MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 172

Statistische schatting en vergelijking van de verhouding van opheldering en vergrauwing bij verschillende wasmiddelen.

(vertrouwelijk)

door

Constance van Eeden

Mei 1955

The Mathematical Centre at Amsterdam, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1. Inleiding

Door het proefstation voor de wasindustrie werden de volgende wasproeven uitgevoerd:

Per wasproef werden 10 zakjes gewassen. Van deze 10 zakjes waren er 6 gemaakt van kunstmatig vuilgemaakt materiaal en 4 van schoon materiaal.
Vòòr het wassen werd van ieder zakje op vier verschillende plaatsen de helderheid gemeten en het gemiddelde van deze vier waarnemingen berekend. Daarna werden de 10 zakjes in een wasmachine gewassen.
Na het wassen werd van ieder zakje weer op vier plaatsen de helderheid gemeten en het gemiddelde van deze vier waarnemingen berekend.
Vervolgens werd voor de kunstmatig vuilgemaakte zakjes het verschil in helderheid na en vòòr het wassen (de opheldering) berekend en voor de schone zakjes het verschil in helderheid vòòr en na het wassen (de vergrauwing).

Deze proeven werden uitgevoerd met drie wasmiddelen, die aangeduid worden met A, S en R. Voor iedere wasproef werden de zakjes zodanig over de drie wasmiddelen verdeeld dat zowel voor de 18 kunstmatig vuilgemaakte als voor de 12 schone zakjes, de gemiddelde helderheid der zakjes vòòr het wassen voor de drie wasmiddelen zoveel mogelijk gelijk was.

Deze proef werd driemaal uitgevoerd, iedere keer met deze drie wasmiddelen. In totaal werden dus 90 zakjes gewassen.
Op grond van deze waarnemingen hebben wij voor ieder der wasmiddelen berekend:

a. een schatting van de verhouding δ tussen opheldering en vergrauwing,
b. een schatting van de spreiding van de onder a genoemde schatting.

2. Beschrijving van de methode en resultaten

De reden dat wij werken met de verhouding δ tussen opheldering en vergrauwing is, dat volgens verstrekte inlichtingen, bij niet te grote verschillen in vuilconcentratie in het sop, deze verhouding vrijwel onafhankelijk is van deze vuilconcentratie en alleen afhangt van het vuilverwijderend en vuildragend vermogen van het wasmiddel.

Stel nu dat de j^e waarneming van de opheldering, voor een bepaald wasmiddel, bij de i^e wasproef een waarneming is van een stochastische grootheid $\underline{x}_{ij}$ [1] ($i = 1,2,3$; $j = 1,2, \ldots,6$) en stel verder

$$\underline{x}_{ij} = \mu + \underline{u}_i + \underline{v}_{ij} ,$$

waarbij $\underline{u}_i$ en $\underline{v}_{ij}$ voor iedere i en j verwachting 0 hebben en waarbij $\sigma^2(\underline{v}_{ij})$ niet van j afhangt. Analoog voor de vergrauwing

$$\underline{y}_{ij} = \mu' + \underline{u}'_i + \underline{v}'_{ij} .$$

Dan houdt dus de bovenstaande onderstelling over de constantheid van δ in dat

$$\frac{\mu + \underline{u}_i}{\mu' + \underline{u}'_i} = \delta$$

is en niet van i afhangt. Strict genomen betekent dit, dat de correlatie tussen $\underline{u}_i$ en $\underline{u}'_i$ gelijk aan 1 zou zijn. Ook als hieraan slechts bij benadering voldaan is, kan onder bepaalde omstandigheden het genoemde quotiënt nog wel als een constante beschouwd worden, b.v. als $\underline{u}_i$ en $\underline{u}'_i$ een kleine spreiding hebben in vergelijking met de grootte van μ en μ', terwijl de correlatie tussen $\underline{u}_i$ en $\underline{u}'_i$ groot is.

Om nu de in de inleiding genoemde schattingen te verkrijgen vergelijken wij eerst, voor iedere wasmiddel apart, de drie wasproeven, zowel voor de opheldering als voor de vergrauwing, wat betreft hun gemiddelde en hun varianties.
Om de varianties te vergelijken passen wij FISHERS's F-toets toe op tweetallen steekproeven [2]; wij krijgen dan voor ieder wasmiddel zes overschrijdingskansen: drie voor de opheldering en drie voor de vergrauwing. De resultaten van deze toets staan vermeld in tabel I.

---

1) Stochastische grootheden worden onderscheiden van getallen (b.v. van de waarden die ze bij een experiment aannemen) door hun symbolen te onderstrepen.
2) In plaats van de F-toets van FISHER, toegepast op paren steekproeven, kan ook de toets van BARTLETT of die van COCHRAN, voor de gelijkheid van twee of meer varianties toegepast worden op drietallen steekproeven.

Tabel I

Overschrijdingskansen gevonden bij de vergelijking van de varianties

wasmiddel	Overschrijdingskansen voor					
	opheldering			vergrauwing		
A	0,72	0,74	1	0,52	0,56	0,92
S	0,40	0,28	0,06	0,12	0,60	0,28
R	0,84	1,00	0,84	0,24	0,84	0,32

Hieruit zien wij dus dat er geen reden is om aan te nemen dat de varianties verschillend zijn.
Om te onderzoeken of de gemiddelden verschillen passen wij de toets van STUDENT toe op tweetallen steekproeven. Wij vinden dan

Tabel II

Overschrijdingskansen gevonden bij de vergelijking der gemiddelden

wasmiddel	Overschrijdingskansen voor					
	opheldering			vergrauwing		
A	0,04	0,04	0,95	1	0,40	0,50
S	0,70	0,05	0,06	0,20	0,04	1,00
R	0,15	0,30	0,70	0,07	0,04	0,30

Hieruit zien wij dat er wel aanwijzingen zijn dat de gemiddelden verschillen.
Om een schatting voor δ te krijgen zullen wij dus, in verband met het verschil tussen de gemiddelden der drie wasproeven, voor ieder der wasproeven apart een schatting voor δ berekenen en dan het gemiddelde van drie schattingen nemen.
Om deze schattingen te vinden maken wij gebruik van het volgende:

Als $\underline{u}$ en $\underline{v}$ onderling onafhankelijk verdeeld zijn en $\underline{z} = \frac{\underline{u}}{\underline{v}}$ dan geldt:

(1) $$\mathcal{E}\,\underline{z} \approx \frac{\mathcal{E}\,\underline{u}}{\mathcal{E}\,\underline{v}} \left\{ 1 + \frac{\sigma^2(\underline{v})}{(\mathcal{E}\,\underline{v})^2} \right\}.$$

Indien $\frac{\sigma(\underline{v})}{\mathcal{E}\underline{v}} \ll 1$ (wat bij ons het geval zal blijken te zijn) dan is $\underline{z}$ dus bij benadering een zuivere schatting van $\frac{\mathcal{E}\underline{u}}{\mathcal{E}\underline{v}}$. Verder geldt:

$$(2) \qquad \sigma^2(\underline{z}) \approx \frac{\sigma^2(\underline{u})(\mathcal{E}\underline{v})^2 + \sigma^2(\underline{v}).(\mathcal{E}\underline{u})^2 + 3\sigma^2(\underline{v}).\sigma^2(\underline{u})}{(\mathcal{E}\underline{v})^4}.$$

Een schatting van deze variantie verkrijgt men door in (2) schattingen in te vullen voor de gemiddelden en varianties van $\underline{u}$ en $\underline{v}$.

Als nu

$$(3) \qquad \bar{x}_i = \frac{1}{6}\sum_j x_{i,j} \quad , \quad \bar{y}_i = \frac{1}{4}\sum_j y_{i,j}$$

dan is dus

$$(4) \qquad d_i = \frac{\bar{x}_i}{\bar{y}_i}$$

een schatting van δ voor de i^e wasproef. Als schatting voor δ uit de drie wasproeven tezamen nemen wij dus

$$(5) \qquad d = \frac{1}{3}\sum_i d_i .$$

Deze schattingen staan vermeld in tabel III

Tabel III
Schattingen voor δ

wasmiddel	schatting
A	3,62
S	9,02
R	10,55

Uit (5) volgt nu

$$(6) \qquad \sigma^2(\underline{d}) = \frac{1}{9}\sum_i \sigma^2(\underline{d}_i);$$

een schatting voor $\sigma^2(\underline{d})$ vinden wij dus uit schattingen voor $\sigma^2(\underline{d}_i)$ die wij weer vinden door in (2) voor $\underline{u}$ en $\underline{v}$ resp. $\bar{\underline{x}}_i$ en $\bar{\underline{y}}_i$ in te vullen. Schattingen voor $\mathcal{E}\bar{\underline{x}}_i$ en $\mathcal{E}\bar{\underline{y}}_i$ zijn $\bar{x}_i$ resp. $\bar{y}_i$ en schattingen voor $\sigma^2(\bar{\underline{x}}_i)$ en $\sigma^2(\bar{\underline{y}}_i)$ zijn

$$(7) \qquad s^2(\bar{\underline{x}}_i) = \frac{1}{2}\sum_i (\bar{x}_i - \bar{x})^2 \quad . \quad s^2(\bar{\underline{y}}_i) = \frac{1}{2}\sum_i (\bar{y}_i - \bar{y})^2 ,$$

waarin

(8) $$\bar{x} = \frac{1}{3} \sum_i \bar{x}_i \ , \quad \bar{y} = \frac{1}{3} \sum_i \bar{y}_i .$$

Bij de afleiding der schattingen (7) is gebruik gemaakt van de onderstelling dat $\sigma^2(x_{ij})$ niet van i afhangt. Deze onderstelling is boven (zie tabel I) getoetst en kon op grond van het waarnemingsmateriaal niet verworpen worden. Berekenen wij op deze wijze de schatting voor $\sigma^2(\underline{d})$ dan vinden wij als schatting voor $\sigma(\underline{d})$:

Tabel IV

Schattingen voor $\sigma(\underline{d})$

wasmiddel	schatting
A	0.0209
S	0.0941
R	0.1430

Uit de tabellen III en IV blijkt direct dat de wasmiddelen duidelijk verschillende waarden van δ bezitten.

Opmerking: Wij wijzen erop, dat de beschreven methode alleen dan van toepassing is, als de proeven zo ingericht zijn, dat δ inderdaad voor ieder wasmiddel als een (ongeveer) constant getal beschouwd kan worden.

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 $x_1, x_2, \ldots, x_n$, 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 $\leq \alpha$ is.

De toetsingstheorie biedt in het algemeen geen mogelijkheid om tot aanvaarding 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 nevenvoorwaarden 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 overschrijdingskans 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. Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leq \alpha$ is.

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.

Litteratuur:

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 4 (1950) p.54-66.

De toets van Student voor 2 steekproeven [1]

Deze toets van Student wordt gebruikt voor het toetsen van de hypothese $\mathcal{H}_0$, dat 2 steekproeven $x_1, x_2, \ldots, x_n$ en $y_1, y_2, \ldots, y_m$ beschouwd kunnen worden als steekproeven uit dezelfde normale verdeling.

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

$$\underline{t} = \frac{\underline{m}_x - \underline{m}_y}{\underline{S}} \sqrt{\frac{nm}{n+m}}$$

Hierin zijn $\underline{m}_x$ en $\underline{m}_y$ de gemiddelden van de 2 steekproeven terwijl S een schatting is van de spreiding σ van de normale verdeling, waaruit volgens $\mathcal{H}_0$ de steekproeven getrokken zijn. De bij de gevonden steekproeven behorende waarde van $\underline{m}_x$, $\underline{m}_y$ en S zijn op de volgende wijze uit de waarnemingen te berekenen:

$$m_x = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) \qquad m_y = \frac{1}{m}(y_1 + y_2 + \ldots + y_m)$$

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

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 $\underline{t}$ berekend, welke behoren bij de onbetrouwbaarheidsdrempels $\alpha = 0{,}01$ en $\alpha = 0{,}05$. [2]

Is $\mathcal{H}_0$ 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 α, 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 $\underline{t} \geqq t_1$.

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. (In deze tabel worden tweezijdige overschrijdingskansen gegeven).

---

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

THEMATISCH CENTRUM
e Boerhaavestr. 49
M.S.T.E.R.D.A.M 0.
atistische Afdeling
S 53 (M 24)

z-toets van Fisher[1)]

Stel we hebben twee onafhankelijke steekproeven

$x_1 \dots x_{n_1}$ met gemiddelde $\bar{x}$
$y_1 \dots y_{n_2}$ 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 toetsingsgrootheid 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 <u>z</u> 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.