R 664.6 GR 15
C
232
m
15

BIBLIOTHEEK
CENTRALE ORGANISATIE
T. N. O.
's-GRAVENHAGE

Overgedrukt uit „Bakkerijwetenschap", 3e jaargang, 1950

Mededeling no 15 van de Afd. Graan-, Meel- en Broodonderzoek van het Centraal Instituut voor Voedingsonderzoek T.N.O.

PROEF-SCHEMA'S VOOR ORGANOLEPTISCH ONDERZOEK

(WITH SUMMARY)

IR H. DE MIRANDA

Centr. Inst. v. Voedingsonderzoek,
Afd. Graan-, Meel- en Broodonderzoek, Wageningen

I

Bij de beoordeling van bakkerij-producten wordt een ruim gebruik gemaakt van zgn. „organoleptische" waarneming, d.w.z. directe waarneming door de zintuigen van het uiterlijk, de smaak, het aroma, e.d. Tegenwoordig ziet men veelal een streven om de elementen van willekeur hierbij zoveel mogelijk uit te schakelen.

Als de beoordeling het routinewerk is van bepaalde personen, kan met voordeel gebruik worden gemaakt van waarderings-cijfers volgens een van te voren vastgestelde schaal. Elke waarneming afzonderlijk wordt dan in een cijfer uitgedrukt, zo mogelijk onafhankelijk van de daaraan voorafgaande en daarop volgende waarnemingen (1). De waarneming krijgt dan meer het karakter van een meting.

Hier willen wij echter slechts aandacht schenken aan de gevallen, waarin een dergelijke waarderings-schaal niet kan worden toegepast. De waarneming is dan een vergelijking, gevolgd door rangschikking.

Wij zullen hier spreken van een aantal *objecten* 0_1, 0_2, ... enz., die volgens een bepaald gezichtspunt worden beoordeeld door een aantal *personen* P_1, P_2, ... enz. Stel n is het aantal objecten, en m het aantal personen.

Een eenvoudig proef-schema is dan het volgende: De proefpersonen geven onafhankelijk van elkaar, elk een rangschikking aan. De objecten krijgen dan *rangnummers*, die voor elke persoon verschillend kunnen zijn. Voorbeeld:

	0_1	0_2	0_3	0_4	0_5	Totaal	Gemidd.
P_1	2	3	1	5	4	15	3
P_2	3	1	2	4	5	15	3
P_3	2	3	1	4	5	15	3
P_4	4	2	1	3	5	15	3
Totaal	11	9	5	16	19	60	12
Gemidd.	2,75	2,25	1,25	4,00	4,75		

n = 5 m = 4

Doordat elke proefpersoon dezelfde rangnummers toekent (zij het aan verschillende objecten), zijn de totalen van de horizontale rijen steeds gelijk.

De totalen van de vertikale kolommen zullen echter des te meer uiteenlopen, naarmate de personen in hun oordeel overeenstemmen. Dit uiteenlopen van de totaalcijfers kan als volgt in een getal worden uitgedrukt:

BIBLIOTHEEK
CENTRALE ORGANISATIE
T. N. O.
's-GRAVENHAGE

TNO
16765

S is de spreiding van de totaalcijfers, uitgedrukt als de som van de kwadraten van de afwijkingen van het gemiddelde. Wij vinden in ons voorbeeld:
$S = (11-12)^2 + (9-12)^2 + (5-12)^2 + (16-12)^2 + (19-12)^2 = 124.$

S kan variëren van nul tot een maximale waarde, gelijk aan $\frac{1}{12} m^2 (n^3 - n)$.

Een „coëfficiënt van overeenstemming" („coëfficient of concordance") wordt nu als volgt gedefinieerd:

$$W = \frac{S}{\frac{1}{12} m^2 (n^3-n)}$$

W kan variëren van 0 tot 1.

In het voorbeeld vinden wij: $W = \frac{124}{160} = 0{,}774.$

afgerond: $W = 0{,}77$

Verder kan nog een grootheid F worden berekend, aldus:

$$F = \frac{(m-1)\ W}{1-W}$$

Hiermede zijn wij op het terrein gekomen van de normale variatie-analyse („analysis of variance"). Een grote waarde voor F wijst op de aanwezigheid van duidelijke verschillen tussen de objecten. Bij de beoordeling van een gevonden waarde voor F kan de bekende F-tabel (2) worden gebruikt, waarbij dan het aantal vrijheidsgraden als volgt moet worden gekozen:

$$v_1 = (n-1) - \frac{2}{m}$$
$$v_2 = (m-1) \,.\, v_1$$

Alleen als m en n zeer klein zijn, geeft deze methode geen goede uitkomsten en moet een speciale tabel worden gebruikt (3).
Wij vinden in ons voorbeeld:

$$F = 10,\ v_1 = 3\tfrac{1}{2},\ v_2 = 10\tfrac{1}{2}$$

Uit de F-tabel blijkt, dat deze waarde voor F wijst op duidelijke verschillen tussen de objecten.
Er zijn dus duidelijke verschillen, maar dat wil nog niet zeggen, dat al de gevonden verschillen duidelijk zijn. Hoe is het bijv. met het verschil tussen 0_1 en 0_2? Om een dergelijke vraag te beantwoorden is het nuttig de middelbare fout van de gemiddelde rangcijfers te kennen. Deze kan als volgt worden berekend:

$$\text{middelb. fout} = \sqrt{\frac{1}{F} \cdot \frac{S}{m^2 n}}$$

Wij vinden in ons voorbeeld:

$$\text{middelb. fout} = \sqrt{\frac{1}{10} \cdot \frac{124}{80}} = 0{,}4$$

De gemiddelde rangcijfers van 0_1 en 0_2, nl. 2,75 en 2,25, zijn dus niet duidelijk verschillend.
De resultaten kunnen het best kort worden samengevat door opgave van

de gemiddelde rangcijfers, de waarden voor W en F, het resultaat van de F-test, en – als duidelijke verschillen aanwezig blijken te zijn – door de opgave van de middelbare fout.

Soms worden twee objecten gelijk beoordeeld. Zij krijgen dan gelijke rangnummers, waarbij men echter moet zorgen, dat de som van de rangcijfers niet verandert. Zou in ons voorbeeld de persoon P_3, geen verschil hebben gevonden tussen 0_1 en 0_2, dan had dit geleid tot de volgende nummering: $2\frac{1}{2}$, $2\frac{1}{2}$, 1, 4, 5, totaal 15, en dus niet: 2, 2, 1, 3, 4, totaal 12. De Engelse term hiervoor is „tie", i.c. van omvang 2 ($t = 2$). In de reeks 1, 2, 4, 4, 4, 6 komt een „tie" voor van omvang 3 ($t = 3$), enz.

Een gering aantal „ties" heeft op de berekening practisch geen invloed. Komen vele „ties" voor, dan berekent men voor elke „tie" een grootheid T als volgt: $T = \frac{1}{12}(t^3-t)$. De som van alle T's wordt dan met m vermenigvuldigd en de noemer in de formule voor W met het aldus gevonden getal verminderd. De berekening verloopt dan verder, zoals werd aangegeven. Alleen als het aantal „ties" zeer groot is met grote waarden voor t, is deze correctie niet meer voldoende.

Het hier behandelde schema zal in sommige gevallen minder goed voldoen. Zo is het bij de beoordeling op smaak van producten met weinig uitgesproken smaak, zoals brood, zeer moeilijk een reeks van objecten te rangschikken. Als men het derde object proeft, is men de smaak van het eerste object alweer vergeten! Beter voldoet hier een systeem, waarbij telkens slechts twee objecten worden vergeleken. Bij een dergelijk systeem van „paarsgewijze vergelijking" is het tevens mogelijk na te gaan, of bij een eventueel gebrek aan overeenstemming moet worden opgemerkt: „over smaken valt niet te twisten", of dat de smaakverschillen zo gering waren, dat de proevers geen reproduceerbaar oordeel konden geven. Wij stellen ons voor in een volgend artikel het systeem van „paarsgewijze vergelijking" te behandelen.

II

Bij de beoordeling op kleur, uiterlijk, e.d. is het mogelijk een aantal objecten te overzien en elk object direct te vergelijken met alle andere. Het is dan mogelijk een aantal objecten naar volgorde van voorkeur te rangschikken en als een aantal personen hun oordeel op deze wijze hebben weergegeven, kunnen de resultaten worden bewerkt zoals beschreven in het vorige artikel.

Bij de beoordeling op smaak is het veel moeilijker om een aantal objecten te rangschikken, omdat het „overzicht" ontbreekt. Vooral als de smaakverschillen weinig uitgesproken zijn, kan een goed oordeel slechts worden gegeven bij vergelijking van een zeer klein aantal, liefst twee, objecten. Door echter een aantal objecten in allerlei combinaties in paren te vergelijken, kan men toch tot conclusies komen betreffende de rangschikking.

De situatie is enigszins te vergelijken met die van een schaakcompetitie. De methode van de „paarsgewijze vergelijking" is een nadere uitwerking van het principe van een dergelijke competitie.

Een practisch schema voor de bewerking van de resultaten zal aan de hand van een voorbeeld worden beschreven. Het aantal objecten zal weer n worden genoemd en het aantal proefpersonen m.

Stel 4 objecten (nr 1–4) worden beoordeeld door 5 personen (A–E), dus n = 4, m = 5. Het aantal mogelijke paren van objecten is dan $\frac{1}{2}$ n (n–1) = 6.

Men rangschikt de objecten, bijv. monsters brood, volgens deze paren, en plaatst de paren in een willekeurige, aan de proefpersonen onbekende volgorde. Men noteert deze volgorde in een schema, zoals aangegeven in tabel I. De met streepjes verbonden getallen zijn de nummers van de tot paren verenigde objecten. Men laat nu door elke proefpersoon achtereenvolgens de zes paren beoordelen. Telkens wordt 1 punt gegeven aan het best beoordeelde object, of in geval van gelijke beoordeling aan beide objecten $\frac{1}{2}$ punt. Men noteert deze punten in het schema, zoals aangegeven in tabel I. Men leest hier bijv. uit af, dat bij de combinatie 4–2 alleen de persoon E aan object no 4 de voorkeur gaf enz. Bij de combinatie 1–3 kon persoon B geen keuze doen.

Tabel II geeft de rangschikking van de punten volgens personen en objecten. Achtereenvolgens gaat men de 12 kolommen van tabel I na. Men vindt dan in de eerste kolom boven het getal 4 naast E een 1, en plaatst dus een 1 in tabel II, rij E, kolom 4. Vervolgens komt een 1 op de plaatsen A2, B2, D2, E2 enz.

Tabel III verschilt slechts van tabel II door een eenvoudiger notatie. In de rijen van deze tabel vindt men telkens voor een bepaalde persoon hoeveel keer hij aan elk object de voorkeur gaf. Achter D vindt men de cijfers 0, 1, 2, 3, d.w.z. persoon D heeft volledig een rangschikking kunnen aangeven. Bij A vindt men echter het cijfer 1 voor de objecten 1, 3 en 4. Uit tabel I blijkt, dat volgens A object 1 beter zou zijn dan 3, 3 beter dan 4, maar toch 4 beter dan 1! M.a.w. persoon A is bij de beoordeling niet consequent geweest. Dergelijke inconsequenties hebben tot gevolg, dat de *spreiding* van de cijfers in een rij kleiner wordt. Hoe beter de proefpersoon zich een oordeel heeft kunnen vormen, des te groter zal, in het algemeen, de spreiding zijn van de door hem toegekende aantallen punten.

In elke kolom van tabel III zijn verenigd de bij een bepaald object behorende cijfers. Onder elke kolom vindt men de som en het gemiddelde. Deze gemiddelden zijn te vergelijken met de in het vorige artikel genoemde gemiddelde rangcijfers. Zijn de verschillen tussen de gemiddelden duidelijk? Om deze vraag te kunnen beantwoorden moet de spreiding van de kolom-gemiddelden worden vergeleken met de spreiding van de cijfers binnen de kolommen.

TABEL I

A		1	1			
B		1	$\frac{1}{2}$		1	
C				1	1	
D		1	1		1	
E	1	1	1		1	
	4–2	2–1	1–3	4–3	1–4	3–2
A	1			1	1	1
B	1		$\frac{1}{2}$	1		1
C	1	1	1			1
D	1			1		1
E				1		1

TABEL II

	1	2	3	4
A	1	111	1	1
B	1$\frac{1}{2}$	111	1$\frac{1}{2}$	
C	11	11	1	1
D	11	111	1	
E	11	11	1	1

TABEL III

	1	2	3	4	
A	1	3	1	1	
B	1$\frac{1}{2}$	3	1$\frac{1}{2}$	0	
C	2	2	1	1	
D	2	3	1	0	
E	2	2	1	1	
Totaal, SX	8,5	13	5,5	3	
Gemidd. SX/m . . .	1,7	2,6	1,1	0,6	
$(SX)^2/m$ . .	14,45	33,80	6,05	1,80	56,10

$\frac{1}{4}$ mn $(n-1)^2$ . . . = 45

Obj. som d. kwadr. = 11,10

n = 4, m = 5

Het is dus van belang de spreiding te kennen zowel binnen de rijen als binnen de kolommen. Wij maken hiertoe eerst een tabel van de kwadraten, tabel IV, en gebruiken nu de volgende notatie:

X = de cijfers van tabel III (in tabel IV dus de waarden voor X^2).
S = som-teken, bijv.:
SX = de som der cijfers X voor een rij of kolom.
$\bar{x}$ = gemiddelde van de cijfers X voor een rij of kolom.
Voor de kolom-gemiddelden is $\bar{x} = SX/m$.

Een maat voor de spreiding is $S(X-\bar{x})^2$. Volgens een bekende statistische formule is $S(X-\bar{x})^2 = SX^2 - (SX)^2/a$. De factor $(SX)^2/a$ heet „correctieterm". Voor de rijen is a = n, voor de kolommen is a = m.

Door optelling van de getallen van tabel IV, horizontaal of verticaal, vinden wij SX^2. De correctieterm voor de kolommen is $(SX)^2/m$, waarbij voor SX de onder tabel III genoemde waarden worden gebruikt. Voor de rijen is de correctieterm steeds $\frac{1}{4}n(n-1)^2$. Hier is $\frac{1}{4}n(n-1)^2 = 9$.

Wij vinden aldus de spreiding binnen de kolommen, $S(X-\bar{x})^2$, met als totaalcijfer 3,40. Dit getal houdt verband met de spreiding van de afzonderlijke (individuele) cijfers rond de kolom-gemiddelden, d.w.z. met de „waarnemingsfout" bij de bepaling van de gemiddelde cijfers voor de objecten.

Wij moeten dit vergelijken met de spreiding van de kolom-gemiddelden $\bar{x}$ onderling. Het blijkt, dat het betreffende getal wordt verkregen door de som van de zo juist als correctie-termen gebruikte factoren $(SX)^2/m$ te verminderen met een nieuwe correctieterm, gelijk aan $\frac{1}{4}mn(n-1)^2$. In tabel III is de berekening opgenomen, waarbij wordt verkregen het getal 11,10. Dit getal houdt dus verband met de spreiding van de kolom-gemiddelden, d.w.z. met de verschillen tussen de gemiddelde cijfers voor de objecten.

Wij maken nu een voor variatie-analysen gebruikelijke samenvatting. Zie tabel V. De „som der kwadraten" wordt gedeeld door het aantal vrijheidsgraden. Men vindt dan de „gemiddelde kwadraten", met als verhouding het getal F. (Betreffende de F-test zie het vorige artikel.) De standaardafwijking is de wortel uit het „gemiddelde kwadraat" voor de ind. cijfers. Deling door $\sqrt{m}$ levert tenslotte de middelbare fout van de gemiddelde cijfers voor de objecten.

TABEL IV

	1	2	3	4	Tot.	Corr.	T	z
A	1	9	1	1	12	9	3	0,5
B	$2\frac{1}{4}$	9	$2\frac{1}{4}$	0	$13\frac{1}{2}$	9	$4\frac{1}{2}$	0,88
C	4	4	1	1	10	9	1	0
D	4	9	1	0	14	9	5	1,0
E	4	4	1	1	10	9	1	0
Totaal, SX^2	15,25	35,00	6,25	3,00				
Uit tab. III: $(SX)^2/m$ . . .	14,45	33,80	6,05	1,80				
$S(X-\bar{x})^2$ . . .	0,80	1,20	0,20	1,20	3,40 (= ind., som d. kwadr.)			

TABEL V

	Vrijheidsgr.	Som der kwadr.	Gem. kwadr.	Stand afw.	Midd. fout van obj. gem.
Objecten	(n–1) = 3	11,10	3,7		
Ind. cijfers	(m–1) (n–1) = 12	3,40	0,28	0,53	0,24

F = 3,7 : 0,28 = 13,2

Nu moeten wij nog aandacht besteden aan de spreiding binnen de rijen. De getallen, die worden gevonden na aftrekking van de correctieterm, i.c. 9 (zie tabel IV), worden T genoemd. T is maximaal (T max.) als in de betreffende rij in tabel III de hele getallen van 0 tot n–1 voorkomen (bijv. bij D). Wij berekenen nu getallen z volgens een van onderstaande formules.

Als n een oneven getal is: $z = \frac{T}{T\ max}$ Als n een even getal is: $z = \frac{T - \frac{1}{2}n}{T\ max - \frac{1}{2}n}$

In het voorbeeld gebruiken wij de tweede formule. De waarden voor z (d.i de „coëfficiënt of consistence") kunnen variëren van 0 tot 1, en zijn een maat voor het onderscheidingsvermogen van de proefpersonen. Op één proef mag men echter niet af gaan. De z-waarden voor verschillende proefseries zijn echter direct vergelijkbaar. Op den duur kan men dus nagaan welke personen over het algemeen hoge of lage z-waarden verkrijgen.

Tabel VI geeft een samenvatting van de resultaten.

TABEL VI. SAMENVATTING VAN DE RESULTATEN

Objecten no	2	1		3	4
Gemidd. aant. punten	2,6	1,7		1,1	0,6
Middelbare fout			0,24		
F-test: F = 13,2. Er zijn duidelijke verschillen					
Proefpersonen	A	B	C	D	E
Waarde voor z	0,5	0,9	0	1,0	0

Opgemerkt moet worden dat bij kleine waarden voor m en n de F-test slechts bij benadering mag worden toegepast. Bij de beoordeling geven F en de middelbare fout waardevolle aanwijzingen, mits zij met de nodige voorzichtigheid en niet automatisch als feilloze, statistische test worden toegepast. Verder wordt bij dit proef-schema aangenomen, dat de objecten niet op andere wijze dan door de smaak direct zijn te herkennen, of althans dat de proef-personen zich niet door iets anders laten beinvloeden.

Dit schema kan op verschillende manieren worden gewijzigd. Als het aantal objecten groot is, is het wenselijk niet alle combinaties op dezelfde dag ter beoordeling te geven. Ook kan men het principe loslaten, dat elke persoon alle combinaties moet beoordelen. Men kan hierbij vele personen inschakelen, waarbij dan de berekening van z vervalt. Steeds zal men moeten trachten het schema aan de omstandigheden aan te passen.

LITERATUUR

O.a. KENDALL, M. A., Rank Correlation Methods. – London, 1948; Ch. 6: The problem of m rankings; Ch. 11: Paired comparisons.

1. Bakkerijwetenschap, *2* (1949), nr 8, 116.
2. SNEDECOR, G. W., Statistical Methods. – 1946, blz. 222.
3. KENDALL, M. A., Rank Corr. Meth., loc. cit. 146 e.v.

SUMMARY

EXPERIMENTAL SCHEMES FOR THE ORGANOLEPTIC EVALUATION OF BAKERY PRODUCTS

Two schemes are described for the organoleptic evaluation of bakery- or other food products. In the first scheme several persons are asked to grade a set of objects, and the gradings are combined. In the second scheme each person is asked to compare the objects in all the possible combinations. This method of „paired comparisons" is suitable to evaluate the taste, especially when the taste differences are not very pronounced, e.g. in the case of bread.

DRUK: H. VEENMAN EN ZONEN, WAGENINGEN