MATHEMATISCH CENTRUM
AMSTERDAM

Statistische Afdeling
Rapport S 63
door
Mej.C.van Eeden.

Groeiproeven met ratten, gevoed met boter en met plantenvetten, waaraan vitamine A en D is toegevoegd.

1. Inleiding:

Door Dr J.Boer en Prof.Dr B.C.P.Jansen en later door de heren Groot en Nieman werden onderzoekingen gedaan naar het verschil in voedingswaarde tussen boter en plantenvetten waaraan vitamine A en D was toegevoegd.

Het waarnemingsmateriaal werd ons toegezonden van beide onderzoekingen, met het verzoek de vraag te beantwoorden of uit dit materiaal een verschil in voedingswaarde tussen genoemde voedingsmiddelen blijkt.

In § 2 behandelen we de proeven van Boer en Jansen; in § 3 die van Groot en Nieman. In de appendix zijn een aantal memoranda toegevoegd, waarin de gebruikte toetsingsmethoden kort worden uiteengezet. In het eerste hiervan (S 47 (M6)) wordt de algemene gang van zaken bij het toetsen van een hypothese behandeld. Deze wordt in het volgende bekend verondersteld.

2. Proeven van Boer en Jansen.

2.1 De proefopzet.

Deze proeven werden gedaan met mannelijke ratten, die vier weken na de geboorte ~~uit~~ in gezet werden en gevoed werden met:

A. 10% zomerboter, of
B. 10% plantenolie, waaraan werd toegevoegd 70 γ caroteen en 14 I.E. Calciferol per week, of
C. 10% plantenolie, waaraan werd toegevoegd 70 γ caroteen + 0,54 I.E. Calciferol per week.

In de gevallen A en B werd ons opgegeven de groei na 7 en 12 weken; in geval C alleen de groei na 7 weken.

MATHEMATISCH CENTRUM
AMSTERDAM

2.2 Onderzoek van normaliteit en verdere toetsingsonderstellingen.

De eerste vraag, die zich bij het onderzoek naar verschil in voedingswaarde voordoet is of de proeven die in verschillende jaargetijden met dezelfde voeding gedaan zijn[1], tezamen genomen mogen worden, d.w.z. of ze beschouwd kunnen worden als steekproeven uit eenzelfde collectie, en in de tweede plaats, of deze collectie als een normale beschouwd kan worden.[2]. Is dit het geval, dan kunnen we alle waarnemingen met één bepaalde voeding vergelijken met alle waarnemingen met een andere voeding, zonder verder op een seizoen te letten.

Dit kan op drie manieren getoetst worden:

I. Eerst is het materiaal op normaliteit onderzocht. In de figuren I en II vinden we voor de verschillende voedingen histogrammen. Voor de voedingen A (bij 7 en 12 weken) en B (bij 7 weken) is de normaliteitstoets van Geary en Pearson toegepast (zie bijlagen S 47 (M 16)). Het resultaat hiervan is

Tabel I:
Normaliteitstoets voor ieder dergroepen afzonderlijk

Voeding	Groei na	Aantal waarnemingen	a	Overschrijdingskans
A	7 weken	83	0,80	1,00
A	12 "	45	0,79	0,80
B	7 "	48	0,78	0,~~99~~ 55
Voeding	**Groei na**	**Aantal waarnemingen**	$\sqrt{b_1}$	**Overschrijdingskans**
A	7 weken	83	0,3	0,40
A	12 "	45	0,6	0,08
B	7 "	48	-0,3	0,50

---

1) Wij zullen deze in het volgende aanduiden met "subgroepen"

2) Anders uitgedrukt: of de groei gedurende een bepaald aantal weken geacht kan worden verdeeld te zijn volgens de normale verdeling ("verdeling van Gausz")en wel volgens dezelfde verdeling onafhankelijk van het seizoen.

MATHEMATISCH CENTRUM
AMSTERDAM

De conclusie luidt dus: geen verwerping van de hypothese van normaliteit, daar geen der overschrijdingskansen kleiner is dan men redelijkerwijze kan verwachten.

Hoewel uit dit resultaat niet tot normaliteit van de verdelingen besloten kan worden ("niet-verwerpen" staat niet gelijk met "aanvaarden") kan men er wel uit op maken, dat eventueel toch aanwezige afwijkingen van normaliteit niet van ernstige aard zijn, zodat verdere toetsingen, waarbij van normaliteit gebruik gemaakt wordt, een voldoende mate van betrouwbaarheid bezitten.

II. Om te toetsen of er een verschil in de spreidingen der subgroepen is, is de L_1-toets toegepast. (zie bijlage S 53 (M 25)) en we vinden:

Tabel II

L_1-toets voor ieder der groepen afzonderlijk.

Voeding	Groei na		Aantal subgroepen	M	
A	7	weken	11	8,9	niet significant
A	12	"	7	6,4	" "
B	7	"	5	1,1	" "
B	12	"	4	7,5	" "
C	7	"	4	1,8	" "

III Verder is nog een variantieanalyse[1)] toegepast om de gelijkheid der gemiddelden te toetsen. Het resultaat hiervan is (zie ook bijlage I):

Tabel III

Variantieanalyse voor ieder der groepen afzonderlijk.

Voeding	Groei na	z	Overschrijdingskans
A	7 weken	0,18	0,40
A	12 "	0,09	> 0,40
B	7 "	-0,10	> 0,40
B	12 "	-0,23	> 0,40
C	7 "	0,11	> 0,40

---

1) Zie: M.G.Kendall: The advanced Theory of Statistics; deel II, pag. 175-180.

Uit het voorgaande blijkt dus dat er geen reden is om aan te nemen dat de waarnemingen (met één bepaalde voeding) niet uit één normale verdeling komen en we kunnen dus, bij het vergelijken der voedingen, alle subgroepen van één bepaalde voedihg tezamen nemen.

2.3 Onderzoek der verschillen met behulp van de toets van Student.

Om een antwoord te geven op de in § 1 gestelde vraag, zijn de volgende voedingen vergeleken.

a. A met B bij 7 weken.
b. A met B bij 12 weken.
c. A met C bij 7 weken.
d. B met C bij 7 weken.

Hiervoor is de toets van Student voor twee steekproeven toegepast (zie bijlage S 47 (M 9)), waarbij dus de hypothese getoetst wordt, dat de gemiddelden gelijk zijn.

Daar bij deze toets echter ondersteld wordt, dat de spreidingen gelijk zijn, is eerst de z-toets van Fischer toegepast.(Zie bijlage S 53 (M 24)). Dit geeft (zie ook bijlage I):

Tabel IV:
z-toets van Fischer voor alle tweetallen groepen.

		s^2	ν	z	Overschrijdingskans:
a	A	390	82	0,03	> 0,40
	B	412	47		
b	A	622	44	0,07	> 0,40
	B	543	34		
c	A	390	82	0,07	> 0,40
	C	342	26		
d	B	412	47	0,09	> 0,40
	C	342	26		

Dus in geen van de vier gevallen verwerping van de hypothese, dat de spreidingen gelijk zijn en de toets van Student kan dus zonder bezwaar worden toegepast:

Tabel V:

Toets van Student voor alle tweetallen groepen.

		m	t	ν	Overschrijdingskans
a	A B	157 146	2,97	129	0,003
b	A B	218 198	3,58	78	0,000
c	A C	157 120	8,50	108	0,000
d	B C	146 120	5,59	73	0,000

De conclusie luidt hier dus: verwerping van de hypothese, dat de gemiddelden gelijk zijn.

2.4 Conclusies, wat betreft de proeven van Boer en Jansen:

Uit tabel V blijkt dat er een verschil in voedingswaarde is tussen:

I Zomerboter en plantenolie, waaraan 70 γ caroteen en 14 I.E. calciferol per week is toegevoegd, zowel voor ratten, die zeven, als voor ratten, die twaalf weken hiermee gevoed zijn.

II Zomerboter en plantenolie, waaraan 70 γ caroteen en 0,54 I.E. calciferol per week is toegevoegd voor ratten, die hiermee 7 weken zijn gevoed.

III Plantenolie, waaraan 70 γ caroteen en 14 I.E. calciferol en plantenolie, waaraan 70 γ caroteen en 0,54 I.E. calciferol is toegevoegd, voor ratten, die hiermee 7 weken zijn gevoed.

In alle gevallen heeft het eerstgenoemde voedingsmiddel de hoogste voedingswaarde.

3. Proeven van Groot en Nieman:

3.1 De proefopzet.

Deze proeven werden grotendeels gedaan met mannelijke ratten, die gevoed werden met:

A : Zomerboter, of

B_1: Botervet, of

MATHEMATISCH CENTRUM
AMSTERDAM

B 2: Botervet + od, of
C 1: Botervetzuur+ od, of
2: Botervetzuur+ ov, of
3: Botervetzuur + ov + od, of
D 1: Slaolie + ov, of
2: Slaolie + od, of
3: Slaolie + ov + od, of
E 1: Slaolievetzuur + od, of
2: Slaolievetzuur + ov, of
3: Slaolievetzuur + ov + od.

In alle gevallen werd ons opgegeven de groei na 7 weken; in sommige gevallen ook die na 12 weken.

Bovendien werd ons opgegeven welke ratten uit dezelfde nesten kwamen.

3.2 De methode van onderzoek.

De hiertoegepaste methode is, dat wij voor twee voedingen, die we wilden vergelijken, het verschil in groei berekenden voor paren ratten, die uit hetzelfde nest kwamen. De ratten, waarvoor dergelijke paren , bij een bepaalde toets niet konden worden gevormd, werden buiten beschouwing gelaten.

Op deze verschillen werd de symmetrietoets T_2 toegepast.(zie bijlage S 47 (M 10)).
Het voordeel van de toepassing van deze toets boven die van Student is, dat voor de toepassing ervan veel minder onderstellingen behoeven te worden gemaakt. Zo is o.a. de onderstelling van normaliteit overbodig, evenals de onderstelling, dat de spreidingen der subgroepen gelijk zijn. Uit het waarnemingsmateriaal blijkt duidelijk, dat lang niet alle subgroepen als gelijkwaardig kunnen worden beschouwd, zodat het inderdaad gewenst is deze onderstellingen te vermijden.

De volgende voedingen zijn vergeleken:

a. Bij zeven weken:

1. B en C samen met D en E.
2. A met D_1
3. A met D_2
4. B_1 met D_2
5. B_2 met D_2
6. C_1 met E_1
7. C_2 met E_2

8. C_3 met E_3
9. D_1 met D_2.

b. Bij twaalf weken:

1. A met D_1
2. A met D_2
3. B_1 met D_2
4. B_2 met D_2
5. D_1 met D_2.

Andere vergelijkingen waren niet mogelijk omdat òf geen paren gevormd konden worden, òf ~~omdat~~ het aantal paren te klein was.

Het resultaat van het toepassen van de symmetrietoets is:

Tabel VI

Symmetrietoets T_2 toegepast op een aantal tweetallen groepen.

	n	r	u	v	Overschrijdingskans
a. 1	63	33	29	24	0,000
2	25	13	10	4	0,053
3	~~12~~ 13	~~6~~ 7	~~6~~ 7	4	~~0,014~~ 0,007
4	30	16	14	7	0,006
5	18	11	11	5	0,000
6	29	16	13	8	0,018
7	9	5	5	4	0,004
8	7	4	4	3	0,015
9	21	11	9	6	0,053
b. 1	15	8	7	4	0,07
2	7	4	4	3	0,015
3	14	7	7	4	0,009
4	8	4	4	3	0,04
5	13	7	6	3	0,09

3.3 Een complicatie.

Uit tabel VI blijkt, dat voor de vergelijken a_1, a_9, b_1 en b_5 de overschrijdingskansen weliswaar klein zijn, maar niet zo klein, dat we op grond hiervam mogen concluderen, dat er een verschil in voedingswaarde is.

Tijdens het onderzoek bleek ons, dat dit (althans bij 7 weken) wellicht veroorzaakt werd, door-dat er bij de proeven met slaolie + ov één was (n.l. subgroep Ax), waarbij de groei der ratten veel groter was, dan bij de andere proeven met dezelfde voeding.

De vraag is nu of de waarnemingen ~~nu~~ inderdaad zoveel hoger liggen, dat we deze verder buiten beschouwing moeten laten. Dit is op twee manieren onderzocht:

I. Bij de eerste methode is onderzocht of bij de vergelijking van zomerboter met slaolie + ov het aantal positieve en negatieve verschillen in groei voor paren ratten uit één nest, in de subgroep Ax te rijmen valt met het aantal positieve en negatieve verschillen in groei in de vier andere subgroepen met dezelfde voedingen. Dit is getoetst met behulp van de methode der 2x2-tabel (zie S 53 (M 23)) en dit geeft:

Tabel VII

Vergelijking van het aantal positieve en negatieve verschillen in groei, bij de vergelijking van zomerboter en slaolie + ov, in subgroep Ax en de andere subgroepen.

	Aantal positieve verschillen	Aantal negatieve verschillen	Totaal
Subgroep Ax	1	7	8
Andere subgroepen	~~13~~ 12	~~4~~ 5	17
Totaal	~~14~~ 13	~~11~~ 12	25

Exacte berekening der overschrijdingskans geeft: ~~0,007~~ 0,01. Deze overschrijdingskans is zo klein, dat wij zelfs indien wij rekening houden met het feit, dat de subgroep Ax uit de 5 groepen van deze proef gekozen is, omdat in deze subgroep het kleinste aantal positieve verschillen gevonden werd (hiermee kan men rekening houden door de gevonden overschrijdingskans met 5 te vermenigvuldigen) het weglaten van deze subgroep gerechtvaardigd kunnen achten. De hypothese nl., dat deze 5 subgroepen uit éénzelfde collectie genomen zijn, kan op grond van dit resultaat verworpen worden.

II. Bij de tweede methode is onderzocht of de subgroep Ax bij voeding met slaolie + ov hoger ligt dan de andere twee subgroepen met deze voeding. Hiervoor is de toets van Wilcoxon (zie bijlage § 47 (M 7)) toegepast en dit geeft op subgroep Ax in vergelijking met de overige subgroepen tezamen:

Tabel VIII:

Vergelijking van subgroep Ax bij voeding met slaolie + ov met de andere subgroepen met dezelfde voeding, met behulp van de toets van Wilcoxon.

n	m	s	$\frac{d}{\sigma}$	k
24	15	288	3,1	0,002

De kleine overschrijdingskans, die we hier vinden (en die met 3 vermenigvuldigd moet worden om rekening te houden met het feit, dat we de subgroep Ax uit drie subgroepen gekozen hebben, omdat de gemiddelde groei hier groter was dan in de beide andere subgroepen) versterkt de conclusie van de eerste methode.

3.4 Definitieve vergelijking van de voedsels boter, slaolie + ov en slaolie + cd.

Op grond van hetgeen gevonden is in § 3, 3.3, is de subgroep Ax met slaolie + ov verder buiten beschouwing gelaten en is bij de vergelijkingen a_1 en a_9 nogmaals de symmetrietoets toegepast. Dit geeft:

Tabel IX

Symmetrietoets

	n	r	u	v	Overschrijdingskans
a 2	17	9	9	2	0,004
9	14	8	6	2	> 0,10

Daar parametervrije toetsen, zoals de hier gebruikte, vaak een geringer onderscheidingsvermogen bezitten dan de klassieke methoden, waarbij van de onderstelling van normaliteit en gelijke spreidingen gebruik wordt gemaakt, hebben wij voor die vergelijkingen, die geen significantie gaven, nl. a_2, a_9, b_1 en b_5 nog Student's toets voor het gemiddelde van een normale verdeling toegepast, op dezelfde verschillen waarop eerst de symmetrietoets toegepast was. (Zie bijlage S 47 (M 8)).

Tabel X

Student's toets.

		m	t	ν	Overschrijdingskans
a	2	19	3,3	16	0,004
	9	7	1,0	13	0,34
b	1	20	2,8	14	0,014
	5	22	2,2	12	0,046

We krijgen nu het volgende overzicht voor de overschrijdingskansen, die gevonden zijn voor de vergelijkingen a_2, a_9, b_1 en b_5:

Tabel XI:

		Met subgroep Ax	Zonder subgroep Ax	
		Symmetrietoets		Student
a	2	0,053	0,004	0,004
	9	0,053	0,10	0,34
		Symmetrietoets	Student	
b	1	0,07	0,014	
	5	0,09	0,046	

Inderdaad zijn dus de overschrijdingskansen bij toepassing van de toets van Student kleiner dan bij de toepassing van de symmetrietoets. Daar het volgens § 3.3 verantwoord is subgroep Ax buiten beschouwing te laten, terwijl het gebruik van Student's toets alleen niet geheel verantwoord moet worden geacht, trekken wij de volgende conclusies:

I. (uit a_2 en b_1): dat er een verschil in voedingswaarde is tussen zomerboter en slaolie + ov,

II. (uit a_9 en b_5): dat er slechts een uiterst zwakke aanwijzing is voor een verschil in voedingswaarde tussen slaolie + ov en slaolie + cd.

Opmerking: Wellicht zou een groter aantal waarnemingen hier een duidelijker resultaat hebben gegeven.

3.5 Conclusies, wat betreft de proeven van Groot en Nieman:

MATHEMATISCH CENTRUM
AMSTERDAM

Uit de tabellen VI en XI blijkt, dat er een verschil in voedingswaarde is tussen:

I. Botervet(zuur) en slaolievet(zuur), waarbij aan beide voedingen ov en /of od is toegevoegd. (Zie a_1, tabel VI).

II. Zomerboter en slaolie + ov (a_2 en b_1, tabel XI).

III. Zomerboter en slaolie + od (a_3 en b_2, tabel VI).

IV. Botervet en slaolie + od (zie a_4 en b_3, tabel VI).

V. Botervet + od en slaolie + od (zie a_5 en b_4, tabel VI).

VI. Botervetzuur + od en slaolievetzuur + od (zie a_6, tabel VI).

VII. Botervetzuur + ov en slaolievetzuur + ov (zie a_7, tabel VI).

VIII. Botervetzuur + ov + od en slaolievetzuur + ov +od (zie a_8, tabel VI).

In alle gevallen heeft het eerstgenoemde voedingsmiddel de hoogste voedingswaarde.
Verder is er slechts een uiterst zwakke aanwijzing voor een verschil in voedingswaarde tussen slaolie+ ov en slaolie + od.

Een gedeelte der berekeningen werd door de Rekenafdeling van het Mathematisch Centrum uitgevoerd.

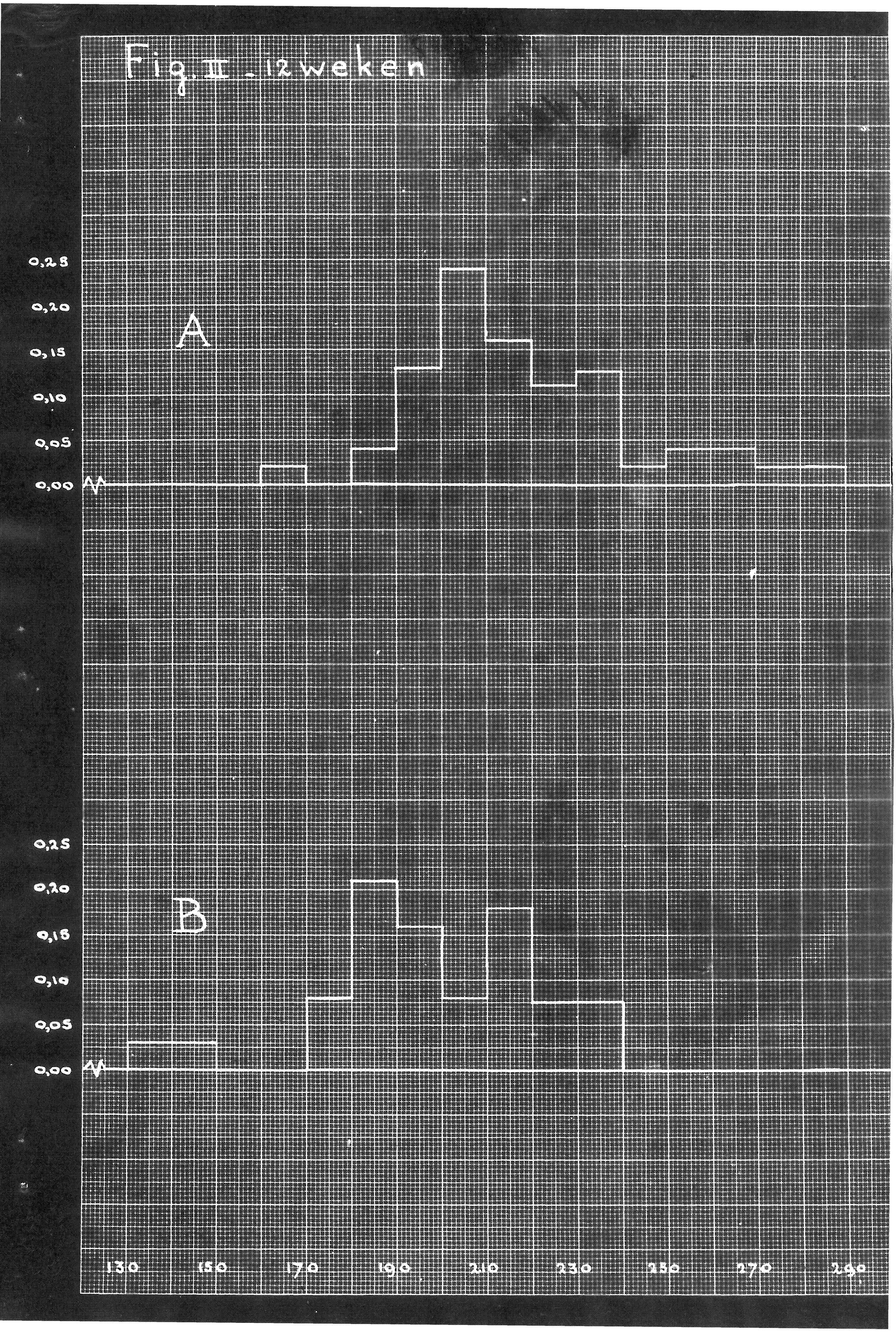
Fig. II - 12 weken
A
B
0,25
0,20
0,15
0,10
0,05
0,00
130
150
170
190
210
230
250
270
290

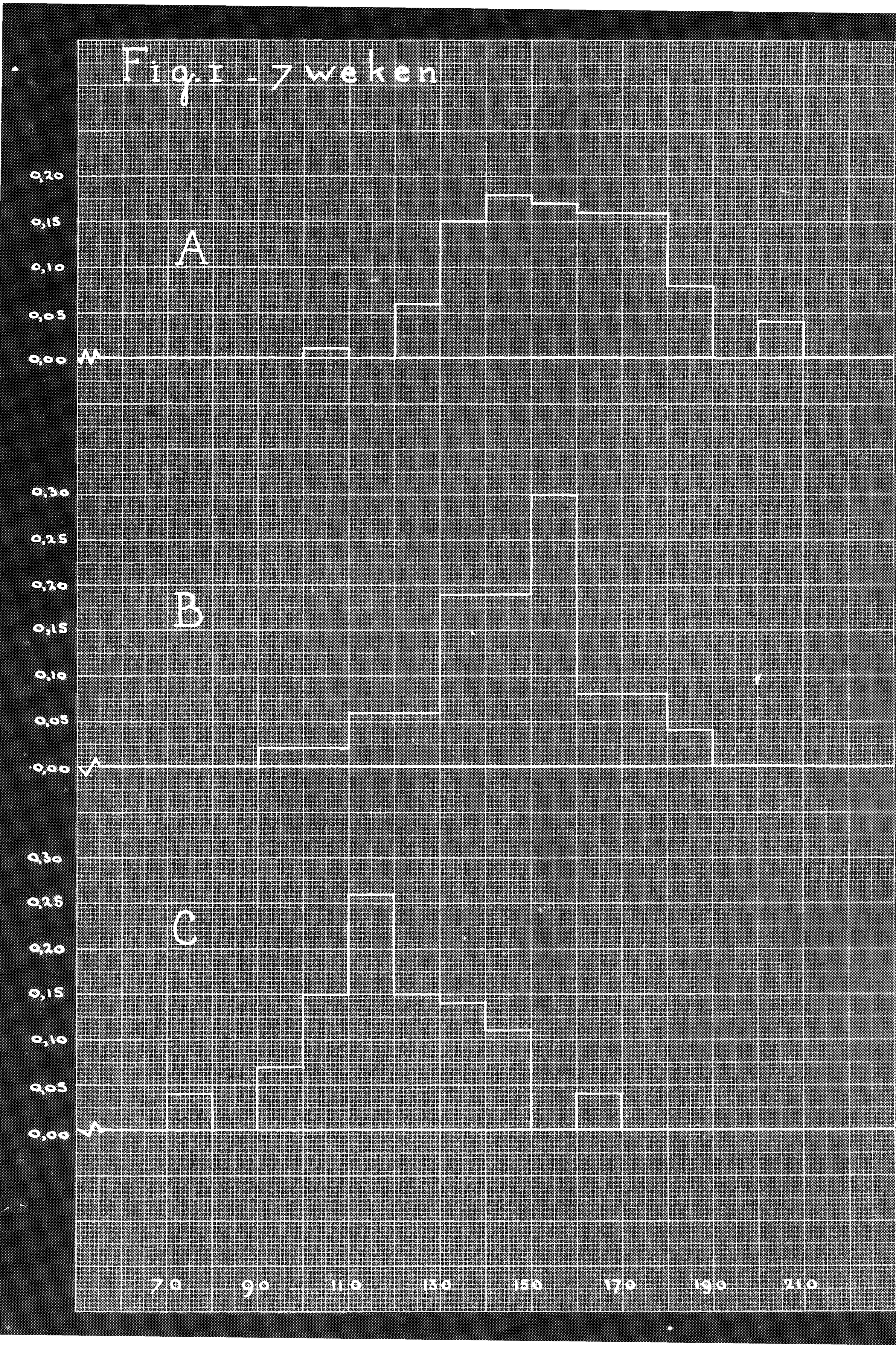
Fig. I – 7 weken
A
B
C
0,20
0,15
0,10
0,05
0,00
0,30
0,25
0,20
0,15
0,10
0,05
0,00
0,30
0,25
0,20
0,15
0,10
0,05
0,00
70
90
110
130
150
170
190
210

Bijlage I

Variantieanalyse:

Voeding

A 7 weken

		Aantal vrijheidsgraden	Quotiënt
$\sum_j n_j(\bar{x}_j-\bar{x})^2$	5265	10	527
$\sum_{ij}(x_{ij}-\bar{x}_j)^2$	26.682	72	371
$\sum_{ij}(x_{ij}-\bar{x})^2$	31.947	82	390

$$z=\tfrac{1}{2}\lg\frac{527}{371}=0{,}18$$

A 12 weken

		Aantal vrijheidsgraden	Quotiënt
$\sum_j n_j(\bar{x}_j-\bar{x})^2$	4372	6	729
$\sum_{ij}(x_{ij}-\bar{x}_j)^2$	22999	38	605
$\sum_{ij}(x_{ij}-\bar{x})^2$	27371	44	622

$$z=\tfrac{1}{2}\lg\frac{729}{605}=0{,}09$$

B 7 weken

		Aantal vrijheidsgraden	Quotiënt
$\sum_j n_j(\bar{x}_j-\bar{x})^2$	1369	4	342
$\sum_{ij}(x_{ij}-\bar{x}_j)^2$	17.998	43	419
$\sum_{ij}(x_{ij}-\bar{x})^2$	19.367	47	412

$$-z=\tfrac{1}{2}\lg\frac{419}{342}=0{,}10$$

B 12 weken

		Aantal vrijheidsgraden	Quotiënt
$\sum_j n_j(\bar{x}_j-\bar{x})^2$	1064	3	355
$\sum_{ij}(x_{ij}-\bar{x}_j)^2$	17395	31	561
$\sum_{ij}(x_{ij}-\bar{x})^2$	18459	34	543

$$-z=\tfrac{1}{2}\lg\frac{561}{355}=0{,}23$$

C 7 weken

		Aantal vrijheidsgraden	Quotiënt
$\sum_j n_j(\bar{x}_j-\bar{x})^2$	1243	3	414
$\sum_{ij}(x_{ij}-\bar{x}_j)^2$	7661	23	333
$\sum_{ij}(x_{ij}-\bar{x})^2$	8904	26	342

$$z=\tfrac{1}{2}\lg\frac{414}{333}=0{,}11$$

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.

S47(M6)

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

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

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

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

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

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

De toetsingstheorie biedt in het algemeen <u>geen</u> mogelijkheid om tot <u>aanvaarding</u> 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 <u>nevenvoorwaarden</u> 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 <u>overschrijdingskans</u> k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van H_0 zou zijn overgegaan; anders gezegd: de kleinste α, waarvoor de gevonden waarde der toetsingsgrootheid nog juist in de (bij α behorende) kritieke zône Z ligt. <u>Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leq \alpha$ is.</u>

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijke men bv. de tweede hieronder gegeven litteratuurplaats. 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.

<u>Litteratuur</u>:

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 <u>4</u> (1950) p.54-66.

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.

S47 (M7)

De toets van Wilcoxon.[1]

Deze methode dient tot het toetsen van de hypothese $\mathcal{H}_0$, 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 $\mathcal{U}$ is het aantal paren (x_i, y_j) waarvoor $x_i > y_j$ is (het aantal "inversies"). Daar er $n.m$ dergelijke paren zijn, kan $\mathcal{U}$ alle gehele waarden van 0 tot en met $n.m$ aannemen. Is $\mathcal{U}$ groot, dan liggen er veel waarden x_i verder naar rechts dan waarden y_j, is $\mathcal{U}$ klein, dan juist weinig.

De kritieke zône $\mathcal{Z}$ neemt nu daarom de kleine en de grote waarden van $\mathcal{U}$ 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 $\mathcal{Z}_1$, resp. $\mathcal{Z}_2$, die geheel bestaan uit kleine, resp. grote waarden van $\mathcal{U}$.

Verwerping van $\mathcal{H}_0$ ten gevolge van het vinden van een grote (resp. kleine) waarde van $\mathcal{U}$ 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 ≦ 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 ≦ 10, Rapport S32 (M_4)(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.

MATHEMATISCH CENTRUM
Amsterdam.
Statistische Afdeling.
S47(MB)

Toets van "Student" voor het gemiddelde van een normale verdeling.[1]

Gegeven: de steekproef $x_1, x_2, \ldots, x_n$ uit een normale collectie. Anders gezegd: $x_1, \ldots, x_n$ zijn onafhankelijke waarnemingen van de stochastische grootheid $\underline{x}$, die normaal verdeeld is (de zgn. waarschijnlijkheidsverdeling van Gauss bezit)[2].

H_0 (te toetsen hypothese): Het gemiddelde van $\underline{x}$ bezit de waarde μ; μ is hierin een gegeven getal, b.v.0.

Toetsingsgrootheid :

$$t = (\underline{m} - \mu)/\underline{s}'$$

waarin de bij de steekproef behorende waarden van $\underline{m}$ en $\underline{s}'$ gegeven worden door $m = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ en $s' = \sqrt{\dfrac{(x_1-m)^2 + \ldots (x_n-m)^2}{n-1}}$ is.

Indien H_0 juist is, zullen dicht bij 0 gelegen waarden vaker voorkomen dan ver van 0 gelegen waarden. Is echter het gemiddelde van $\underline{x}$ verschillend van μ, dan zullen verder van 0 af liggende waarden vaker voorkomen dan indien H_0 juist is. Als kritieke zône Z kiest men daarom voor tweezijdige toetsing een gebied van de vorm

$$|\underline{t}| \geqq t_0$$

en voor éénzijdige toetsing

linker-toetsing $\underline{t} \leqq -t_1$ rechter-toetsing $\underline{t} \geqq t_2$

De waarden t_0, t_1 en t_2 zijn getabelleerd voor verschillende waarden van de onbetrouwbaarheidsdrempel α.

Litteratuur:

M.G. Kendall, The Advanced Theory of Statistics, London 1946, Vol.II, p.98-102; tabellen in deel I, p.440-41. Opmerking: Het bij de tabellen vermelde aantal vrijheidsgraden (aangegeven door ν) is gelijk aan $n-1$

A.M. Mood, Introduction to the theory of Statistics, London 1950, p.425.

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

2) d.w.z., dat de kans, dat $\underline{x} \leqq x$ is, gegeven wordt door:

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

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
Amsterdam-O.

Statistische Afdeling
S 47 (M9)

De toets van Student voor 2 steekproeven [1].

Deze toets van Student wordt gebruikt voor het toetsen van de hypothese H_0, dat 2 steekproeven $x_1, x_2, \ldots\ldots x_n$ en $y_1, y_2, \ldots\ldots y_m$ beschouwd kunnen worden als steekproeven uit eenzelfde normale verdeling.

De grootheid $\underline{t}$, welke hierbij als toetsingsgrootheid gebruikt wordt en waarvan Student de verdelingsfunctie berekend heeft, is de volgende:

$$\underline{t} = \frac{m_x - m_y}{S} \sqrt{\frac{nm}{n+m}}$$

Hierin zijn m_x en m_y de gemiddelden van de 2 steekproeven terwijl S een schatting is van de spreiding σ van de normale verdeling, waaruit volgens H_0 de steekproeven getrokken zijn. De bij de gevonden steekproeven behorende waarde van m_x, m_y en S zijn op volgende wijzw uit de waarnemingen te berekenen:

$$m_x = \frac{1}{n}(x_1 + x_2 + \cdots\cdots + x_n)$$

$$m_y = \frac{1}{m}(y_1 + y_2 + \cdots\cdots + y_m)$$

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - m_x)^2 + \sum_{j=1}^{m}(y_j - m_y)^2}{n+m-2}$$

Bovengenoemde grootheid $\underline{t}$ bezit als verdeling de verdeling van "Student" met $\nu = n+m-2$ vrijheidsgraden.

Deze verdeling is getabelleerd voor $\nu = 1$ t/m 20 [1] voor $\nu = 20$ t/m 100 zijn waarden van t berekend, welke behoren bij de onbetrouwbaarheidsdrempels $\alpha = 0{,}01$ en $\alpha = 0{,}05$. [2]

Is H_0 juist, dan zullen dicht bij nul gelegen waarden vaker voorkomen, dan ver van nul gelegen waarden. De kritieke zône kiezen we dus zodanig, dat ze gevormd worden door grote en kleine waarden van $\underline{t}$. De kritieke zône bestaat dus, bij gegeven waarde van , uit twee stukken: $\underline{t} < -t_0$ en $\underline{t} > t_0$, waarin t_0 de bij α en ν behorende kritieke waarde is, die in [1] en [2] getabelleerd is. Voor éénzijdige toetsing is de kritieke zône van de vorm $\underline{t} \leqq -t_1$ of $\underline{t} \geqq t_1$

---

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

Literatuur:

[1]. M.G.Kendall, The Advanced Theory of Statistics, London 1946, Vol. II p.109, tabellen in deel I p. 440-41;
Opmerking: het bij de tabellen vermelde aantal vrijheidsgraden (aangegeven door ν) is gelijk aan n + m - 2.

[2]. Elizabeth M.Baldin, Table of Percentage Points of the t-distribution, Biometrika 33 (1943), p. 362. (Deze tabel worden tweezijdige overschrijdingskansen gegeven).

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - O.
Statistische Afdeling.
S 47 (M 10)

Symmetrietoets[1].

Hypothese H_0: de waarnemingen $z_1,\ldots,z_{n'}$ zijn afkomstig van n' onafhankelijk verdeelde stochastische grootheden, die alle symmetrisch ten opzichte van 0 verdeeld zijn[2]. Van deze toets bestaan meerdere versies $T_1,\ldots,T_2''$. We bespreken eerst T_1 en T_2.

Toetsingsgrootheden. Deze worden als volgt uit $z_1,\ldots,z_{n'}$ afgeleid:

1e. de waarnemingen, die gelijk aan 0 zijn worden weggelaten. Stel er blijven over: $z_1,\ldots,z_n$.

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

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

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

5e. De groep waarden $w_1,\ldots,w_n$ wordt verdeeld in twee groepen $w_1,w_2,\ldots,w_r$ en $w_{r+1},\ldots,w_n$, 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 $w_1,\ldots,w_r$ 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 B[3].

---

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 $x_1-a,\ldots,x_{n'}-a$.

3) In de oorspronkelijke publicaties over deze toets (zie de litteratuurverwijzingen 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 $x_1,\ldots,x_n$ die in A voorkomen noemen wij u.

De toetsingsgrootheden zijn $\underline{n_1}$ en $\underline{u}$, r is een hulpgrootheid.

<u>V.B.</u> 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/-4,6/-4,6/-4,8/-6,3/-7,0/-7,9/-8,0/-8,7.

$\therefore\ r = 11,\quad n_1 = 8$
$\qquad\qquad\quad u = 2$

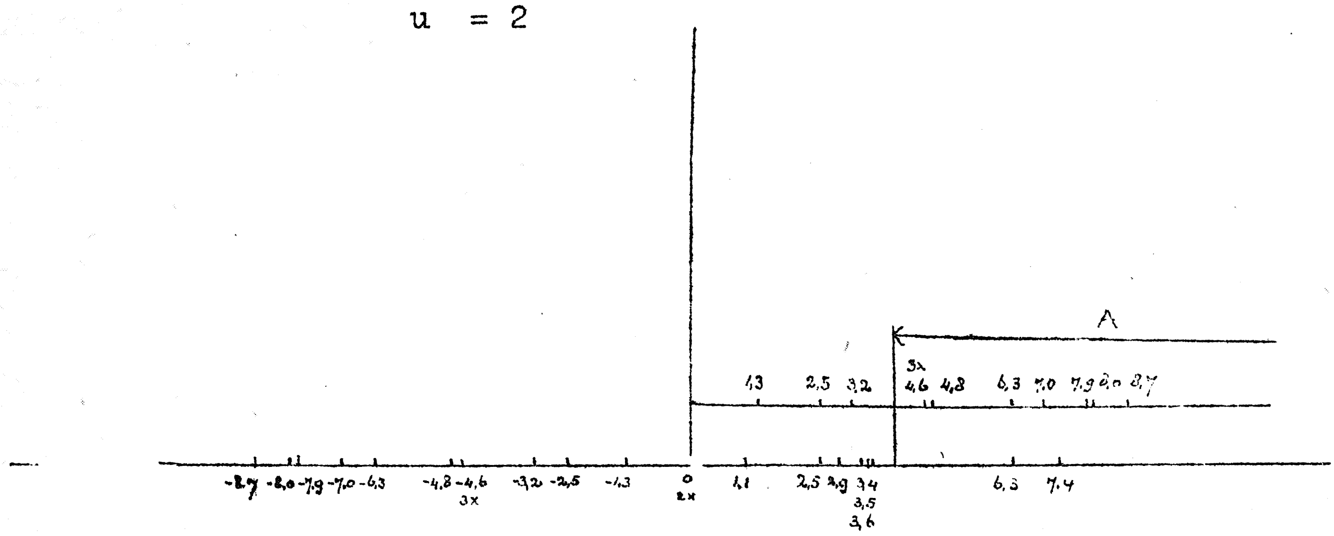

fig. 1

<u>Kritieke zônes</u>. 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 <u>positieve</u> (resp, negatieve) waarden zijn, die <u>verder</u> van 0 verwijderd liggen dan de <u>negatieve</u> (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.

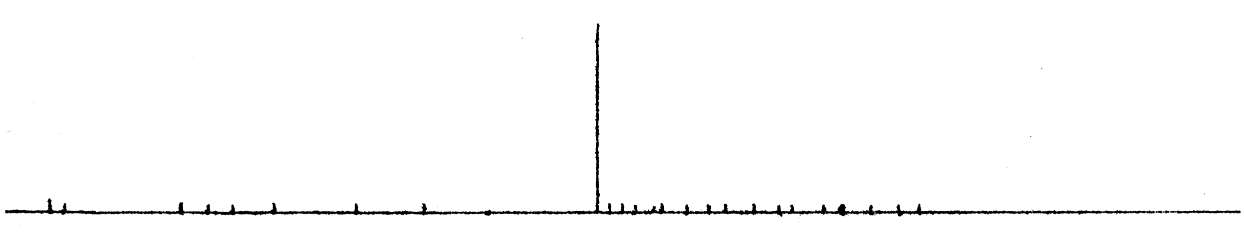

fig. 2

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

<u>T_1' en T_2'</u>.

<u>Toetsingsgrootheden</u>.

1e, 2e en 3e als boven. Eerste toetsingsgrootheid: $\underline{n}_1$.

4e: op $x_1,\ldots,x_{n_1}$ en $y_1,\ldots,y_{n_2}$ wordt de toets van Wilcoxon toegepast (vgl. S 47 (M 8)). De toetsingsgrootheden zijn $\underline{n}_1$ en de $\underline{U}$ van deze toets van Wilcoxon.

<u>Kritieke zônes</u>. 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'.

<u>Opmerkingen</u>: 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 $\underline{u}$ in plaats van $\underline{U}$ wordt gebruikt (vgl. b.v. [2], blz. 77, § 6.4.5).

<u>Litteratuur</u>:

[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. <u>53</u> (1950), p.945-955. Indagationes Mathematicae <u>12</u> (1950), p. 340-350.

[2] - " - , Symmetrietoetsen, Diss., Den Haag 1950, Excelsior.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
Amsterdam-O.

Statistische Afdeling
S 47 (M 16)

Normaliteitstoetsen van Geary en Pearson[1].

Om de hypothese $\mathcal{H}_0$ te toetsen, dat een reeks waarnemingen $x_1, x_2, \ldots\ldots x_n$ 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 $\mathcal{H}_0$, berekend van enkele statistische grootheden, welke betrekking hebben op de vorm van de kromme $f(x)$.

De eerste grootheid is de volgende:

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

Deze grootheid 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".

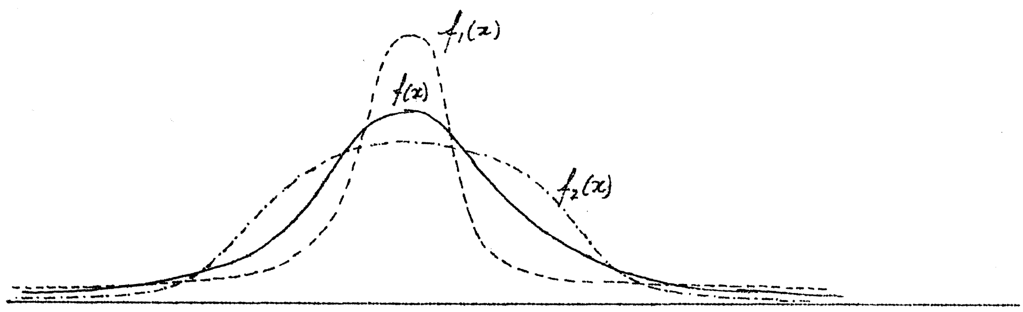

fig. 1

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

---

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

Als tweede toetsingsgrootheid wordt gebruikt:

$$(2) \qquad \sqrt{b_1} = \sqrt{n}\,\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^3}{\left\{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{3}{2}}}$$

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.

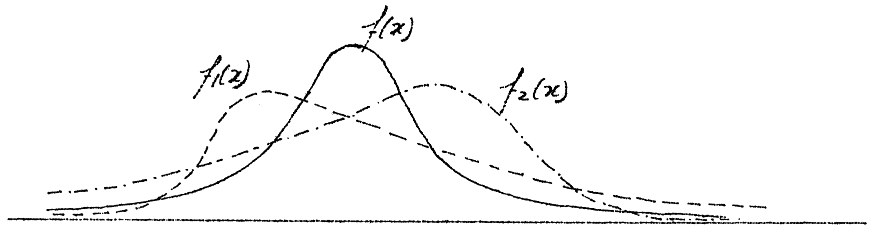

fig. 2

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

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

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.

<u>Nomogrammen der 3 grootheden:</u>

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 $10 \leqq n \leqq 1000$.

2e. Voor de grootheid $\sqrt{b_1}$ zijn nomogrammen berekend voor de onbetrouwbaarheidsdrempels 0,01 en 0,05, $25 \leqq n \leqq 1000$.

3e. Voor de grootheid $\sqrt{b_2}$ zijn nomogrammen berekend voor de onbetrouwbaarheidsdrempels 0,01 en 0,05. Hierbij is $100 \leqq n \leqq 1000$.

<u>Literatuur</u>:

R.C.Geary, Moments of the ratio of the mean deviation to the standard deviation for normal samples, Biometrika <u>28</u> (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 <u>éénzijdig</u>.

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

Statistische Afdeling.
S 53 (M 23).

Toetsing van de hypothese $p_1 = p_2$ met behulp van een 2 x 2-tabel[1].

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:

$$H_0: \quad p_1 = p_2.$$

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 toetsingsgrootheid wordt n_1, het aantal malen A in de eerste reeks waarnemingen, gebruikt. Indien H_0 juist is bezit deze grootheid 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 éénzijdige 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 éénzijdige 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_o 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_o juist is):

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

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.

<u>Litteratuur</u>:

R.A.Fisher, Statistical Methods for Research Workers, London 1948, p. 96. <u>Opmerking</u>: Fisher gebruikt hier de éénzijdige overschrijdingskans.

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

MATHEMATISCH CENTRUM
2de Boerhaavestr. 49
A M S T E R D A M O.
Statistische Afdeling
S 53 (M 24)

z-toets van Fisher[1)]

Stel we hebben twee onafhankelijke steekproeven

$$x_1 \ldots x_{n_1} \quad \text{met gemiddelde } \bar{x}$$
$$y_1 \ldots y_{n_2} \quad \text{met gemiddelde } \bar{y},$$

beide uit een normale verdeling.

De te toetsen hypothese is dan, dat de spreidingen van deze normale verdelingen gelijk zijn.

Is

$$s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i-\bar{x})^2$$

en

$$s_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i-\bar{y})^2$$

dan is de toetsingsgrootheid:

$$z = \tfrac{1}{2} \lg \frac{s_1^2}{s_2^2}$$

waarbij van s_1^2 en s_2^2 de grootste in de teller van de breuk genomen wordt.

Als H_0 juist is zal z in het algemeen klein zijn en de kritieke zône bestaat dus uit grote waarden van z. Men kan als toetsinggrootheid ook nemen

$$F = e^{2z} = \frac{s_1^2}{s_2^2} .$$

De kritieke zône bestaat dan uit grote waarden van e^{2z}. De toets wordt dan gewoonlijk de F-toets van Snedecor genoemd.

In de verdelingen van z en van e^{2z} komen twee parameters voor: $\nu_1 = n_1 - 1$ en $\nu_2 = n_2 - 1$, die het aantal vrijheidsgraden in de teller resp. in de noemer genoemd worden.

Litteratuur: M.G.Kendall: The advanced theory of statistics: deel II pag.115.
P.G.Hoel: Introduction to mathematical statistics; pag. 152-154.

---

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

Vervolg litteratuuropg.:

Tabellen: M.G.Kendall: The advanced theory of statistics; deel I pag. 442-443.

P.G.Hoel: Introduction to mathematical statistics; pag. 250-253.

ν_1: 1 tot 500; ν_2: 1 tot 1000; onbetrouwbaarheidsdrempels 0,05 en 0,01.

R.A.Fisher en F.Yates: Statistical tables for biological, agricultural and medical research; pag. 34-43.

ν_1: 1 tot 24; ν_2: 1 tot 120; onbetrouwbaarheidsdrempels:

0,20; 0,10; 0,05; 0,01; 0,001.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - O.
Statistische Afdeling.
S 53 (M 25).

L_1-toets[1].

Stel we hebben N waarnemingen:

$$\begin{array}{cccc} x_{11} & x_{21} & \dots & x_{n_1 1} \\ x_{21} & x_{22} & \dots & x_{n_2 2} \\ \vdots & \vdots & & \vdots \\ x_{1p} & x_{2p} & \dots & x_{n_p p} \end{array}$$

die verdeeld zijn in p subgroepen, terwijl het aantal waarnemingen in de j^e subgroep n_j is.

Veronderstel verder dat ieder der subgroepen een steekproef is uit een normale verdeling. De te toetsen hypothese H_0 is dan dat de spreidingen der subgroepen gelijk zijn.

Is $\bar{x}_j$ het gemiddelde van de waarnemingen in de j^e subgroep, $\nu_j = n_j - 1$, $N' = \sum_{j=1}^{p} \nu_j = N - p$

$$s_j^2 = \frac{1}{\nu_j} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

$$s^2 = \sum_{j=1}^{p} \frac{\nu_j}{N'} s_j^2$$ dan definiëren wij:

$$L_1 = \prod_{j=1}^{p} \left(\frac{s_j^2}{s^2} \right)^{\frac{\nu_j}{N'}}$$

L_1 is dus de verhouding van het geometrisch en het rekenkundig gemiddelde van de varianties der subgroepen en is dus ≤ 1.

Als toetsingsgrootheid wordt nu genomen

$$M = -N' \lg L_1 = N' \lg \sum_{j=1}^{p} \frac{\nu_j}{N'} s_j^2 - \sum_{j=1}^{p} \nu_j \lg s_j^2$$

en M is dus ≥ 0.

Als H_0 juist is zal M in het algemeen klein zijn en de kritieke zône bestaat dus uit grote waarden van M.

Litteratuur: H.O.Hartley: Testing the homogeneity of a set of estimated variances. Biometrika 31 (1940) p.249.

Tabellen: H.O.Hartley en E.S.Pearson: Tables for testing the homogeneity of a set of estimated variances. Biometrika 33 (1946) p. 296.

---

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