

STICHTING
MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

S 5

Dantzig, D. van, Theil, H.

Tweede voorlopig verslag betreffende groepproeven met Wistar-ratten.

1948



**Tweede voorlopig verslag betreffende groeiproeven
met Wistar-ratten.**

I

De opdrachtgever heeft schriftelijk het eerste voorlopig verslag beantwoord; naar aanleiding hiervan moeten de volgende opmerkingen gemaakt worden:

ad 4 Niet alleen voor de derde week wijkt β , (misschien) significant van 0 af; een duidelijke significante afwijking van 0 treedt op voor de collectie van de eerste week: b_1 is voor deze collectie 0.13; de afwijking van dit getal van nul is 5 maal zo groot als de afwijking, die behoort bij een overschrijdingkans van 1%. We moeten dus zeggen, dat β voor de eerste week significant van 0 afwijkt, voor de derde week dubieus, voor de overige weken niet significant.

ad 6 Het is dus gebleken, dat de 6 grote collecties niet geacht kunnen worden normaal verdeeld te zijn. Maar zelfs al zou dit het geval zijn, dan nog mag men zeker niet concluderen, dat elk der kleinere collecties normaal verdeeld is. Het is bijv. denkbaar, dat van deze sub-collecties een gedeelte links-steil verdeeld is en een ander deel rechts-steil, zodanig dat de samenvoeging van deze collecties toch weer normaal verdeeld is. Aan elke willekeurige verdeling kan men een daartoe geschikte andere verdeling toevoegen, zodanig, dat de samenvoeging normaal is.

De opdrachtgever heeft in § 4 van zijn "statistische analyse" gepeogd een inductieve verificatie te geven van de hypothese, dat de sub-collecties hetzelfde verdelingstype hebben als de grote collecties. Deze beschouwing is onjuist, hetgeen als volgt kan worden aangetoond.

Beschouw de gewichtstoename in een zekere week. Stel x_{ij} de gewichtstoename van de j^{de} rat in de i^{de} sub-collectie. Stel er zijn n sub-collecties; de subcollectie i telt N_i ratten; De volledige collectie telt dus $\sum_{i=1}^n N_i = N$ ratten. Het groeps-gemiddelde voor de sub-collectie i is dus

$$\frac{1}{N_i} \sum_j x_{ij} = m_i$$

en het algemene gemiddelde voor de volledige collectie

$$\frac{1}{N} \sum_i \sum_j x_{ij} = \sum_i \frac{N_i}{N} m_i = m$$

Het kwadraat van de spreiding voor de sub-collectie i is

$$S_i^2 = \frac{1}{N_i} \sum_j x_{ij}^2 - m_i^2$$

Het kwadraat van de algemene spreiding voor de gehele collectie is

$$\begin{aligned}
 S^2 &= \frac{1}{N} \sum_i \sum_j x_{ij}^2 - m^2 = \\
 &= \frac{1}{N} \sum_i (N_i s_i^2 + m_i^2) - m^2 = \\
 &= \sum_i \frac{N_i}{N} s_i^2 + \sum_i \frac{N_i}{N} m_i^2 - m^2 = \\
 &= \sum_i \frac{N_i}{N} s_i^2 + \sum_i \frac{N_i}{N} (m_i - m)^2
 \end{aligned}$$

De opdrachtgever heeft op de gegevens een dussdanige bewerking toegepast, dat alle groepsgemiddelden gelijk nul zijn en dus ook het algemene gemiddelde. We krijgen dus in dit geval:

$$S^2 = \sum_i \frac{N_i}{N} s_i^2$$

Indien we stellen, dat alle sub-collecties ongeveer even groot zijn, wordt dit:

$$S^2 = \frac{1}{n} \sum_i s_i^2$$

Deze gelijkheid is een identiteit, d.w.z. zij geldt steeds, hoe groot x_{ij} ook mag zijn, m.a.w. zij geldt voor elke verdeling van de gewichtsvermeerderingen. Elke inductieve verificatie aan de hand van deze identiteit moet dus als resultaat opleveren, dat de variance van de grote collectie ongeveer gelijk is aan het gemiddelde van de variances van de sub-collectie (ongeveer gelijk, omdat niet alle sub-collecties even groot zijn). Een dergelijke "verificatie" kan dus nooit leiden tot een dieper inzicht omtrent de aard van de betreffende verdelingsfuncties.

In werkelijkheid heeft de opdrachtgever niet gewerkt met het kwadraat van de spreiding, maar met de spreiding zelf: hij heeft vergeleken de spreiding van de grote collectie (s) met het gemiddelde van de spreidingen van de sub-collecties (indien het aantal ratten per greep niet te sterk varieert: $\frac{1}{n} \sum_i s_i$). Eenvoudig kan worden aangetoond (bijv. door middel van de ongelijkheid van Cauchy-Schwarz), dat het gemiddelde van de spreidingen van de sub-collecties steeds

kleiner dan of gelijk aan de spreiding van de grote collectie is, hetgeen in de betreffende paragraaf is gebleken.

Hetgeen de opdrachtgever in zijn § 4 heeft verricht, blijkt dus te zijn een empirische verificatie van een benadering van een identiteit. Conclusie hieruit te trekken is niet verantwoord.

ad 7 Op deze punten moet nog nader worden ingegaan.

ad 8

II

De opdrachtgever stelt enige wegen voor, die kunnen worden ingeslagen bij het onderzoek van de niet normaal gebleken verdelingen. Dit zal door ons nader worden bezien.

III

De opdrachtgever besluit zijn brief met het maken van enkele opmerkingen.

ad 1 Wanneer men de variabele y met de variabele x correleert, stelt men vaak 2 regressievergelijkingen op. Welke van beide vergelijkingen men als "de meest juiste" interpreteert, hangt af van de veronderstellingen, die men omtrent de variabelen maakt. Neemt men aan, dat de x -waarden volkomen nauwkeurig zijn en de y -waarden aan stochastische invloeden onderhevig zijn, dan komt men tot de eerste regressievergelijking; neemt men aan, dat de y -waarden nauwkeurig zijn en de x -waarden niet, dan gebruikt men de tweede regressievergelijking. Iedere regressievergelijking met een regressiecoëfficient tussen die van de eerste en de tweede regressievergelijking heeft betrekking op een bepaalde onnauwkeurighedsverhouding van de beide variabelen.

In het onderhavige geval zijn de correlatiecoëfficienten weinig van de eenheid verwijderd (zie ook sub IV), zodat de eerste en de tweede regressiecoëfficient weinig verschillen; het bovengenoemde probleem is hier dus van betrekkelijk gering belang.

ad 2 Deze punten moeten nog nader onderzocht worden.

ad 3

IV

Tenslotte de resultaten van de berekeningen van onze Rekenafdeling.

1. Voor de 24 soorten proefvoedsel zijn bepaald de correlatie- en regressiecoëfficienten en de spreidingen van de correlatiecoëfficienten voor het verband tussen de gewichtsvermeerderingen in de eerste en tweede week, tweede en derde week, ..., vijfde en zesde week. In onderstaande tabel zijn de resultaten van de opdrachtgever (ODG), vergeleken met die van de Rekenafdeling (RA)

Weken	1/2		2/3		3/4		4/5		5/6	
	ODG	RA	ODG	RA	ODG	RA	ODG	RA	ODG	RA
b_{xy}	0.42	0.44	0.69	0.68	0.76	0.83	0.80	0.84	0.92	0.88
b_{yx}	2.09	2.13	1.31	1.39	1.23	1.17	1.14	1.16	1.06	1.06
r	0.93	0.97	0.95	0.98	0.96	0.98	0.96	0.99	0.99	0.97
σ_r	0.004		0.003		0.002		0.001		0.004	

De betrekkelijk geringe afwijkingen van de resultaten van de opdrachtgever en van de Rekenafdeling kunnen verklaard worden door de omstandigheid, dat de opdrachtgever een collectie van geringere uitgebreidheid bewerkte.

2. Vervolgens heeft de Rekenafdeling door middel van een lineaire correlatieberekening tussen E 20 Z b en E 0 de gecorrigeerde E 20 Z b-waarden berekend, van de gewichtsvermeerderingen behorende bij de verschillende doseringen (10, 20, ..., 70) deze gecorrigeerde E 20 Z b waarden afgetrokken, de 6 aldus gevonden getallen steeds door het weeknummer gedeeld en daarvan het gemiddelde bepaald over de eerste 4 weken en over de eerste 6 weken. Aldus zijn aan elke soort van proefvoedsel en aan elke dosering 2 gemiddelden toegevoegd. Voor elke soort proefvoedsel zijn deze gemiddelden afgezet als functie van de dosis, waarbij de vier-weekse gemiddelden door een stippellijn en de zes-weekse door een getrokken lijn verbonden zijn.

De aanwijzing van de opdrachtgever, dat het niet nodig is het verband tussen dosis en remming grafisch weer te geven, kwam te laat in ons bezit; de Rekenafdeling had het resultaat reeds in deze vorm gegeven. Het aflezen van de grafische resultaten is echter eenvoudig.

V

Met uitzondering van de resultaten van de Rekenafdeling draagt dit tweede voorlopige verslag een enigszins negatief karakter. Het is echter niet mogelijk binnen zo korte tijd resultaten van nieuw ingeslagen wegen te geven. Wij hopen te zijner tijd een meer volledig verslag te kunnen verzenden.

Bijlagen: 22 grafieken.

18 November 1948

Den Heer Thomasson
Unilever Laboratories
ZWIJNDRECHT.

Geachte Heer,

Ingesloten treft U aan het tweede voorlopig verslag betreffende
groeiproeven met Wistar-ratten.

Naar aanleiding hiervan zult U binnenkort een schrijven
ontvangen van Prof. van der Corput.

Hoogachtend,

J.

secretaresse.

Tweede voorlopig verslag betreffende groeiproeven
met Wistar-ratten.

I

De opdrachtgever heeft schriftelijk het eerste voorlopig verslag beantwoord; naar aanleiding hiervan moeten de volgende opmerkingen gemaakt worden:

ad 4 Niet alleen voor de derde week wijkt β , (misschien) significant van 0 af; een duidelijke significante afwijking van 0 treedt op voor de collectie van de eerste week: b_1 is voor deze collectie 0.13; de afwijking van dit getal van nul is 5 maal zo groot als de afwijking, die behoort bij een overschrijdingskans van 1%. We moeten dus zeggen, dat voor de eerste week significant van 0 afwijkt, voor de derde week dubieus, voor de overige weken niet significant.

ad 6 Het is dus gebleken, dat de 6 grote collecties niet gesocht kunnen worden normaal verdeeld te zijn. Maar zelfs al zou dit het geval zijn, dan nog mag men zeker niet concluderen, dat elk der kleinere collecties normaal verdeeld is. Het is bijv. denkbaar, dat van deze sub-collecties een gedeelte links-steil verdeeld is en een ander deel rechts-steil, zodanig dat de samenvoeging van deze collecties toch weer normaal verdeeld is. Aan elke willekeurige verdeling kan men een daartoe geschikte andere verdeling toevoegen, zodanig, dat de samenvoeging normaal is.

De opdrachtgever heeft in § 4 van zijn "statistische analyse" gepoogd een inductieve verificatie te geven van de hypothese, dat de sub-collecties hetzelfde verdelingstype hebben als de grote collecties. Deze beschouwing is onjuist, hetgeen als volgt kan worden aangetoond.

Beschouw de gewichtstoename in een zekere week. Stel x_{ij} de gewichtstoename van de j^{de} rat in de i^{de} sub-collectie. Stel er zijn n sub-collecties; de subcollectie 1 telt N_1 ratten; De volledige collectie telt dus $\sum_{i=1}^n N_i = N$ ratten. Het groeps-gemiddelde voor de sub-collectie 1 is dus

$$\frac{1}{N_1} \sum_j x_{1j} = m_1$$

en het algemene gemiddelde voor de volledige collectie

$$\frac{1}{N} \sum_i \sum_j x_{ij} = \sum_i \frac{N_i}{N} m_i = m$$

Het kwadraat van de spreiding voor de sub-collectie 1 is

$$S_1^2 = \frac{1}{N_1} \sum_j x_{1j}^2 - m_1^2$$