MATHEMATISCH CENTRUM
AMSTERDAM
Statistische Afdeling,
Rapport S 55,
door
Dr J.Hemelrijk Jr

Februari 1951.

# Zevende verslag omtrent groeiproeven met Wistarratten.

## 1. Inleiding.

Om na te gaan, of bij de gebruikte wijze van verwerking der proefresultaten geen systematische veranderingen van de spreiding der gewichtstoenamen over het hoofd zijn gezien, werd nagegaan:

1e. Of er een systematisch verschil te constateren is tussen de genoemde spreiding bij 3 opeenvolgende worpen van de ratten en

2e. Of er een systematisch verschil geconstateerd kan worden tussen de spreidingen 's winters en 's zomers,

3e. Of er een afhankelijkheid geconstateerd kan worden tussen de per subgroep berekende gemiddelden $\bar{x}$ en spreidingen $s$ der gewichtstoenamen van de ratten.

Ook bij de beantwoording van de vragen 1e en 2e wordt steeds gebruik gemaakt van de spreidingen van de waarnemingen binnen de subgroepen.

Verder zijn op verzoek van Drs Thomasson enkele tabellen berekend resp. uitgebreid, die aan het eind van dit rapport zijn toegevoegd. De berekeningen zijn uitgevoerd door de Rekenafdeling van het Mathematisch Centrum.

De voor de toetsingen gebruikte methoden zijn uiteengezet in memoranda, die aan dit verslag zijn toegevoegd. Tevens is memorandum S 47 (M 6) bijgevoegd, waarin de algemene gang van zaken bij het toetsen van hypothesen wordt toegelicht.

## 2. Vergelijking van de 3 worpen.

Van ieder der 3 worpen waren series waarnemingen van de 4 spreidingen $s_1$, $s_2$, $s_3$, $s_4$ (voor de groei in de eerste, eerste twee, eerste drie en eerste vier weken).
Deze series werden voor ieder der spreidingen $s_1, \ldots, s_4$ afzonderlijk vergeleken met behulp van de toets van Wilcoxon (zie S 47 (M 7)), aan het eind van dit rapport). Deze toets werd gekozen, daar de verdeling der spreidingen niet normaal geacht kan worden. De resultaten zijn samengevat in

tabel I.[1]

Tabel I

Vergelijking der worpen.

| Vergeleken worpen | Overschrijdingskans bij vergelijking van: $s_1$ | $s_2$ | $s_3$ | $s_4$ |
|---|---|---|---|---|
| 1e en 2e | 0,19 | 0,65 | 0,45 | 0,50 |
| 1e en 3e | 0,75 | 0,16 | 0,75 | 0,78 |
| 2e en 3e | 0,09 | 0,55 | 0,25 | 0,35 |

Daar geen van deze twaalf vergelijkingen een kleine overschrijdingskans geeft, is er geen reden om,voor zoverre het de spreiding betreft, een verschil tussen de drie worpen te vermoeden. De waarnemingsreeksen bevatten bovendien vrij veel waarnemingen (54, 64 resp. 63 voor de eerste, tweede resp. derde worp), zodat een eventueel toch aanwezig verschil vrij gering moet zijn.

3. Vergelijking van de spreidingen in zomer en winter.

Eveneens voor ieder de spreidingen $s_1$, $s_2$, $s_3$ en $s_4$ zijn de waarnemingen gesplitst naar het seizoen, waarbij als zomer genomen zijn de maanden April tot en met Septemver en als winter de maanden October tot en met Maart. Toepassing van de toets van Wilcoxon gaf de in tabel II samengevatte resultaten

Tabel II

Onderzoek naar seizoen invloed.

zomer - winter

| spreiding | $s_1$ | $s_2$ | $s_3$ | $s_4$ |
|---|---|---|---|---|
| overschrijdingskans | 0,88 | 0,34 | 0,22 | 0,18 |

Ook hier is geen der overschrijdingskansen klein, zodat wij geen verschil tussen zomer en winter vinden. Wel zijn in alle 4 gevallen de spreidingen in doorsnee 's zomers iets groter dan 's winters, terwijl dit verschil bovendien met het aantal weken, dat de proef duurt, toeneemt, zoals men aan de dalende overschrijdingskansen kan zien. Dit wekt het ver-

---

1) Alle overschrijdingskansen in tabel I en II zijn tweezijdige overschrijdingskansen.

moeden, dat er misschien toch een verschil is, dat men bij nog langer voortzetten van de proef wellicht zou kunnen ontdekken. Het onderhavige materiaal, dat vrij uitgebreid is (114 waarden van ieder der spreidingen $s_1, \ldots, s_4$ voor de zomer en 67 voor de winter) biedt echter geenszins de mogelijkheid dit vermoeden te bevestigen. Integendeel, nog nadrukkelijker dan in § 2 kan men hier, vanwege de nog grotere aantallen waarnemingen per serie, concluderen, dat een eventueel aanwezig verschil zeker een gering verschil moet zijn.

### 4. Onderzoek der onafhankelijkheid van $\bar{x}$ en s.

Voor dit onderzoek konden, weer voor $s_1$, $s_2$, $s_3$ en $s_4$ afzonderlijk, alle beschikbare waarden worden gebruikt. Dit waren er in totaal 181. De onafhankelijkheid van $\bar{x}$ en s is getoetst voor $s_1, \ldots, s_4$ apart met behulp van de in het memorandum S 47 (M 15) beschreven methode der dubbele dichotomie, waarbij de op één van de deellijnen vallende punten bij de verdere berekeningen weggelaten zijn. De toets is in de achter aan dit rapport bijgevoegde grafieken in beeld gebracht. De ongunstigste week, d.w.z. die, waarvoor de in de 4 vakken liggende aantallen punten het meest van de verwachting afwijken, is de derde. Voor deze week (voor $s_3$ en de bijbehorende $\bar{x}$ dus) is de overschrijdingskans berekend. Deze bedraagt ruim 0,06. Voor de overige 3 weken is de overschrijdingskans aanzienlijk groter. Derhalve is er ook hier geen reden om tot verwerping van de hypothese, in dit geval de hypothese van onafhankelijkheid van $\bar{x}$ en s, over te gaan.

### 5. Samenvatting van de toetsingsresultaten.

Er werd geen systematisch verschil gevonden tussen de spreidingen voor de 3 worpen en evenmin tussen de spreidingen 's zomers en 's winters. Er is geen afhankelijkheid gebleken tussen $\bar{x}$ en s. Daar de aantallen waarnemingen groot zijn, is er geen reden de mogelijkheid van aanzienlijke verschillen resp. sterke afhankelijkheid aanwezig te achten, zodat bij de analyse der waarnemingsresultaten de getoetste hypothese van een niet van worp, seizoen of gemiddelde afhankelijke spreiding, zonder bezwaar kan worden gebruikt.

### 6. Enkele tabellen.

A. Contrôletabel voor de constantheid der spreiding bij toekomstige waarnemingsgroepen. Deze tabel bevat de grenzen, waarbinnen de som van de varianties van 10,5 resp. 1

s~~ecte~~ubgroep behoudens een wh 0,05 zullen vallen, indien gelijk is aan de opgegeven waarde (deze opgegeven waarde is de uit het tot nu toe verzamelde materiaal verkregen schatting).

Tabel A

m = aantal subgroepen
n = aantal waarnemingen per subgroep.

| m | n | grenzen voor $\Sigma s_2^2$ | | grenzen voor $\Sigma s_3^2$ | | grenzen voor $\Sigma s_4^2$ | |
|---|---|---|---|---|---|---|---|
| 10 | 12 | 494.5 | 841.3 | 890.4 | 1514. | 1702. | 2895. |
| | 11 | 487.4 | 850.6 | 877.2 | 1531. | 1677. | 2928. |
| | 10 | 478.9 | 861.8 | 862.0 | 1551. | 1648. | 2966 |
| | 9 | 469.0 | 875.0 | 844.5 | 1575. | 1614. | 3012. |
| | 8 | 457.4 | 891.1 | 823.1 | 1604. | 1574. | 3067. |
| | 7 | 443.0 | 911.3 | 797.3 | 1641. | 1525. | 3137. |
| | 6 | 424.8 | 938.0 | 764.8 | 1688. | 1462. | 3228. |
| 5 | 12 | 217.2 | 461.8 | 391.1 | 831.5 | 747.6 | 1589. |
| | 11 | 212.4 | 469.0 | 382.4 | 844.2 | 731.2 | 1614. |
| | 10 | 206.9 | 477.1 | 372.5 | 858.9 | 712.2 | 1642. |
| | 9 | 200.5 | 487.0 | 361.0 | 876.8 | 690.0 | 1676. |
| | 8 | 192.9 | 498.9 | 347.3 | 898.3 | 664.0 | 1717. |
| | 7 | 183.7 | 514.0 | 330.7 | 925.3 | 632.3 | 1769. |
| | 6 | 172.2 | 533.6 | 310.1 | 960.8 | 592.8 | 1837. |
| 1 | 12 | 22.77 | 130.8 | 41.00 | 235.5 | 78.38 | 450.2 |
| | 11 | 21.32 | 134.5 | 38.38 | 242.1 | 73.37 | 462.8 |
| | 10 | 19.70 | 138.8 | 35.46 | 249.8 | 67.80 | 477.6 |
| | 9 | 17.89 | 143.9 | 32.20 | 259.0 | 61.56 | 495.2 |
| | 8 | 15.85 | 150.2 | 28.53 | 270.4 | 54.54 | 516.9 |
| | 7 | 13.54 | 158.1 | 24.37 | 284.6 | 46.60 | 544.0 |
| | 6 | 10.92 | 168.5 | 19.65 | 303.3 | 37.57 | 580.0 |

Hierbij is: $\sigma_2'^2 = 65,6545$

$\sigma_3'^2 = 118,1911$

$\sigma_4'^2 = 225,9534$

Voor de berekening van deze tabel vergelijke men het "Derde verslag betreffende groeiproeven met Wistarratten" (Rapport S 14, Mei 1949).

B. Aanvullende tabel van $x.\sigma' \sqrt{\dfrac{n_1+n_2}{n_1 n_2}}$ , waarin x gedefinieerd is door

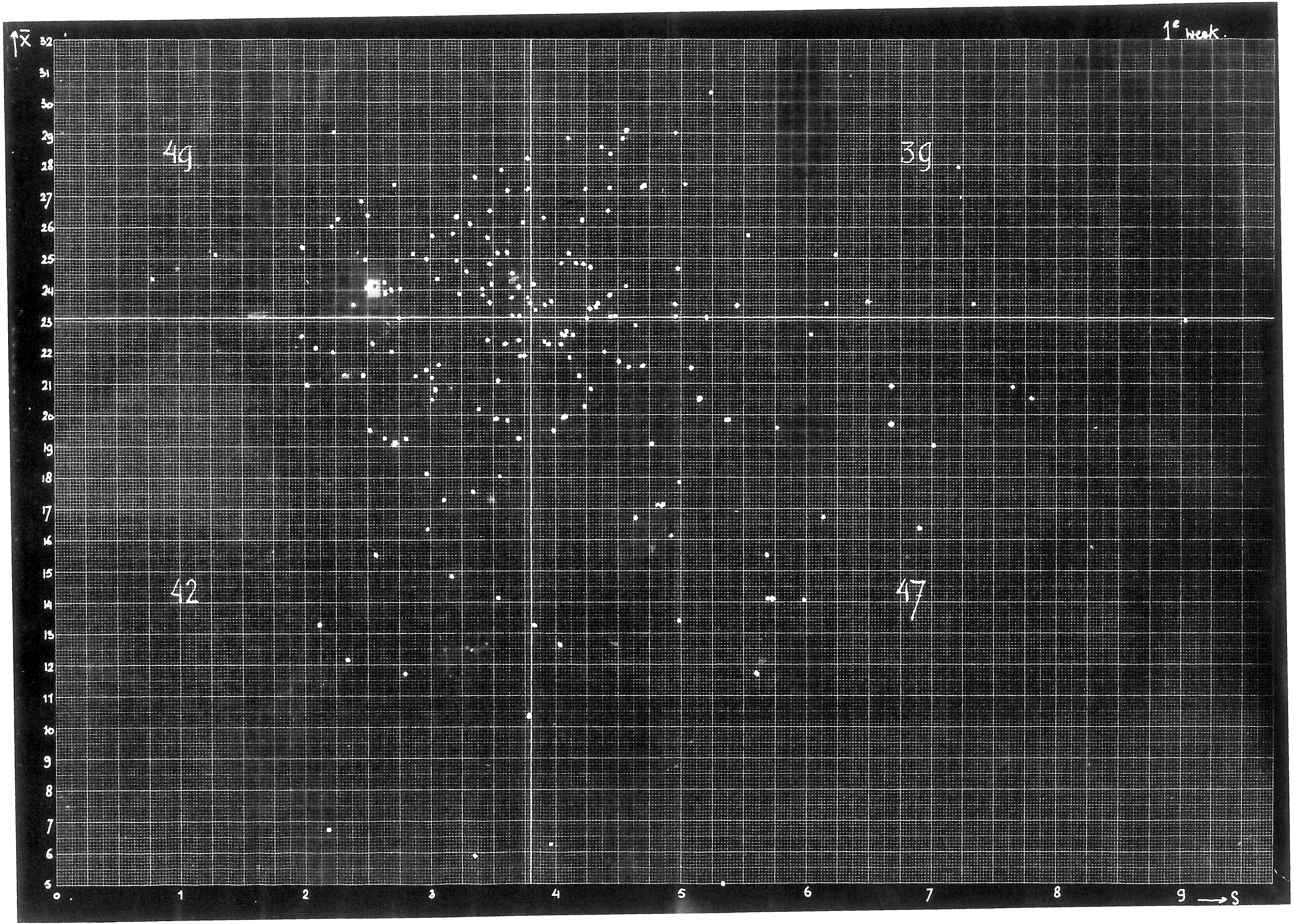
$$P = \frac{2}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{1}{2}u^2}\, du,$$

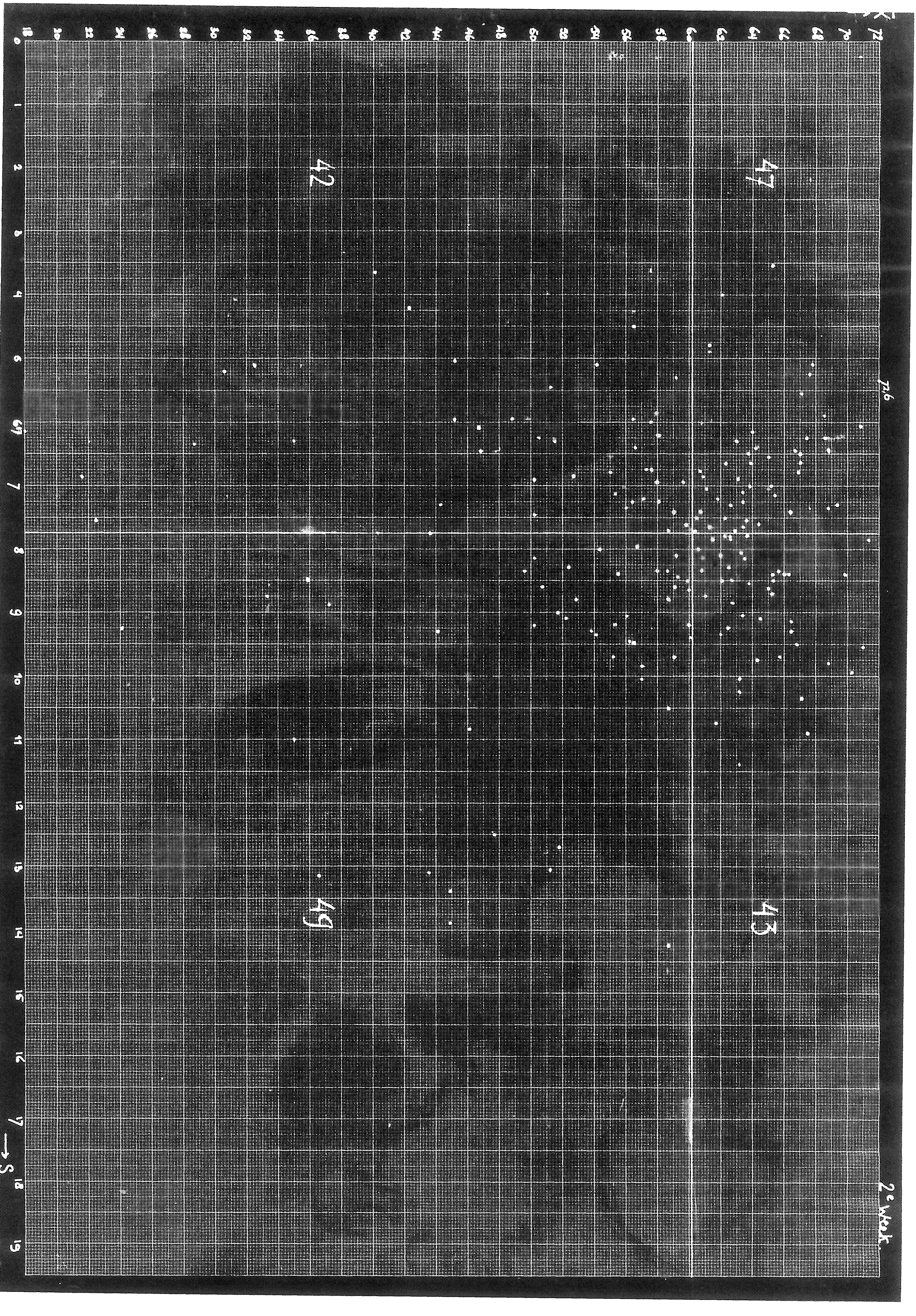
terwijl $\sigma'$ de drie op de vorige bladzijde aangegeven waarden aanneemt.

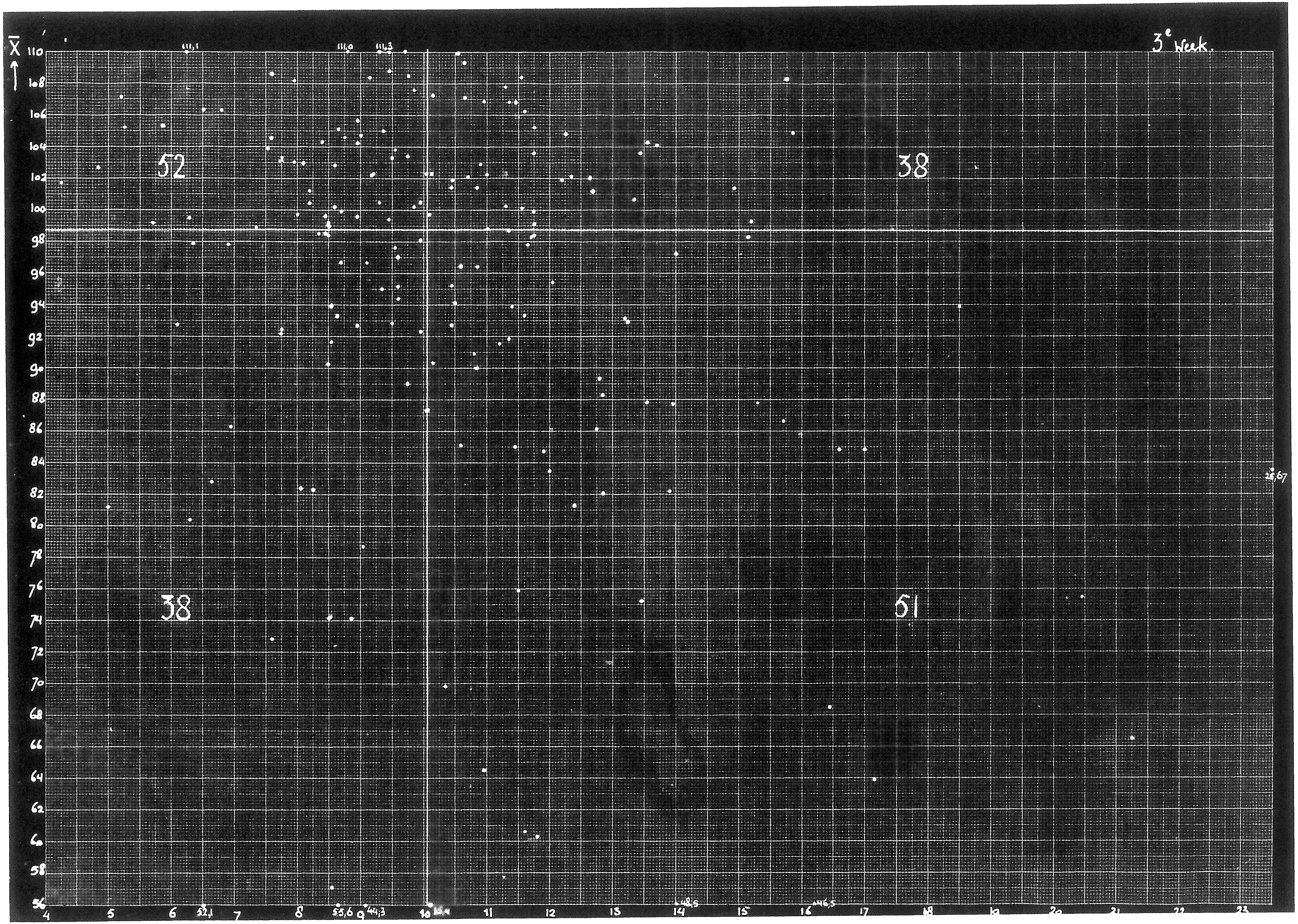
Tabel B

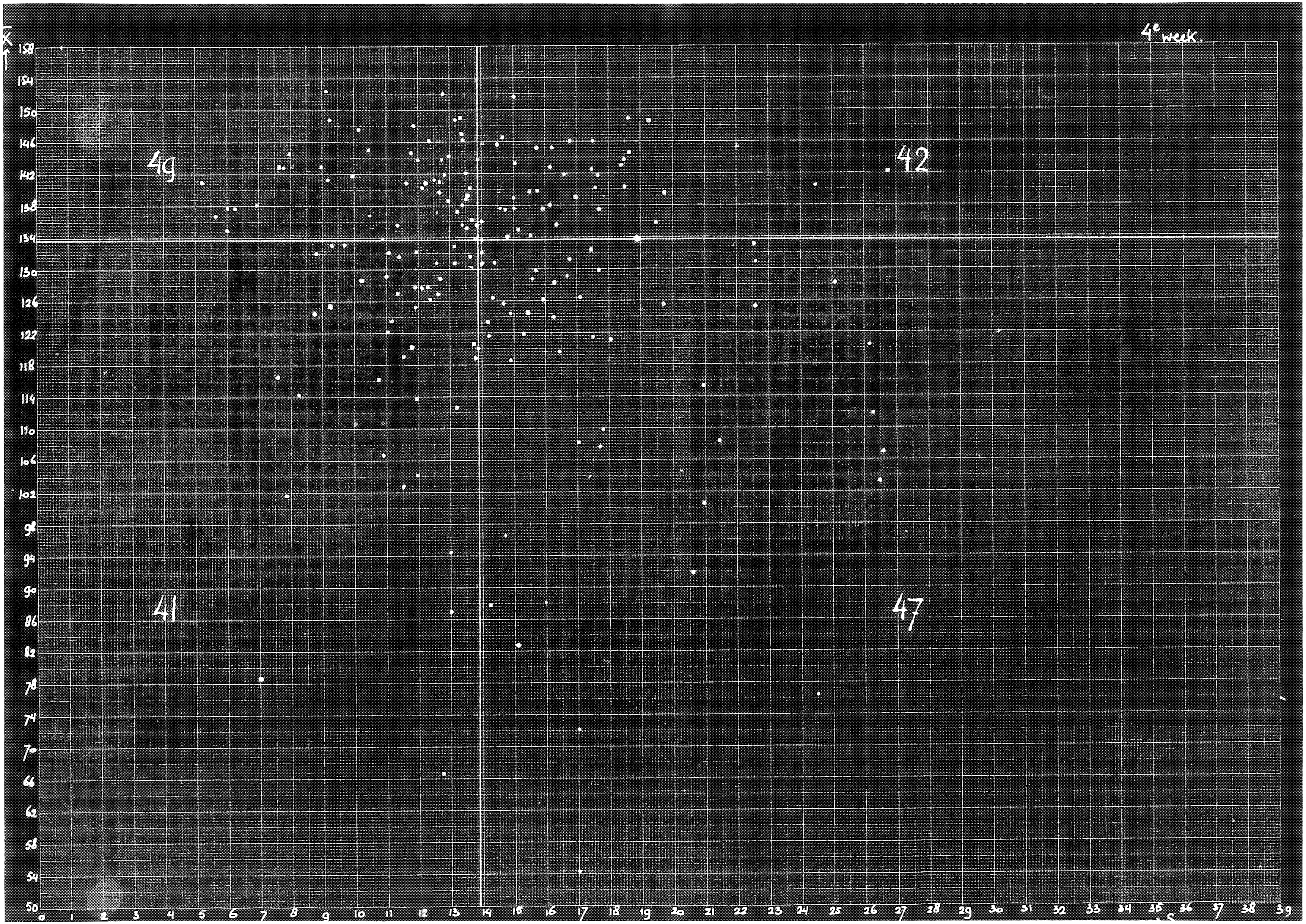
| | | 2 weken | | | 3 weken | | | 4 weken | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$ | $n_2$ | P=0,10 | P=0,05 | P=0,01 | P=0,10 | P=0,05 | P=0,01 | P=0,10 | P=0,05 | P=0,01 |
| 12 | 7 | 6,34 | 7,55 | 9,93 | 8,50 | 10,13 | 13,32 | 11,76 | 14,01 | 18,41 |
| 12 | 6 | 6,66 | 7,94 | 10,44 | 8,94 | 10,65 | 14,00 | 12,36 | 14,73 | 19,36 |
| 11 | 7 | 6,44 | 7,68 | 10,09 | 8,65 | 10,30 | 13,54 | 11,95 | 14,24 | 18,72 |
| 11 | 6 | 6,76 | 8,06 | 10,59 | 9,08 | 10,81 | 14,21 | 12,55 | 14,95 | 19,65 |
| 10 | 7 | 6,57 | 7,83 | 10,29 | 8,81 | 10,50 | 13,80 | 12,19 | 14,52 | 19,08 |
| 10 | 6 | 6,88 | 8,20 | 10,78 | 9,23 | 11,00 | 14,46 | 12,77 | 15,21 | 19,99 |
| 9 | 7 | 6,72 | 8,00 | 10,52 | 9,01 | 10,74 | 14,11 | 12,46 | 14,85 | 19,51 |
| 9 | 6 | 7,02 | 8,37 | 11,00 | 9,43 | 11,23 | 14,76 | 13,03 | 15,53 | 20,41 |
| 8 | 7 | 6,90 | 8,22 | 10,80 | 9,26 | 11,03 | 14,49 | 12,80 | 15,25 | 20,04 |
| 8 | 6 | 7,20 | 8,58 | 11,27 | 9,66 | 11,51 | 15,12 | 13,35 | 15,91 | 20,91 |
| 7 | 7 | 7,12 | 8,49 | 11,16 | 9,56 | 11,39 | 14,97 | 13,22 | 15,75 | 20,70 |
| 7 | 6 | 7,42 | 8,84 | 11,61 | 9,95 | 11,85 | 15,58 | 13,76 | 16,39 | 21,54 |
| 6 | 6 | 7,70 | 9,17 | 12,05 | 10,32 | 12,30 | 16,17 | 14,28 | 17,01 | 22,35 |

Deze tabel bevat alleen die waarden, die niet reeds door Drs Thomasson waren berekend.


49
39
42
47


42
47
49
43
2e week


3e Week.
$\bar{X}$
52
38
38
51


4e week.
x
s
158
154
150
146
142
138
134
130
126
122
118
114
110
106
102
98
94
90
86
82
78
74
70
66
62
58
54
50
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
4g
42
41
47

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 $\alpha$, zodat Z dus van $\alpha$ afhankelijk is[3)]. Z heet de kritieke zône van de toets, $\alpha$ de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor $\alpha$ 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 $\alpha$ moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien $\mathcal{H}_0$ juist is, gelijk aan $\alpha$. Derhalve is $\alpha$ 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 <u>geen</u> mogelijkheid om tot <u>aanvaarding</u> van een hypothese te komen. Indien een bepaalde hypothese $\mathcal{H}_0$ 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 $\alpha$ wordt vaak bij de uitslag van een toetsing de <u>overschrijdingskans</u> $k$ opgegeven; dit is de kleinste waarde van $\alpha$, waarbij in het betrokken geval, nog tot verwerping van $\mathcal{H}_0$ zou zijn overgegaan; anders gezegd: de kleinste $\alpha$, waarvoor de gevonden waarde der toetsingsgrootheid nog juist in de (bij $\alpha$ behorende) kritieke zône Z ligt. <u>Wordt dus de waarde $k$ opgegeven en werkt men met onbetrouwbaarheidsdrempel $\alpha$, 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 $\mathcal{H}_0$ leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

<u>Litteratuur:</u>

J.Neyman, <u>First</u> 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 $H_0$, inhoudend dat twee steekproeven $x_1, \ldots, x_n$ en $y_1, \ldots, y_m$ 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 $n\,m$ dergelijke paren zijn, kan $U$ alle gehele waarden van 0 tot en met $n.m$ 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 $Z$ neemt nu daarom de kleine en de grote waarden van $U$ en wel van beide zoveel, dat de gekozen onbetrouwbaarheidsdrempel $\alpha$ niet overschreden wordt.

Voor éénzijdige toetsing, te onderscheiden in linker- en rechter-éénzijdige toetsing, gebruikt men kritieke zônes $Z_1$, resp. $Z_2$, 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 $x_1, \ldots, x_n$ en $y_1, \ldots, y_m$ 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,
2de Boerhaavestraat 49,
A m s t e r d a m - O.
Statistische Afdeling.



S 47 (M 15).

## Toets voor onafhankelijkheid van 2 grootheden met behulp van de methode der Dubbele Dichotomie ')

Deze methode dient o.a. om de onafhankelijkheid te toetsen van twee continue grootheden. $(x_1, y_1)$, $(x_2, y_2)$ --- $(x_n, y_n)$ zijn waarnemingsparen van de stochastische grootheden $\underline{x}$ en $\underline{y}$. Deze getallenparen worden als punten in een vlak getekend.

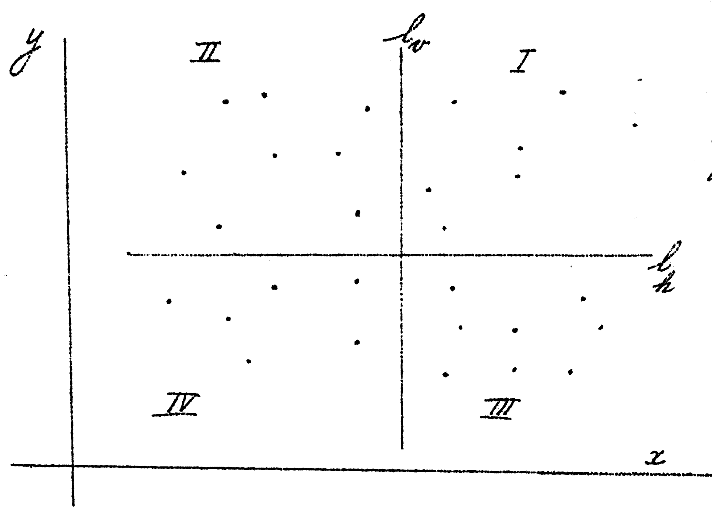



fig 1

We verdelen de puntenwolk door een verticale en een horizontale rechte ($l_v$ resp. $l_h$) zo, dat links en rechts van $l_v$, en ook boven en onder $l_h$ evenveel punten liggen.*) Het vlak is nu verdeeld in vier gebieden. Is er geen afhankelijkheid dan verwachten we dat in alle vier gebieden ongeveer evenveel punten zullen liggen. Een exacte afleiding, (geschikt voor kleine aantallen waarnemingen) verkrijgen we als volgt:

We beschouwen de puntne in ons vlak als waarnemingen van elementen die ieder tegelijk twee kenmerken bezitten; A of $\bar{A}$ '') (rechts resp. links van verticale streep) en B of $\bar{B}$ (boven resp. onder horizontale streep). Er ontstaan dus vier groepen elementen met de kenmerken

| $AB$ | $A\bar{B}$ | $\bar{A}B$ | $\bar{A}\bar{B}$ |
|---|---|---|---|
| vak I | vak III | vak II | vak IV |

We tekenen dit als volgt in een 2 x 2 tabel: (Zie blz 2)

----------

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

'') $\bar{A}$ betekent: "non-A"

*) Komen eenige punten op een der lijnen te liggen, dan laten wij deze verder buiten beschouwing

| | $A$ | $\bar{A}$ | |
|---|---|---|---|
| $B$ | $a$ | $n-a$ | $n$ |
| $\bar{B}$ | $b$ | $m-b$ | $m$ |
| | $r$ | $s$ | $N$ |

Het totale aantal elementen is $N$. Het aantal elementen dat kenmerk $B$ bezit, is $n$, het aantal dat kenmerk $\bar{B}$ bezit, is $m$. Hetzelfde geldt voor de kenmerken $A$ en $\bar{A}$ met de aantallen $r$ resp. $s$.
In het bovenstaande voorbeeld werd de puntenwolk zo verdeeld, dat $n = m = r = s = \frac{1}{2}N$
Dit principe wordt vaak toegepast, maar is niet noodzakelijk. De toets kan ook worden gebruikt met andere waarden van $m, n, r, s$ mits deze steeds op grond van de proefopstelling, niet naar aanleiding van de uitkomst van het experiment worden gekozen. 1)

Onafhankelijkheid van de kenmerken wil zeggen, dat de kans dat een element het kenmerk B bezit even groot is, indien het tevens het kenmerk A bezit, als het tevens het kenmerk Ā bezit. Wij toetsen nu deze hypothese van onafhankelijkheid.

Daar de rand-totalen gegeven zijn, is er nog één vrijheidsgraad, d.w.z. indien één van de waarden in één der vier binnenvakjes bepaald is, volgen daaruit de andere waarden.
Wij beschouwen het getal $a$ in het linker-boven vakje. Dit kan verschillende waarden aannemen en is dus een stochastische grootheid $\underline{a}$, waarvan de waarschijnlijkheidsverdeling, onder aanname van de te toetsen hypothese, een z.g. hypergeometrische verdeling is, die gegeven wordt door

$$P[\underline{a} = a] = \frac{\binom{n}{a}\binom{m}{r-a}}{\binom{m+n}{r}} = \frac{n!\,m!\,r!\,s!}{a!\,b!\,(n-a)!\,(m-b)!\,N!}$$

Het gemiddelde van deze verdeling is $\mathcal{E}\underline{a} = \frac{nr}{N}$ en het spreidings kwadraat is $\sigma^2_{\underline{a}} = \frac{n\,m\,r\,s}{N^2(N-1)}$

Als voorbeeld nemen wij $N=20$, $r=s=n=m=10$
De waarschijnlijkheidsverdeling van $\underline{a}$ is nu in fig. 2 en tabel I weergegeven.

[illegible] het begin kunnen wij de horizontale en de verticale deellijn zó trekken, dat er geen punten op liggen en dat het aantal punten ter weerszijden van elk van deze lijnen zo weinig mogelijk van elkaar verschilt. [illegible] indien het aantal punten klein is, [illegible] punten op de deellijnen [illegible] getallen v.b.

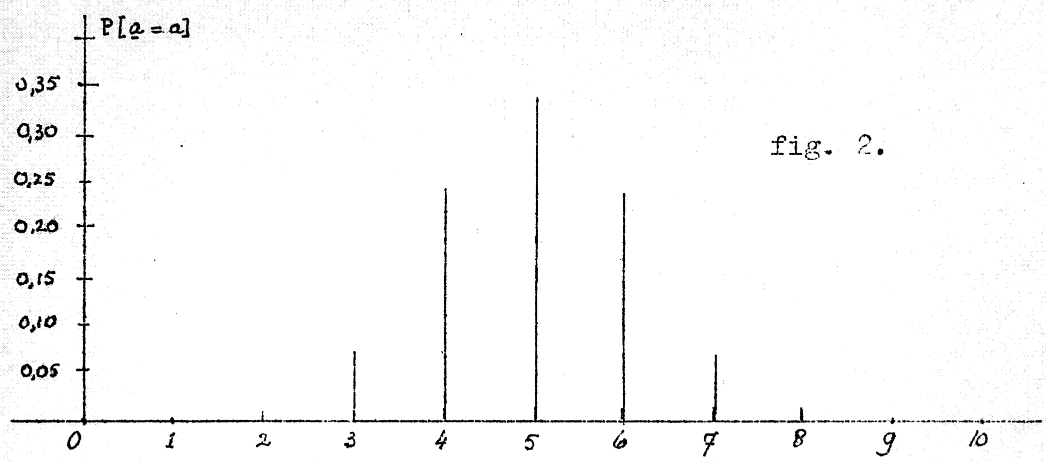



fig. 2.

mogelijkheden:

0|10 / 10|0 — 1|9 / 9|1 — 2|8 / 8|2 — 3|7 / 7|3 — 4|6 / 6|4 — 5|5 / 5|5 — 6|4 / 4|6 — 7|3 / 3|7 — 8|2 / 2|8 — 9|1 / 1|9 — 10|0 / 0|10 —

Tabel I

| $a$ | $P[\underline{a} = a]$ |
|---|---|
| 0 | 0,00001 |
| 1 | 0,00054 |
| 2 | 0,01096 |
| 3 | 0,07794 |
| 4 | 0,24356 |
| 5 | 0,34372 |
| 6 | 0,24356 |
| 7 | 0,07794 |
| 8 | 0,01096 |
| 9 | 0,00054 |
| 10 | 0,00001 |

Stel nu dat de volgende waarden gevonden zijn:

| | $A$ | $\bar{A}$ | |
|---|---|---|---|
| $B$ | 2 | 8 | 10 |
| $\bar{B}$ | 8 | 2 | 10 |
| | 10 | 10 | 20 |

We berekenen

$$P[\underline{a} = 2] = \frac{(10!)^4}{20!(8!)^2(2!)^2} = 0,01096$$

en zoeken verder alle waarden van $a$ bijeen, waarvoor $P[\underline{a} = a] \leq P[\underline{a} = 2]$ is

Vervolgens tellen we de daarbij behorende waarschijnlijkheden $P[\underline{a} = a]$ op. (In ons voorbeeld dus voor $\underline{a}$ = 0,1,2,8,9,10). Deze som is per definitie de overschrijdingskans, behorende bij het gevonden resultaat, (in ons voorbeeld 0,02302) [1])

[1]) De bij deze definitie van de overschrijdingskans behorende kritieke zône bestaat uit alle waarden voor $a$ met overschrijdingskans $\leq \alpha$ ,in ons voorbeeld 0,1,2,8,9 en 10.

Is deze $< \alpha$ ($\alpha$ heet de onbetrouwbaarheidsdrempel), dan wordt de hypothese van onafhankelijkheid verworpen (In ons voorbeeld treedt dus, als men $\alpha = 0{,}05$ neemt, verwerping op).

Voor grote aantallen waarnemingen maken we gebruik van het feit dat $\frac{a - \bar{a}}{\sigma_a}$ bij benadering normaal verdeeld is, met gemiddelde 0 en spreiding 1.

Continuïteitscorrectie

Deze behoeft alleen bij kleine aantallen toegepast te worden en bestaat daarin, dat men alle getallen $a, b, n-a, m-b$ met $\frac{1}{2}$ vermeerdert of vermindert, zodanig dat de randtotalen dezelfde blijven $|a - \bar{a}|$ kleiner wordt.

Literatuur:

M.G.Kendall, The advanced theory of Statistics, Vol.I, (London 1947), p.303;

E.S.Pearson, The choice of statistical tests illustrated on the interpretation of data classed in a 2 x 2 table, Biometrika 34 (1949) p.139-167.