STATISTISCHE AFDELING
MATHEMATISCH CENTRUM
AMSTERDAM

September 1950.

Rapport S 45
door J. Hemelrijk.

Zesde verslag omtrent groeiproeven met Wistar-ratten.

1. Inleiding.

De in dit verslag vermelde berekeningen sluiten aan op het "Derde verslag omtrent groeiproeven met Wistarratten", d.d. 18 Mei 1949. De in dat verslag beschreven analyse van de spreidingen van de voedseltoename gedurende de eerste 5 weken is nu voor analoge gegevens van de eerste 4, 3, 2 en 1 week (weken) uitgevoerd. Terwille van de duidelijkheid geven wij een samenvatting van de in vorige verslagen vermelde resultaten, voor zoverre deze van belang zijn voor de in dit rapport beschreven analyse.

De proeven zijn steeds van het volgende type: een aantal "sub-groepen", gewoonlijk ieder van 12 ratten, wordt terzelfdertijd ingezet, met verschillend dieet voor iedere subgroep. De gewichtstoename g_1 gedurende de eerste week, g_2 gedurende de eerste twee weken, enz., wordt voor iedere rat waargenomen. De proef duurt tegenwoordig 4 weken; gewoonlijk zijn er dan in de sub-groepen, door ziekte of andere oorzaken, enkele ratten uitgevallen. De sub-groepen van één dergelijke proef tezamen genomen zullen wij een "proef-~~gr~~groep" van ratten nomen. Analoog zullen wij spreken van "subproeven" en "proeven". Het voor dit rapport gebruikte waarnemingsmateriaal is afkomstig van de proeven met de volgende nummers: 476 A; 477 A; 478 A; 479 A; 4710 A; 4711 A; 4713 A; 4714 A; 4715 A; 4716 A; 4717 ; 4718 A; 4719 A; 4721; 4724 A; 4725; 4726; 482 A; 484 A; 487 A; 4815 A; 4817 B.

De statistische analyse is uitgevoerd voor ieder der grootheden g_1, g_2, g_3 en g_4 afzonderlijk en behoeft dus slechts voor één hiervan besproken te worden. Wij zullen daarom bij de beschrijving van de analyse en de onderstelling die daaraan ten grondslag liggen, de letter g, zonder index, gebruiken voor de gewichtstoename gedurende de eerste i weken (zodat men i = 1,2,3 of 4 ~~maar~~ behoefte kan substitueren).

2. Onderstellingen.

Omtrent deze grootheid g worden nu, bij de verwerking der waarnemingsresultaten, zoals deze wordt uitgevoerd, de volgende onderstellingen gemaakt.

1^e. De binnen een subgroep verrichte waarnemingen zijn onafhankelijke waarnemingen van een normaal verdeelde stochastische variabele g .
2^e. De spreiding (standaarddeviatie) σ van deze stochastische variabele is bij alle proeven en sub-proeven dezelfde

Toelichting: Deze spreiding is wèl verschillend voor i = 1, 2, 3 en 4, dus voor g_1, g_2, g_3 en g_4 . Verder betekent het constant zijn van σ bij de sub-proeven niet, dat de experimenteel gevonden spreidingen steeds gelijk aan σ zullen zijn. Zij zullen uiteraard om σ heen variëren. Hier komen wij nog op terug.

De bedoeling van de proeven is nu met behulp van deze onderstellingen uit de waarnemingsresultaten der sub-groepen conclusies te trekken omtrent de voedingswaarde der diëten. Dit geschiedt door met behulp van de toets van Student of met behulp van een door Drs Thomasson ingevoerde vereenvoudigde methode het verschil van de sub-groep-gemiddelden te toetsen.

Deze toetsingsmethoden zijn slechts geldig, indien aan bovenstaande onderstellingen is voldaan. Daarbij dient echter opgemerkt te worden, dat de toets van Student slechts gebruik maakt van de onderstelde gelijkheid der spreidingen van de grootheid g voor die twee subgroepen, waarvan de gemiddelden tegen elkaar getoetst worden, terwijl bij de vereenvoudigde methode ondersteld wordt, dat alle spreidingen gelijk zijn. Juist dit verschil maakt de in het onderhavige verslag vermelde analyse belangrijk, daar de vereenvoudigde methode in de resultaten van deze analyse voldoende steun vindt, om in plaats van de toets van Student te worden gebruikt. Dit spaart zeer veel berekeningen.

3. Beschrijving van de toegepaste toetsingsmethode.

In het eerste en tweede verslag van het M.C. betreffende deze proeven is reeds gebleken, dat de onderstelling der normaliteit van g niet houdbaar is. Dit tast de geldigheid van de Student-toets en van de vereenvoudigde methode in gelijke mate aan. Wij houden ons in dit rapport niet bezig met het zoeken naar een rechtvaardiging voor het desondanks te gebruiken van deze methoden. Het is vrij algemeen gebruikelijk, dergelijke afwijkingen van normaliteit te negeren. De beslissing ligt in deze bij het Unilever Resear Laboratorium. Desgewenst zouden wij methoden aan de hand kunnen doen, waarbij de onderstelling van normaliteit niet wordt gebruikt.

Indien men echter besluit, de afwijkingen van de normaliteit te verwaarlozen, is het van belang na te gaan, of de hypothese, dat de spreiding σ voor alle proeven en subproeven dezelfde is, houdbaar is. Tegelijkertijd is het van belang de invloed van de afwijking van normaliteit op het gedrag van de experimenteel te vinden spreidingen na te gaan. Wij hebben deze twee aspecten op de volgende wijze in één toetsing verenigd:

De waarnemingen van g in één sub-groep geven wij aan met $x_1, \ldots, x_n$. Daar deze waarnemingen onafhankelijk zijn ondersteld (een onderstelling, die voortdurend wordt gehandhaafd en die, bij een behoorlijke uitvoering van de proef zeker verantwoord is), kunnen wij ieder der waarnemingen $x_1, \ldots, x_n$ beschouwen als één waarneming van een stochastische variabele, die we voor x_j zullen aangeven door X_j, waarbij dan al deze stochastische variabelen dezelfde waarschijnlijkheidsverdeling bezitten als g en bovendien stochastisch onafhankelijk zijn.

Het experimenteel gemiddelde

$$m = \frac{x_1 + \cdots + x_n}{n} \tag{1}$$

is dan een schatting van het theoretisch gemiddelde μ, dat voor alle X_j hetzelfde is als van g; en de experimenteel gevonden grootheid s', gedefinieerd als

$$s' = \sqrt{\frac{1}{n-1} \sum (x_j - m)^2} \tag{2}$$

is een schatting van σ (de spreiding van g en dus ook van ieder der X_j). Deze grootheid s' is echter tevens een waarneming van de overeenkomstige stochastische grootheid:

$$S' = \sqrt{\frac{1}{n-1} \sum (X_j - M)^2} \tag{3}$$

waarin

$$M = \frac{1}{n} \sum X_j \tag{4}$$

is. Indien nu de onderstellingen 1^e en 2^e, vermeld in § 2, juist zijn, volgt daaruit de vorm van de waarschijnlijkheidsverdeling van S'. En wel is dan de grootheid

$$T' = \frac{1}{2}(n-1) \frac{S'^2}{\sigma^2} \tag{5}$$

verdeeld volgens een z.g. Γ-verdeling, dat wil zeggen, dat de waarschijnlijkheidsdichtheid (de "frequentiekromme") gegeven wordt door de vergelijking

(6) $$f(T') = C.\, e^{-T'}(T')^{\frac{1}{2}(n-3)}$$

waarin C een bekende constante is. De waarde van σ is hierin nog onbekend. Hiervoor vullen wij een schatting in, verkregen uit de waarnemingen van g uit alle subgroepen tezamen, waarbij deze waarnemingen met het subgroep-gemiddelde verminderd zijn.
In formule:

(7) $$\sigma' = \sqrt{\frac{1}{N-\nu} \sum (x_j - m')^2}$$

waarin N = totaal aantal waarnemingen
ν = aantal subgroepen
x_j = waarneming van g
m' = gemiddelde van de waarnemingen van de subgroep, waartoe x_j behoort.

Het $\sum$-teken duidt aan, dat over de waarnemingen van alle sub-groepen gesommeerd is. Deze schattingen σ' van σ werden ons door Drs Thomassen verstrekt. Zij bedragen voor de vier perioden:

Tabel I
Schattingen van de spreiding voor ieder der grootheden $g_1, \ldots, g_4$.

eerste week	σ'^2 =	17,5519
eerste twee weken	σ'^2 =	~~17,5519~~ 65,6545
eerste drie weken	σ'^2 =	~~65,6545~~ 118,1911
eerste vierweken	σ'^2 =	225,9534

Na ~~Als~~ invulling van deze schatting kunnen de gevonden waarden van s' (waarnemingen van S'), worden omgerekend in waarden t' (waarnemingen van T') volgens formule (5), nadat daarin S' door s' en T' door t' vervangen is. Vervolgens kunnen deze gevonden waarden met behulp van de χ^2-toets voor "goodness of fit" vergeleken worden met de theoretische verdeling, die door (6) gegeven wordt. Een kleine moeilijkheid, die hierbij wordt veroorzaakt door de ongelijkheid der aantallen waarnemingen in iedere subgroep, kan op een reeds in het Derde Verslag vermelde wijze worden overwonnen. We zullen daar momenteel niet opnieuw op in gaan.
In principe verloopt deze χ^2-toetsing aldus: wij verdelen de T'-as (en daarmee dan tevens de S'-as) in 10 delen en wel zo, dat de mathematische verwachting van het aantal waarnemingen t', dat in ieder dezer vakjes valt, gelijk is aan 1/10 van het totale aantal waarnemingen van t' (deze mathematische verwachting wordt in dit geval ook wel de "theoretische frequentie" van het bijbehorende vakje ge-

noemd). Daar er 10 vakjes zijn en één parameter (nl. σ) door een uit de waarnemingen verkregen schatting (nl. σ') vervangen is, blijven er 8 vrijheidsgraden over voor .

Op deze wijze wordt dus de volgende hypothese H_0 getoetst: H_0: de experimenteel gevonden waarden s' van de spreiding in de subgroepjes bezitten een experimentele frequentieverdeling, die in grote lijnen niet in tegenspraak is met de onderstelling, dat g normaal verdeeld is met dezelfde spreiding voor alle subgroepen.

Indien deze hypothese niet voor verwerping in aanmerking komt, betekent dit niet, dat g inderdaad normaal verdeeld is met gelijke spreiding voor alle sub-groepen, maar alleen dat de gevonden spreidingen in de subgroepen van de afwijki[ng] van normaliteit van g geen sterke invloed ondervinden. Tevens kan men dan concluderen, dat het gebruik van de vereenvoudigde methode, indien men zekere in het Derde Verslag reeds beschreven voorzorgen in acht neemt, gerechtvaardigd kan worden geacht.

4. Resultaten der statistische analyse.

De in punt 3 beschreven χ^2 -toets is voor ieder der grootheden g_1, g_2, g_3 en g_4 apart toegepast. De nummering van de 10 vakjes van de s' -as is zodanig gekozen, dat met stijgende nummering toenemende waarden van s' corresponderen. Daar er voor ieder der perioden 181 sub-groepen zijn, dus evenzovele spreidingen berekend zijn, is de verwachtingswaarde voor ieder der vakjes 18,1. De resultaten der berekening zijn samengevat in tabel II.

Tabel II

χ^2 - toets voor de spreidingen van de grootheden g_1, g_2, g_3 en g_4.

vak no:	aantal spreidingen, dat in een vakje ligt			
	g_1	g_2	g_3	g_4
1	37	17	24	26
2	19	17	17	20
3	17	22	27	17
4	18	19	17	22
5	14	20	15	18
6	16	14	18	15
7	14	23	21	14
8	9	23	10	14
9	13	11	12	16
10	24	15	20	19
$\chi^2 =$	29,9	8,1	13,3	7,2
$P =$	0,0003	0,42	0,10	0,52

8 vrijheidsgraden.

P = overschrijdingskans ("P -waarde").

Conclusie.

1. Voor de eerste week moet de hypothese H_0 verworpen worden. In het bijzonder zijn er voor deze week veel te veel sub-groepen met zeer geringe spreiding. Deze conclusie sluit aan bij de ondervindingen van Drs Thomasson, die ook in andere opzichten een abnormaal gedrag van ~~de~~ g_1 constateerde.

2. Voor de overige drie perioden komt de hypothese H_0 niet voor verwerping in aanmerking. Dit houdt een rechtvaardiging in voor het gebruik van de "vereenvoudigde methode" van Drs Thomasson voor de toetsing van het verschil van twee gemiddelden.

De in tabel I vermelde waarden de vier 181-tallen spreidingen werden ons door Drs Thomasson verstrekt. De overige berekeningen zijn uitgevoerd door de rekenafdeling van het Mathematisch Centrum.

Bijlage bij rapport S 45.

De Student-toets voor het verschil van twee gemiddelden.

Indien X en Y twee stochastische variabelen zijn, beide normaal verdeeld met gemiddelden μ_1 en μ_2 en gelijke spreiding (standaarddeviatie) σ , kan de hypothese $\mu_1 = \mu_2$ met behulp van de Studnet-toets getoetst worden op grond van een aantal onafhankelijke waarnemingen $x_1, \ldots, x_{n_1}$ van X en $y_1, \ldots, y_{n_2}$ van Y .

Daartoe berekent men

$$m_1 = \frac{1}{n_1} \Sigma x_i \qquad \text{en} \qquad m_2 = \frac{1}{n_2} \Sigma y_i ,$$

de steekproef-gemiddelden, die schattingen zijn van de parameters μ_1 en μ_2 . Verder

$$S_1^2 = \Sigma (x_i - m_1)^2 \quad \text{en} \quad S_2^2 = \Sigma (y_i - m_2)^2$$

De grootheid

$$t = \frac{m_1 - m_2}{\sqrt{S_1^2 + S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \tag{1}$$

bezit dan een "t-verdeling" met $n_1 + n_2 - 2$ vrijheidsgraden, zodat de overschrijdingskans (de "P-waarde") in een tabel van de Student-verdeling kan worden opgezocht.

Deze toets wordt in de litteratuur hier en daar wel eens verkeerd beschreven. Dit is b.v. het geval Paterson p. 29, Bok p. 148, van der Laan p. 3. De methode, die op deze plaatsen aangegeven wordt berust op het volgende:

f onafhankelijk verdeelde

Voor het verschil van twee/stochastische variabelen u en v , met spreidingen σ_1 en σ_2 geldt:

$$\sigma_{u-v}^2 = \sigma_u^2 + \sigma_v^2 .$$

Daar verder de spreiding van het gemiddelde van n onafhankelijke waarnemingen van X (de z.g. "middelbare fout") gelijk is aan $\sigma/\sqrt{n_1}$ en van Y aan $\sigma/\sqrt{n_2}$, geldt voor de spreiding van $m_1 - m_2$:

$$\sigma_{m_1 - m_2}^2 = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} ,$$

of, als de spreidingen van X en Y niet gelijk zijn, maar σ_1 en σ_2 bedragen:

$$\sigma_{m_1 - m_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} . \tag{2}$$

Deze formule is juist; de waarde van σ (of van σ_1 en σ_2) zijn echter onbekend. De schrijvers substitueren nu voor deze onbekende parameters schattingen, die uit de waarneming verkregen worden, en wel voor σ_1^2 : $\frac{1}{n_1-1}\sum(x_i-m_1)^2$

en voor σ_2^2 : $\frac{1}{n_2-1}\sum(y_i-m_2)^2$.

Vervolgens behandelen zij de grootheid

$$(3) \qquad t = \frac{m_1-m_2}{\sigma_{m_1-m_2}}$$

alsof deze een Student-verdeling bezit met n_1+n_2-2 vrijheidsgraden. Dit is in het algemeen niet het geval. Daar voor $n_1 = n_2$ deze laatste formule voor t overgaat in de op de vorige pagina vermelde vorm (1), geldt dit wèl als $\sigma_1 = \sigma_2$ ondersteld wordt en $n_1 = n_2$ is (en x en y normaal verdeeld zijn). Indien niet te veel van deze voorwaarden afgeweken wordt, zal de verdeling ook nog wel ongeveer een Student-verdeling blijven. Er is echter geen enkele reden om, in plaats van de op de vorige pagina vermelde juiste formule, deze onjuiste te gebruiken.

Fisher, p. 116 geeft formule (2). Op pag. 122 echter formule (1) Peters and Voorhis, p. 162 formule (2); pag. 178 formule (1). Alleen bij de bovengenoemde drie schrijvers Paterson, Bok en van der Laan en in een boek over landbouw-experimenten van Pearson en Bennett konden wij tot nu toe de foutieve formule (3) vinden.

Nageslagen litteratuur met nummer der gevonden formule:

S.T. Bok, Gedachtegang der Statistica, 1946, p. 148 (3).

R.A. Fisher, Statistical Methods for Research Workers, 1948, p. 116 (2), 122 (1).

E. van der Laan, De bewerking der uitkomsten van biologische experimenten volgens de methode Fisher, 1943, p. 3 rn 8 (3).

D.D. Paterson, Statistical Technique in Agricultural Research, 1939, p. 29 (3).

F.A. Pearson and W.R. van Voorhuis, Statistical procedures, 1940, p. 178 (1).

G.W. Snedecor, Statistical methods, =948, p. 82 (1).

L.H.C. Tippett, The methods of statistics, 1948, p. 115 (1)

S.S. Wilks, Mathematical Statistics, 1946, p. 112 (1).

G.U. Yule and M.G. Kendall, Introd. to the theory of Statistics, =946, p. 442 (1).