Beoordeeling van het verschil tusschen twee variëteiten op grond van een waargenomen opbrengstverschil, door Prof. M. J. van Uven.

Wil men nagaan, of van een zeker gewas de variëteit A beter is dan de variëteit B, dan neemt men een reeks veldproeven, waarbij elk van beide variëteiten wordt uitgezaaid op een zeker aantal perceelen, die in grootte, bodemkwaliteit, enz. gelijkwaardig kunnen worden geacht.

Stel, dat de variëteit A over M perceelen is verdeeld. Van de opbrengsten daarvan moge het gemiddelde a, en de middelbare afwijking α bedragen, zoodat de middelbare afwijking van het gemiddelde a het bedrag $\varepsilon_a = \frac{\alpha}{\sqrt{M}}$ heeft. Evenzoo moge de variëteit B over N perceelen zijn verdeeld, waarvan de opbrengsten het gemiddelde b en de middelbare afwijking β hebben, zoodat van dit gemiddelde b de middelbare afwijking $\varepsilon_b = \frac{\beta}{\sqrt{N}}$ bedraagt.

Het verschil $V = a - b$ tusschen beide gemiddelden heeft dan de middelbare afwijking $\varepsilon_V = \sqrt{\varepsilon_a^2 + \varepsilon_b^2} = \sqrt{\frac{\alpha^2}{M} + \frac{\beta^2}{N}}$.

We onderstellen, dat $a > b$, dus $V > 0$. De vraag is nu met welke kans we mogen verwachten, dat de variëteit A ook bij een volgend onderzoek hooger opbrengst zal geven dan de variëteit B, m. a. w. dat het verschil $V = a - b$ opnieuw positief zal zijn.

We hebben dus naast elkaar te beschouwen de werkelijk *geconstateerde* waarden (M, a, α), (N, b, β), met de daaruit berekende waarden $V = a - b$, $\varepsilon_V = \sqrt{\frac{\alpha^2}{M} + \frac{\beta^2}{N}}$ eenerzijds, en de bij een tweede proef mogelijk te verkrijgen waarden (M', a', α'), (N', b', β'), $V' = a' - b'$ anderzijds.

Hoe sterk V' vermoedelijk zal afwijken van V wordt beoordeeld uit ε_V, en wel aldus:

Bij gebrek aan verdere inlichtingen achten we V de waarschijnlijkste waarde van V'. Volgens de exponentieele foutenwet van GAUSS is de kans, dat V' tusschen V'_1 en V'_2 in gelegen is:

$$W_{V'_2}^{V'_1} = \frac{h}{\sqrt{\pi}} \int_{V'_1}^{V'_2} e^{-h^2 (V'-V)^2} \, dV', \quad \text{waarbij } h = \frac{1}{\varepsilon_V \sqrt{2}} > 0.$$

De kans, dat de variëteit A opnieuw beter zal blijken dan de variëteit B, is de kans, dat $a' > b'$, of dat $V' = a' - b'$ positief is, d. w. z. in ligt tusschen 0 en $+\infty$. De kans op: A beter dan B $(A > B)$ is dus:

$$W(A > B) = W_0^{+\infty} = \frac{h}{\sqrt{\pi}} \int_{V'=0}^{V'=+\infty} e^{-h^2 (V'-V)^2} \, dV', \text{ met } h = \frac{1}{\varepsilon_V \sqrt{2}} > 0.$$

Stellen we $V' = V + u$, of $u = V' - V$, dan worden de grenzen van u : $u_1 = 0 - V = -V$ en $u_2 = +\infty$, terwijl $dV' = du$; derhalve:

$$W(A > B) = \frac{h}{\sqrt{\pi}} \int_{-V}^{+\infty} e^{-h^2 u^2} \, du.$$

Om deze integraal met behulp van een tabel te kunnen berekenen, stellen we eerst nog

$$u = \frac{-t}{h} \text{ of } t = -hu;$$

de grenzen van t zijn dan $t_1 = -hu_1 = +hV$, $t_2 = -hu_2 = -h \times +\infty = -\infty$, terwijl $du = -\frac{dt}{h}$.

Er komt dus:

$$W(A > B) = \frac{-1}{\sqrt{\pi}} \int_{+hV}^{-\infty} e^{-t^2} dt = \frac{+1}{\sqrt{\pi}} \int_{-\infty}^{+hV} e^{-t^2} dt = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{T} e^{-t^2} dt,$$

wanneer $T = hV = \frac{V}{\varepsilon_V \sqrt{2}}$ wordt gesteld.

De integraal $\frac{1}{\sqrt{\pi}} \int_{-\infty}^{T} e^{-t^2} dt$ wordt, als funcfie van haar bovenste grens T, aangeduid door $\Theta(T)$. We hebben dus

$$W(A > B) = \Theta(T). \text{ waarbij } T = \frac{V}{\varepsilon_V \sqrt{2}}.$$

Uitgaande van de gegeven waarden (M, a, α), (N, b, β), berekent men eerst V en ε_V, en vervolgens T, waarna de tabel van de functie Θ de waarde voor $W(A > B)$ levert.

Wil men, om de een of andere reden, dat A niet alleen een grooter opbrengst per perceel levert dan B, maar zelfs minstens een bedrag V_m meer levert, dus dat $V' = a' - b' \geq V_m$, dan verlangt men blijkbaar, dat V' tusschen V_m en $+\infty$ in ligt. Stellen we de kans hierop voor door $W(V_m)$, dan vinden we:

$$W(V_m) = W_{V_m}^{+\infty} = \frac{h}{\sqrt{\pi}} \int_{V'=V_m}^{V'=+\infty} e^{-h^2(V'-V)^2}\, dV', \text{ met } h = \frac{1}{\varepsilon_V\sqrt{2}}.$$

De substitutie $V' = V + u$ geeft dan als onderste grens voor u:

$$u_m = V_m - V = -(V - V_m),$$

terwijl de bovenste grens $+\infty$ blijft.

De substitutie $u = -\frac{t}{h}$ geeft dan als onderste grens voor t:

$$T_m = +h\,(V - V_m),$$

zoodat

$$W(V_m) = \frac{-1}{\sqrt{\pi}} \int_{T_m}^{-\infty} e^{-t^2}\, dt = \frac{+1}{\sqrt{\pi}} \int_{-\infty}^{T_m} e^{-t^2}\, dt = \Theta(T_m), \text{ waarbij}$$

$$T_m = \frac{V - V_m}{\varepsilon_V \sqrt{2}}.$$

Hieronder volge een tabel van de functie Θ *(T)*.

T	Θ	T	Θ	T	Θ	T	Θ
−2,5	0,0002	−1,2	0,045	0	0,500	+1,3	0,967
−2,4	0,0003	−1,1	0,060	+0,1	0,556	+1,4	0,976
−2,3	0,0006	−1,0	0,079	+0,2	0,611	+1,5	0,983
−2,2	0,0009	−0,9	0,101	+0,3	0,664	+1,6	0,9882
−2,1	0,0015	−0,8	0,129	+0,4	0,714	+1,7	0,9919
−2,0	0,0023	−0,7	0,161	+0,5	0,760	+1,8	0,9945
−1,9	0,0036	−0,6	0,198	+0,6	0,802	+1,9	0,9964
−1,8	0,0055	−0,5	0,240	+0,7	0,839	+2,0	0,9977
−1,7	0,0081	−0,4	0,286	+0,8	0,871	+2,1	0,9985
−1,6	0,0118	−0,3	0,336	+0,9	0,899	+2,2	0,9991
−1,5	0,017	−0,2	0,389	+1,0	0,921	+2,3	0,9994
−1,4	0,024	−0,1	0,444	+1,1	0,940	+2,4	0,9997
−1,3	0,033	0	0,500	+1,2	0,955	+2,5	0,9998

Voorbeeld: $M = 20 \quad a = 10{,}4 \quad \alpha = 0{,}3$
$N = 25 \quad b = 10{,}1 \quad \beta = 0{,}6$

$$V = 10{,}4 - 10{,}1 = +0{,}3;\ \varepsilon_V = \sqrt{\frac{0{,}3^2}{20} + \frac{0{,}6^2}{25}} =$$

$$= \sqrt{0{,}0045 + 0{,}0144} = \sqrt{0{,}0189} = 0{,}14.$$

$$T = \frac{0{,}3}{0{,}14\sqrt{2}} = 1{,}5;\quad \Theta\,(T) = \Theta\,(1{,}5) = 0{,}983.$$

Men heeft dus een kans 0,983, of 58 tegen 1, dat A beter is dan B.

Verlangt men, dat de opbrengst per perceel van A minstens 0.2 meer is dan van B, dan is $V_m = 0{,}2$, dus $V - V_m = 0{,}3 -$

$$- 0{,}2 = 0{,}1;\ T_m = \frac{0{,}1}{0{,}14\sqrt{2}} = 0{,}5;\quad \Theta\,(T_m) = \Theta\,(0{,}5) = 0{,}760.$$

Men heeft dus ongeveer een kans 3 tegen 1, dat A minstens 0,2 per perceel meer opbrengt dan B.

$V_m = 0{,}5$ geeft evenzoo $V - V_m = 0{,}3 - 0{,}5 = -0{,}2$, $T_m =$

$$= \frac{-0{,}2}{0{,}14\sqrt{2}} = -1{,}0;\quad \Theta\,(T_m) = \Theta\,(-1.0) = 0{,}079.$$

Er is dus slechts een kans 0,097 tegen 1—0,097 = 0,921, of ongeveer 3 tegen 35, of (nog ruwer) 1 tegen 12, dat A minstens 0,5 per perceel meer opbrengt dan B.

Had men een waarde voor V gekregen, juist gelijk aan 3 maal de middelbare afwijking, dus $V = 3\,\varepsilon_V$, dan zou men gevonden hebben $T = \frac{3}{\sqrt{2}} = 2{,}12$, dus $\Theta\,(T) = \Theta\,(2{,}12) =$ $= 0{,}999$ (nauwkeuriger 0.9986).

Als men dus $V = 3\,\varepsilon_V$ heeft waargenomen, dan is er een kans 9986 tegen 14 of 713 tegen 1, dat A beter is dan B. Deze kans is zoo groot, dat men zich in de meeste gevallen wel veilig zal gevoelen.

Is V nog grooter dan $3\,\varepsilon_V$, dan is die kans eveneens grooter.

Acht men een kans 100 tegen 1 voldoende gelijkwaardig met zekerheid, dan verlangt men $W\,(A > B) = 0{,}9901$, of

$$T = +1.65, \text{ dus } \frac{V}{\varepsilon_V} = +1{,}65\sqrt{2} = 2{,}33 \text{ of } V = 2{,}33\,\varepsilon_V.$$

In den regel eischt men, als kenmerk voor de wezenlijkheid van het verschil tusschen de variëteiten, dat het geconstateerde verschil V minstens gelijk is aan $3\,\varepsilon_V$ en wel op grond van de volgende redeneering:

Een *toevallig* verschil v, waarvan de volstrekte waarde hoogstens v_m bedraagt, heeft een kans

$$W_{-v_m}^{+v_m} = \frac{h}{\sqrt{\pi}} \int_{-v_m}^{+v_m} e^{-h^2 v^2}\, dv, \quad \text{waarbij } h = \frac{1}{\varepsilon_V \sqrt{2}},$$

V = geconstateerd verschil.

Nu is

$$W_{-v_m}^{+v_m} = 1 - \frac{h}{\sqrt{\pi}} \int_{-\infty}^{-v_m} e^{-h^2 v^2}\, dv - \frac{h}{\sqrt{\pi}} \int_{+v_m}^{+\infty} e^{-h^2 v^2}\, dv =$$

$$= 1 - \frac{2h}{\sqrt{\pi}} \int_{-\infty}^{-v_m} e^{-h^2 v^2}\, dv = 1 - 2\,\Theta(-hv_m) = 1 - 2\,\Theta(-t_m),$$

als we

$$t_m = hv_m = \frac{v_m}{\varepsilon_V \sqrt{2}}$$

stellen.

Hoe grooter v_m wordt gekozen, hoe kleiner $\Theta(-h\,v_m)$ is, hoe dichter dus $W_{-v_m}^{+v_m}$ bij 1 ligt. Constateert men derhalve, dat de volstrekte waarde van V grooter is dan v_m, dan is de kans, dat deze V *toevallig* is, zeer klein, en wel W (V toev.) = $= 1 - W_{-v_m}^{+v_m} = 2\,\Theta(-h\,v_m)$. De kans, dat V *niet toevallig*, dus *wezenlijk* is, is dan

W (V wez.) $= 1 - W$ (V toev.) $= 1 - 2\,\Theta(-h\,v_m) = W_{-v_m}^{+v_m}$.

Kiest men $v_m = 3\,\varepsilon_V$, dan komt er volgens deze redeneering

$$W\,(V \text{ wez.}) = 1 - 2\,\Theta(-3\,h\,\varepsilon_V) = 1 - 2\,\Theta\left(-\frac{3}{\sqrt{2}}\right) =$$

$= 1 - 2\,\Theta(-2{,}12) = 1 - 2 \times 0{,}0014 = 0{,}9972$, of 356 tegen 1.

Deze kans is kleiner dan die, welke overeenkomt met $V = 3\,\varepsilon_V$ volgens dé eerste methode.

De fout van de laatste methode bestaat daarin, dat men werkt met de *volstrekte* waarde van V, terwijl het op de *algebraïsche* waarde aankomt.

$W_{-v_m}^{+v_m}$ is de kans, dat de volstrekte waarde van v kleiner is dan v_m, dus dat òf $v > + v_m$, òf $v < - v_m$. Is er nu een *positief* bedrag V gevonden, dan interesseert ons alleen het geval $v > + v_m$. Alleen de wetenschap, dat de kans op een toevallig verschil $v > + v_m$ klein is, geeft ons de overtuiging dat een gevonden V grooter dan $+ v_m$ niet op toeval berust. Dat de kans op een *toevallig negatief* verschil $v < - v_m$ klein zou zijn, draagt niet bij tot een oordeel over een *werkelijk geconstateerd positief* verschil.

Constateeren we dus, dat $V \geqq 3 \varepsilon_V$, dan is de kans op de realiteit van dit verschil *niet*: minstens $1 - 2\,\Theta\,(-2,12) = 1 - 2 \times 0{,}0014$ (of minstens 356 tegen 1), maar *wel* $1 - \Theta\,(-2,12) = \Theta\,(+2,12) = 1 - 0{,}0014$ (of minstens 713 tegen 1).

In 't algemeen is het geven van een universeel recept. als $V \geqq 3\,\varepsilon_V$, verwerpelijk, afgezien daarvan, dat het meestal op een onjuisten grondslag rust. Waar het op aankomt is het beoordeelen van de *kans*, dat — na een eenmaal geconstateerd verschil — een eventueel later waar te nemen verschil in dezelfde richting zal uitvallen, of een zeker minimum zal te boven gaan, Voor deze kans hebben we gevonden:

$W\,(A > B) = \Theta\,(T)$ met $T = \dfrac{V}{\varepsilon_V \sqrt{2}}$, resp. $W\,(V_m) = \Theta\,(T_m)$ met

$T_m = \dfrac{V - V_m}{\varepsilon_V \sqrt{2}}$.

Het beoordeelen van deze kans is een zaak van persoonlijke waardeering. Deze taxatie moet worden overgelaten aan hem, wien het opbrengstverschil ter harte gaat. Elk bijzonder geval stelt daarbij aan de te verlangen kans zijn eigen eischen.