

Vurdering af to metoder til lugtanalyse

med særligt fokus på præstationsprøvning

Arne Oxbøl
Marts 2010

Indhold

| | | |
|------|--|----|
| 1. | Indledning | 3 |
| 2. | Konklusion | 4 |
| 3. | Undersøgelsen | 5 |
| 3.1. | Deltagere i undersøgelsen | 5 |
| 3.2. | Formål | 5 |
| 3.3. | Metode | 6 |
| 4. | Præstationsprøvning - beregningsmetoder | 7 |
| 4.1. | Bestemmelse af repeterbarhed og reproducerbarhed | 7 |
| 4.2. | Beregning af z-score for hvert laboratorium og for hvert panel | 8 |
| 5. | Præstationsprøvning - resultater | 9 |
| 6. | Præstationsprøvning - udførte test | 11 |
| 6.1. | Vurdering af repeterbarhed og reproducerbarhed | 11 |
| 6.2. | Sammenligning af metodernes repeterbarhed, reproducerbarhed og nøjagtighed | 13 |
| 6.3. | Beregning af z-score | 16 |
| 7. | Vurdering af metoderne | 17 |
| 7.1. | Beskrivelse af metoderne | 17 |
| 7.2. | Tilkendegivelser om metoderne | 18 |
| 7.3. | Sammenligning af resultater i Olfatec-test | 19 |
| 8. | Diskussion | 20 |
| 8.1. | Generelt | 20 |
| 8.2. | Betydning af brug af flere paneler | 21 |

1. Indledning

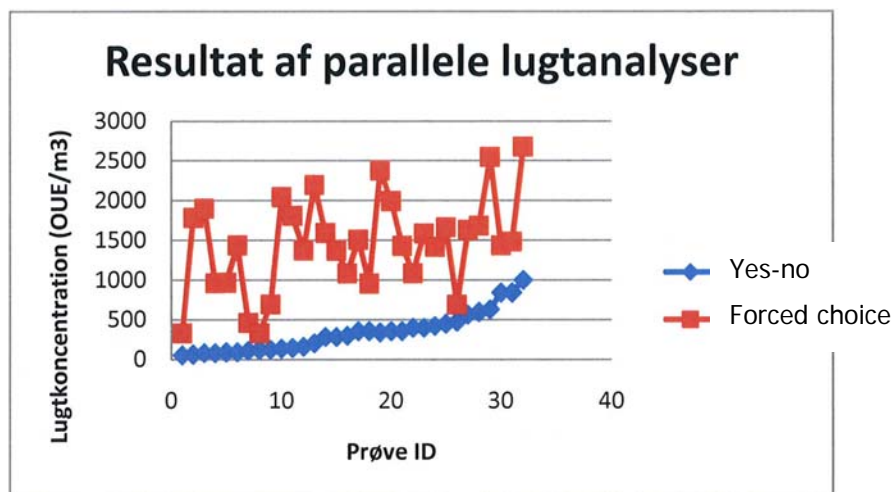
Referencelaboratoriet under den danske Miljøstyrelse har blandt sine opgaver at bidrage til kvaliteten i akkrediterede emissionsmålinger, der udføres af danske målefirmaer. Referencelaboratoriets styregruppe besluttede at udføre en sammenlignende prøvning i 2009 blandt danske laboratorier, der er akkrediteret til måling af lugt i strømmende gas.

Det var tanken, at et antal laboratorier skulle mødes på en virksomhed og hver udtage samtidige prøver af luften fra to afkast med særligt fokus på prøvetagning i våde afkast, hvor fortynding er nødvendig. Prøvningen skulle således vise laboratoriernes formåen inklusive prøvetagning og analyse. Inden prøvningen blev endeligt planlagt blev der imidlertid rejst tvivl om, hvorvidt de to metoder til lugtanalyse (som begge er omfattet af EN 13.725) giver samme resultat. Et mindre antal eksempler antydede, at den ene metode giver tydeligt lavere resultater.

Referencelaboratoriets styregruppe besluttede derfor at lave et lille projekt med det formål at undersøge, om der er generel erfaring for, at den ene metode giver lavere resultater. Hvis det er tilfældet, giver det ikke umiddelbart mening at udføre en prøvning, hvor også prøvetagning indgår. Det blev derfor videre besluttet, at den planlagte præstationsprøvning kun skulle omfatte analyse og, at resultatet skulle indgå i vurderingen af de to metoder.

DANAK har været behjælpelig med at modtage måleresultater fra de enkelte deltagende firmaer, samt at sende dem videre i anonymiseret form til Referencelaboratoriet for beregning.

I forbindelse med vurdering af et system til lugtreduktion har det danske firma, Skov ventilation, fået analyseret parallelprøver hos et dansk laboratorium, der anvender forced choice metoden, og hos et tysk laboratorium, der bruger yes-no metoden. Yes-no resultaternes størrelse var i nogle tilfælde kun ca. 25% af forced choice resultaterne, og samtlige yes-no resultater var lavere. Der er således en systematisk afvigelse. Samtlige de opnåede resultater er vist i figur 1.



Figur 1. Sammenhørende lugtresultater for to metoder¹

I 2009 gik ét af de danske laboratorier fra forced choice til yes-no og foretog i den forbindelse en række sammenlignende tests, hvor størrelsen af yes-no metodens resultater var ca. 30% af resultaterne med forced choice metoden. Der blev ved denne test anvendt samme gruppe af panelister i begge metoder.

¹ Brev fra SKOV A/S til Referencelaboratoriet

Disse to observationer skabte en del uro både hos lugtlaboratorier og hos deres kunder. Muligheden for, at man kan opnå lavere resultater ved at vælge laboratorium efter metode, kan både give anledningen til konkurrenceforvridning og til, at lugtresultaterne bruges uhensigtsmæssigt. For lave lugtresultater kan måske dokumentere overholdelse af grænseværdier, hvor der reelt er problemer. Tilsvarende kan for høje resultater give anledning til etablering af rensningsforanstaltninger, hvor det reelt ikke er nødvendigt.

Miljøstyrelsen vurderer derfor, at det er meget vigtigt at søge dokumentation for evt. forskelle mellem metoderne for at blive i stand til at give de rette anbefalinger i en kommende lugtvejledning.

For at skabe et bedre vurderingsgrundlag i prøvningen er et norsk og et svensk laboratorium inviteret til at supplere de tre danske laboratorier.

2. Konklusion

Der er to formål med undersøgelsen:

- Gennem en præstationsprøvning at undersøge fem laboratoriers formåen inden for analyse af tilsendte lugtprøver
- At vurdere, om yes-no metoden og forced choice metoden giver signifikant forskellige resultater

De fem deltagende laboratorier præsterer ens inden for rammerne af lugtanalysers store variation. Ingen laboratoriers enkeltresultater skiller sig afgørende ud fra gennemsnittet for hver enkelt af fem prøver.

Der er ingen signifikant forskel på gennemsnitsresultaterne for de to metoder. Der er mindre forskelle, og i to tilfælde er gennemsnittet for forced choice højest, i tre tilfælde er gennemsnittet for yes-no højest.

De fem laboratoriers gennemsnitlige lugttærskel for en prøve med n-butanol er meget fint i overensstemmelse med den forventede værdi, og kun to af 19 enkeltresultater ligger lige uden for det accepterede interval.

Uanset hvilken metode, der anvendes, kan gentagelser udføres lige godt (ens reproducerbarhed mellem laboratorier og paneler).

Undersøgelsen viser, at det er fordelagtigt at analysere lugtprøver på så mange paneler som muligt i stedet for at gentage analyser på samme panel.

I afsnit 8 ses en uddybende diskussion af resultater og beregninger.

3. Undersøgelsen

3.1. Deltagere i undersøgelsen

De fem deltagende laboratorier er:

DMRI - Danish Meat Research Institute
En division i Teknologisk Institut
Lugtlaboratoriet
Maglegaardsvej 2
4000 Roskilde
Danmark

ÅF-kontroll
Kvarnbergsgatan 2
411 05 Göteborg
Sverige

Molab as
Kjelsåsveien 174
0884 Oslo
Norge

Eurofins A/S
Smedeskovvej 38
8464 Galten
Danmark

FORCE Technology
Park Allé 345
2605 Brøndby
Danmark

3.2. Formål

Der er to formål med undersøgelsen:

- Gennem en præstationsprøvning at undersøge fem laboratoriers formåen inden for analyse af tilsendte lugtprøver
- At vurdere, om yes-no metoden og forced choice metoden giver signifikant forskellige resultater

Det primære formål med prøvningen er at vurdere lugtlaboratoriernes formåen og – om muligt - at give laboratorierne et grundlag for evt. forbedringer af analyserne. Da den "sande værdi" ikke er kendt, kan præstationsprøvningen kun vanskeligt benyttes til at fastlægge, at en eventuel fejl ligger hos et bestemt laboratorium. Det samlede resultat og en vurdering af enkeltresultater i forhold til gennemsnittet kan benyttes af hvert enkelt laboratorium til en bedømmelse af egen præstation.

Hvis resultaterne giver inspiration til forbedringer eller relevante korrigerende handlinger på eventuelle afvigelser er det op til det laboratorium at iværksætte disse.

Resultaterne giver mulighed for at vurdere, om der er systematiske forskelle mellem de to analyseprincipper hvad angår nøjagtighed og kvalitet i gentagelser.

Er der begået fejl, som først opdages efter, at resultaterne er sendt ind til DANAK, eller er der andres særlige forhold, har laboratorierne fået lov til at informere herom/rette, således at der kan udarbejdes en rapport kun med "fejlfri" resultater. Sådanne justeringer fremgår af den endelige rapport. På den måde vil eksempelvis en regnefejl eller en enhedsfejl ikke skygge over det samlede resultat, som udelukkende bør afspejle laboratoriernes analyseevne. Samtidig vil eventuelle konklusioner vedr. de to metoder ikke vil blive forplumret af fejl, der kan rettes.

3.3. Metode

Præstationsprøvning

Præstationsprøvningen er gennemført som følger:

FORCE Technology udtog én stor luftprøve i en Tedlarpose fra hver af to forskellige industrier: tobaksfremstilling og asfaltfremstilling. Fra hver af disse store prøver blev en del af luften fordelt til mindre poser af Nalophan. Der blev fremstillet så mange mindre poser af hver prøve, at hvert laboratorium kunne modtage én eller to poser (afhængigt af hvert laboratoriums ønsker).

Den resterende mængde luft i hver stor pose blev fortyndet med nitrogen til en væsentligt lavere koncentration, hvorefter denne fortyndede luft blev fordelt i mindre poser af Nalophan.

En femte prøve bestod af n-butanol fordampet i nitrogen i en stor pose af Nalophan. Efter tilstrækkelig opblanding blev også denne luft fordelt i mindre poser af Nalophan.

Luftprøverne blev udtaget og prøveposerne fremstillet tirsdag 10. november og sendt til laboratorierne samme dag, således at alle laboratorierne skulle have prøverne senest torsdag 12. november kl. 9. Der er derfor gået mere end 30 timer mellem prøvefremstilling og analyse, hvilket normalt ikke er acceptabelt i henhold til lugtstandarden². I denne prøvningsammenhæng er det imidlertid kun vigtigt, at alle laboratorier får ens prøver og analyserer dem nogenlunde samtidig. Dermed udelukkes tidsfaktorens indflydelse på resultatet.

Gennem udtagning og analyse af prøver fra de to industrier i ugen inden testen tilstræbtes det at lave lugtkoncentrationer i følgende størrelsesorden, tabel 1.

Tabel 1. Oversigt over vejledende prøvekoncentrationer

| Prøve | <5.000 OUE/m ³ | 3.000 – 12.000 OUE/m ³ | >15.000 OUE/m ³ |
|---------------|---------------------------|-----------------------------------|----------------------------|
| 1 – Asfalt | | | X |
| 2 – Tobak | | | X |
| 3 – n-butanol | | X | |
| 4 – Asfalt | X | | |
| 5 - Tobak | X | | |

Lugtkoncentrationen i asfaltfremstillingen var på prøvedagen meget højere end det erfaringsmæssigt forventede, og derfor blev koncentrationerne i prøvningsposerne væsentligt højere end tilstræbt.

Laboratorierne analyserede hver prøve to gange på hvert af to lugtpaneller (formiddag og eftermiddag).

² DS/EN 13.725 Air quality – Determination of odour concentration by dynamic olfactometry, Januar 2006

For at minimere evt. systematiske variationer mellem f.eks. første og sidste prøve og mellem forskellig rækkefølge på laboratorierne blev laboratorierne bedt om at analysere prøverne i den rækkefølge, som angives i tabel 2.

Tabel 2. Analyserækkefølge

| | |
|-------------|---------------------------------------|
| Formiddag | 1 – 2 – 3 – 4 – 5 – 3 – 5 – 1 – 4 – 2 |
| Eftermiddag | 5 – 2 – 4 – 1 – 3 – 2 – 5 – 4 – 3 – 1 |

Laboratorierne sendte derefter resultaterne til DANAK, som sendte dem videre til Referencelaboratoriet i anonymiseret form. Her blev resultaterne bearbejdet med statistiske metoder i henhold til ISO 5725-2³, ISO 5725-6⁴ og ISO 43-1⁵. Alle beregninger sker på basis af logaritmerede resultater, idet lugtopfattelsen er logaritmisk.

Vurdering af de to analysemetoder

De to analysemetoder beskrives kort med fokus på forskelle i principperne og svartider for panelisterne.

Metoderne er vurderet på baggrund af nærværende præstationsprøvnings resultater og resultater fra en tysk præstationsprøvning fra 2005⁶. Den tyske test er udført ved fremsendelse af fire forskellige lugtprøver i trykflasker til en række laboratorier i Europa. Rapporten har angivet, hvilke olfaktometre de forskellige laboratorier har anvendt.

Endelig har vi søgt kommentarer til metoderne hos ledende lugtlaboratorier i Holland og Tyskland.

4. Præstationsprøvning - beregningsmetoder

Statistikken bag metoderne beskrives ikke i detaljer, da det ligger uden for denne rapport's formål. Læserne må – efter behov – orientere sig i de bagvedliggende formelapparater i de anvendte dokumenter.

I den aktuelle udførelse af testen er den ene metode blevet anvendt af to laboratorier og den anden af tre andre laboratorier. I det ideelle testdesign skulle begge metoder anvendes af samme laboratorier for derved at eliminere evt. forskelle mellem laboratorier, som ikke stammer fra metodevalget.

Resultaterne kan derfor ikke bruges til éntydigt at "dømme" metoderne, men påvises en signifikant forskel mellem yes-no- og forced choice laboratorier, er det en god indikation af metodeforskel. Resultaterne giver under alle omstændigheder brugerne af olfaktometri mulighed for at vurdere den kvalitet, de får ved valg af det ene eller det andet laboratorium, som anvender den ene eller den anden metode.

4.1. Bestemmelse af repeterbarhed og reproducerbarhed

Resultater og varianser testes først ved hjælp af Grubb's test hhv. Cochran's test for stragglers⁷ og outliers⁸. Ved hjælp af metoderne i ISO 5725-2 beregnes repeterbarhed og reproducerbarhed i lugt-

³ ISO 5725-2: Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method, 1994

⁴ ISO 5725-2: Accuracy (trueness and precision) of measurement methods and results – Part 6: Use in practice of accuracy values, 1994

⁵ ISO 43-1: ISO 43-1 Proficiency testing by interlaboratory comparisons. Part 1: Development and operation of proficiency testing schemes

⁶ Inspection report: Olfactometrical Interlaboratory Test 2005 – RV Standard, OLFAtec GmbH, 2005

⁷ En straggler er en værdi, som med 95% sandsynlighed ikke tilhører samme fordeling som de øvrige værdier. Den er dermed mindre usikker end en outlier.

analyse, dels for alle laboratorier samlet, dels for paneler og endelig for hver metode. De beregnede størrelser er

s_{rj}^2 : repeterbarhedsvariansen for prøve j
 s_{Lj}^2 : variansen mellem laboratorier for prøve j
 s_{Rj}^2 : reproducerbarhedsvariansen for prøve j
 m_j : middelværdien for prøve j

Ved hjælp af metoderne i ISO 5725-6 sammenlignes metodernes repeterbarhed, reproducerbarhed og selve resultatet. De beregnede størrelser er:

F_r er en teststørrelse for, om de to metoder har forskellig repeterbarhed (variation indenfor det enkelte laboratorium). Den beregnes ved at dividere de to repeterbarhedsvarianser med hinanden. Størrelsen testes tosidet på 5% niveau ved at sammenligne med tabellerede værdier for F på hhv. 97,5% og 2,5% niveau.

F_R er en teststørrelse for, om metoderne har forskellig total variation (reproducerbarhed) – dvs. både indenfor laboratorier og mellem laboratorier. Størrelsen testes på samme måde som F_r .

Bemærk, at der i rapporten både anvendes begreberne spredning og varians, hvor spredning er kvadratroden af varians.

Til test af metodernes gennemsnitlige resultater (nøjagtighed) beregnes

$$\frac{ABS(Y_A - Y_B)}{S_{fælles}}$$

Hvor $ABS(Y_A - Y_B)$ er forskellen mellem de beregnede gennemsnitsværdier (altid positiv)
 $S_{fælles}$ er en poollet spredning på prøveniveau.

Hvis den derved opnåede teststørrelse er mindre end 2, er der ikke signifikant forskel på de opnåede gennemsnitsværdier.

4.2. Beregning af z-score for hvert laboratorium og for hvert panel

z-scorer beregnes for resultater, hvor den nominelle værdi ikke kendes (f.eks. målinger i skorstene og ventilationskanaler).

z-scoren beregnes for hvert enkelt måleresultat ud fra følgende formel⁵:

$$z = \frac{|x_i - X|}{s}, \text{ hvor}$$

- x_i er laboratoriets resultat (gennemsnit af alle værdier og gennemsnit for hvert panel)
- X er gennemsnittet af alle laboratoriers værdier (ekskl. outliers)
- s er spredningen på måleværdierne (1xRSD) eller en på forhånd fastsat spredning⁵

⁵ En outlier er en værdi, som med 99% sandsynlighed ikke tilhører samme fordeling som de øvrige værdier.

Det er i beregningerne forudsat, at $(x_i - X)$ er normalfordelt omkring 0. På baggrund heraf angiver ISO 43-1 følgende vurderingskriterier for resultaterne:

| Z | Vurdering |
|---------------|------------------------|
| $ z \leq 2$ | Tilfredsstillende |
| $2 < z < 3$ | tvivlsom |
| $ z \geq 3$ | ikke tilfredsstillende |

Denne testmetode er almindeligt anvendt ved de præstationsprøvninger, som laves på foranledning af Miljøstyrelsen og DANAK. Typisk deltager 4-6 laboratorier og gennemfører 4-5 måleserier. Inden for hver måleserie vurderes ved beregning af z-scoren, om et eller flere laboratorier afviger fra gennemsnittet. Der er således ikke en kendt "sand værdi", men gennemsnittet betragtes som den sande værdi.

Beregningerne er gennemført for laboratoriernes gennemsnitsresultater og for de enkelte panelers gennemsnitsresultater. Hvor et panel kun har analyseret prøven én gang er panelets resultat ikke medtaget.

5. Præstationsprøvning - resultater

På trods af den tidlige fremsendelse af prøveposerne nåede prøverne til laboratorium 4 først frem midt på formiddagen torsdag den 12. november. Laboratoriet kunne derfor kun nå én analyse af hver af de fire prøver på formiddagspanelet. Kun én prøve blev analyseret to gange. Laboratoriet har en øvre analysegrænse på 64.000 OU/m³, hvorfor den overraskende høje koncentration i prøve 1 skulle fortyndes. Dette nåede laboratoriet pga. forsinkelsen ikke for panelet om formiddagen, og der er derfor intet resultat for denne prøve om formiddagen.

Prøve 5 til laboratorium 3 var gået i stykker ved modtagelsen, hvorfor analyse af denne var umulig.

Beregningsformlerne i 5725-2 kan håndtere forskellige antal resultater for de enkelte laboratorier og prøver. Derfor er det muligt at kassere enkelte resultater, hvis de vurderes at være stragglers⁷ eller outliers⁸.

Laboratorium 2 blev overrasket over den høje koncentration i prøve 1 og har anført, at det afvigende resultat for første analyse sandsynligvis også er påvirket af fortyndingen.

Laboratorium 5 oplyste, at samtlige prøveposer tilsyneladende var utætte ved stutsen og, at utætheden i et vist omfang måske kontaminerede prøvningslokalerne trods grundig udluftning. Man tager derfor forbehold for, om utæthederne kan påvirke resultatet.

For at kunne belyse evt. forskelle blev laboratorierne bedt om at oplyse visse detaljer om forholdene på analysedagen (metodedetaljer). Oplysningerne ses i nedenstående tabel 3.

Tabel 3. Oversigt over laboratoriernes metodedetaljer

| Laboratorienummer | 1 | 2 | 3 | 4 | 5 |
|---------------------------------|---------------|------------|-------------|--------|---------------|
| Analysemetode | Forced choice | yes-no | yes-no | yes-no | Forced choice |
| Antal panelister i hvert panel | 6 | 4 | 4 | 4 | 6 |
| Antal gengangere i panel 2 | 0 | 1 | 3 | 1 | 1 |
| Responstid (sekunder) | 15 | 4,4 | Ikke oplyst | 2,2* | 15 |
| Usikkerhed (% interval/faktor) | -48 - +110 | -59 - +142 | $\pm 1,9$ | 1,8 | 1,5 |
| Enhed for usikkerhed | % | % | dB (ouE) | faktor | faktor |
| Sikkerhedsfaktor for usikkerhed | 2 | 2 | 2 | 2** | 1 |

* Panelisterne har 2,2 sekunder til at trykke svar, efter at man har sniffet ind og ud

** Laboratoriet har angivet at bruge reproducerbarheden for butanoltest og at der ikke er angivet k-værdi i denne formel. I henhold til standarden er der tale om $k = 2$.

Samtlige laboratoriers resultater angivet som OU_E/m^3 ses i tabel 4. I denne og alle følgende tabeller er resultaterne angivet uden afrunding. Antallet af betydende cifre må ikke tages som udtryk for nøjagtigheden i analyse og beregninger, men for at undgå fejl i forbindelse med afrundinger er de regnet med alle cifre. For nemheds skyld og for at undgå overførselsfejl er alle tabeller kopieret ind fra EXCEL med deraf følgende mange cifre.

Tabel 4. Resultater fra alle laboratorier (OU_E/m^3)

| Laboratorium | Panel | | Prøve 1 | Prøve 2 | Prøve 3 | Prøve 4 | Prøve 5 |
|---------------|-------------|-----------|----------------|---------|---------|---------|---------|
| Lab 1 | Formiddag | Analyse 1 | 121.123 | 16.762 | 5.358 | 29.085 | 3.173 |
| Forced choice | | Analyse 2 | 131.121 | 15.286 | 5.044 | 37.431 | 2.484 |
| | Eftermiddag | Analyse 1 | 77.666 | 9.969 | 3.004 | 13.282 | 2.131 |
| | | Analyse 2 | 85.435 | 8.491 | 3.847 | 18.490 | 2.065 |
| Lab 5 | Formiddag | Analyse 1 | 50.119 | 32.236 | 7.442 | 19.424 | 4.592 |
| Forced choice | | Analyse 2 | 49.927 | 16.533 | 4.536 | 31.867 | 3.388 |
| | Eftermiddag | Analyse 1 | 54.013 | 18.836 | 13.243 | 10.666 | 3.581 |
| | | Analyse 2 | 59.704 | 19.387 | 8.128 | 22.388 | 8.128 |
| Lab 2 | Formiddag | Analyse 1 | 102.400 | 13.777 | 8.192 | 36.491 | 3.251 |
| Yes-no | | Analyse 2 | 64.508 | 16.384 | 6.137 | 32.510 | 2.435 |
| | Eftermiddag | Analyse 1 | 30.444 | 11.585 | 8.679 | 32.510 | 3.444 |
| | | Analyse 2 | 40.637 | 11.585 | 7.298 | 24.355 | 5.161 |
| Lab 3 | Formiddag | Analyse 1 | 275.399 | 5.623 | 6.761 | 47.863 | |
| Yes-no | | Analyse 2 | 295.121 | 7.943 | 6.026 | 47.863 | |
| | Eftermiddag | Analyse 1 | 218.776 | 7.079 | 7.127 | 36.308 | |
| | | Analyse 2 | 231.582 | 7.943 | 8.476 | 40.738 | |
| Lab 4 | Formiddag | Analyse 1 | | 8.000 | 5.187 | 29.344 | 1.414 |
| Yes-no | | Analyse 2 | | | | | 1.834 |
| | Eftermiddag | Analyse 1 | 155.621 | 8.000 | 6.727 | 19.027 | 1.834 |
| | | Analyse 2 | 142.705 | 7.336 | 5.339 | 22.627 | 1.834 |

Fed: outlier i henhold til Cochran's test.

Ved Cochran's test af varianserne for hvert laboratorium er spredningen for laboratorium 2, prøve 1, dømt som outlier. Det er umiddelbart indlysende, at årsagen skal søges i det resultat, som efter laboratoriets egen opfattelse er unødvendigt påvirket af fortyndingen. Set i lyset af, at Referencelaboratoriet "lovede" en koncentration $>15.000 OU/m^3$ (underforstået i denne størrelsesorden), vælger vi at fjerne dette ene resultat og ikke alle laboratoriets resultater for denne prøve. Det vil svække undersøgelsen at fjerne de tre andre nogenlunde ens resultater for denne prøve, som ikke angives at være fejlbehæftet.

I alle de udførte tests er denne ene værdi elimineret. Tabel 5 viser gennemsnit for de to metoder efter fjernelse af resultatet med usikker fortynding.

Tabel 5. Gennemsnitsværdier for de to metoder

| | | | | | |
|--------------------------|---------|--------|-------|--------|-------|
| Gennemsnit forced choice | 73.485 | 15.965 | 5.713 | 21.115 | 3.346 |
| Gennemsnit yes-no | 124.713 | 9.107 | 6.810 | 32.321 | 2.433 |
| Gennemsnit alle | 97.233 | 11.535 | 6.324 | 27.017 | 2.853 |

Det ses, at yes-no metoden giver de højeste resultater for tre af prøverne, mens forced choice metoden giver de højeste resultater for to af prøverne.

Prøve 3 er n-butanol opløst i nitrogen med en omtrentlig analyse på 245 ppm. Den accepterede lugt-tærskelværdi for n-butanol er 40 ppb (0,040 ppm). Det betyder, at gennemsnitsværdien for prøve 3 skal være ca. 6.125 OU/m³. Det er opfyldt. Yes-no laboratorierne ligger i gennemsnit 11% over den forventede gennemsnitsværdi, mens forced choice laboratorierne i gennemsnit ligger 7% under.

Acceptintervallet (95% konfidensinterval) for prøven er en faktor to til hver side², dvs. værdier mellem 3.068 OU/m³ og 12.272 OU/m³ er tilfredsstillende. Én værdi er lavere end nedre grænse (3.004 OU/m³ fra forced choice laboratorium 1), mens én værdi er højere end øvre grænse (13.243 OU/m³ fra forced choice laboratorium 5). De tre yes-no laboratorier varierer mellem 5.187 OU/m³ og 8.679 OU/m³.

Med forbehold for den kun omtrentlige analyse konkluderes det, at de fem laboratorier som helhed ligger godt i forhold til acceptintervallet. Umiddelbart præsterer yes-no laboratorierne lidt bedre end forced choice.

6. Præstationsprøvning - udførte test

6.1. Vurdering af repeterbarhed og reproducerbarhed

Den første beregning omfatter samtlige laboratoriers resultater uden hensyn til hvilken metode, der anvendes. Testen giver et mål for lugtanalysers variation inden for laboratorier, mellem laboratorier og totalt.

Tabel 6 viser for hver metode værdier på prøveniveau for spredning inden for laboratorier, mellem laboratorier og totalt. For hver metode er spredningen inden for laboratorier (repeterbarhed) således beregnet som en poollet værdi for de (to eller tre) laboratorier, der bruger metoden. Værdierne viser, hvordan de respektive laboratorier som gruppe præsterer.

Table 6. Oversigt over spredninger i resultaterne – metodeopdelt (logaritmisk)

| | | | | | | |
|----------|---------------|-------|-------|--------|-------|-------|
| Middel | Forced choice | 4,87 | 4,20 | 3,76 | 4,32 | 3,52 |
| | Yes-no | 5,10 | 3,96 | 3,83 | 4,51 | 3,39 |
| S_{Rj} | Forced choice | 0,083 | 0,135 | 0,158 | 0,199 | 0,137 |
| | Yes-no | 0,105 | 0,063 | 0,064 | 0,075 | 0,103 |
| S_{Lj} | Forced choice | 0,193 | 0,154 | 0,1711 | 0,000 | 0,185 |
| | Yes-no | 0,418 | 0,147 | 0,049 | 0,124 | 0,207 |
| S_{Rj} | Forced choice | 0,210 | 0,205 | 0,233 | 0,199 | 0,230 |
| | Yes-no | 0,431 | 0,160 | 0,080 | 0,145 | 0,231 |

Middel: middelværdien for prøve j

S_{Rj} : repeterbarhedsspredning for prøve j

S_{Lj} : spredning mellem laboratorier for prøve j

S_{Rj} : reproducerbarhedsspredning for prøve j

Kun én poollet repeterbarhedsspredning er større end den i standarden² højest accepterede spredning på 0,172. Standarden stiller kun krav til det enkelte laboratorium, og tabellen viser, hvordan repeterbarheden er for to (forced choice) hhv. tre laboratoriers (yes-no) resultater. Det vurderes derfor, at laboratorierne som grupper har en god repeterbarhed.

Spredningen mellem laboratorier, S_{Lj} , er for én prøve meget stor (0,418). Omsat til usikkerhedsangivelse betyder det, at et resultat fra et vilkårligt laboratorium er en del af et interval, hvis bredde er 4,2 gange til hver side for gennemsnittet af samtlige resultater (ca. 18 gange mellem laveste og højste). Det er formodentlig uacceptabelt for brugerne af lugtanalyser, selv om standarden ikke umiddelbart stiller krav om forskelle mellem laboratorier. Den aktuelle prøve er den meget stærke prøve, som laboratorierne havde forventet var meget svagere, og heri ligger formodentlig en del af forklaringen. En anden forklaring kunne være, at laboratorium 3 (som har de klart højeste resultater) har regnet forkert ved forforyndingen.

Den totale spredning på resultaterne, reproducerbarheden S_{Rj} , er uacceptabelt høj for prøve 1, men for de øvrige er den højest 0,233 svarende til en intervalbredde på 2,9. Det er noget højere end den intervalbredde, der svarer til standardens acceptgrænse (2,2) for ét laboratoriums repeterbarhed, men der er som nævnt ingen krav til forskelle mellem laboratorier.

Table 7 viser værdier på prøveniveau for spredning inden for laboratorier, mellem laboratorier og totalt uden at skelne mellem metoder.

Table 7. Oversigt over spredninger i resultaterne – uden metodeopdeling (logaritmisk)

| | | | | | | |
|----------|-------|-------|-------|--------|-------|-------|
| Middel | Begge | 4,99 | 4,06 | 3,80 | 4,43 | 3,46 |
| S_{Rj} | Begge | 0,095 | 0,101 | 0,114 | 0,142 | 0,121 |
| S_{Lj} | Begge | 0,320 | 0,184 | 0,0970 | 0,117 | 0,176 |
| S_{Rj} | Begge | 0,334 | 0,210 | 0,150 | 0,184 | 0,213 |

Repeaterbarheden, som er baseret på 16 til 19 resultater på de enkelte prøveniveauer, er rigtig god.

Reproducerbarheden for prøveniveau 1 er ikke god, mens den for de øvrige fire niveauer vurderes at være god. Det må forventes, at reproducerbarheden for en gruppe laboratorier er dårligere end repeterbarheden for det enkelte laboratorium. En reproducerbarhedsspredning på 0,213 (prøveniveau 5) er således kun lidt større end standardens krav til repeterbarhedsspredning for ét laboratorium.

Det dårlige resultat for prøveniveau 1 kan ikke forklares med kendte fejl eller outliers. De gode resultater for de øvrige niveauer gør det imidlertid nærliggende at tro på, at der faktisk er en fejl for ét laboratorium på prøveniveau 1. Det kunne f.eks. være en fortyndingsfejl hos laboratorium 3.

6.2. Sammenligning af metodernes repeterbarhed, reproducerbarhed og nøjagtighed

Metodernes repeterbarhed og reproducerbarhed testes ved en F-test på varianserne for hver metode.

Teststørrelsen F_r beregnes ved at dividere de to metoders repeterbarhedsvarianser med hinanden. Størrelsen testes tosidet på 5% niveau ved at sammenligne teststørrelsen med tabellerede værdier for F på hhv. 97,5% og 2,5% niveau. Hvis teststørrelsen er mindre end tabelværdien på 2,5% niveau, har den ene metode signifikant mindre varians (bedre repeterbarhed). Er teststørrelsen større end tabelværdien på 97,5% niveau, har den anden metode signifikant bedre repeterbarhed.

På samme måde testes reproducerbarheden gennem beregning af F_R .

Både repeterbarhed og reproducerbarhed testes først på prøveniveau. Da ingen varianser for de to metoder er outliers eller stragglers (altså i princippet samme varians for hver prøve), kan der beregnes poolede varianser for hver metode og derefter F-teststørrelser for de poolede varianser. Test af disse F-størrelser er stærkere end på prøveniveau, da antallet af frihedsgrader er væsentligt større.

Til brug for testen af, om de to metoder giver samme gennemsnitlige resultat (nøjagtighed), beregnes en poollet spredning (fælles spredning for de to metoder) på prøveniveau. Det vil sige, at for hver prøve beregnes et gennemsnit af spredningerne for de to metoder. Forskellen mellem gennemsnitsresultaterne (altid positiv) for de to metoder beregnes og divideres med den fælles spredning. Hvis den derved opnåede teststørrelse er mindre end 2, er der ikke signifikant forskel på de opnåede gennemsnitsværdier.

Nøjagtighed kan kun testes på prøveniveau, da de enkelte prøver har meget forskellige styrker.

Test på laboratorieniveau

I den første test betragtes alle resultater fra hvert laboratorium ($n = 4$ for de fleste laboratorier/prøver og $n = 3$ eller $n = 2$ for resten).

Tabel 8 viser testen af de to metoder. Metode 1 er forced choice, og metode 2 er yes-no. Tabellen viser beregnede teststørrelser og frihedsgrader og tabelopslag af hensyn til særligt interesserede læsere.

De væsentligste oplysninger er fremhævet med gul baggrund.

Tabel 8. Testresultater for forced choice og yes-no metoderne (laboratorier)

| Testparametre | Niveau (prøve) | | | | | Poollet |
|-------------------------------|----------------|---------|---------|---------|---------|---------|
| | Prøve 1 | Prøve 2 | Prøve 3 | Prøve 4 | Prøve 5 | |
| $S^2_{r,met 1}$ | 0,00689 | 0,0183 | 0,0249 | 0,0398 | 0,0187 | 0,0217 |
| $S^2_{r,met 2}$ | 0,0110 | 0,00394 | 0,00403 | 0,00563 | 0,0106 | 0,0065 |
| F_r (met 2/met 1) | 1,60 | 0,215 | 0,162 | 0,142 | 0,567 | 0,298 |
| $v_{met 1}$ | 6 | 6 | 6 | 6 | 6 | 30 |
| $v_{met 2}$ | 6 | 9 | 9 | 9 | 6 | 39 |
| $F_{0,975} (v_2,v_1)$ | 5,82 | 5,52 | 5,52 | 5,52 | 5,82 | 2,01 |
| $F_{0,025} (v_2,v_1)$ | 0,17 | 0,23 | 0,23 | 0,23 | 0,17 | 0,52 |
| Samme repeterbarhed? | Ja | Nej | Nej | Nej | Ja | Nej |
| $S^2_{R,met 1}$ | 0,0441 | 0,0421 | 0,0541 | 0,0398 | 0,0530 | 0,0466 |
| $S^2_{R,met 2}$ | 0,186 | 0,026 | 0,006 | 0,021 | 0,054 | 0,0590 |
| F_R (met 2/met 1) | 4,58 | 0,799 | 0,096 | 1,678 | 1,171 | 1,785 |
| $v_{met 1}$ | 1 | 1 | 1 | 1 | 1 | 5 |
| $v_{met 2}$ | 2 | 2 | 2 | 2 | 1 | 9 |
| $F_{0,975} (v_2,v_1)$ | 800 | 800 | 800 | 800 | 648 | 6,68 |
| $F_{0,025} (v_2,v_1)$ | 0,02597 | 0,02597 | 0,02597 | 0,02597 | 0,00154 | 0,22321 |
| Samme reproducerbarhed? | Ja | Ja | Ja | Ja | Ja | Ja |
| $y_{met 1}$ | 4,87 | 4,20 | 3,76 | 4,32 | 3,52 | - |
| $y_{met 2}$ | 5,10 | 3,96 | 3,83 | 4,51 | 3,39 | - |
| $s_{fælles}$ | 0,281 | 0,147 | 0,137 | 0,103 | 0,206 | - |
| $ABS(y_A - y_B) / s_{fælles}$ | 0,818 | 1,65 | 0,555 | 1,80 | 0,673 | - |
| Samme middelværdi? | Ja | Ja | Ja | Ja | Ja | - |

De to metoder har i den aktuelle test vist signifikant forskellig repeterbarhed for laboratoriernes resultater for tre prøveniveauer.

Det spiller imidlertid ind, at yes-no laboratorium 4 kun leverede to resultater for én prøve og tre resultater for tre prøver. Hvor der kun er to resultater (ét panel) påvirkes disse overhovedet ikke af variationen mellem forskellige panelister, som vurderes at være den væsentligste delkomponent i usikkerhed på lugtmålinger. Hvor der kun er ét resultat på det ene panel, har variationen mellem forskellige panelister relativt mindre betydning.

Det spiller ligeledes ind, at yes-no laboratorium 3 havde tre gengangere blandt fire panelister for de to paneler. Variationen mellem panelister er derfor af væsentlig mindre betydning. De to øvrige yes-no laboratorier havde én genganger blandt fire panelister.

Det ene forced choice laboratorium havde én genganger blandt seks panelister, mens det andet ingen havde blandt seks panelister. Resultaterne for disse laboratorier repræsenterer derfor 11 hhv. 12 panelisters varierende lugtesans, mens yes-no laboratorierne har brugt fem til syv panelister.

Testens resultat er selvfølgelig gældende, idet der er den konstaterede, signifikante forskel. Forskellen kan bare ikke éntydigt (om overhovedet) relateres til metoden.

Testen af reproducerbarhed er mere éntydig, fordi en væsentlig komponent i variationen her er forskellen mellem laboratorier (herunder især metode). Testen viser på alle prøveniveauer og for den poolede reproducerbarhed, at der ikke er signifikant forskel mellem metoderne. Forced choice har størst variation på tre prøveniveauer, og yes-no på to prøveniveauer og størst poollet variation.

Testen viser, at der ikke på noget prøveniveau er signifikant forskel på gennemsnitsresultaterne af de to metoder. Yes-no giver højest gennemsnit på tre prøveniveauer og force-choice på to.

Test på panelniveau

I denne test betragtes resultater fra hvert panel ($n = 2$ for alle paneler som indgår). Hvor et panel kun har ét resultat udgår panelet af testen. Det betyder, at yes-no laboratorium 2 ikke indgår for prøve 1 om formiddagen, mens yes-no laboratorium 4 ikke indgår for prøverne 1 til 4 om formiddagen. Yes-no laboratorium 3 indgår ikke for prøve 5 på nogen af panelerne.

Tabel 9 viser testen af de to metoder. De væsentligste oplysninger er fremhævet med gul baggrund.

Tabel 9. Testresultater for forced choice og yes-no metoderne (paneler)

| Testparametre | Niveau (prøve) | | | | | Poolet |
|------------------------------------|----------------|---------|---------|---------|---------|---------|
| | Prøve 1 | Prøve 2 | Prøve 3 | Prøve 4 | Prøve 5 | |
| $S^2_{r,met 1}$ | 0,00060 | 0,0113 | 0,0129 | 0,0228 | 0,0195 | 0,0134 |
| $S^2_{r,met 2}$ | 0,0023 | 0,00321 | 0,00396 | 0,00264 | 0,0074 | 0,0038 |
| F_r (met 2/met 1) | 3,89 | 0,283 | 0,307 | 0,116 | 0,381 | 0,285 |
| $^{\circ}_{met 1}$ | 4 | 4 | 4 | 4 | 4 | 20 |
| $^{\circ}_{met 2}$ | 4 | 5 | 5 | 5 | 4 | 23 |
| $F_{0,975} (^{\circ}_{\text{NS}})$ | 9,60 | 9,36 | 9,36 | 9,36 | 9,60 | 2,42 |
| $F_{0,025} (^{\circ}_{\text{NS}})$ | 0,10 | 0,14 | 0,14 | 0,14 | 0,10 | 0,42 |
| Samme repeterbarhed? | Ja | Ja | Ja | Nej | Ja | Nej |
| $S^2_{R,met 1}$ | 0,0327 | 0,0354 | 0,0464 | 0,0373 | 0,0414 | 0,0386 |
| $S^2_{R,met 2}$ | 0,167 | 0,024 | 0,005 | 0,020 | 0,040 | 0,0454 |
| F_R (met 2/met 1) | 5,12 | 0,748 | 0,071 | 0,734 | 1,138 | 1,361 |
| $^{\circ}_{met 1}$ | 3 | 3 | 3 | 3 | 3 | 15 |
| $^{\circ}_{met 2}$ | 3 | 4 | 4 | 4 | 3 | 18 |
| $F_{0,975} (^{\circ}_{\text{NS}})$ | 15,44 | 15,10 | 15,10 | 15,10 | 15,44 | 2,80 |
| $F_{0,025} (^{\circ}_{\text{NS}})$ | 0,06477 | 0,10020 | 0,10020 | 0,10020 | 0,06477 | 0,37453 |
| Samme reproducerbarhed? | Ja | Ja | Nej | Ja | Ja | Ja |
| $y_{met 1}$ | 4,87 | 4,20 | 3,76 | 4,32 | 3,52 | - |
| $y_{met 2}$ | 5,13 | 3,96 | 3,84 | 4,51 | 3,39 | - |
| $S_{fælles}$ | 0,223 | 0,109 | 0,103 | 0,101 | 0,130 | - |
| $ABS(y_A - y_B) / S_{fælles}$ | 1,192 | 2,19 | 0,858 | 1,86 | 1,063 | - |
| Samme middelværdi? | Ja | Nej | Ja | Ja | Ja | - |

Testen viser, at der nu kun er signifikant forskel på repeterbarheden for de to metoder på ét prøveniveau. Det understøtter, at variationen mellem panelister er væsentlig for repeterbarheden. Når der testes for paneler, er det uden betydning, om der er gengangere, og alle resultaterne er gennemsnit af to gentagelser.

Det er imidlertid tydeligt, at den lidt stærkere test af de poolede variationer viser signifikant forskel mellem de to metoder. Yes-no metoden har den bedste repeterbarhed.

Der er ikke signifikant forskel på reproducerbarheden på fire prøveniveauer og for den stærkere, poolede reproducerbarhed. For prøve 3 har yes-no metoden en signifikant bedre reproducerbarhed. For prøve 1 har yes-no metoden en klart dårligere reproducerbarhed, om end den ikke er statistisk signifikant.

Testen viser, at der på ét prøveniveau er signifikant forskel på gennemsnitsresultaterne af de to metoder, mens der ikke er på de andre fire. Den ændrede konklusion i forhold til testen på laboratorieniveau skyldes bl.a., at enkeltresultater er udgået af testen på panelniveau. Gennemsnitsresultaterne er derfor ændrede.

6.3. Beregning af z-score

Beregning af z-scoren inden for hver måleserie giver mulighed for at vurdere, om ét eller flere laboratorier afviger fra gennemsnittet af samtlige resultater. Testen bruges ikke til at skelne mellem metoder.

Laboratorieresultater

I den første test betragtes gennemsnitsresultatet fra hvert laboratorium ($n = 4$ for de fleste laboratorier/prøver og $n = 3$ eller $n = 2$ for resten), dvs. hvert laboratorium leverer ét resultat. Tabel 10 viser resultatet af beregningen.

Spredningerne mellem resultater for hver måleserie er ikke signifikant forskellige (Cochrans test), hvorfor der er beregnet en poollet spredning til testen.

Tabel 10. z-scoring for laboratorieresultater

| | Prøve | 1 | 2 | 3 | 4 | 5 |
|------------|--------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| | Laboratorium | $\log(\text{OU}/\text{m}^3)$ | $\log(\text{OU}/\text{m}^3)$ | $\log(\text{OU}/\text{m}^3)$ | $\log(\text{OU}/\text{m}^3)$ | $\log(\text{OU}/\text{m}^3)$ |
| Resultater | Lab 1 | 5,006 | 4,084 | 3,624 | 4,357 | 3,385 |
| | Lab 5 | 4,727 | 4,322 | 3,890 | 4,292 | 3,664 |
| | Lab 2 | 4,634 | 4,120 | 3,876 | 4,493 | 3,537 |
| | Lab 3 | 5,404 | 3,850 | 3,848 | 4,633 | |
| | Lab 4 | 5,173 | 3,891 | 3,757 | 4,367 | 3,235 |
| z-scoring | Lab 1 | 0,1 | 0,2 | 0,9 | 0,4 | 0,3 |
| | Lab 5 | 1,3 | 1,3 | 0,5 | 0,7 | 1,0 |
| | Lab 2 | 1,8 | 0,3 | 0,4 | 0,3 | 0,4 |
| | Lab 3 | 2,1 | 1,0 | 0,2 | 1,0 | |
| | Lab 4 | 0,9 | 0,8 | 0,2 | 0,3 | 1,1 |

Resultatet for prøve 1 fra yes-no laboratorium 3 er "tvivlsomt" i forhold til gennemsnittet af resten, men kan ikke betragtes som "Ikke tilfredsstillende". Alle andre resultater har tilfredsstillende z-score.

Panelresultater

I denne test betragtes gennemsnitsresultater fra hvert panel ($n = 2$ for alle paneler som indgår). Hvor et panel kun har ét resultat udgår panelet af testen. Det betyder, at yes-no laboratorium 2 ikke indgår for prøve 1 om formiddagen, mens yes-no laboratorium 4 ikke indgår for prøverne 1 til 4 om formiddagen. Yes-no laboratorium 3 indgår ikke for prøve 5 på nogen af panelerne. Tabel 11 viser resultatet af beregningen.

I dette tilfælde er én af spredningerne mellem resultater for hver måleserie en straggler (Cochrans test). Derfor anvendes spredningen for hver måleserie i testen af den enkelte måleserie.

Tabel 11. z-scorer for panelresultater

| | Måling | 1 | 2 | 3 | 4 | 5 |
|------------|--------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | Laboratorium | log(OU/m ³) | log(OU/m ³) | log(OU/m ³) | log(OU/m ³) | log(OU/m ³) |
| Resultater | 1, 1 | 5,100 | 4,204 | 3,716 | 4,518 | 3,448 |
| | 1, 2 | 4,911 | 3,964 | 3,531 | 4,195 | 3,322 |
| | 5, 1 | 4,699 | 4,363 | 3,764 | 4,396 | 3,596 |
| | 5, 2 | 4,754 | 4,281 | 4,016 | 4,189 | 3,732 |
| | 2, 2 | 4,546 | 4,064 | 3,901 | 4,449 | 3,625 |
| | 2, 1 | | 4,177 | 3,851 | 4,537 | 3,449 |
| | 3, 1 | 5,455 | 3,825 | 3,805 | 4,680 | |
| | 3, 2 | 5,352 | 3,875 | 3,891 | 4,585 | |
| | 4, 2 | 5,173 | 3,884 | 3,778 | 4,317 | 3,263 |
| | 4, 1 | | | | | 3,207 |
| z-scorer | 1, 1 | 0,9 | 0,1 | 0,6 | 0,8 | 0,4 |
| | 1, 2 | 0,3 | 1,1 | 1,9 | 1,1 | 1,1 |
| | 5, 1 | 0,3 | 1,0 | 0,2 | 0,1 | 0,4 |
| | 5, 2 | 0,1 | 0,5 | 1,6 | 1,1 | 1,1 |
| | 2, 2 | 0,8 | 0,6 | 0,8 | 0,4 | 0,5 |
| | 2, 1 | | 0,0 | 0,4 | 0,9 | 0,4 |
| | 3, 1 | 2,0 | 1,8 | 0,1 | 1,8 | |
| | 3, 2 | 1,7 | 1,5 | 0,7 | 1,2 | |
| | 4, 2 | 1,1 | 1,5 | 0,1 | 0,4 | 1,4 |
| | 4, 1 | | | | | 1,7 |

Yes-no laboratorium 3's ene panel leverer et resultat, som er "tvivlsomt" i forhold til gennemsnittet, mens alle andre resultater har tilfredsstillende z-score.

7. Vurdering af metoderne

7.1. Beskrivelse af metoderne

For begge metoder gælder, at

- olfaktometeret ved hjælp af pumper, kanyler, kritiske dyser, masseflowmetre og andet på hver sin måde fortynder den udtagne prøve af lugtende luft
- olfaktometrene skal kunne lave fortyndinger i området fra mindre end 2⁷ (128) til mindst 2¹⁴ (16.384) med mindst en faktor 8.192 mellem højeste og laveste fortynding
- panelisterne må højst have 15 sekunder til at vurdere lugten og svare i hvert trin (i det følgende kaldt responstiden)²
- tiden mellem hvert trin skal være lang nok til at undgå tilvænning - hvis lugten præsenteres tilfældigt skal denne tid være 30 sekunder. Der er ingen specifikation af tiden for tilfælde, hvor lugten præsenteres i stigende styrke²
- mindst fire panelister skal afgive korrekte svar og dermed bidrage til resultatet

Den umiddelbart største forskel mellem de to metoder består i, at ved yes-no metoden har lugtpanelisterne ét rør hver, hvori lugten kommer, mens de i forced choice metoden har mindst to, hvoraf det ene indeholder lugtende luft og de/det andre/andet indeholder ren luft. Fra trin til trin varierer det tilfældigt, hvilket rør der er lugt i.

I yes-no metoden skal der præsenteres ikke-lugtende luft i nogle trin i tilfældig rækkefølge. Ved hver præsentation af lugtende eller ikke-lugtende luft skal panelisten blot svare, om han/hun kan lugte noget eller ej.

I forced choice metoden skal panelisten snuse til begge rør og udpege det, han/hun mener, der er lugt i, ledsaget af en indikation af, hvor sikker han/hun er. Kun rigtigt rør og indikationen "sikker" er et rigtigt svar.

Den umiddelbart enklere teknik med kun ét rør giver panelisten færre ting at forholde sig til, og panelisten kan måske derfor koncentrere sig mere om det ene svar, der skal afgives. Dermed kan yes-no metoden operere med kortere responstid og dermed hurtigere gå fra trin til trin. Laboratorier, der bruger yes-no metoden, kan dermed analysere flere prøver per tidsenhed.

At panelisten kun skal forholde sig til én ting modsvares måske af, at flere prøver per tid hurtigere svækker opmærksomheden og trætter næsen. Vi har ikke kendskab til undersøgelser, der belyser dette.

For begge metoder gælder, at hvis responstiden er for kort, er der en risiko for, at panelisten ikke når at opfatte lugten ved lugtstofkoncentrationer omkring panelistens lugttærskel og dermed ikke at afgive rigtigt svar. Dermed er risikoen, at det rigtige svar først kommer ved lugtstofkoncentrationer noget højere end panelistens lugttærskel, og resultatet bliver for lavt (mindre fortynding og dermed mindre lugtkoncentration).

I den aktuelle prøvning har begge forced choice laboratorierne en responstid på 15 sekunder. Denne tid skal bruges på to rør og på svar. Det vil sige, at panelisterne har ca. 10 sekunder til at vurdere luften i de to rør og derefter 5 sekunder til at svare. I praksis kan de svare hurtigere og dermed undlade at gøre brug af hele tiden.

De to yes-no laboratorier, som har oplyst om dette har responstider på hhv. 2,2 sekunder og 4,4 sekunder.

Resultaterne i tabel 5 viser, de to metoder "skiftes" til at have højest resultat, og der er således ingen indikation af, at kortere responstid giver lavere resultater.

I princippet kan begge metoder gøre brug af fire eller flere panelister, men ofte bruger yes-no metoden kun fire panelister, da mange kommercielle yes-no olfaktometre har plads til fire samtidige panelister. I den aktuelle prøvning har yes-no laboratorierne fire panelister og forced choice laboratorierne seks panelister.

Ved brug af flere panelister (seks) er panelet mere repræsentativt, og resultatet beskriver bedre gennemsnittet for en stor gruppe mennesker ("den sande værdi"). Det kan derfor teoretisk forventes, at spredningen mellem panelresultater (reproducerbarhed) er mindre jo flere panelister, der deltager i hvert panel. Det vil sige, at man kunne forvente en bedre reproducerbarhed for forced choice laboratorierne. Tabel 9 viser varianserne for reproducerbarhed mellem paneler. På tre prøveniveauer har yes-no laboratorierne den bedste reproducerbarhed, på to har forced choice den bedste. Forced choice laboratorierne har den bedste poolede reproducerbarhed. Resultaterne er således ikke éntydige, og som før nævnt vil de flere panelistgængere i yes-no laboratorierne fremme mere ens resultater mellem paneler.

7.2. Tilkendegivelser om metoderne

Vi har fortalt to førende eksperter inden for lugtanalyse om, at vi har set indikationer på, at yes-no metoden giver lavere resultater end force-choice metoden. Ingen af dem tror på, at det er et generelt

fænomen⁹. Begge fremhæver, at lugtpanelister uanset laboratoriets metode skal være godkendt i henhold til testen med n-butanol og ligge inden for de rammer, som standarden angiver². Når de gør det, mener eksperterne ikke, at det er sandsynligt, at den ene metode konsekvent vil give lavere resultater end den anden.

7.3. Sammenligning af resultater i Olfatec-test

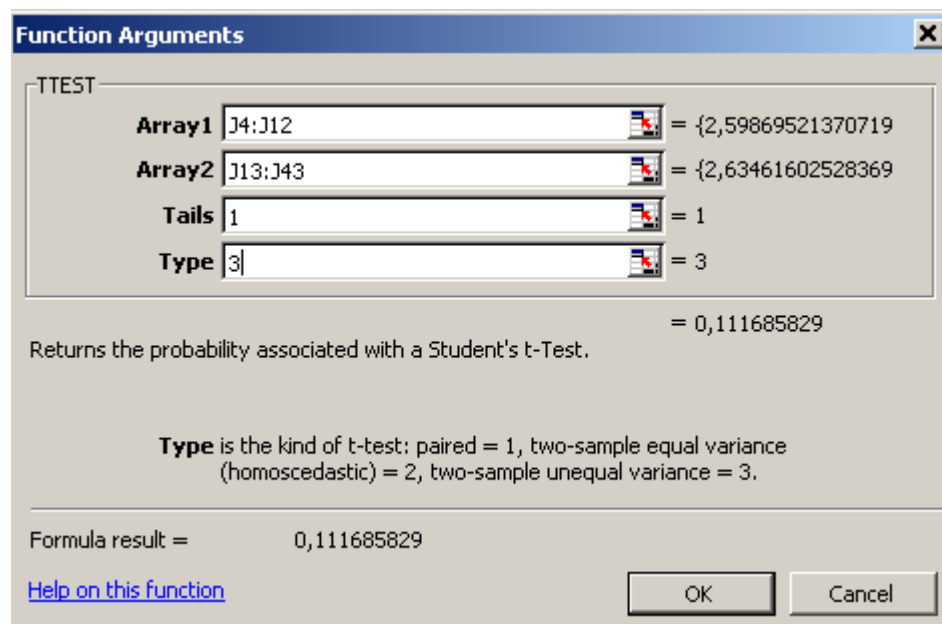
Det tyske firma, OLFatec GmbH, udbyder hvert andet år en test, hvor deltagerne (typisk 70-80 hovedsageligt europæiske laboratorier) modtager fire forskellige lugtprøver i trykflasker. Laboratorierne overfører selv indholdet til lugtposer og foretager analyser i henhold til medsendte instruktioner. I rapporten for 2005⁶ er det angivet, hvilket fabrikat olfaktometer de forskellige laboratorier har anvendt. I senere, tilsvarende tests er fabrikatet ikke nævnt.

Det har ikke været muligt at identificere analyseprincippet i alle fabrikater, idet enkelte fabrikater på deres respektive internetsider omtaler at kunne bruge både forced choice- og yes-no metoderne.

Som repræsentanter for yes-no metoden er 31 sæt resultater med olfaktometrene TO-7 og TO-8 anvendt.

Som repræsentanter for forced choice metoden er ni sæt resultater med olfaktometrene Odormat Aromatrix, Olfaktomat at-n6 og fem selv fremstillede anvendt. Vi kan ikke være sikre på, at de fem selv fremstillede olfaktometre alle er forced choice, men de tre danske laboratorier havde i 2005 selv fremstillede forced choice olfaktometre. For at få dem med i testen er vi nødt til at inddrage alle selv fremstillede. Det giver en fejl mulighed, og testen er dermed ikke nødvendigvis éntydig. Det er imidlertid det bedst mulige, idet Olfatec ikke har kunnet oplyse yderligere om olfaktometertypen.

Testen er udført som en t-test ved hjælp af EXCEL-funktionen TTEST. Den er udført både som en to-sidet og som en én-sidet test. I begge tilfælde er testen udført med de faktiske varianser for de to metoders resultater. Et eksempel på indtastning af tal til beregning ses i figur 2



Figur 2. Indtastning i EXCEL-funktionen TTEST

⁹ Björn Maxeiner, Olfatec, og Ton van Harreveld, Odournet

I bilag 1 ses alle de anvendte resultater og beregninger. Olfatec's rapport angiver resultaterne som lugttaerskler for hvert laboratorium og oplyser desuden de anvendte gassers koncentration. På det grundlag er de enkelte laboratoriers OU/m³ beregnet og omregnet til logaritmiske værdier. t-testen er udført på logaritmiske værdier. Nulhypotesen for t-testen er, at middelværdierne for metoderne er ens.

Tabel 12 viser de væsentligste resultater.

Tabel 12. Sammenligning af Olfatec-resultater for to metoder

| | Gas A | Gas B | Gas C | Gas D | Gas A | Gas B | Gas C | Gas D |
|--|-------------------|-------|-------|-------|-------------------|-------|-------|-------|
| | µg/m ³ | | | | OU/m ³ | | | |
| Antal forced-choice | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |
| Antal yes-no | 31 | 31 | 31 | 31 | 31 | 31 | 31 | 31 |
| Middel forced-choice | 104 | 113 | 1,52 | 0,178 | 453 | 1.291 | 5.657 | 9.055 |
| Middel yes-no | 160 | 142 | 1,32 | 0,189 | 370 | 1.443 | 7.200 | 8.787 |
| t-test tosidet (risiko for fejlagtigt at forkaste nul-hypotesen) | | | | | 0,223 | 0,662 | 0,244 | 0,871 |
| t-test énsidet (risiko for fejlagtigt at forkaste nul-hypotesen) | | | | | 0,112 | 0,331 | 0,122 | 0,435 |

Resultatet af t-testen udtrykker den risiko, man tager, hvis man forkaster nulhypotesen. Hvis denne risiko er lille (typisk 0,05) er der statistisk signifikans mod nulhypotesen, dvs. man kan forkaste den. I dette tilfælde er der i intet tilfælde en lille risiko, og der er ikke statistisk signifikans mod nulhypotesen. De to metoder giver derfor samme middelværdier.

I bilag 1 ses, at yes-no metoden i fire tilfælde har den største spredning, I tabel 12 ses, at forced choice giver det højeste resultat (OU/m³) i to tilfælde (gas A og gas D), mens yes-no metoden giver det højeste resultat i de andre to tilfælde.

Med ovennævnte forbehold for de selvfremstillede olfaktometre er der således intet belæg for at mene, at yes-no metoden giver lavere resultater end forced choice metoden.

8. Diskussion

8.1. Generelt

Den udførte præstationsprøvning viser, at de to metoder generelt præsterer ens.

Det kan ikke udelukkes, at yes-no metoden har en lidt bedre repeterbarhed (laver bedre gentagelser), men resultatet er ikke éntydigt, idet yes-no laboratorierne har flere gengangere på panelerne i denne prøvning, og nogle laboratorieresultater stammer fra kun ét panel.

Ser man på den repeterbarheden uden at skelne mellem metoder er denne kun i ét tilfælde større end den i standarden² højest accepterede spredning på 0,172. Det er umiddelbart tilfredsstillende, og det vurderes, at laboratorierne som gruppe har en god repeterbarhed. Repeterbarheden siger imidlertid ikke noget om, hvordan de enkelte resultater ligger i forhold til hinanden.

Metoderne har samme reproducerbarhed (ved betragtning af alle laboratorieresultater), hvilket vil sige, at en bruger af olfaktometri kan vælge et hvilket som helst laboratorium blandt både yes-no laboratorier og blandt forced choice laboratorier uden at kunne forudsige noget om resultatets størrelse. Når testen udføres for panelresultater er der signifikant forskel på ét prøveniveau, hvor yes-no metoden har dårligere reproducerbarhed.

Ser man på reproducerbarheden uden at skelne mellem metoder (tabel 7), er billedet det samme, hvilket er meget positivt. For én prøve er reproducerbarheden meget dårlig, hvilket skyldes resultatet fra ét yes-no laboratorium. For de øvrige er reproducerbarhedsspredningen højest 0,213 svarende til

et 95% konfidensinterval med en bredde på 2,8 gange til hver side for gennemsnittet (15 frihedsgrader). Det er noget højere end den intervalbredde, der svarer til standardens acceptgrænse (2,2) for ét laboratorium, men ikke umiddelbart uacceptabelt. Det må forventes, at den totale spredning af resultater for en gruppe laboratorier er større end de enkelte laboratoriers spredning.

Metoderne giver samme gennemsnitsresultat (ved betragtning af alle laboratorieresultater), mens der er signifikant forskel på ét prøveniveau, når testen udføres for panelresultater.

Beregningen af z-scorer bringer intet nyt til vurderingen af laboratoriernes præstation, men supplerer den generelle opfattelse, at laboratorierne præsterer ens. Kun ét resultat for laboratorier og ét for paneler er tvivlsomt. Dette tvivlsomme resultat er formodentlig også årsagen til de få, mindre tilfredsstillende spredninger, som er konstateret i de øvrige tests.

Alle testresultaterne viser, at brugere af olfaktometri kan vælge frit blandt alle laboratorier uden at tage hensyn til, om det laboratorierne anvender yes-no metoden eller forced choice metoden. Men denne konklusion betyder ikke, at laboratorierne ikke skal bruge testens resultater til at blive bedre. Således vil det være i brugernes interesse at finde og rette årsagen til, at yes-no laboratorium 3 har så markant høje resultater for prøve 1.

Resultater fra Olfatec's præstationsprøvning i 2005 understøtter nærværende undersøgelses hovedresultat, at der ikke er forskel på gennemsnitsresultater opnået med de to metoder. Denne konklusion støttes yderligere af udtalelser fra eksperter på området, som understreger begge metoders kalibrering over for samme teststof (n-butanol).

8.2. Betydning af brug af flere paneler

Når en virksomhed skal dokumentere overholdelse af lugtgrænseværdier, kræver myndighederne oftest udtagning af tre prøver og analyse for lugt i disse. I nogle tilfælde kræves flere, og i enkelte tilfælde specificeres et tidsrum, hvor prøverne skal tages. Dette tidsrum er oftest således, at der er mindst en halv time mellem hver prøve. Der er imidlertid aldrig andre krav til analysen end, at den skal udføres af et akkrediteret laboratorium.

Når virksomheder skal bestemme lugtemission som grundlag for dimensionering af rensningsanlæg, anlægger man ofte samme strategi, tre prøver udtaget i løbet af få timer og analyseret af et akkrediteret laboratorium.

Der er således normalt ingen overvejelser af betydningen af, at prøverne bliver analyseret af så mange panelister som muligt og af, at resultaterne bliver så repræsentative som muligt.

Nærværende undersøgelse kan bidrage til at belyse betydningen af, at flere paneler analyserer prøverne. Variationen mellem prøver, som er analyseret på samme panel, beskrives ved repeterbarheden inden for paneler, s^2_r . Variationen mellem prøver, som er analyseret på forskellige paneler, beskrives ved reproducerbarheden mellem paneler, s^2_R . Det bedste estimat af de to størrelser opnås ved at se på de poolede værdier.

Tabel 13 viser repeterbarheden og reproducerbarheden for paneler for hver metode og for begge metoder samlet. I alle tilfælde er de poolede værdier anvendt. For begge metoder samlet skal der som nævnt tages forbehold for, at varianserne ikke er ens og, at der derfor strengt taget ikke kan beregnes en pooled værdi.

I tabel 13 ses desuden antal frihedsgrader for hver varians og de dertil hørende værdier fra Students t-fordeling. Faktorer for 95% konfidensintervaller for resultaterne er beregnet henholdsvis for resultater inden for et panel og for resultater mellem paneler efter nedenstående formel:

$$F_{95\% \text{ konfidensinterval}} = 10^{s \cdot t\text{-værdi}}$$

Tabel 13. Beregning af 95% konfidensintervaller

| Forced choice paneler | | Yes-no paneler | | Alle paneler | |
|------------------------|---------|------------------------|---------|------------------------------|---------|
| $S_{r,1}^2$ | 0,01343 | $S_{r,2}^2$ | 0,00383 | $S_{r,begge}^2$ | 0,00829 |
| $S_{r,1}$ | 0,1159 | $S_{r,2}$ | 0,0619 | $S_{r,begge}$ | 0,0911 |
| \bar{r}_1 | 20 | \bar{r}_2 | 23 | \bar{r}_{begge} | 43 |
| Faktor ₁ | 1,73 | Faktor ₂ | 1,34 | Faktor _{begge} | 1,53 |
| $t_{0,975, \bar{r}_1}$ | 2,06 | $t_{0,975, \bar{r}_2}$ | 2,07 | $t_{0,975, \bar{r}_{begge}}$ | 2,02 |
| $S_{R,1}^2$ | 0,0386 | $S_{R,2}^2$ | 0,0454 | $S_{R,begge}^2$ | 0,0479 |
| $S_{R,1}$ | 0,197 | $S_{R,2}$ | 0,213 | $S_{R,begge}$ | 0,219 |
| \bar{r}_1 | 15 | \bar{r}_2 | 18 | \bar{r}_{begge} | 38 |
| Faktor ₁ | 2,62 | Faktor ₂ | 2,80 | Faktor _{begge} | 2,77 |
| $t_{0,975, \bar{r}_1}$ | 2,13 | $t_{0,975, \bar{r}_2}$ | 2,10 | $t_{0,975, \bar{r}_{begge}}$ | 2,02 |

Tabellen viser, at hvis en prøve bliver analyseret mange gange på samme panel med forced choice metoden (metode 1), vil 95% af resultaterne ligge i et interval med bredden **1,73 gange** middelværdien til hver side for middelværdien (f.eks. 5.780 OU/m³ til 17.300 OU/m³).

Blive den samme prøve analyseret mange gange med forskellige forced choice paneler vil intervallet være en faktor **2,62** til hver side for middelværdien (3.820 OU/m³ til 26.200 OU/m³).

For yes-no laboratorierne er forskellen endnu større mellem intervallerne inden for paneler og mellem paneler. Hvis alle laboratorier betragtes under ét, er faktorerne hhv. 1,53 og 2,77.

Usikkerheden inden for et panel kan i princippet kun bruges, hvis to prøver skal sammenlignes (f.eks. før og efter rensning). Er prøverne analyseret på samme panel, kan forskellen mellem dem vurderes på baggrund af den mindre usikkerhed.

Skal et resultat derimod bruges til at dimensionere en skorsten, bør det vurderes i forhold til en større usikkerhed. Ellers risikerer man, at skorstenen underdimensioneres og, at en kontrolmåling efter opførelse af skorstenen viser et så højt tal, at skorstenen ikke er høj nok.

Hvis man har analyseret en prøve på flere paneler, er gennemsnittet af resultatet det bedste estimat for koncentrationen. Usikkerheden på gennemsnittet kan da beregnes med ovenstående faktorformel, men spredningen mellem paneler kan divideres med antallet af paneler, som vist nedenfor.

$$S_{\text{gennemsnit}} = \frac{S_{\text{mellem paneler}}}{n_{\text{paneler}}}$$

Derved opnås et noget mere sikkert estimat for "den sande værdi".

Det vurderes derfor, at den reelle usikkerhed, som en bruger af olfaktometri bør kende, er den usikkerhed der ligger i, at prøven kan analyseres af forskellige paneler på forskellige laboratorier med forskellige metoder. Denne usikkerhed domineres af variationen mellem paneler, og det er således

forskellen i lugtesans hos forskellige mennesker, der har størst betydning for usikkerhed i lugtanalyse.

Det vurderes desuden, at lugtvurderinger af virksomheder i videst muligt omfang skal udføres med flere paneler og, at estimatet skal baseres på gennemsnit og kendskab til variation mellem paneler. Der kan selvfølgelig være nogle praktiske og økonomiske konsekvenser ved at udtage prøver over flere dage, henholdsvis fordele dem over flere paneler. Hvis et laboratorium får seks prøver fra to afkast, er det naturligt at analysere dem alle på samme lugtpanel, men det vil ikke give det bedste resultat inden for de givne rammer. Hvor et laboratorium skal analysere tre prøver fra hver af f.eks. seks kilder, er der mulighed for at fordele prøverne mellem to paneler uden, at det koster ekstra.

Da der er omkostninger forbundet med at sprede en prøveserie fra en skorsten på flere paneler, bør sådanne krav nok i højere grad stamme fra miljøgodkendelser mv., ligesom det enkelte laboratorium/rådgiver bør informere kunden om konsekvensen af at spare penge på analyser i forhold til korrekt dimensionering af dyre skorstene eller rensningsforanstaltninger.

Bilag 1 Resultater fra Olfatec's internationale prøvning, 2005

| Gaskoncentrationer mg/m ³ | 36 | 117,8 | 7,5 | 1 | | | | | | | | | |
|---|-------------------|-------|-------|-------|-------------------|-------|--------|--------|-------------------------|-------|-------|-------|-------|
| | Gas A | Gas B | Gas C | Gas D | Gas A | Gas B | Gas C | Gas D | Gas A | Gas B | Gas C | Gas D | |
| | µg/m ³ | | | | OU/m ³ | | | | log(OU/m ³) | | | | |
| Forced choice | 90,7 | 124,5 | 0,7 | 0,07 | 397 | 946 | 10.714 | 14.286 | 2,599 | 2,976 | 4,030 | 4,155 | |
| | 90 | 84,1 | 1,46 | 0,11 | 400 | 1.401 | 5.137 | 9.091 | 2,602 | 3,146 | 3,711 | 3,959 | |
| | 64,7 | 46,4 | 1,26 | 0,15 | 556 | 2.539 | 5.952 | 6.667 | 2,745 | 3,405 | 3,775 | 3,824 | |
| | 95,2 | 101 | 2,37 | 0,14 | 378 | 1.166 | 3.165 | 7.143 | 2,578 | 3,067 | 3,500 | 3,854 | |
| | 122,4 | 151,8 | 0,98 | 0,04 | 294 | 776 | 7.653 | 25.000 | 2,469 | 2,890 | 3,884 | 4,398 | |
| | 27,9 | 47,6 | 1,14 | 0,53 | 1.290 | 2.475 | 6.579 | 1.887 | 3,111 | 3,394 | 3,818 | 3,276 | |
| | 105,9 | 130,9 | 1,97 | 0,12 | 340 | 900 | 3.807 | 8.333 | 2,531 | 2,954 | 3,581 | 3,921 | |
| | 156,5 | 161,8 | 1,84 | 0,22 | 230 | 728 | 4.076 | 4.545 | 2,362 | 2,862 | 3,610 | 3,658 | |
| | 186,5 | 170,7 | 1,96 | 0,22 | 193 | 690 | 3.827 | 4.545 | 2,286 | 2,839 | 3,583 | 3,658 | |
| | 83,5 | 136,8 | 1,29 | 0,41 | 431 | 861 | 5.814 | 2.439 | 2,635 | 2,935 | 3,764 | 3,387 | |
| | 171,4 | 155 | 1 | 0,09 | 210 | 760 | 7.500 | 11.111 | 2,322 | 2,881 | 3,875 | 4,046 | |
| | 118,4 | 121 | 0,81 | 0,08 | 304 | 974 | 9.259 | 12.500 | 2,483 | 2,988 | 3,967 | 4,097 | |
| | 124,1 | 130,9 | 1,14 | 0,12 | 290 | 900 | 6.579 | 8.333 | 2,463 | 2,954 | 3,818 | 3,921 | |
| | 580,6 | 463,8 | 1,66 | 0,31 | 62 | 254 | 4.518 | 3.226 | 1,792 | 2,405 | 3,655 | 3,509 | |
| | Yes-no | 48 | 27,8 | 1,4 | 0,18 | 750 | 4.237 | 5.357 | 5.556 | 2,875 | 3,627 | 3,729 | 3,745 |
| 183,7 | | 139,3 | 1,77 | 0,2 | 196 | 846 | 4.237 | 5.000 | 2,292 | 2,927 | 3,627 | 3,699 | |
| 54,5 | | 48,3 | 0,64 | 0,15 | 661 | 2.439 | 11.719 | 6.667 | 2,820 | 3,387 | 4,069 | 3,824 | |
| 48 | | 42,1 | 1,25 | 0,03 | 750 | 2.798 | 6.000 | 33.333 | 2,875 | 3,447 | 3,778 | 4,523 | |
| 171,4 | | 210,4 | 1,08 | 0,12 | 210 | 560 | 6.944 | 8.333 | 2,322 | 2,748 | 3,842 | 3,921 | |
| 200 | | 294,5 | 0,93 | 0,12 | 180 | 400 | 8.065 | 8.333 | 2,255 | 2,602 | 3,907 | 3,921 | |
| 200 | | 199,7 | 1,52 | 0,42 | 180 | 590 | 4.934 | 2.381 | 2,255 | 2,771 | 3,693 | 3,377 | |
| 157,9 | | 152,4 | 1,74 | 0,2 | 228 | 773 | 4.310 | 5.000 | 2,358 | 2,888 | 3,635 | 3,699 | |
| 133,8 | | 62,4 | 0,99 | 0,12 | 269 | 1.888 | 7.576 | 8.333 | 2,430 | 3,276 | 3,879 | 3,921 | |
| 99,4 | | 154,2 | 0,99 | 0,1 | 362 | 764 | 7.576 | 10.000 | 2,559 | 2,883 | 3,879 | 4,000 | |
| 156,5 | | 143,7 | 1,05 | 0,09 | 230 | 820 | 7.143 | 11.111 | 2,362 | 2,914 | 3,854 | 4,046 | |
| 94,7 | | 117,8 | 0,62 | 0,26 | 380 | 1.000 | 12.097 | 3.846 | 2,580 | 3,000 | 4,083 | 3,585 | |
| 276,9 | | 302,1 | 2,99 | 0,83 | 130 | 390 | 2.508 | 1.205 | 2,114 | 2,591 | 3,399 | 3,081 | |
| 107,8 | | 157,3 | 2,8 | 0,3 | 334 | 749 | 2.679 | 3.333 | 2,524 | 2,874 | 3,428 | 3,523 | |
| 300 | | 157,1 | 2,2 | 0,13 | 120 | 750 | 3.409 | 7.692 | 2,079 | 2,875 | 3,533 | 3,886 | |
| 292,7 | | 122,5 | 1,36 | 0,11 | 123 | 962 | 5.515 | 9.091 | 2,090 | 2,983 | 3,742 | 3,959 | |
| 60,5 | | 66,1 | 2,49 | 0,06 | 595 | 1.782 | 3.012 | 16.667 | 2,775 | 3,251 | 3,479 | 4,222 | |
| 473,7 | | 87,3 | 1,04 | 0,22 | 76 | 1.349 | 7.212 | 4.545 | 1,881 | 3,130 | 3,858 | 3,658 | |
| 98,9 | | 198,3 | 0,33 | 0,26 | 364 | 594 | 22.727 | 3.846 | 2,561 | 2,774 | 4,357 | 3,585 | |
| 74,5 | | 121,8 | 1,62 | 0,12 | 483 | 967 | 4.630 | 8.333 | 2,684 | 2,985 | 3,666 | 3,921 | |
| 76,6 | | 87,4 | 1,06 | 0,05 | 470 | 1.348 | 7.075 | 20.000 | 2,672 | 3,130 | 3,850 | 4,301 | |
| 35,2 | | 16,1 | 0,57 | 0,26 | 1.023 | 7.317 | 13.158 | 3.846 | 3,010 | 3,864 | 4,119 | 3,585 | |
| 327,3 | | 27,4 | 0,74 | 0,06 | 110 | 430 | 10.135 | 16.667 | 2,041 | 2,633 | 4,006 | 4,222 | |
| 31,3 | | 24,2 | 1,45 | 0,05 | 1.150 | 4.868 | 5.172 | 20.000 | 3,061 | 3,687 | 3,714 | 4,301 | |
| 88,7 | | 102,5 | 0,64 | 0,12 | 406 | 1.149 | 11.719 | 8.333 | 2,608 | 3,060 | 4,069 | 3,921 | |
| 93,8 | | 96,7 | 1,62 | 0,3 | 384 | 1.218 | 4.630 | 3.333 | 2,584 | 3,086 | 3,666 | 3,523 | |
| Antal forced-choice | | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |
| Antal yes-no | | 31 | 31 | 31 | 31 | 31 | 31 | 31 | 31 | 31 | 31 | 31 | 31 |
| Middel forced-choice | | 104 | 113 | 1,52 | 0,178 | 453 | 1.291 | 5.657 | 9.055 | 2,587 | 3,059 | 3,721 | 3,856 |
| Middel yes-no | | 160 | 142 | 1,32 | 0,189 | 370 | 1.443 | 7.200 | 8.787 | 2,463 | 3,018 | 3,804 | 3,836 |
| Spredning forced-choice | - | - | - | - | - | - | - | - | 0,239 | 0,216 | 0,171 | 0,319 | |
| Spredning yes-no | - | - | - | - | - | - | - | - | 0,314 | 0,327 | 0,213 | 0,316 | |
| t-test tosided | - | - | - | - | - | - | - | - | 0,223 | 0,662 | 0,244 | 0,871 | |
| t-test ensided | - | - | - | - | - | - | - | - | 0,112 | 0,331 | 0,122 | 0,435 | |