 en 0,10, waarbij $10 \leq n \leq 1000$.
- 2e. Voor de grootheid $\sqrt{b_1}$ zijn nomogrammen berekend voor de onbetrouwbaarheidsdrempels 0,01 en 0,05, $25 \leq n \leq 1000$.
- 3e. Voor de grootheid $\sqrt{b_2}$ zijn nomogrammen berekend voor de onbetrouwbaarheidsdrempels 0,01 en 0,05. Hierbij is $100 \leq n \leq 1000$.

Literatuur:

R.C.Geary, Moments of the ratio of the mean deviation to the standard deviation for normal samples, Biometrika 28 (1936) p.295.

R.C.Geary and E.S.Pearson, Tests of Normality, Biometrika Office, London 1938.

Bovenstaande nomogrammen en tabellen komen in deze beide publicaties voor.

Alle daarin genoteerde overschrijdingskansen zijn éénzijdig.