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 195
(Vertrouwelijk)

Oriënterend rapport over de samenstelling en dosering van een zelfwerkend wasmiddel.

door

Prof. Dr J. Hemelrijk

Maart
1956

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. Probleemstelling

Hieronder volgt, letterlijk overgenomen, de probleemstelling, zoals deze aan ons werd voorgelegd.

Een zelfwerkend wasmiddel is een mengsel van twee poedervormige stoffen nl.:

1) zeeppoeder
2) bleekmiddel.

Het zeeppoeder heeft een zeker gehalte aan vetzuren.

Het mengsel van zeeppoeder en bleekmiddel (d.i. dus het zelfwerkend wasmiddel) wordt verpakt in eenheden, waarvan het minimum gewicht 250 gram per eenheid moet bedragen ($g_{min.}$).

De standaard-afwijking van de verpakkingsgewichten is bekend (s_g).

Behalve de eis van een minimum gewicht van 250 gram per verpakte eenheid worden aan het zelfwerkend wasmiddel als zodanig nog twee eisen gesteld en wel:

a) het <u>gehalte</u> aan bleekmiddel in het mengsel moet tenminste 6,5% bedragen.
b) het <u>gehalte</u> aan vetzuren in het mengsel moet tenminste 40% bedragen.

Het eerste ingrediënt (het zeeppoeder) levert het vetzuur. De vetzuurfractie in dit zeeppoeder bedraagt x en varieert met een standaard-afwijking s_x, welke bekend is.

De zeeppoeder-fractie in het zelfwerkend wasmiddel is y met standaard-afwijking s_y.

De bleekmiddel-fractie in het eindproduct is $(1-y)$ met standaard-afwijking $s_{(1-y)}$.

Bij de controle in het bedrijf wordt de bleekmiddel-fractie geregeld bepaald, waardoor automatisch de zeeppoeder-fractie bekend is.

De standaard-afwijking van de bleekmiddel-fractie ($s_{(1-y)}$) is bekend en gelijk aan de standaard-afwijking van de zeeppoeder-fractie (s_y).

Zeeppoeder-fractie en bleekmiddel-fractie zijn gecorreleerd ($\rho = -1$).

<u>Vraagstelling</u>:

Hoeveel moet, om aan de drie minimum eisen te voldoen (met een bepaalde over- resp. onderscheidingskans) bedragen:

1) het gemiddelde gewicht per pakje;
2) het gemiddelde vetzuurgehalte van het zeeppoeder;
3) het gemiddelde bleekmiddel-gehalte van het mengsel.

2. Opmerkingen over de probleemstelling en de gegevens

Wat de probleemstelling betreft valt op te merken, dat deze niet geheel ondubbelzinnig is, in het bijzonder wat de eisen a) en b) betreft. Beschouwen wij b.v. eis a): het gehalte aan bleekmiddel in het mengsel moet tenminste 6,5% bedragen. Hier is niet duidelijk aangegeven wat met "het mengsel" bedoeld wordt: het mengsel in ieder (of vrijwel ieder) pakje afzonderlijk, ieder (of vrijwel ieder) monster van bepaalde grootte uit een willekeurig pakje, of een grote hoeveelheid wasmiddel, die nog over een aantal pakjes verdeeld moet worden. In dit rapport zal de eerstgenoemde interpretatie aanvaard worden: berekeningen omtrent gehalten zullen steeds betrekking hebben op één gehele verpakkingseenheid als monstergrootte. Dit geldt dus zowel voor eis a) als b).

In overeenstemming hiermee zullen ook de in de probleemstelling genoemde fracties op deze wijze geïnterpreteerd worden. Zo is dus [1)]

$\underline{x}$ = de vetzuurfractie van het zeeppoeder uit een aselect uit de productie gekozen pakje,

en

$\underline{y}$ = de zeeppoeder-fractie van het wasmiddel van een aselect uit de productie gekozen pakje.

De standaard-afwijkingen van $\underline{x}$ en $\underline{y}$, behorende bij deze monstergrootte, geven wij aan met $\sigma_{\underline{x}}$ en $\sigma_{\underline{y}}$ en wij nemen aan dat deze bekend zijn. Of zij precies overeenstemmen met de in de probleemstelling bedoelde $s_{\underline{x}}$ en $s_{\underline{y}}$ is ons niet bekend en blijve voor het moment buiten beschouwing. Onze beschouwingen kunnen, indien dit wenselijk zou blijken, ook uitgewerkt worden voor andere monstergrootten dan verpakkingseenheden. Zijn de waarnemingen verricht aan meerdere kleine monsters uit een aantal pakjes, dan worden $\sigma_{\underline{x}}$ en $\sigma_{\underline{y}}$ geschat door de wortels uit de z.g. "tussen-pakjes"-variantie. Analoog hiermee geven wij met $\sigma_{\underline{g}}$ de standaard-afwijking van het gewicht van één verpakkingseenheid aan.

Naast de standaard-afwijkingen der beschouwde grootheden zijn hun populatie-gemiddelden (verwachtingen) van belang. Deze geven wij aan met $\mu_{\underline{g}}$, $\mu_{\underline{x}}$ en $\mu_{\underline{y}}$. In tegenstelling tot de σ's wordt van de μ's niet ondersteld dat zij bekend zijn, doch wel dat de machines op een bepaalde grootte der μ's ingesteld kunnen worden.

---

1) De beschouwde grootheden zijn stochastisch, hetgeen wij aangeven door de desbetreffende symbolen te onderstrepen.

De antwoorden op de gestelde vragen zullen gegeven worden in de vorm van aan deze μ's op te leggen voorwaarden.

In dit oriënterende rapport wordt eenvoudigheidshalve ondersteld, dat alle beschouwde en ook de verderop nog in te voeren stochastische grootheden bij (voldoende) benadering normaal verdeeld zijn. Of dit zo is, en zo neen, in hoeverre dan een verfijning van het onderzoek en van de uitkomsten nodig is, valt op grond van de verstrekte probleemstelling niet na te gaan. Dit punt blijve daarom voor later bewaard.

De bij het vullen der pakjes gevolgde procedure is zodanig, dat eerst een grote hoeveelheid van het wasmiddel gemaakt wordt en dat deze vervolgens over een aantal pakjes wordt verdeeld. Dit betekent, dat de samenstelling van het wasmiddel eerst beschouwd kan worden en daarna de gewichtsverdeling der verpakkingseenheden. Wij zullen dus eerst de vragen 2) en 3) bespreken en daarna vraag 1).

3. Beantwoording van de vragen 2) en 3)

Bij de beantwoording van de vragen 2) en 3) hebben wij te maken met de volgende grootheden.

$\underline{x}$ = de vetzuurfractie van het zeeppoeder in een aselect gekozen pakje, met standaard-afwijking $\sigma_{\underline{x}}$.

$\underline{y}$ = de zeeppoeder-fractie in het wasmiddel van hetzelfde pakje, met standaard-afwijking $\sigma_{\underline{y}}$.

$\underline{z}$ = $\underline{x} \cdot \underline{y}$ = de vetzuurfractie van het wasmiddel in het aselect gekozen pakje. De standaard-afwijking $\sigma_{\underline{z}}$ van $\underline{z}$ moet, ter oplossing van het probleem, berekend worden uit de gegevens omtrent $\underline{x}$ en $\underline{y}$.

Onderstelling

Het ligt voor de hand de vereenvoudigende onderstelling te maken, dat $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld zijn. Practisch gezien betekent dit, dat de fractie zeeppoeder in een pakje niet afhangt van de vetzuurfractie van dit zeeppoeder. Mocht deze onderstelling, die ons plausibel toeschijnt, niet vervuld zijn, dan moeten de verderop te geven formules herzien worden. Technische gegevens over de wijze van productie en - uiteraard - experimenteel onderzoek kunnen over de juistheid van de onderstelling inlichtingen verschaffen.

Zijn $\mu_{\underline{x}}$ en $\mu_{\underline{y}}$ de populatie-gemiddelden van $\underline{x}$ en $\underline{y}$, dan volgt uit de onderstelling van onafhankelijkheid:

(1) $$\mu_{\underline{z}} = \mu_{\underline{x}} \cdot \mu_{\underline{y}} \,.$$

Verder is, zoals hieronder bewezen wordt:

(2) $$\sigma^2_{\underline{z}} = \sigma^2_{\underline{x}}\ \sigma^2_{\underline{y}} + \sigma^2_{\underline{x}}\ \mu^2_{\underline{y}} + \sigma^2_{\underline{y}}\ \mu^2_{\underline{x}} \,.$$

<u>Bewijs</u>:

Laat $\mathcal{E}$ het symbool voor de mathematische verwachting voorstellen, dan is

$$\sigma^2_{\underline{z}} = \mathcal{E}\{\underline{z} - \mu_{\underline{z}}\}^2 = \mathcal{E}\{\underline{x}\,\underline{y} - \mu_{\underline{x}}\mu_{\underline{y}}\}^2 = \mathcal{E}\,\underline{x}^2\underline{y}^2 - 2\mu_{\underline{x}}\mu_{\underline{y}}\,\mathcal{E}\,\underline{x}\,\underline{y} + \mu^2_{\underline{x}}\mu^2_{\underline{y}} =$$
$$= \mathcal{E}\,\underline{x}^2 \cdot \mathcal{E}\,\underline{y}^2 - \mu^2_{\underline{x}}\mu^2_{\underline{y}} = (\sigma^2_{\underline{x}} + \mu^2_{\underline{x}})(\sigma^2_{\underline{y}} + \mu^2_{\underline{y}}) - \mu^2_{\underline{x}}\mu^2_{\underline{y}} =$$
$$= \sigma^2_{\underline{x}}\ \sigma^2_{\underline{y}} + \sigma^2_{\underline{x}}\ \mu^2_{\underline{y}} + \sigma^2_{\underline{y}}\ \mu^2_{\underline{x}} \,.$$

De eisen a) en b) kunnen nu als volgt geformuleerd worden:

a) Behoudens een kans α_a geldt voor een aselect gekozen pakje:
$1 - \underline{y} \geqq 0{,}065$, dus $\underline{y} \leqq 0{,}935$.

b) Behoudens een kans α_b geldt voor een aselect gekozen pakje:
$\underline{z} \geqq 0{,}4$.

Beschouwen wij eerst deze eisen apart, dan is aan de eerstgenoemde voldaan, als $\mu_{\underline{y}}$ voldoet aan

(3) $$\mu_{\underline{y}} \leqq 0{,}935 - \xi_a\,\sigma_{\underline{y}} \,,$$

waarin ξ_a de bij α_a behorende waarde van een standaard-normale verdeling voorstelt, dus die waarde waarvoor geldt:

(4) $$\frac{1}{\sqrt{2\pi}} \int_{\xi_a}^{\infty} e^{-\frac{1}{2}u^2}\, du = \alpha_a \,.$$

Aan de tweede is, volgens (1) en (2) voldaan, indien

(5) $$\mu_{\underline{z}} = \mu_{\underline{x}}\mu_{\underline{y}} \geqq 0{,}4 + \xi_b\,\sigma_{\underline{z}} = 0{,}4 + \xi_b\sqrt{\sigma^2_{\underline{x}}\,\sigma^2_{\underline{y}} + \sigma^2_{\underline{x}}\,\mu^2_{\underline{y}} + \sigma^2_{\underline{y}}\,\mu^2_{\underline{x}}} \,,$$

waarin ξ_b analoog volgens (4) bepaald wordt.

Noemen wij de rechterleden van (3) en (5) A resp. B, dus

(6) $$A = 0{,}935 - \xi_a\,\sigma_{\underline{y}} \quad ; \quad B = 0{,}4 + \xi_b\sqrt{\sigma^2_{\underline{x}}\,\sigma^2_{\underline{y}} + \sigma^2_{\underline{x}}\,\mu^2_{\underline{y}} + \sigma^2_{\underline{y}}\,\mu^2_{\underline{x}}} \,,$$

dan moet dus tegelijkertijd gelden:

$$\mu_{\underline{y}} \leqq A \qquad \text{en} \qquad \mu_{\underline{x}}\mu_{\underline{y}} \geqq B \,,$$

zodat wij als voorwaarden verkrijgen

(7) $$\mu_{\underline{y}} \leqq A \qquad \text{en} \qquad \mu_{\underline{x}} \geqq \frac{B}{\mu_{\underline{y}}} \,.$$

Kiest men de instelwaarden μ_x en μ_y zo, dat aan deze beide voorwaarden voldaan is, dan heeft men dus bij een aselect gekozen pakje hoogstens een kans α_a om er minder dan 6,5% bleekmiddel in aan te treffen en tevens een kans van hoogstens α_b, dat er minder dan 40% vetzuur in zit.

De eerste der beide ongelijkheden: $\mu_y \leqq A$, geeft geen moeilijkheden, daar A zonder meer uit de gegevens berekend kan worden. De tweede: $\mu_x \geqq \frac{B}{\mu_y}$, moet nog herleid worden, daar B nog de onbekende instelwaarde μ_x bevat. Kiezen wij eerst een waarde $\mu_y \leqq A$, dan kan de tweede eis als volgt herleid worden. Deze eis luidt:

$$\mu_x \mu_y \geqq 0{,}4 + \xi_b \sqrt{\sigma_x^2 \sigma_y^2 + \sigma_x^2 \mu_y^2 + \sigma_y^2 \mu_x^2},$$

waarin nu dus alleen de μ_x nog als onbekend beschouwd dient te worden. Of ook

$$(\mu_x \mu_y - 0{,}4)^2 \geqq \xi_b^2 (\sigma_x^2 \sigma_y^2 + \sigma_x^2 \mu_y^2 + \sigma_y^2 \mu_x^2),$$

hetgeen na groepering naar μ_x^2, μ_x en bekende term overgaat in:

$$(\mu_y^2 - \xi_b^2 \sigma_y^2)\mu_x^2 - 0{,}8\mu_y \mu_x + \{0{,}16 - \xi_b^2 \sigma_x^2 (\mu_y^2 + \sigma_y^2)\} \geqq 0.$$

Stellen wij het linkerlid gelijk aan 0, dan wordt dit een vierkantsvergelijking in μ_x, die twee wortels heeft. De grootste van deze twee is de grens voor μ_x, die wij zoeken. Dit geeft

(8) $$\mu_x \geqq \frac{0{,}4\mu_y + \xi_b \sqrt{\sigma_x^2 (\mu_y^2 + \sigma_y^2)(\mu_y^2 - \xi_b^2 \sigma_y^2) + 0{,}16\sigma_y^2}}{\mu_y^2 - \xi_b^2 \sigma_y^2}.$$

In de practijk zullen verschillende termen hierin wel zo klein blijken te zijn in vergelijking met andere, dat zij gevoegelijk weggelaten kunnen worden. De formule is echter niet zo ingewikkeld, dat het nodig is dit nu reeds te doen.

Wij vinden zo bij iedere μ_y een ondergrens voor μ_x en wel een lagere naarmate μ_y groter is. Het hangt nu van de kostenverhoudingen der verschillende bestanddelen af, welke instelwaarden de goedkoopste oplossing geven. Hierover zijn geen gegevens verstrekt en wij bespreken daarom hier alleen het geval, dat men μ_y graag zo groot mogelijk wil nemen, b.v. omdat men er (om welke reden dan ook) nietop gesteld is meer bleekmiddel in het wasmiddel te hebben dan strikt nodig is om aan eis a) te voldoen.

In dat geval zal men dus

(9) $$\mu_y = A = 0{,}935 - \xi_a \sigma_y$$

nemen en de ongelijkheid (8) gaat dan over in

$$(10) \qquad \mu_{\underline{x}} \geqq \frac{0{,}4A + \xi_b \sqrt{\sigma_{\underline{x}}^2 (A^2 + \sigma_{\underline{y}}^2)(A^2 - \xi_b^2 \sigma_{\underline{y}}^2) + 0{,}16\sigma_{\underline{y}}^2}}{\mu_{\underline{y}}^2 - \xi_b^2 \sigma_{\underline{y}}^2}.$$

Wil men zo weinig mogelijk vetzuur toevoegen, voor zoverre eis b) dit toelaat, dan zal men in (10) het $\geqq$ -teken vervangen door het = teken.

Men kan zich nu nog afvragen, hoe groot de kans is, dat bij één aselect gekozen pakje zowel aan eis a) als aan eis b) voldaan is. Voor deze kans geldt:

$$(11) \qquad P[a \text{ en } b] > 1 - (\alpha_a + \alpha_b).$$

<u>Bewijs</u>:

Laat het symbool a resp. b voorstellen, dat voorwaarde a resp. b vervuld is, dan is

$$P[a] \geqq 1 - \alpha_a \qquad \text{en} \qquad P[b] \geqq 1 - \alpha_b .$$

Nu geldt, als $\bar{a} = non - a$ en $\bar{b} = non - b$:

$$P[\bar{a} \text{ of } \bar{b}] = P[\bar{a}] + P[\bar{b}] - P[\bar{a} \text{ en } \bar{b}] < P[\bar{a}] + P[\bar{b}],$$

daar $P[\bar{a} \text{ en } \bar{b}] > 0$ is. Dus is

$$P[\bar{a} \text{ of } \bar{b}] < \alpha_a + \alpha_b$$

en

$$P[a \text{ en } b] = 1 - P[\bar{a} \text{ of } \bar{b}] > 1 - (\alpha_a + \alpha_b).$$

De totale kans, dat een pakje niet aan beide eisen zal voldoen, is dus $< \alpha_a + \alpha_b$.

Men kan nog opmerken, dat $\underline{y}$ en $\underline{x}$ positief gecorreleerd zijn, hetgeen gemakkelijk na te rekenen is. Dit betekent, dat het niet voldaan zijn aan één der twee eisen de kans op het wel voldaan zijn van de andere vergroot, maar ook, dat het wel voldaan zijn van één der eisen de kans verkleint, dat de andere het ook is. Ondanks deze "wederzijdse tegenwerking" der beide eisen, is (11) geldig.

4. <u>Beantwoording van vraag 1)</u>

De beantwoording van vraag 1) kan nu min of meer onafhankelijk van die van 2) en 3) plaats vinden, althans indien aangenomen mag worden, dat $\sigma_{\underline{z}}$ niet afhangt van de samenstelling van het wasmiddel, d.w.z. van de gekozen instelwaarden $\mu_{\underline{x}}$ en $\mu_{\underline{y}}$.

Mocht dit wel zo zijn, dan zou σ_g bepaald moeten worden voor de gekozen waarden van μ_x en μ_y . Is deze σ_g bekend, dan is de oplossing eenvoudig:

(12) $$\mu_g \geqq 25a + \xi_g \sigma_g ,$$

met α_g , behorende bij ξ_g volgens formule (4), als onbetrouwbaarheidsdrempel. Voor de kans, dat aan alle drie eisen voldaan is, geldt dan, dat deze minstens gelijk aan

(13) $$1 - (\alpha_a + \alpha_b + \alpha_g)$$

is.

5. Opmerkingen

In formule (2) zal vermoedelijk de term $\sigma_x^2 \sigma_y^2$ veel kleiner zijn dan de beide andere. Laten wij deze term weg en delen wij links en rechts door $\mu_z^2 = \mu_x^2 \mu_y^2$, dan ontstaat de formule

(4') $$\left(\frac{\sigma_z}{\mu_z}\right)^2 \approx \left(\frac{\sigma_x}{\mu_x}\right)^2 + \left(\frac{\sigma_y}{\mu_y}\right)^2 ,$$

die in de bijlage ("Oplossing van ... ")van de probleemstelling wordt gebruikt. Het verschil tussen de daar en hier gegeven oplossing is, dat de vragen, die met gehalten en met gewicht samenhangen hier afzonderlijk behandeld zijn, hetgeen beter bij de practische aspecten van de vraagstelling aansluit.

In par. 3 is verder reeds gewezen op het feit, dat de gestelde eisen niet zonder meer tot een ondubbelzinnige bepaling der instelwaarden μ_x en μ_y leiden. In het bijzonder kunnen b.v. kostenverhoudingen hier nog invloed uitoefenen. Indien het niet zonder meer vaststaat, dat $\mu_y = A$ de beste keuze is, ligt hier nog een optimaliseringsprobleem.

Tenslotte valt nog op te merken, dat problemen samenhangende met experimentele keuring van één of meer pakjes door het nemen van monsters hier niet zijn behandeld, omdat de vraagstelling hierover niets vermeldt. Voor de practijk zullen deze echter ook nog van belang zijn, daar het er niet alleen om gaat, of behoudens een zekere kans aan de gestelde eisen voldaan is, maar ook of men bij de controle inderdaad tot deze conclusie zal komen. Daarbij spelen de factoren als uitdroging, waarnemingsfouten en steekproefmethode een belangrijke rol. Wij wijzen hierop, om de indruk te vermijden, dat met het bovenstaande een volledige oplossing van het problem gegeven zou zijn. Alvorens echter verdere vragen beantwoord kunnen worden, zou eerst de precieze formulering daarvan overwogen moeten worden.

6. Samenvatting der resultaten

Onder bepaalde voorwaarden die in de paragrafen 2, 3 en 4 vermeld zijn, luidt het antwoord op de drie gestelde vragen als volgt.

1) $$\mu_{\underline{g}} \geqq 250 + \xi_g \sigma_{\underline{g}} .$$

3) $$\mu_{\underline{y}} \leqq A = 0{,}935 - \xi_a \sigma_{\underline{y}} .$$

2) $$\mu_{\underline{x}} \geqq \frac{0{,}4\,\mu_{\underline{y}} + \xi_b \sqrt{\sigma_{\underline{x}}^2 (\mu_{\underline{y}}^2 + \sigma_{\underline{y}}^2)(\mu_{\underline{y}}^2 - \xi_b^2 \sigma_{\underline{y}}^2) + 0{,}16\,\sigma_{\underline{y}}^2}}{\mu_{\underline{y}}^2 - \xi_b^2 \sigma_{\underline{y}}^2} ,$$

waarin $\mu_{\underline{y}}$ de voor $\underline{y}$ gekozen instelwaarde is.

Hierin zijn ξ_g, ξ_a en ξ_b de drie waarden van een gestandaardiseerde normale verdeling met overschrijdingskans α_g, α_a en α_b en de kans, dat een willekeurig pakje aan alle drie in de vraagstelling vermelde eisen voldoet, is minstens gelijk aan

$$1 - (\alpha_g + \alpha_a + \alpha_b) .$$