



METODER FOR OBJEKTIV BEDØMMELSE



Foto: Mari Svenningsen





METODER FOR OBJEKTIV BEDØMMELSE

Marit Rødbotten og Mats Carlehög, Nofima

Det er etter hvert blitt utviklet mange metoder som er mer eller mindre standardiserte og som gir svar på bestemte problemstillinger vedrørende produkters sensoriske egenskaper. Som for kjemiske og mikrobiologiske analyser brukes det forskjellige metoder også for sensorisk analyse, alt etter hva som skal undersøkes. En kjemiker velger bestemte kjemiske analysemetoder for å undersøke surhetsgrad i et produkt, og en sensoriker velger en bestemt sensorisk metode for å analysere om det er en sensorisk målbar forskjell i surhet mellom to produkter.

I dette kapitlet skal vi gjennomgå de metodene som oftest blir brukt for analyse av objektive egenskaper ved et produkt.

Det finnes forskjellige måter å klassifisere sensoriske metoder på, som for eksempel etter bruksområde, paneltyper, metodetype. Vi kan her klassifisere dem i følgende tre hovedgrupper:

- forskjellstester
- kvantitative tester
- forbrukertester

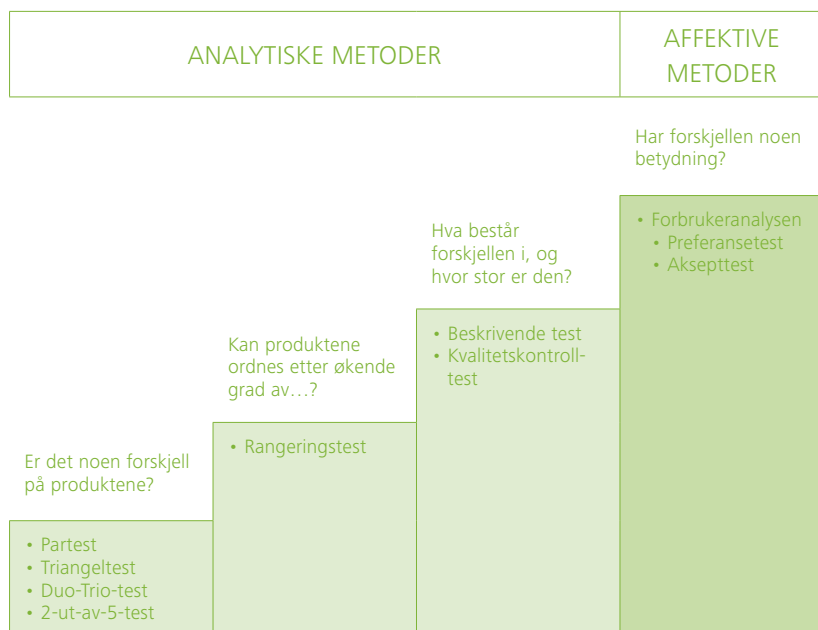
Forskjellstester og kvantitative metoder kalles også laboratoriemetoder eller analytiske metoder, og er objektive metoder som utføres av et trent panel (se kapittel 3.2.1). Metoder for forbrukeranalyser kalles affektive (= følelsesmessig) eller hedoniske (= lykkefølelse) metoder og er subjektive analyser utført av forbrukere som skal gi informasjon om hvilken prøve de liker best eller på annen måte gi uttrykk for sine subjektive meninger (se kapittel 5).





Vi finner ofte at ordene «test» og «metode» brukes om hverandre. «Metode» blir brukt om å undersøke eller teste noe. Det betyr at når vi for eksempel snakker om en forskjellstest burde vi riktigere snakke om en forskjellsmetode eller kanskje en forskjellstestmetode. På engelsk heter det en «sensory difference test», og det er antakelig forklaringen på at den norske oversettelsen er blitt «sensorisk forskjellstest». Vi velger i denne boka å fortsette bruken av ordene test og metode slik det er blitt en vane innen vitenskapen sensorikk, vel vitende om den språklige unøyaktigheten.

Det er hva en ønsker å finne ut om produkters sensoriske kvaliteter som avgjør hvilken metode en skal velge. Metoden avgjør hva slags panel en skal bruke for å få riktig analyseresultat. Analytiske metoder krever som oftest et trent dommerpanel mens utrente dommere (forbrukere) gjennomfører analyser med affektive metoder, se figur 4.1.



Figur 4.1 Sensoriske problemstillinger løst med bestemte metoder





4.1. Skala

Når en eller flere av våre sanser får et inntrykk eller en stimulanse (for eksempel vi spiser et eple) vil det bli oppdaget som en følelse eller sansing (vi kjenner sprøhet, søt smak). Personen som spiser eplet må bruke en metode for å gi uttrykk for sin oppfatning av denne følelsen, en respons på følelsen, og ofte angi grad eller intensitet av den stimulus som er opprinnelsen til følelsen.

Stimulans — følelse — respons

Det er først når observasjoner kan relateres til en skala at vi kan snakke om måling. Data kan forekomme på fire forskjellige skalaer: nominal-, ordinal-, intervall- eller ratioskala.

- Den enkleste måten å anvende tall på er til gruppering av informasjon ved opptelling av antall tilfeller i bestemte kategorier. Kategoriene kan være for eksempel ja/nei, mann/kvinne eller eple/pære/banan. Selv om dette ikke representerer noen skala og derfor ikke er noen måling slik vi vanligvis forstår ordet, snakker vi likevel om måling med nominalskala (nomen=navn). Eksempel på sensorisk analyse på nominalnivå er forskjellstester.
- Måling forutsetter at objektene lar seg ordne etter en rekkefølge fra «minst» til «mest» med hensyn på en definert egenskap. Hvis vi antar at det dreier seg om en naturlig ordning mellom kategoriene, kalles det en ordinalskala. Slik måling kan for eksempel være en konsistensmåling av majones som «fast», «fastere», «fastest». Det dreier seg da om en rangordning eller en rangering. Dersom vi bruker tall for å betegne de forskjellige kategoriene (fast=1, fastere=2, fastest=3), er dette bare en betegnelse for rangordning. Tallene sier ikke noe om innbyrdes avstand mellom kategoriene, og beregning av gjennomsnittsverdier er uten mening.

Det er først når observasjoner kan relateres til en skala at vi kan snakke om måling.





- Målinger som er utført slik at tallverdiene formidler informasjon om avstanden mellom punktene på en skala med hensyn på en bestemt egenskap, kalles måling på intervallnivå eller måling med intervallskala. I tillegg til rangordning kan det her også defineres meningsfylte differanser. Differansen mellom 3 og 6 er den samme som mellom 5 og 8. Ved bruk av intervallskala er det meningsfylt å beregne gjennomsnittsverdier av målingene. Tester som beskrivende test og kvalitetskontrolltest forutsetter data på intervallnivå.
- Ved bruk av fysiske målemetoder får vi data på ratioskala (meter, liter, gram). Her har vi måleenheter med absolutt nullpunkt, og først på dette nivået kan vi snakke om forhold mellom størrelser, for eksempel at 20 gram er dobbelt så mye som 10 gram.

Innen forbrukeranalyser brukes måling av preferanse og aksept. Ved måling av preferanse angis valg mellom to eller flere prøver, men her får vi ikke svar på om prøvene likes eller mislikes, bare rekkefølgen i liking av prøvene. En akseptmåling viser om produktet blir oppfattet som godt eller ikke, og i hvor stor grad prøven likes eller mislikes. Forbrukere kan altså foretrekke (preferere) produkt A fremfor produkt B, men mislike begge produktene (lav aksept av begge to).

En forbruker kan også bli presentert forskjellige påstander relatert til produkter og bli bedt om å angi grad av enighet i påstandene. En slik skala blir kalt en Likert-skala og den kan deles inn som en tallskala (for eksempel 5-, 7-, 9-punkt) med endepunkter enig og ikke enig.

Kroppsspråk og ansiktsuttrykk avslører ofte om vi liker eller misliker produkter og påstander. Dette ligger til grunn for en skala som ofte brukes når barn er forbrukere i sensoriske tester. En skala som viser blide og sure ansikter er hensiktsmessig for barn som ikke kan lese verbale skjema (figur 4.2).

Poengbedømmelser

Til å bedømme den sensoriske kvaliteten eller deler av den, er poengbedømmelse (ofte kalt numerisk gradering eller scoring) mest brukt.





Figur 4.2 Skala med ansikter hensiktsmessig ved analyser hvor barn er forbrukere

Egenskapene ved varen som skal bedømmes blir karakterisert ved hjelp av en tallskala (poengskala). Dette er måling på intervallnivå.

Ved kvalitetskontrolltest er det vanlig at den ene enden av skalaen angir riktig kvalitet og den andre enden angir avvikende kvalitet. Det er med andre ord et karaktersystem. Metoden er illustrerende og anvendelig, men den krever at dommerne må ha en klart definert spesifika-sjon å arbeide etter.

Ved beskrivende test blir intensiteten av enkelte egenskaper målt ved at den ene enden (oftest venstre siden) av skalaen betegner «ingen» eller «lite» av den angitte egenskapen, mens den andre enden (høyre siden) av skalaen betegner «mye» eller «tydelig» intensitet av den angitte egenskapen.

De ord som velges i bedømmelse av et produkt må være beskrivende for produktet, entydige og klart definerte.

Unipolar eller dipolar skala?

På en intensitetsskala måles styrken av en gitt egenskap. En unipolar skala går fra «ingen» til «mye» av den enkelte egenskapen, for eksempel søthet. En dipolar skala får vi ved å sette sammen to unipolare skalaer for egenskaper som vi tror er de absolutte motpoler til hverandre, for eksempel tørr og saftig.

De ord som velges i bedømmelse av et produkt må være beskrivende for produktet, entydige og klart definerte. For svært mange egenskaper er det vanskelig å si med sikkerhet hvilke som er nøyaktige motsetninger til hverandre, og det er en viktig grunn for å bruke unipolare skalaer. Både de sensoriske dommerne og leserne av resultatene får en enklere



oppgave på denne måten. Dersom en prøve er for eksempel tydelig tørr, vil den ikke samtidig være saftig. Det kan derfor være bortkastet å bruke begge disse egenskapene ved beskrivelsen av et produkt, velg det ordet som er mest naturlig å bruke for produktet (for eksempel saftighet i kjøtt og tørrhet i kjeks).

Det er vanlig å merke endene av skalaen med ord som for eksempel «ingen» og «tydelig» eller «noe» og «mye». For mange egenskaper kan det være vanskelig å tenke seg en prøve som inneholder «ingen» intensitet, eksempelvis gjelder dette egenskaper som har med tekstur eller konsistens å gjøre. Det er enklere å tenke seg egenskaper som dreier seg om farge, lukt og smak med «ingen» intensitet i en prøve. Det blir en trenings- og definisjonssak å forholde seg til ordene på skalaen. Skalaen må i alle fall trenes inn i forhold til de prøvene som skal bedømmes.

Det kan knyttes beskrivende ord til alle punktene på skalaen. Problemet er da at det kan diskuteres om betydningen av ordene gir en skala som har lik avstand mellom punktene. Eksempel på en bipolar skala med ni punkter ved måling av mørhet eller seighet:

9. ekstremt mør
8. meget mør
7. moderat mør
6. litt mør
5. verken seig eller mør
4. litt seig
3. moderat seig
2. meget seig
1. ekstremt seig





Eksempel på en ikke-symmetrisk intensitetsskala som også er unipolar:

- 0 ingen søt smak
- 1 merkbar søt smak
- 2,5 svært svak søt smak
- 5 svak søt smak
- 7,5 svak/moderat søt smak
- 10 moderat søt smak
- 12,5 tydelig søt smak
- 15 svært tydelig søt smak

Skalaer av denne typen kan både forlenges og forkortes. I disse to eksemplene er det brukt adjektiver som er logiske ved beskrivelse av økende/minkende intensitet av egenskaper. Men er det like stor opplevd avstand mellom adjektivene?

Ved bruk av disse skalatypene kan data vanligvis bli analysert med parametriske statistiske teknikker (se kapittel 11) dersom vi kan anta at lik numerisk avstand tilsvarer lik sensorisk forskjell. Hvis vi ikke har like intervaller mellom punktene på skalaen, bør dataene behandles som data fra kategorien ordinalskala og dermed bli analysert med ikke-parametriske teknikker. Det er vanskelig å påstå at vi kan måle nøyaktige avstander av sensoriske stimuli. I praksis er likevel en variansanalyse såpass robust at vi får et resultat som vi kan stole på.

De fleste skalaer går fra null til et positivt tall. I enkelte tilfeller benyttes skalaer med nullpunkt på midten av skalaen og for eksempel fem positive og fem negative tallverdier. Skalaen kan være anvendelig når en har en referanseprøve (standardprøve) og skal måle om de andre prøvene har mer eller mindre intensitet av de angitte egenskapene. Skal en bedømme egenskapen «salt smak» i et produkt, kan en avmerke referanseprøven (R) ved null-punktet, og analysere om den andre prøven er mer eller mindre salt enn denne referansen (R).





Skalaen kan da se slik ut:

-5 -4 -3 -2 -1 0 1 2 3 4 5

ikke salt R mye salt

Det samme kan uttrykkes ved en skala fra 0 til for eksempel 11, hvor 0=ikke salt og 11=mye salt. Det kan da opplyses til dommerne om at referanseprøven har poeng 6. En fordel ved denne skalaen er at den ikke har negative tall som kompliserer behandlingen av tallmaterialet.

Ved en forbrukertest der aksept for det ene eller det andre produktet skal måles mot en referanse (for eksempel det markedsledende produktet), kan det være hensiktsmessig å velge en skalatype der referansen er plassert på midten av skalaen. Det blir da målt om en prøve er bedre eller dårligere enn referanseprøven (R), Just-about-right eller JAR-skala (kapittel 5).

Kontinuerlig eller diskret skala?

Når vi skal konstruere en skala for måling av en sensorisk egenskap, kan vi velge mellom en kontinuerlig skala som er en linje uten avmerkinger, eller en diskret skala som er en linje med oppmerkede delstreker.

En kontinuerlig skala (A):

En diskret skala (B):



Fordelen med den kontinuerlige skalaen (A) er at dommerne kan avmerke prøven hvor som helst på skalaen, de er ikke bundet av noen merker eller tall på skalalinjen.





Mange av data-innregistreringssystemene som finnes på markedet i dag, gir muligheten til å bruke kontinuerlig skala. Datamaskinen omregner markeringene til tallverdier for videre databehandling. Ved manuell lesing av bedømmelse utført på kontinuerlig skala, kan en benytte linjal med millimeteravlesning for korrekt registrering.

Når en benytter den diskrete skalaen (B), kan en risikere å miste informasjon om forskjeller mellom prøver som er svært like. Ofte vil dommerne ønske å gi karakterer mellom de vertikale linjene eller boksene, men det medfører selvsagt at panellederen burde utarbeide en annen skala med flere vertikale streker eller bokser på bedømmelseslinjen. Det betyr altså en forlengelse av skalaen. I slike tilfeller ser en da ofte at dommerne bare benytter deler av skalaen (den ene halvdel eller bare midten), slik at i realiteten ingenting blir vunnet ved å «utvide» skalaen.

Det blir brukt skalaer av forskjellig lengde, fra 3 til 100 skalaenheter. Målet er at skalaen skal være stor nok til å gi mulighet for å skille mellom små nyanser i prøvematerialet, men ikke så stor at dommerne bare benytter deler av skalaen. Det har vist seg at ved bruk av skala på 100 enheter, blir skalaen redusert ved at dommerne praktisk omarbeider skalaen ved bare å bruke hver femte eller tiende del. I dag brukes mest 5-, 7-, 9- og 15-punkts skala i sensorisk analyse.

4.2. Forskjellstester

Lene Waldenstrøm, Høgskolen i Sør-Trøndelag

Forskjellstester brukes for å undersøke om det er sensoriske forskjeller mellom prøver. Noen tester egner seg hvis den forventede forskjellen antas å være svært liten mens andre igjen benyttes der man ønsker å bekrefte en antatt forskjell eller finne eventuelle preferanseforskjeller.

Forskjellstester, eller differansetester kan deles inn i to grupper, generelle forskjellstester og spesifikke forskjellstester.

Det man kan få svar på ved bruk av de generelle forskjellstestene triangeltest, duo-trio-test og to-av-fem-test er om det beviselig er forskjell





mellom *to* prøver, gjerne kalt prøve A og prøve B. Hvor mange ganger den enkelte prøven serveres bestemmes av metoden. Med de generelle forskjellstestene stilles spørsmålet: Eksisterer det en forskjell? De oftest brukte metodene er triangeltest, duo-trio-test, to-av-fem-test og A- ikke A-test. Disse testene gir bare et generelt svar på om det er sensorisk forskjell mellom prøvene. Ved mindre endringer i ingredienser eller produksjonsmetoder er disse testene gode å bruke. Man vet altså at det for eksempel er en teknologisk eller kjemisk forskjell mellom prøvene, men lurer på om forskjellen kan oppdages sensorisk. Det kan en generell forskjellstest gi svar på.

Spesifikke forskjellstester brukes når det er spørsmål om forskjell mellom prøvene for en konkret egenskap, eller om preferanse for en enkelt prøve. Metoder som brukes er partest og rangeringstest. Disse testene gir oss svar på om det er en merkbar forskjell i én angitt egenskap, eller om prøvene kan rangeres etter økende grad av preferanse.

Disse sensoriske metodene brukes også ved utvelgelse og trening av dommere i den hensikt å finne ut om de aktuelle personene har sensorisk evne til å skille mellom prøver som beviselig bare er litt forskjellige. Da vil panelleder alltid ha en fasit, og dommerens svar vil være enten feil eller riktig.

Felles prinsipper for metodene

Tvunget valg

Dommerne må avgi svar. Det er ikke lov å svare “kjenner ikke forskjell”. På denne måten vil selv forskjeller som er så små at dommeren nesten ikke registrerer det selv, bli fanget opp. Statistikken som ligger bak tabellene som er utarbeidet for beregning av sikkerhet i analysesvaret, tar høyde for at dommerne må velge en av prøvene selv om han eller hun ikke egentlig kjenner forskjell.

Variert servering

Det er viktig å sikre at serveringsrekkefølgen av prøvene blir variert mellom dommerne. De ulike metodene krever ulike oppsett som styres av panelleder ved hjelp av en serveringsplan (se 16. vedlegg, a). Hvordan serveringsplanene skal utformes avhenger av metode og antall





serveringsomganger. Dette er nærmere beskrevet under den enkelte metode. For å sikre at dommerne vurderer prøvene i den rekkefølgen serveringsplanen forteller så må det opplyses om rekkefølgen for bedømming (vanligvis fra venstre mot høyre) på bedømmelseskjemaet.

Koding

Forskjellstestene utføres i all hovedsak som blindtester der dommerne skal vite minst mulig om prøven eller produktet. Dette sikres ved å kode prøvene. Det er vanlig å benytte nøytrale tretalls-koder (se kapittel 3). Hvis man utfører tester i flere omganger med de samme dommerne så må prøvene få nye kodesett til hver omgang.

Studier viser at det å servere en eller to prøver som eksempler på de prøvene som skal inngå i analysen, altså en oppvarmingsprøve (eng. warm-up-sample), vil styrke resultatet av analysen ved at dommerne blir mer fokuserte på de aktuelle egenskapene ved prøvene. Det noteres ikke resultater fra en slik oppvarmingsprøve, den inngår ikke i den reelle analysen.

Hypotesetesting

Alle metoder baserer seg på hypotesetesting. Ved fastsettelse av 0-hypotesen, skal man gå ut fra at panelet ikke klarer å skille prøvene fra hverandre generelt eller for en angitt egenskap. Resultatet av testen vil gi oss svar på om 0-hypotesen skal forkastes eller ikke. For statistisk behandling av de ulike forskjellstester, se kapittel 11

Statistisk signifikans

Et signifikansnivå gir uttrykk for hvor stor sjanse man ønsker å ta for feilaktig å forkaste nullhypotesen. Vanligvis brukes 5 %, 1 % eller 0,1 % signifikansnivå. Ved signifikansnivå 5 % kan vi være 95 % sikre på at vi ikke forkaster nullhypotesen på feil grunnlag. Hvis vi ikke får forkastet nullhypotesen, betyr ikke det at nullhypotesen er sann. Dommernes sensitivitet for eksempel kan være avgjørende, lite sensitive dommere finner kanskje ikke forskjellen mellom prøvene, mens mer sensitive dommere finner en forskjell. Forkaster vi nullhypotesen, kan vi si at alternativ hypotese er sann med 95 % sikkerhet (ved 5 % signifikansnivå).





4.2.1. Triangeltest

Triangeltest brukes for å påvise om det er forskjell mellom to produkter, og er en av de mest benyttede forskjellstestene. Testen kan også brukes til å teste likhet, men det kommer vi ikke inn på i dette kapitlet. Testen brukes i de tilfeller hvor vi mener at forskjellen mellom produktene er svært liten og kanskje ikke er merkbar. Antall testomganger er begrenset fordi metoden kan være mentalt og sensorisk utmattende å utføre.

I praksis vil vi ofte ha flere enn to produkter vi vil teste. I slike tilfeller er det relativt ressurskrevende å gjennomføre en triangeltest, fordi det kreves mange triangler. Noen produkter skal kanskje smakes og luktes på, mens andre bare skal sees på eller føles på. De forskjellige sansene våre trenger varierende tid for restitusjon (se kapittel 2, sansene) så avhengig av produkt er det derfor en forskjell i hvor mange prøver som kan analyseres i løpet av en gitt tid. Matvarer skal som oftest smakes på og da vil dommerne fort kunne bli smakstrette, noe som ødelegger testen. Ofte kan det da lønne seg å bruke andre tester, som for eksempel poengbedømmelse.

Prinsipp og gjennomføring

Hver dommer presenteres for tre prøver samtidig (et triangel), hvorav to er identiske og én er forskjellig fra de to andre. Hver prøve er gitt tilfeldige koder, se eksempel i kapittel 16, vedlegg a) hvor prøve A har fått to forskjellige koder og prøve B har to andre tilfeldige koder. Dommerens oppgave er å peke ut den ulike prøven. Hvis dommeren ikke merker noen forskjell, må han likevel foreta et valg. Dette betyr at dommeren da må velge én av prøvene. Sannsynligheten for å velge den ulike prøven selv om man egentlig ikke merker forskjell på de tre prøvene, er $p=1/3$ (33,33 %).

Siden forskjellen mellom prøvene er svært liten kan det være fristende for en dommer å svare «ingen forskjell» dersom han er litt usikker og slett ikke ønsker å svare «feil». Tabellene som brukes for å se om dommerpanelets svar viser at prøvene er merkbart forskjellige eller ikke, er utarbeidet med utgangspunkt i at dommerne må velge en prøve som den forskjellige (se kapittel 11, statistikk). Framgangsmåten for





gjennomføring av triangeltest hvor dommerne kan få mene «ingen forskjell» er imidlertid beskrevet i ISO-standarden for triangeltest (ISO 4120:2004) med den statistiske beregningen som da kreves for rett konklusjon.

Når man skal lage triangler med de to prøvene A og B, vil det være seks mulige kombinasjoner:

ABB BAB BBA BAA ABA AAB

Disse seks trianglene utgjør ett kombinasjonssett med ulik prøvepresentasjon til 6 dommere. Hvis dommerantallet ikke er delelig med 6, kreves en tilpasset framgangsmåte. Når dommerantallet for eksempel er 16, tenker vi oss tre kombinasjonssett, altså 18 prøver. To sett (12

Eksempel på gjennomføring av triangeltest:

1. Er det sensorisk forskjell mellom A (lagret tomatpuré) og B (ny type tomatpuré)?
2. Panelet består av 24 dommere. Det lages 4 hele sett med de 6 serveringskombinasjonene ABB, BAB, BBA, BAA, ABA, AAB. I denne testen er det kun en prøveomgang noe som gjør at vi kan dele de 24 dommerne i to grupper à 12, der den ene gruppen får prøve A (ordinær tomatpure) som den ulike prøven, mens den andre gruppen får B (ny type) som ulike prøve. Vurderingsrekkefølgen sikres ved at det lages individuelle bedømmelseskjema til hver dommer.

Hvis vi her hadde ønsket flere prøveomganger med de samme dommerne måtte vi ha sikret tilfeldig fordeling av prøvene da det ikke må være noen form for systematikk i serveringsrekkefølgen.

3. Antall riktige besvarelser (de dommerne som har klart å plukke ut den ulike prøven) summeres. Ved å gå inn i tabell for triangeltest for 24 dommere (kapittel 16 Vedlegg b, del a binomisk fordeling) finner vi det antall besvarelser i den ulike prøvens favør, som er nødvendig for å kunne forkaste nullhypotesen ved de ulike signifikansnivåene.

Resultatet fra en triangeltest kan vise at det er en forskjell mellom de to analyserte prøvene. Ofte er det et tilfredsstillende svar, men ikke dersom det også er behov for å vite hvilken egenskap som er årsak til forskjellen. Vi kan for eksempel benytte en partest og spørre panelet hvilken prøve som er mest.....(søt, rød, likt) eller en beskrivende test som gir en intensitetsbeskrivelse av flere av produktets egenskaper (se 4.3.).





triangler) lages og deles ut i sin helhet til 12 dommere, fra det tredje settet velges fire prøver, to med A som ulik og to med B som ulik.

Det er viktig å sikre at prøvene bedømmes i angitt rekkefølge (ABB/BAB osv.) fordi den ene prøven ikke skal ha noen fordel eller ulempe av å bli bedømt som første, andre eller tredje prøve. A-prøven og B-prøven skal serveres som singelprøve like mange ganger. Dersom produktet har en sterk smak vil det kanskje være en mulighet at dommeren ikke har klart å skylle ut smaken fra en prøve før den neste smakes på (carry-over effect). Det skal instrueres i nødvendigheten av å skylle munnen mellom hver prøve og vente med å smake på neste prøve til smaken fra den forrige er helt borte.

Uavhengig av om vi benytter papir eller databasert besvarelse kan vi sikre en variert og kontrollert serveringsrekkefølge enten ved å:

a) lage individuelle bedømmelsesskjema med koder og angitt vurderingsrekkefølge. Det er viktig å passe på at prøvene som blir servert står i samme rekkefølge som på bedømmelsesskjemaet. Denne fremgangsmåten bør velges ved bruk av utrente dommere og ved utvelgelse og trening av dommere.

b) skrive på skjemaet at dommeren skal smake fra venstre mot høyre og sikre at rekkefølgen blir riktig for hver enkelt dommer idet vi serverer prøvene. Denne fremgangsmåten bør kun benyttes ved bruk av trent panel.

Samtidig må vi, ved flere prøveomganger med de samme dommerne, sikre randomisert servering. Det må ikke være noen form for systematikk i serveringsrekkefølgen.

Eksempel på bedømmelsesskjema vises i kapittel 16. vedlegg a. Det er lov å be dommerne kommentere prøvoforskjellen på skjemaet, men ikke stille ledende spørsmål. Imidlertid er det ingen verdi i de kommentarene som blir gitt av en dommer som ikke har klart å plukke ut den single prøven i triannglet. Kommentarer gir ikke grunnlag for statistisk behandling.





Bearbeiding av resultatet:

Først må man undersøke om noen av skjemaene er feil og mangelfullt utfylt. Det kan skje når dommerne gir sine svar på papirskjema, men svært sjelden når besvarelsen registreres elektronisk fordi slike programmer sikrer fullstendig besvarelse før dommeren kan avslutte bedømmelsen.

Når vi må plukke ut et skjema på grunn av feil utfylling, kan det bli skjeheter i fordelingen av prøve A og B. Dette kan rettes opp ved å forkaste et tilfeldig valgt riktig utfylt skjema med motsatt serveringsrekkefølge. Ut fra de riktig utfylte skjemaene teller man opp hvor mange som har klart å plukke ut den ulike prøven. Deretter går man inn i tabellen for triangeltester i kapittel 16 b, statistiske tabeller, del



Figur 4.3 Presentasjon av kodete prøver til henholdsvis partest, triangeltest, duo-triotest og rangering.

Foto: Kjell Merok





a binomisk fordeling. I rubrikken for antall dommere finner vi fram det antallet dommere som var med i vår test. Antall riktig utfylte skjema tilsvarende antall dommere. Tabellen viser det minimum antall bedømmelser som har gitt det forventede svaret i den ulike prøvens favør. Vi kan finne ut om vi med en gitt sikkerhet kan påstå at prøvene er forskjellige.

4.2.2 Partest

Partest brukes for å undersøke om det er merkbare forskjeller mellom to prøver (A og B) når det gjelder en bestemt sensorisk egenskap. Dommerens oppgave er å velge den ene prøven framfor den andre på grunnlag av nærmere angitt kriterium, f.eks. lydstyrke, fargenyanse, fruktig lukt, syrlighet, hardhet eller sprøhet. Her er det svært viktig at panellederen forsikrer seg om at alle i dommerpanelet forstår hvilken egenskap prøvene skal bedømmes etter og at dommerne har samme egenskapsforståelse. Metoden kan også brukes til å avgjøre preferanse for den ene prøven framfor den andre, innenfor en gitt egenskap, eller generelt.

Testen brukes også til å velge ut og trene dommere. Da benyttes prøver med kjent styrkeforskjell av angitt egenskap, f.eks. søttest. Testen har altså en fasit, og panelleder vet hvilken prøve som er søttest og om dommeren har svart riktig eller feil.

Prinsipp og gjennomføring

Dommeren presenteres for to prøver og stilles spørsmålet: Hvilken av disse to prøvene oppfatter du som for eksempel søttest/rødest/surest/sprøest/best?

Hver dommer serveres to kodete prøver samtidig (et par). Dommerens oppgave er å peke ut den prøven han mener har mest av den egenskapen det spørres etter. Hvis dommeren ikke kjenner noen forskjell når det gjelder angitt egenskap, må han likevel avgi et svar. Sannsynligheten for å velge den ene prøven framfor den andre hvis man ikke kjenner forskjell på de to prøvene når det gjelder intensitet av angitt egenskap, er $p=1/2$ (50 %).





Vi tilbereder likt antall prøvesett av kombinasjonen AB og BA. Disse fordeles tilfeldig mellom dommerne. Antall sluttbedømmelser må være delelig med 2, og like mange AB som BA skal serveres. Det er også her viktig å sikre at dommerne bedømmer prøvene i riktig rekkefølge (se også triangeltest). Samtidig må vi, ved flere prøveomganger med de samme dommerne, sikre variert servering. Det må ikke være noen form for systematikk i serveringsrekkefølgen. Bare én egenskap kan prøves om gangen – det er altså ikke lov å be dommeren avgi svar på både søthet og konsistens på samme skjema og med samme prøver.

Bearbeiding av resultatet:

Man skiller mellom ensidig og tosidig partest avhengig av kunnskap om prøvene og problemstillingen for undersøkelsen (se kapittel 11.3). Det er kun i bearbeidingen av resultatene de to typene partest skiller seg fra hverandre. Både bedømmelsesskjemaer og måten man gjennomfører testen på er lik.

Ut fra de riktig utfylte bedømmelsesskjemaene finner man den prøven som flest dommere mener inneholder mest av egenskapen det spørres etter. Kapittel 16b viser tabell for ensidig- og tosidig partest. Tabellene forteller oss det minimum antall bedømmelser som kreves for at vi med en bestemt sikkerhet kan si at det er forskjell mellom prøvene, altså om 0-hypotesen kan forkastes ved de ulike signifikansnivåene.

4.2.3 Duo-trio-test

Duo-trio-testen kan sies å være en mellomting mellom partest og triangeltest. Her blir dommerne presenterte for tre prøver, hvorav to er like og den tredje er ulik. Her blir en av de to like prøvene merket som en standard eller referanse (R), og de to andre prøvene kodes etter de samme prinsipper som nevnt i kapittel 3. Her er det naturlig å tenke seg at bedriftens vanlige produkt bør fungere som referanseprøve, altså merkes R. Dommerens oppgave er å finne hvilken av de to kodede prøvene som er ulik standarden/referansen.

Ved bruk av duo-trio-metoden er det mulig å servere prøvene på fire forskjellige måter i forhold til hverandre:





AR-A-B (Referansen er en prøve fra A)

AR-B-A (Referansen er en prøve fra A)

BR-A-B (Referansen er en prøve fra B)

BR-B-A (Referansen er en prøve fra B)

Det er to måter å organisere denne testen på; ved en balansert referanseteknikk eller en konstant referanseteknikk. De to prøvene A og B kan begge inngå som referanseprøve (balansert referanseteknikk) eller bare den ene prøven kan inngå som referanse (konstant referanseteknikk).

Ved konstant referanseteknikk er den ene prøven alltid referanseprøve, og det er bare to mulige serveringskombinasjoner (AR-A-B eller AR-B-A, der A er kontrollprøven eller altså referanseprøven). Denne situasjonen er hensiktsmessig ved kvalitetskontroll der et trent panel er kjent med bedriftens standardprodukt, og dommernes oppgave er for eksempel å avgjøre om den løpende produksjonen avviker fra denne standarden eller referansen. I de tilfeller der det er et lite produksjonsparti (for eksempel ny resept) av den ene prøven i forhold til den andre, er det hensiktsmessig å velge en duo-trio-test med en konstant referanse som metode. Den prøven som det finnes mest av, velges følgelig som referanseprøve (R).

Ved balansert referanseteknikk sikrer vi lik serveringsbehandling for begge prøvene. Har vi et panel bestående av tolv personer, vil hver av de fire serveringskombinasjonene bli servert tre ganger. Dersom panelet består av ti personer, skal et sett der A er referanse og et sett der B er referanse, fjernes fra det tilsvarende oppsettet til tolv dommere.

En duo-trio-test er ikke egnet som metode for å måle preferanse eller akseptanse. Den er heller ikke egnet for å analysere hvilken egenskap som skiller prøvene eller intensiteten av den egenskapen som varierer mellom de målte prøvene.

Resultatbehandling se kapittel 11





4.2.4. To-av-fem-test

En annen metode som kan brukes for å sammenlikne to prøver er to-av-fem-test. Dommerne blir presentert for fem prøver der henholdsvis tre (A) og to (B) er like. Prøvene er kodet etter prinsipper for sensorisk analyse, slik at de serveres for dommeren som fem forskjellige prøver. Dommernes oppgave her er å avgjøre hvilke tre prøver som er like og følgelig hvilke to prøver som er like. Prosedyren og gjennomføringen av en denne analysen er i hovedsak lik det som er skrevet om triangeltesten. Den eneste forskjellen er at de to prøvene A og B som skal analyseres ved to-av-fem metoden må deles i henholdsvis to og tre prøver mens ved triangelmetoden er det bare den ene av A eller B som må deles i to prøver.

Denne metoden er mer krevende for dommerne enn en test med triangelmetoden i og med at i alt fem prøver må vurderes for å kunne gi svar på hvilke prøver som hører til i hver gruppe. Fordelen med metoden sammenlignet med en triangeltest er at det kreves færre dommere som deltagere i analysen for å oppnå et signifikant sikkert svar på oppgaven. Panallederen bør likevel være klar over at denne testen er vanskelig for dommerne, særlig dersom det er snakk om prøver med sterk lukt eller smak. Metoden egner seg kanskje best ved testing av produkter som ikke skal smakes på, men for eksempel bare fargebedømmes.

Denne testen kan serveres med 20 forskjellige prøvekombinasjoner:

AAABB BBBAA AABAB BBABA ABAAB.
BABBA BAAAB ABBBA AABBA BBAAB
ABABA BABAB BAABA ABBAB ABBAA
BAABB BABAA ABABB BBAAA AABBB

Består panelet av færre enn 20 dommere, bør serveringsordningen velges tilfeldig blant de 20 mulige kombinasjonene, men det må sørges for like mange triple A-prøver som triple B-prøver.

Sannsynligheten for å velge en av prøvene fremfor en annen hvis man ikke kjenner forskjell på de fem prøvene er $p=1/10$ (10%).





4.2.5 A – ikke A test

Dommerne får presentert to prøver for trening i å gjenkjenne de sensoriske kvalitetene i hver av dem. Deretter skal de ukjente prøvene bedømmes. Det er altså to prøver; A og B, som her kalles hhv. A og ikke-A.

Det er flere muligheter for serveringsprosedyre ved A- ikke A metoden. Enten kan vi servere en og en prøve som skal sammenlignes med den innøvde A-prøven, eller vi kan servere to prøver (et par) som skal sammenlignes med A-prøven.

Den første varianten kan sammenlignes med en partest, med den forskjell at den ene prøven her bare eksisterer i hukommelsen til dommeren. Flere prøver som er kodet, noen som er lik «A-prøven» og noen som er lik «ikke-A-prøven», blir presentert for sammenligning med den innlærte «A-prøven». Dommernes oppgave er å avgjøre om de ukjente prøvene er lik den innlærte A-prøven.

I den andre varianten, som er den mest brukte kan vi se en likhet med en duo-trio-test, men her med den forskjellen at A-prøven som er den kjente prøven bare eksisterer i hukommelsen mens i duo-trio-test er den merket som Standard eller Referanse. Et par (2) kodete prøver skal bedømmes for å avdekke om en eller begge eller ingen i paret er lik med hukommelsen om A-prøven. Det er fire forskjellige serveringskombinasjoner, AA, BB, AB, BA. Disse parkombinasjonene skal serveres i tilfeldig rekkefølge til dommerne.

Statistisk analyse av data følger samme prinsipper som for partest.

A – ikke A metoden brukes gjerne når det ikke er mulig å presentere helt identiske kopier av en prøve, for eksempel to spekeskinker med litt forskjellig fettmarmorering eller to treplater med litt forskjellig kvistmønster som kan forårsake prøvevariasjon. Forskjellen som følge av variasjon i utseende eller fargenyanse må imidlertid være så liten at den kun kan oppdages når prøvene plasseres ved siden av hverandre. Derfor er det i denne metoden vesentlig at A-prøven skal studeres nøye og lagres i hukommelsen før den fjernes.





4.1.6 Rangering

Når flere prøver skal sammenlignes for intensitet av en spesiell egenskap og man ønsker å vite rekkefølgen av prøvene med tanke på denne egenskapen, er det hensiktsmessig å velge rangeringsmetode. Metoden brukes også som en sortering av prøver før anvendelse av andre sensoriske metoder. Dessuten brukes rangeringstester til å velge ut og trene dommere.

Prinsipp og gjennomføring

Hver dommer presenteres for minst tre prøver, som skal rangeres etter et bestemt kriterium, f.eks. utseende, søthet, konsistens, farge eller preferanse. Antall prøver som skal rangeres bør ikke bli for stort fordi dommerne fort kan bli «smakstrette». Dommerne må informeres om hvilken grad av egenskapen som skal settes som nr. 1 (f.eks. minst søt) og hvilken som skal settes sist (f.eks. mest søt). Alle prøver skal rangeres, ingen får altså settes lik hverandre.

Det er viktig å unngå at prøvene serveres alle dommerne i samme rekkefølge. Dette sikres ved å lage individuelle bedømmelsesskjema med koder og angitt vurderingsrekkefølge (eksempel i kapittel 16 a). Samtidig må vi, ved flere prøveomganger med de samme dommerne, sikre at prøvene blir tilfeldig fordelt. Det må ikke være noen form for systematikk i serveringsrekkefølgen. Bare én egenskap kan prøves om gangen.

Beregning av resultater fra rangeringstest er vist i kapittel 11

4.3. Beskrivende tester

Marit Rødbotten og Mats Carlehög, Nofima

Kvantitative metoder som også kan kalles beskrivende – eller deskriptive metoder, er metoder som beskriver produktets sensoriske profil. Vi kan bedømme produktet som helhet eller bare deler av det. Hvor detaljert en skal gjøre undersøkelsen er avhengig av problemstillingen.

Dersom det er forventet at det kan avsløres en sensorisk forskjell, vil det være mest hensiktsmessig å benytte en kvantitativ beskrivende





metode istedenfor en forskjellsmetode. Ved en forskjellsmetode får vi svar på om to prøver kan oppleves som forskjellige. Dersom resultatet sier at det er forskjell mellom prøvene, er det rimelig å forvente at oppdragsgiveren vil ønske å få svar på hvilken egenskap, eventuelt hvilke egenskaper som varierer mellom prøvene og om forskjellen er betydelig eller minimal. Tid og penger kan spares ved at panellederen får detaljert informasjon om prøvene og undersøkelsens problemstilling, slik at riktig metode blir valgt med en gang.

Det er en selvfølge at det sensoriske personalet har taushetsplikt angående prøvenes problemstilling og eventuelt hvem som er oppdragsgiver. Ikke noe av denne informasjonen blir gitt videre til dommerpanelet fordi informasjon om problemstillingen som ligger bak undersøkelsen kan ubevisst eller bevisst påvirke resultatet fra dommerne.

Metoder som oftest blir brukt er kvalitetskontrolltest og beskrivende test (forskjellige profileringsmetoder). Andre metoder som ikke så ofte brukes, men som er velegnet i spesielle situasjoner er kvalitetsindeksmetode (QIM) og metoder hvor intensitet av sensoriske egenskaper måles over en bestemt tidsperiode, som for eksempel time-intensity metoden (TI) og Temporal Dominance of Sensation (TDS).

4.3.1 Beskrivende test (profilering)

Innen den sensoriske vitenskapen er beskrivende tester de som gir mest informasjon om det analyserte produktet. Her kan vi få detaljert informasjon om blant annet produktegenskaper, om ingredienser i produktet kan identifiseres eller om hvilke egenskaper som er av betydning for forbrukervalg.

Dette kapitlet vil først omhandle faktorer som er viktige for et best mulig resultat av en profileringsmetode som på engelsk ofte blir betegnet som en DA-metode, en Descriptive Analysis metode. Fra ca 1940 er det utarbeidet flere varianter av beskrivende tester som alle er i bruk i dag, og noen av disse blir beskrevet senere i dette kapitlet. Selv med noen vesentlige forskjeller er de alle opptatt av å gi en detaljert beskrivelse av alle eller enkelte av produkters sensoriske egenskaper.



ISO 13299:2003 beskriver trinnvis prosedyren ved profilering. Her vises de forskjellige trinnene som i etterkant blir kommentert.

Tabell 4.1 ISO 13299:2003. Trinnvis prosedyre ved profilering.

Trinn nr	Oppgave	Relevant ISO-dokument
1: Innrede sensorisk lokale	Innrede bedømmelsesbåser, prøveprepareringsrom, osv.	8589
2: Velge prøver som illustrerer de egenskapene som skal bedømmes	En eller to eksperter (produktspert, sensorisk leder) velger egnede prøver.	11035
3: Velge og trene dommere for prosjektet	Panellederen velger en egnet gruppe mennesker, trener dem ved bruk av valgte produkter.	8586 5496
4: Velge egenskaper egnet for beskrivelse av de aktuelle prøvene. Kan kombineres med trinn 3.	Panellederen velger blant eksisterende egenskaper, eller dommerne bedømmer produkter fra trinn 2 og foreslår egenskaper. Valget av egenskaper gjøres enten av panellederen eller i fellesskap med dommerne. Egnede referanseprøver velges der egenskapene blir demonstrert.	5492 6564 11035 11036
5: Bestemme rekkefølgen av de valgte egenskapene	Panellederen arbeider sammen med dommerne under trinn 3 og 4.	11035
6: Velge skala for måling av intensitet av de valgte egenskapene.	Panellederen velger den mest aktuelle skalaen.	4121 11056
7: Gjennomføre testen.	Dommerne bedømmer prøvene.	6658 6564
8: Rapportere resultatene.	Analysere resultatene ved hjelp av statistiske metoder. Presentere resultatene i form av tabeller og diagrammer. Trekke konklusjoner.	6564 13299



Kommentarer til de åtte trinnene i tabellen:

Trinn 1

Betingelsene for et godt resultat av en sensorisk analyse ligger for en stor del på egnetheten av de sensoriske lokalene og de omgivelsene der prøveprepareringen skal foregå (se kapittel 3). Avhengig av hva som skal analyseres må det legges til rette for en objektiv analyse (unngå forstyrrelser, lyder, lukter, variasjon i lystilgang og lignende).

Trinn 2

For å kunne gjøre en grundig undersøkelse av de aktuelle prøvene er det viktig å kjenne til produkter med tilsvarende egenskaper som de som er med i undersøkelsen. Dette vil ofte føre til at vi raskere kan velge egenskaper som skal bedømmes i de aktuelle prøvene. Kjennskap til resepten kan antyde hvilke egenskaper som bør bedømmes. Men vær oppmerksom på at selv om resepten og instrumentelle undersøkelser dokumenterer innhold av et stoff som forventes å gi en bestemt sensorisk effekt, for eksempel brix-måling av sukkerinnhold i erter, er det ikke dermed sagt at et panel som ikke finner den forventede søtsmaken, har gjort en dårlig jobb. Andre smaker kan overdøve eller kamuflere søtsmaken.

Trinn 3

Se kapittel 3. Ved valg av dommere som er ansatt internt i bedriften er det viktig å ta i betraktning kunnskapen dommerne kan ha om prosjektet/produktene som skal analyseres. Forhåndskunnskap om spesiell behandling av prøvene vil lett forstyrre den objektive bedømmelsen og dommers detaljkunnskap må tas til følge når analyseresultatet skal vurderes.

Trinn 4

Dommerne får servert forskjellige prøver som er representative for oppdraget. I denne fasen er det ønskelig å få navngitt alle de egenskapene som er med på å gi prøvene den bestemte sensoriske profilen. Her er det aktuelt å la dommerne nevne alle ord som de finner i beskrivelse av produktet, gjerne kalt «hjernestorm» eller «brainstorming». Den





forhåpentligvis lange listen av egenskaper blir deretter bearbeidet til å inneholde de ønskede og viktigste sensoriske egenskapene. Det viser seg ofte at det er bedre å ha med litt for mange ord enn for få ord på den endelige listen av egenskaper, i alle fall når det er viktig å studerer detaljer ved produktet. På den andre side er det uhensiktsmessig å bruke tid på bedømmelse av egenskaper som ikke forventes å variere mellom produktene og heller ikke er av interesse for oppdragsgiveren.

Sensoriske egenskaper kan klassifiseres i forskjellige nivåer av detaljinformasjon, som eksempel kan vi tenke oss det mer overordnede ordet bærsmak i forhold til ordet jordbærsmak som er mer spesifikt. Det er nødvendig at alle dommerne er enige om forståelsen og definisjonen av hvert egenskapsord som skal bedømmes. Hvor detaljert (i dette eksemplet: bær eller jordbær) egenskapene skal beskrive produktet må avklares på forhånd i samråd med oppdragsgiver/produktansvarlig. Se Aromahjul (figur 4.4) for eksempel på gradert produktbeskrivelse

Gjenkjenning av egenskaper er avhengig av dommerens sensitivitet og tidligere erfaring med produkter med de aktuelle egenskapene. Trening i gjenkjenning av egenskaper i forskjellig intensitet gjøres ved først å bli presentert for det sterke og tydelige, for deretter å bli presentert for den samme egenskapen i svakere intensitet. Læring av gjenkjenning av en smak skjer ved en parsammenligning mellom den sterke smaken sammen med en stadig svakere smak.

Panelet skal være trent i de egenskapene som skal bedømmes. Panelet og panellederen blir enige om en definisjon av hver av de aktuelle egenskapene. Denne definisjonen følger også i rapporten, slik at oppdragsgiveren får samme tolking av egenskapene. Det advares mot å la et trent panel bedømme «totalinntrykk» av en prøve. Det er en hedonisk vurdering og skal følgelig utføres av forbrukere.

Trinn 5

Det er ikke likegyldig i hvilken rekkefølge egenskapene bedømmes. For matvarer er det oftest riktig å bedømme farge og utseende på prøven før lukt, smak og tekstur bedømmes. Selv om vi anstrenger oss til det ytterste for å oppnå en objektiv bedømmelse kan vi ikke





utelukke at dommere blir påvirket av tidligere erfaringer med produktene som skal bedømmes. For eksempel er det ofte en mindre intens jordbærsmak på et «brunt», gammelt jordbærsyltetøy enn et ferskt, rødt syltetøy. For bedømmelse av smak i dette produktet er det hensiktsmessig å presentere prøvene under rødt lys som kamouflerer den rette fargen, for at smaken skal bli mer objektivt bedømt (se kapittel 2). Smaksbedømmelsen i rødt lys bør følgelig gjennomføres før fargebedømmelsen i vanlig lys.

Trinn 6

Samtidig som det trenes på å kjenne igjen og vurdere intensitet av de valgte egenskapene, er det aktuelt å trene på bruk av intensitet på en gitt skala. Denne fasen med trening, ofte kalt for-forsøk eller treningsforsøk, er svært avgjørende for resultatet av hele analysen. Nøyaktigheten og presisjonen av bedømmelsen avhenger av paneldeltakernes felles oppfatning av sensoriske egenskaper, med definisjon og intensitet. Tid og penger er dessverre ofte begrensende faktorer for denne treningsperioden med de konsekvenser som følger for nøyaktigheten av analysen.

Ved for-forsøkene er det en fordel at alle prøvene kommer fra samme prøveuttak eller batch, slik at muligheten for enighet mellom dommerne blir størst mulig. Dommerne skal her «kalibreres» som et instrument. En eller gjerne flere prøver som er identisk fra dag til dag, og som er forskjellige fra hverandre med hensyn til en eller flere sensoriske egenskaper, vil være egnet som treningsobjekt. Disse prøvene forenkler kalibreringen av panelet mot skalaen.

Det er ikke nødvendig at alle dommerne gir den samme tallkarakteren for intensitet av prøven. Men det er svært viktig at den dommeren som bruker lave tall på skalaen alltid gjør det ved presentasjon av samme intensitet av en stimulus. Og tilsvarende for de som bruker de høye tallene på skalaen.

Det er gjennomsnittet av alle dommenes besvarelser som utgjør analyse-resultatet, ikke enkeltdommernes.

Bedømmelsesskjemaet, egenskapsforklaringen og framgangsmåten for bedømmelsen presenteres for dommerne. Alle dommerne serveres



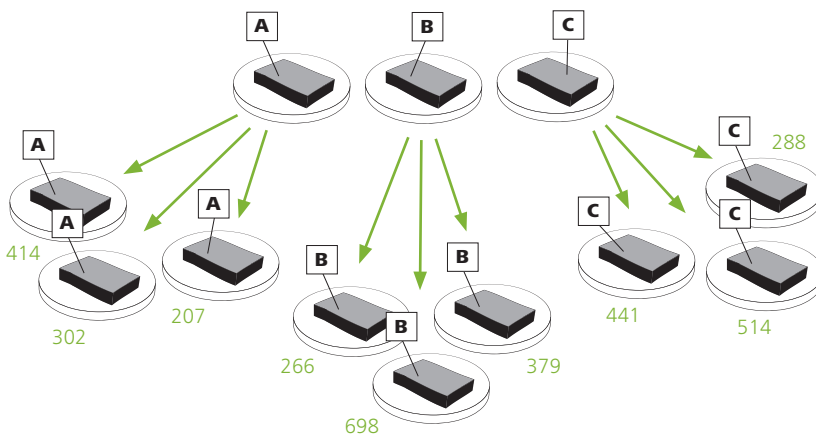


to eller flere prøver som er forskjellige, og testen gjennomføres. Ikke alltid er dommerne helt samkjørte ved første gangs testing av prøvene. Derfor bør panelet kalles sammen for en diskusjon for eventuelt justering av egne bedømmelser.

Følgende punkter kan ligge til grunn for diskusjonen: dommerne har ikke lik forståelse av egenskapene, er uenige om rekkefølgen på prøvene med tanke på intensitet av en egenskap, eller den valgte referanseprøven er uegnet.

Gjenta for-forsøket (treningen) til dommerne er samstemte ved bedømmelse. Under for-forsøkene bør en være oppmerksom på om noen av de planlagte egenskapene kan sløyfes fordi de ikke gir noen informasjon om variasjon mellom prøvene (alle er like søte), eller om egenskaper viser samvariasjon (enten tørr eller saftig).

En måte å øke sikkerheten i resultatet fra en sensorisk analyse på er å gjenta analysen en eller flere ganger. Det forventes at en dommer skal kunne gjenta sin karaktergivning likt for prøver som er like, for eksempel saft fra samme flaske servert i tre glass med forskjellige koder bør få samme karakterer (scoring) på innholdet i hvert glass.



Figur 4.4 Gjentak av prøver

Illustrasjon: Julie Gjermundsen





Ved bestemmelse av hvor mange repetisjoner som bør gjennomføres av en analyse, er det riktig å velge repetisjonen der variasjonen er størst. Forventes det større variasjon mellom to produkter (innen en produksjon, en behandling eller søsken i en flokk griser) enn variasjonen i et panels bedømmelse av ett og samme produkt, ja da er det større behov for repetisjon av prøvemeditet (flere griser, flere fisk, o.l.) enn av antall sensoriske gjentak av bedømmelser. Dette gjelder særlig for beskrivende tester, men til dels også for kvalitetskontrolltester.

Trinn 7

Se kapittel 3, praktisk gjennomføring. Panelansvarlig må ha kontroll med prepareringen av prøvene, serveringsrekkefølgen og kodingen av prøvene. Randomisering av prøveservering er viktig og det er kun i spesielle tilfeller det kan anbefales å servere prøver gruppevis med hensyn på intensitet av egenskaper. Panellederen bør være bevisst på effekten av å servere prøver med lav intensitet av egenskaper før de sterkere prøvene serveres. En slik serveringsrekkefølge vil gi dommerne en forventning om at intensiteten skal øke og dermed ikke nødvendigvis bedømme slik de sanser eller opplever prøven. Eksempler kan være pølser med lite krydder servert før pølser med mye krydder, eller drikke med lite/mye alkohol, eller lyder med høy/lav tone. I alle tilfeller er det viktig med nok av tid mellom hver prøve slik at smak og lukt fra den ene prøven er borte før den neste prøven skal bedømmes.

Trinn 8 Data fra den sensoriske analysen skal behandles med egnede statistiske metoder og resultatene rapporteres lett forståelig og med grundig henvisning til rådata. Rapporten bør beskrive prøvematerialet, dato for bedømmelse, valg av sensorisk metode, paneltreningen, liste med definisjoner av alle bedømte egenskapene, serveringsrekkefølgen av prøvene i tillegg til resultatet med konklusjon.

Språkutvikling, sensorisk leksikon og aromahjul

Wenche Emblem Larssen, Møreforskning AS

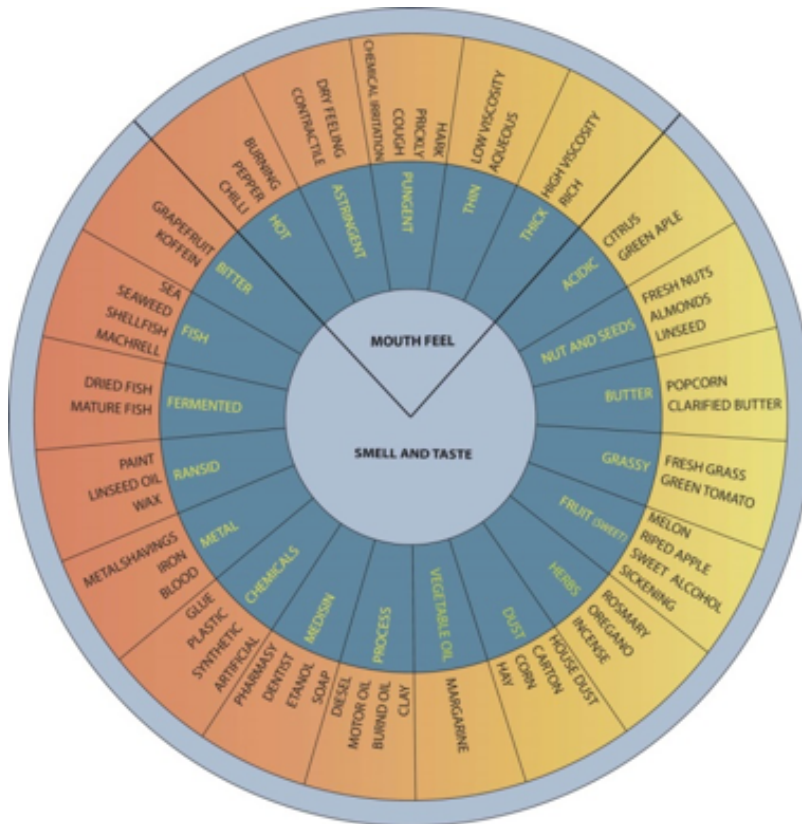
Før en kan gjennomføre en sensorisk vurdering av en produktgruppe må en samle inn ord og uttrykk som kan benyttes til en systematisk beskrivelse av de sensoriske egenskapene. Dette kan gjøres



Praktisk eksempel: Marine oljer (Larsen og Hersleth 2014)

Assosiasjonstest gjennomført på 44 marine oljer av ulikt råstoff og ulike kvaliteter.

Gjennomført av et profesjonelt sensorisk panel og av 6 bedriftspanel. Testen genererte 184 ulike ord og uttrykk der 100 var unike. Sortering og gruppering reduserte antallet til 60 ulike nøkkelord gruppert i 21 ulike kategorier. Kategoriene representeres av de nøkkelordene som gikk mest igjen i assosiasjonstesten. Grafisk fremstilt som et aromahjul.



Figur 4.5 Aromahjul for marine oljer

Illustrasjon: Larsen og Hersleth, 2014



ved å gjennomføre en assosiasjonstest. En assosiasjonstest ledes av en panelleder og deltakerne kan være sensoriske dommere eller utvalgte fokusgrupper (konsumenter, næringsaktører ol.) Dommerne blir presentert for et utvalg av produktene som de skal vurdere for å bli kjent med de sensoriske variasjonene som finnes. Det kan være nyttig å gjennomføre flere runder med innsamling av ord og uttrykk, på et vidt spekter av produkt, med store individuelle forskjeller og ved bruk av ulike dommerpanel for å få en bredes mulig tilnærming (Lyon og Watson 1994, Monteleone og Langstaff 2014)

Den endelige listen med egenskapene (ord og uttrykk) generert gjennom assosiasjonstest blir deretter normalt redusert slik at den kun inneholder egenskaper som gir en klar og nøyaktig beskrivelse. Reduksjon kan gjøres ved at en først sorterer egenskapene alfabetisk og fjerner de som overlapper hverandre, før en deretter grupperer de uttrykkene som hører sammen (Lyon og Watson 1994, Mojet og de Jong 1994, Gawel et al. 2000, Koch et al. 2012). Det er også mulig å vekte de egenskapene som går mest igjen mellom dommerne og eventuelt mellom panelene som de mest viktige. Egenskapslisten kan presenteres som et sensorisk leksikon der de ulike grupperingene defineres eller grafisk som et aromahjul.

Andre beskrivende metoder

Marit Rødbotten og Mats Carlehög, Nofima

Andre metoder som også hører til i gruppen av beskrivende metoder er grundig beskrevet i Lawless og Heymann, 2010. Noen av metodene er kort beskrevet her.

Smaksbeskrivelse. (Flavor Profile®, FP)

Metoden beskriver smak i produkter i fem kategorier; 1) egenskaper, 2) intensitet av egenskapene, 3) egenskapenes rekkefølge med hensyn på når de oppdages i munnen, 4) ettersmak og 5) totalinntrykk. Til å begynne med ble egenskapene målt etter følgende fem kategorier: Ikke merkbar, så vidt merkbar, litt, moderat, sterk. Denne skalaen har etter hvert blitt utvidet til en 17-punkt skala med bruk av halve poeng og pluss og minus for de fem kategoriene. Et godt trenet panel på 4-5





personer bedømmer prøvene og deretter diskuterer de seg frem til et svar de alle er enige om. Panellederen er aktivt med i bedømmelsen. Altså ikke et gjennomsnitt av dommernes bedømmelse, men en felles enighet om svaret.

Kvantitativ beskrivende analyse.
(*Quantitative Descriptive Analysis*®, QDA)

Denne metoden ble utviklet for å korrigere punkter som ble sett på som et stort problem for objektiviteten ved FP-metoden. QDA godtar ikke at resultatet fra analysen kommer som følge av konsensusdiskusjon mellom dommerne, men det skal regnes gjennomsnitt av alle dommernes karakterer for hver egenskap. I tillegg skal ikke panellederen være aktivt med i bedømmelsen og det skal brukes en ustrukturert linjeskala for hver egenskap.

Antall dommere bør være 10-12 og de velges med bakgrunn i gode resultater fra forskjellstester og evne til muntlig å karakterisere produkter. Egenskapslisten blir bestemt etter at dommerne er blitt presentert for mange prøver/produkter av lignende karakter som det som skal bedømmes, og de egenskapene som er mest egnet for prosjektets problemstilling blir valgt for hovedprosjektet.

Det legges vekt på at dommene bedømmer prøvene individuelt i bedømmelsesbåser. Det er ikke påkrevd at dommerne gir samme karakter på prøvene som sine meddommere, men det er viktig at hver dommer kan reprodusere sine egne bedømmelser ved å gi samme karakterer til to prøver som er like, men er kodet forskjellig.

Data kan behandles med statistiske modeller og presenteres med grafiske bilder.

Teksturbeskrivelse. (*Texture Profile*®, TP)

FP metoden ble raskt brukt i modifiserte varianter, som etter hvert endte opp som en spesiell metode egnet til bedømmelse av tekstur av matvarer, nemlig Textur Profile metoden. Metoden gir mulighet til å studere produkter i alle stadier av en tyggeprosess. Produkters geometriske- og mekaniske egenskaper blir målt i tillegg til saftighet, fethet, og det kan måles i hvilken rekkefølge de forskjellige





egenskapene er mest tydelige. Intensiteten av egenskapene blir angitt på en 15-punkt skala (med halve tall i tillegg) og hvert punkt på skalaen har en referanse i et kjent produkt. For eksempel har egenskapen hardhet følgende referansepunkter:

2,5=kokt eggehvite i fem minutter, 9,5=peanøtt og 14,5=hardt sukkertøy. Det er helt avgjørende for riktig bruk av TP at alle dommerne får den samme og grundig trening i bruk av skalaen med de samme referanseproduktene.

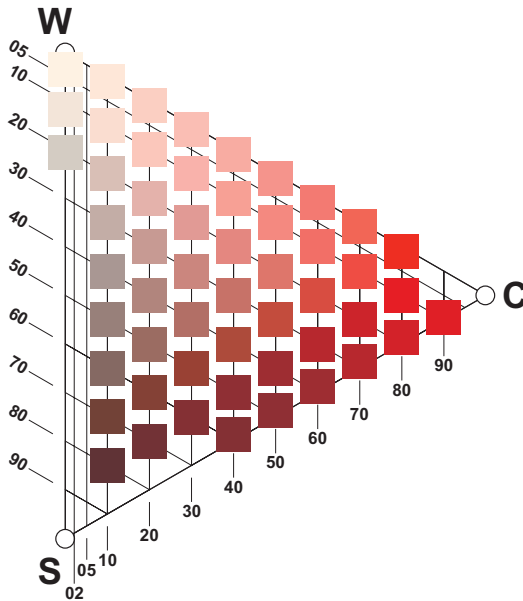
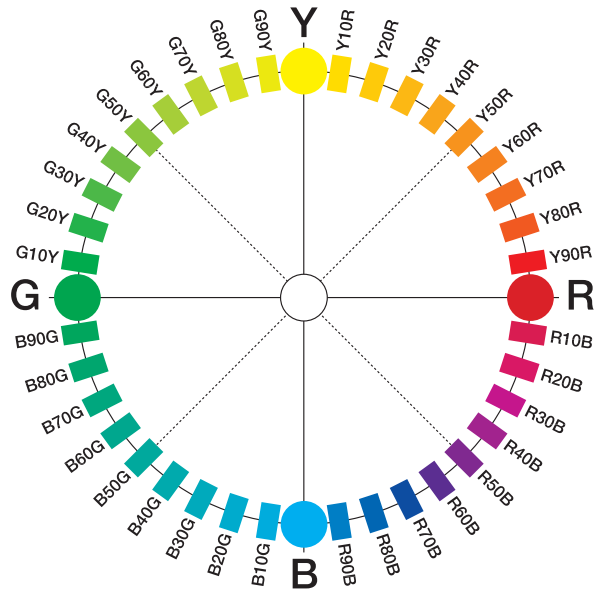
Spectrum metode. (Sensory Spectrum ®)

Denne metoden bygger på en kombinasjon av FP- og TP-metodene. Her kan alle sensoriske sanser brukes for bedømmelse av utseende, lukt, smak, tekstur og følelse. Ikke bare mat og drikke kan bedømmes med Spektrummetoden, men også hygieneartikler som kremer og papir. Skala brukes fra 0-15 med referanseprodukter for nesten alle punkter på skalaen for «alle» sensoriske egenskaper, inklusive grunnsmaker. Eksempel på salt smak: 5 poeng=«salted soda crackers, Premium», 9,5 poeng = «salted potato chips, Frito-lay». Alle egenskapsordene finnes i et leksikon som brukes for de forskjellige produktene og dommerne er ikke delaktige i utvikling av egenskapslisten som skal brukes i prosjektene. Dommerne trenger lang trening for å gjenkjenne de forskjellige egenskapsintensiteten i de forskjellige referansene.

4.3.2. Kvalitetskontrolltest

Serieproduserte varer som forventes å være like bør kontrolleres nettopp med tanke på likhet. Den sensoriske kvaliteten måles opp mot en standard eller referanse og metoden som oftest blir brukt er en poengbedømmelse, også kalt numerisk gradering eller «scoring». Den eller de egenskapene som skal bedømmes, blir kvantifisert ved hjelp av en tallskala (poengskala). Det vanligste er at den ene enden av skalaen angir riktig kvalitet, og den andre enden angir avvikende eller dårlig kvalitet. Metoden er illustrerende og anvendelig, men den stiller store krav til dommerne og panellederen. Grundig trening i bedømmelse av produktet og bruk av skalaen er viktig. Til denne treningen er det nødvendig å arbeide med produktteksempler med riktig kvalitet og





Figur 4.6 og 4.7 Fargesirkel og fargetriangel

Illustrasjon: NCS

110 Sensorikk – Måling med menneskelige sanser





eksempler med avvik fra riktig kvalitet. Et syltetøy kan ha riktig farge, lukt og konsistens, men ha feil intensitet av søtsmak.

I produktutviklingsfasen blir det bestemt hvordan produktet skal være (se kapittel 6) og senere kvalitetskontroll skal dokumentere grad av likhet eller ulikhet med det valgte referanseproduktet (se kapittel 8).

Den produktansvarlige bør være delaktig i prosessen når det sensoriske panelet skal trenes i gradering av avvik i den avvikende prøven, han/hun bør være delaktig i hvor «strengt» de forskjellige avvikene skal markeres.

Som oftest vil aksepttester fra forbrukere gi antydning om hvor grensen bør gå for salgsvare eller ikke, og en eller flere egenskaper kan være avgjørende her. Det kan for eksempel være viktig å angi avvik i fargen på prøven, for eksempel om jordbærsyltetøyet har fått en avvikende farge enten på grunn av lagring eller spesielle ingredienser. Å huske grad av farge er ikke alltid like lett. Som et hjelpemiddel kan en for eksempel benytte fargeprøver fra NCS Natural Colour System®©. NCS er et fargesystem som beskriver farger slik vi ser dem. I 1984 ble NCS en Norsk Standard innen fargebeskrivelse og systemet er brukt i hele Europa, noe som gjør det egnet som referansesystem for sensoriske analyser (www.ncscolour.no). Ved hjelp av NCS fargesirkel (figur 4.6) og NCS fargetriangel (figur nr 4.7) kan de fargene som representerer produkter med riktig og med avvikende farge lett illustreres som hjelpemiddel for dommerne til å gi korrekt angivelse av fargen på prøvene. For trening og kanskje også til bruk under selve bedømmelsen kan enkeltfarger i fargetriangler settes sammen etter økende grad av avvikende farge for produktet, som en rangering av fargene. Hver farge angir et markeringsnivå på bedømmelseskalaen. Det er mulig å finne farger som gjenspeiler alle typer produkter i NCS-systemet.

Det viser seg at dommerne ofte ikke benytter seg av hele skalaen. Poenggivingen er vanskelig å reproducere og skalaen har lett for å «gli» etter en tid. For å motvirke at skalaen glir, bør analysen også inneholde referanseprøver eller standarder som er kodet på lik linje med de ukjente testprøvene. Et veltrent panel bør kunne gi den kodete referanseprøven samme karakter gjennom hele prosjektet. Det kan i noen tilfeller være aktuelt å presentere referanseprøven merket





som «referanse» sammen med de kodede prøvene slik at dommerne kan oppfriske sin hukommelse av de forskjellige egenskapene og intensiteten av disse.

Det bør foreligge en skriftlig sensorisk spesifikasjon for hvert produkt, og denne spesifikasjonen skal i hovedsak beskrive det riktige produktet. En spesifikasjon må beskrive produktet med objektive termer. En beskrivelse av et produkt med for eksempel begrepene «passe salt» eller «riktig farge», er uegnet i denne sammenheng. For visuelle egenskaper som for eksempel mørke eller lyse flekker i produktet kan det være aktuelt å angi flekkens størrelse ved hjelp av standardiserte mål som for eksempel mm.

Ofte kan det være lettere å beskrive hvordan produktet ikke skal være. En slik negativ beskrivelse kan forekomme i produktspesifikasjonen. Da vil det være naturlig å beskrive kvalitetsfeil som har vist seg oftest i kontroll av produktet. Som eksempel kan vi tenke oss antall kvistmerker i et panelgulv eller mengde hull i en ost. En nomenklatur er en liste med aktuelle sensoriske egenskaper for produktet.

I kvalitetskontrollbedømmelsen må en på forhånd ha klart for seg hvilke krav en stiller til presisjonen av bedømmelsen. Hvor mange dommere som skal være med i panelet vil bli en avveining mellom behov for nøyaktighet i analysen, tid og økonomi. Det viktigste er at hver dommer er testet for sensitivitet for aktuelle produktegenskaper.

Vær oppmerksom på at et stort antall dommere ikke nødvendigvis betyr at presisjonen av bedømmelsen er stor. Dommernes sensoriske terskelverdi og deres trening i produktbedømmelsen er av avgjørende betydning for resultatet.

Inndelingen av skala til bruk i kvalitetskontroll er ofte enten 3-, 5- eller 9 poeng. Nordisk Metodikkomite for næringsmidler (NMKL) viser i prosedyre 16 (2005), Sensorisk kvalitetskontroll en sammenstilling av skalalengder.



Tabell 4.3 Sammenligning, 3-, 5- og 9-poengsskala

Kvalitetskategori	Skalalengde		
	3-poeng	5-poeng	9-poeng
Samsvar med spesifikasjon	3	5	9
			8
			7
Avvik fra spesifikasjon, men handeksvare	2	4	6
			5
		3	4
Vesentlig avvik, ikke handelsvare	1	2	3
			2
		1	1

4.3.3. Kvalitetsindeksmetode, QIM

Ferskhet på fisk er et viktig kriterium ved kjøp og salg av fisk og det er ikke alltid like lett å avgjøre fiskens ferskhet. En metode hvor forskjellige deler av rå hel, sløyet fisk blir vurdert i grad av ferskhet, kalles QIM-metoden, (Quality Index Method).

Metoden er basert på at man systematisk har fulgt utviklingen under tid av hel rå fisk lagret på is og dokumenterer viktige endringer av fiskens ferskhet. Viktige kvalitetsparametere som vurderes er skinn, øyne, gjeller, fiskekjøtt og blod. Det gis poeng fra 0 opptil 3 for hver parameter som vurderes. 0 angir mest fersk og 3 angir minst fersk (dårlig). Maksimumsummen (kvalitetsindeksen) av alle parameterne (24 for oppdrettslaks/*Salmo Salar*) viser en ubrukelig fisk for konsum, mens 0 poeng gis til en helt fersk fisk. Denne metoden er utviklet til bruk for de mest benyttede fiskearter og også annen sjømat. Metoden stiller store krav til dommers kunnskap til produktet og grundig trening i bedømmelse av produktet og bruk av skalaen er viktig. Vi viser til nærmere beskrivelse av metoden i www.qim-eurofish.com.



4.3.4. Tidsavhengige målinger

De mest brukte beskrivende metodene er metoder som måler maksimum intensitet av en egenskap. Det er imidlertid også av interesse å undersøke for eksempel når en smaksopplevelse oppstår og smakens endring i munnen med tanke på varighet og styrke.

Time-Intensity

Med Time-Intensity-metoden (TI) måles intensiteten av en egenskap over lengre tid. Bittersmaken i en sjokolade kan måles i et tidsintervall på for eksempel ett til tre minutter etter at prøven er puttet i munnen eller etter spyting/svelging. Det kan også være viktig å undersøke varigheten av smeltetiden av sjokolade i munnen, eller hvor lenge sjokoladens søt smak beholdes. De sensoriske dommerne skal her sanse intensiteten av en sensorisk stimulus, og samtidig angi intensiteten over et gitt tidsrom.

En utfordring ved TI-metoden er at den er best egnet ved analyse av bare en egenskap om gangen.

Temporal Dominance of Sensation

Temporal Dominance of Sensation (TDS) ble utviklet for å hjelpe på problemene som fulgte med TI-metoden, nemlig at TI bare kunne undersøke en egenskap om gangen. TDS er fortsatt under utvikling og det er ennå ikke etablert en bestemt standard prosedyre for gjennomføring av en analyse med bruk av denne metoden. Både trente og utrente dommeren kan bruke denne metoden. I Kapittel 5, forbrukeranalyser, beskrives denne metoden nærmere.





Referanser

- Amerine M.A., Pangborne R.M., Roessler, (1995). *Principals of Evaluation of Food*, Acad. Press, New York
- Ennis, Daniel M., (1993). The Power of Sensory Discrimination Methods, *Journal of Sensory Studies* 353-370
- ISO 10399, Sensory analysis – Methodology – Duo-trio-test
- ISO 13299, Sensory analysis – General guidance for establishing a sensory profile
- ISO 4120, Sensory analysis - Triangular test
- ISO 5495, Sensory analysis – Methodology – Paired comparison
- ISO 6564, Sensory analysis – Methodology – Flavour profile methods
- ISO 6658, Sensory analysis – Methodology – General guidance
- ISO 8587, Sensory analysis – Methodology – Ranking
- Kroll B.J., (1990). Evaluating rating scales for sensory testing with children, *Food Technology*. Vol. 11, pp. 78-86
- Land D.G., (1983). What is sensory quality? I: Sensory quality in foods and beverages: definition, measurement and control, Editor: Williams & Atkin, Chichester: Ellis Horwood
- Lawless H., Heymann H. (2010). *Sensory Evaluation of Food. Principles and practices*. Springer. ISBN 978-1-4419-6487-8
- Lawless H., Klein B.B., (1991). *Sensory Science Theory and Practice*, Ellis Horwood, Chichester, England
- Lawless H., Malone G., (1986). The discriminative efficiency of common scaling methods, *J. Sensory Studies*, 1, pp.85-98
- Lyon D., Francombe M.A., Hasdell T.A., Lawson K. (1992). *Guidelines for Sensory Analysis in Food Product Development and Quality Control*, Chapman & Hall, London
- Meilgaard M.C., Civile B., Carr T. (1991). *Sensory Evaluation Techniques*, (2 ed.) Boca Raton CRC Press, Inc.
- Munoz A.M., Civile G.V., Carr B.T. (1992). *Sensory Evaluation in Quality Control*, VanNostrandReinhold, New York
- Fortsetter neste side





Stone H., Sidel J.L. (1993). *Sensory Evaluation Practices*, Academic Press, London

www.qim-eurofish@com.

Referanser (Aromahjulet)

Gawel, R., A. Oberholster og I. L. Francis (2000). "A 'Mouth feel Wheel': terminology for communicating the mouth feel characteristics of red wine." *Australian Journal of Grape and Wine Research* 6(3): 203-207.

Koch, I. S., M. Muller, E. Joubert, M. van der Rijst og T. Næs (2012). "Sensory characterization of rooibos tea and the development of a rooibos sensory wheel and lexicon." *Food Research International* 46(1): 217-228.

Larssen, W. E. og M. Hersleth (2014). "Sensorisk kvalitet på marine oljer." *Møreforskning*. MA 14/16 51

Lyon, D. H. og M. P. Watson (1994). "Sensory profiling: a method for describing the sensory characteristics of virgin olive oil." *Grasas y aceites* 45(1-2): 20-25.

Mojet, J. og S. de Jong (1994). "The sensory wheel of virgin olive oil." *Grasas y aceites* 45(1-2): 42-47.

Monteleone, E. og S. A. Langstaff (2014). *Olive Oil Sensory Science*. First edition. Wiley Blackwell.

