MATHEMATISCH CENTRUM
AMSTERDAM

Amsterdam, 7 September 1949.

Archief.

Vierde verslag betreffende groeiproeven met Wistar-ratten [1)]

I.

1. De Heer Thomasson suggereert een correctie aan te brengen op de berekening der varianties. Deze verbetering is correct; naar aanleiding hiervan kunnen enige opmerkingen worden gemaakt (zie sub.2).

2. De gebruikte notatie is de volgende:

x_{ij}: de gewichtstoename van rat j uit groep (sub-collectie) i

N_i: het aantal ratten in groep i

n : het aantal groepen

N : het totaal aantal der ratten:

$$N = \sum_{i=1}^{n} N_i$$

$\bar{x}_i$: de gemiddelde gewichtstoename in groep i:

$$\bar{x}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$$

m_i: de gemiddelde gewichtstoename in groep i, naar boven of naar beneden afgerond:

$$m_i = x_i + \xi_i$$

waarbij $-\frac{1}{2} < \xi_i \leq \frac{1}{2}$.

De ongecorrigeerde spreiding σ' van de grote collectie (dus over alle N ratten) is gedefinieerd als

$$\sigma' = \left\{ \frac{\sum_{i=1}^{n} \sum_{j=1}^{N_i} (x_{ij} - m_i)^2}{N-n} \right\}^{\frac{1}{2}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

terwijl de gecorrigeerde spreiding σ'' wordt gegeven door

$$\sigma'' = \left\{ \frac{\sum_{i=1}^{n} \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i)^2}{N-n} \right\}^{\frac{1}{2}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Uit (1) en (2) volgt:

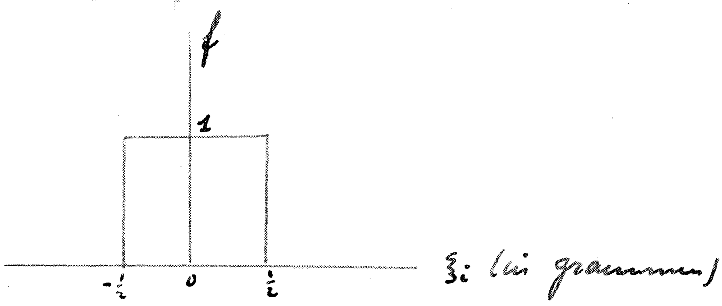
$$(N-n)\sigma''^2 = (N-n)\sigma'^2 - \sum_{i=1}^{n} N_i \xi_i^2 \quad \ldots\ldots\ldots\ldots\ldots (3)$$

De identiteit (3) is gelijkwaardig met de door de Heer Thomasson geformuleerde correctie. Deze correctie is echter klein in vergelijking met de meetnauwkeurigheid; het is daarom zinvol te volstaan met een approximatie van deze correctie, die sub.3 gegeven wordt.

1) Door J. Hemelrijk en H. Theil

3. Stel, dat ξ_i in zijn interval $(-\frac{1}{2}, \frac{1}{2})$ homogeen verdeeld is, d.w.z. dat ξ_i met gelijke waarschijnlijkheid in even grote sub-intervallen voorkomt. De verdelingsdichtheid van ξ_i wordt dan weergegeven door de in fig.1 getekende frequentie-curve.

Fig. 1



In plaats van de grootheid

$$X = \sum_{i=1}^{m} N_i \xi_i^2$$

beschouwen we haar mathematische verwachting:

$$\mathcal{E} X = \mathcal{E} \sum_{i=1}^{m} N_i \xi_i^2 = \sum_{i=1}^{m} N_i \, \mathcal{E} \xi_i^2 =$$

$$= (N_1 + \ldots + N_n) \int_{-\frac{1}{2}}^{\frac{1}{2}} \xi^2 \, d\xi =$$

$$= \frac{N}{12} .$$

Vervanging X door $\mathcal{E}X$ in (3) geeft

$$\sigma''^2 \approx \sigma'^2 - \frac{N}{12(N-n)}$$

Indien N = 1350 en m = 126, krijgt men:

$$\sigma''^2 \approx \sigma'^2 - 0{,}092$$

$$\approx \sigma'^2 \left(1 - \frac{0{,}092}{\sigma'^2}\right)$$

Dus geldt:

$$\sigma'' \approx \sigma' \left(1 - \frac{0{,}046}{\sigma'^2}\right)$$

$$\approx \sigma' - \frac{0{,}046}{\sigma'}$$

II.

4. In Tabel I worden σ', σ'' en σ''/t (t uitgedrukt in weken) voor ieder der 6 weken gegeven.

Tabel I

t	σ'	σ''	σ''/t
1	3,812	3,800	3,800
2	7,919	7,913	3,957
3	11,032	11,028	3,676
4	15,175	15,172	3,794
5	19,339	19,337	3,867
6	21,332	21,330	3,555

De Heer Thomasson stelt de vraag of de grootheden σ''/t als constant beschouwd kunnen worden, of dat hun onderlinge verschillen als significant moeten worden aangemerkt. Het blijkt dat voor twee weken inderdaad significante verschillen moeten worden vastgesteld. De betreffende berekeningen volgen sub.5.

5. We beschouwen nu (voor elke week afzonderlijk) de grote groep van N ratten als een steekproef uit een nog grotere collectie, waarvan de spreiding σ_0 is. We gaan verder uit van de hypothese, dat σ_0/t voor alle weken gelijk is.
De spreiding σ'' van de grote groep is dan (weer voor iedere week afzonderlijk) een stochastische variabele (een steekproef-grootheid), waarvan de verdeling moet worden nagegaan. Echter kan de verdeling van de nog grotere collecties niet als normaal beschouwd worden, omdat Pearson's grootheden β_1 resp. β_2 ongelijk 0 resp. 3 gebleken zijn. Dit bemoeilijkt het houden van waarschijnlijkheidsbeschouwingen omtrent σ''. Gezien het zeer grote aantal waarnemingen (1300 à 1400) is het echter niet onredelijk aan te nemen, dat de verdeling van σ'' wèl ongeveer normaal zou zijn; dit in tegenstelling met de oorspronkelijke verdeling waaruit σ'' berekend is. De (normale) verdeling van σ'' heeft dan als gemiddelde σ_0 en als spreiding

$$u = \left\{ \frac{\sigma_0^2 \, (\beta_2 - 1)}{4N} \right\}^{\frac{1}{2}}$$

waarbij β_2 Pearson's kurtosis-coëfficient van de oorspronkelijke verdeling is. Bij wijze van benadering stellen we $\sigma_0 \approx \sigma''$ en $\beta_2 = b_2$, waarbij b_2 de kurtosis-coëfficient is, die in de "Statistische analyse" van de Heer Thomasson is berekend. Thans kunnen betrouwbaarheidsintervallen voor σ_0 resp. σ_0/t worden berekend. Dusdanige intervallen, berekend uit steekproef-grootheden, bevatten de te schatten parameter (in casu σ_0) met een van tevoren opgegeven waarschijnlijkheid. In dit geval bevat het betrouwbaarheidsinterval ($\sigma'' - 2u$, $\sigma'' + 2u$) met 95,5% waarschijnlijkheid de grootheid σ_0, en dus het interval

$\left(\frac{\sigma'' - 2u}{t}, \frac{\sigma'' + 2u}{t} \right)$ met deze waarschijnlijkheid de grootheid σ_0/t.
In Tabel II worden deze intervallen voor ieder der 6 weken gegeven.

Tabel II

t	u	$\frac{\sigma'' - 2u}{t}$	$\frac{\sigma'' + 2u}{t}$
1	0,094	3,612	3,988
2	0,154	3,803	4,111
3	0,267	3,498	3,854
4	0,366	3,610	3,976
5	0,438	3,692	4,043
6	0,457	3,403	3,707

Het gemiddelde van σ''/t over de 6 weken bedraagt 3,775. Dit gemiddelde wil de Heer Thomasson gebruiken als een approximatie voor σ_0/t. Uit Tabel II volgt nu, dat de waarde 3,775 zowel buiten het betrouwbaarheidsinterval van de tweede als buiten dat van de zesde week ligt. De hypothese, dat σ_0/t voor alle 6 weken constant is, moet dus verworpen worden.
De Heer Thomasson laat dan ook de zesde week in zijn "Statistische analyse" verder buiten beschouwing en bepaalt zich tot de eerste vijf weken. Het gemiddelde van deze weken is 3,819; dit ligt binnen elk der eerste 5 betrouwbaarheidsintervallen. De hypothese, dat σ_0/t voor de eerste 5 weken constant is en wel gelijk aan 3,819, komt dus niet voor verwerping in aanmerking. Het verwerpen van de zesde week (en niet b.v. van de tweede week) moet op biologische gronden berusten en valt dus buiten onze competentie. Aangezien echter de Heer Thomasson ons mededeelde, dat het groeiverloop in de latere weken van karakter pleegt te veranderen, moet het buiten beschouwing laten van de laatste week als niet onredelijk beschouwd worden.

6. In zijn "Statistische analyse" (§5) geeft de Heer Thomasson een andere significantie-berekening voor σ''/t. Deze berust echter op de veronderstelling, dat de oorspronkelijke verdeling normaal is. Daarom moet de sub.5 gegeven methode geprefereerd worden.