# Bijlage 1

## Overzicht normen voor sensorische analyse

Het volgende overzicht toont de nummers en namen van een aantal ISO-normen voor sensorisch onderzoek. Normen zijn afspraken die belanghebbende partijen maken over een product, dienst of systeem, en normalisatie is het proces om tot een norm te komen. Het Nederlands Normalisatie-instituut, NEN, beheert en publiceert een omvangrijke collectie van duizenden internationale en nationale normen. Door zijn lidmaatschap op Europees en mondiaal niveau heeft NEN een centrale positie in het web van internationale normalisatie.

NB

- Hier is alleen – niet-uitputtend – een aantal min of meer algemene normen opgenomen. Er bestaan ook normen met betrekking tot het testen van specifieke productgroepen en bijvoorbeeld verpakkingsmateriaal.
- Op diverse normbladen zijn (en worden nog steeds) aanvullingen en verbeteringen uitgebracht. Deze zijn hier niet genoemd.
- Wees erop bedacht dat normen geregeld kunnen worden vernieuwd.
- Check steeds alle informatie. Raadpleeg bijvoorbeeld de website nen.nl.
- Hoewel de nodige zorg aan dit lijstje is besteed, kunnen fouten en onvolledigheden niet worden uitgesloten. nen, auteur, uitgever noch anderen aanvaarden enige aansprakelijkheid.
- In de tabel hieronder geven we de titels in het Engels. Het NEN hanteert weliswaar Nederlandse titels, maar de documenten zelf zijn bijna nooit in het Nederlands geschreven.
- Voor informatie over normen en bestellingen: de klantenservice van NEN is bereikbaar op nummer +31 15 2690391 en via klantenservice@nen.nl.

Uitleg van de gebruikte afkortingen:

- ISO: internationale norm.
- NEN-ISO: ISO-norm die is overgenomen als Nederlandse norm.
- NEN-EN-ISO: ISO-norm die Europees is overgenomen en ook binnen Nederland.
- IEC: International Electrotechnical Commission.

| Norm | Titel |
|---|---|
| NEN-ISO 3972:2011 | Sensory analysis – Methodology – Method of investigating sensitivity of taste |
| NEN-EN-ISO 4120:2007 | Sensory analysis – Methodology – Triangle test |
| NEN-ISO 4121:2003 | Sensory analysis – Guidelines for the use of quantitative response scales |
| NEN-EN-ISO 5492:2009 | Sensory analysis – Vocabulary |
| NEN-EN-ISO 5495:2007 | Sensory analysis – Methodology – Paired comparison test |
| NEN-ISO 5496:2006 | Sensory analysis – Methodology – Initiation and training of assessors in the detection and recognition of odours |
| NEN-ISO 5497:2012 | Sensory analysis – Methodology – Guidelines for the preparation of samples for which direct sensory analysis is not feasible |
| ISO 6658:2017 | Sensory analysis – Methodology – General guidance |

| | |
|---|---|
| NEN-EN-ISO 8586-2014 | Sensory analysis – General guidance for the selection, training and monitoring of assessors |
| NEN-ISO 8587:2006 | Sensory analysis – Methodology – Ranking |
| NEN-ISO 8588:2017 | Sensory analysis – Methodology – 'A' – 'not A' test |
| NEN-EN-ISO 8589:2010 | Sensory analysis – General guidance for the design of test rooms |
| NEN-EN-ISO 10399:2018 | Sensory analysis – Methodology – Duo-trio test |
| NEN-ISO 11035:2012 | Sensory analysis – Identification and selection of descriptors for establishing a sensory profile by a multidimensional approach |
| NEN-ISO 11036:2012 | Sensory analysis – Methodology – Texture profile |
| NEN-ISO 11037:2011 | Sensory analysis – General guidance and test method for assessment of the colour of foods |
| NEN-ISO 11056:2000 | Sensory analysis – Methodology – Magnitude estimation method |
| NEN-EN-ISO 11132:2017 | Sensory analysis – Methodology – Guidelines for monitoring the performance of a quantitative sensory panel |
| NEN-EN-ISO 11136:2017 | Sensory analysis – Methodology – General guidance for conducting hedonic tests with consumers in a controlled area |
| NEN-EN-ISO 13299:2016 | Sensory analysis – Methodology – General guidance for establishing a sensory profile |
| NEN-ISO 13300-1:2006 | Sensory analysis – General guidance for the staff of a sensory evaluation laboratory - Part 1: Staff responsibilities |
| NEN-ISO 13300-2:2006 | Sensory analysis – General guidance for the staff of a sensory evaluation laboratory – Part 2: Recruitment and training of panel leaders |
| NEN-ISO 13301:2018 | Sensory analysis – Methodology – General guidance for measuring odour, flavour and taste detection thresholds by a three-alternative forced-choice (3-AFC) procedure |
| NEN-ISO 13302:2003 | Sensory analysis – Methods for assessing modifications to the flavour of foodstuffs due to packaging |
| NEN-ISO 16820:2019 | Sensory analysis – Methodology – Sequential analysis |
| NEN-ISO 16779:2015 | Sensory analysis – Assessment (determination and verification) of the shelf life of foodstuffs |
| NEN-EN-ISO/IEC 17025:2018 nl | Algemene eisen voor de bekwaamheid van beproevings- en kalibratielaboratoria (Supplement hierbij: EA-4109 G: 2003 / Accreditation For Sensory Testing Laboratories) |
| NEN-ISO 29842:2011 | Sensory analysis – Methodology – Balanced incomplete block designs |

### *Normen met betrekking tot markt-, opinie- en sociaalwetenschappelijk onderzoek*

| Norm | Titel |
|---|---|
| NEN-ISO 20252:2019 | Market, opinion and social research – Vocabulary and service requirements (waarin normen voor de steekproeftrekking bij de toepassing van onder andere accesspanels zijn opgenomen in een bijlage) |

### *Norm met betrekking tot de tetradtest (van de American Society for Testing and Materials)*

| Norm | Titel |
|---|---|
| ASTM E3009 – 15e1 en | Standard Test Method for Sensory Analysis – Tetrad Test |

# Bijlage 2

## Statistische toetsing: werkwijze, toetsen, formules, toepassing

In dit boek wordt kennis van statistiek en statistische (hypothese)toetsing in principe bekend verondersteld. Niettemin geven we hier – ter opfrissing – een aantal belangrijke punten kort weer. Wie zich opnieuw (of alsnog) in de statistiek wil verdiepen wordt het boek *Cijfers spreken* (6e druk, Noordhoff Uitgevers, 2016) van Joep Brinkman en Hilbrand Oldenhuis aanbevolen. Die uitgave behandelt met dezelfde symbolen, formules, principes en uitgangspunten de methodologie en de statistiek waar in *Proeven van succes* van wordt uitgegaan.

### *1. Statistische (hypothese)toetsing in hoofdlijnen*

***Zes stappen***

In zes stappen uitgedrukt is de algemene wijze van denken en doen bij statistische (hypothese)toetsing als volgt:

1. Je wilt nagaan of (of bewijzen dat) er meer speelt dan alleen maar toeval.
2. Ga er vervolgens van uit dat dat juist níet het geval is. Met andere woorden: *ga ervan uit dat er alleen maar toeval speelt*.
3. Kies het significantieniveau α (doorgaans 1% of 5%). Dit is het risico *ten onrechte* te concluderen dat er meer dan alleen toeval speelt. Als je dit risico wilt beperken kies je een lage α (en omgekeerd).
4. Verzamel gegevens. Bereken met behulp van een toepasselijke theoretische kansverdeling hoe groot de kans is op een zo extreme of nog extremere waarde als in deze ‘steekproef’ is gevonden. Deze kans is p, de *overschrijdingskans*. Hierbij ga je er steeds vanuit dat stap 2 waar is en er dus louter toevalsomstandigheden zijn.
5. Vergelijk deze overschrijdingskans met α. Als p kleiner is dan α, is het gestelde bij stap 2 (alléén toeval) waarschijnlijk niet waar. In dat geval spreek je van *significantie* en wordt stap 6 gezet. Als p groter is dan α, ga er dan voorlopig vanuit dat het gestelde bij stap 2 waar is, en er dus alleen maar toeval heeft gespeeld.
6. Als het gestelde bij stap 2 onjuist is, moet het tegendeel waar zijn. Dan is de gevonden uitkomst niet alleen door het toeval ontstaan en speelt er meer. Dan is het vermoeden van stap 1 dus juist.

***Hypothesen***

De tweede stap, waarin wordt aangenomen dat er alleen maar toeval speelt, verdient bijzondere aandacht. Op die aanname is namelijk het verdere rekenwerk gebaseerd. Deze aanname heet de *nulhypothese*, $H_0$. De nulhypothese heeft altijd de vorm ‘er is niets aan de hand, er speelt alleen maar toeval’. Het toetsen vindt steeds plaats door met betrekking tot een soort ‘nul’situatie de theoretische kansverdeling (zoals een binomiale verdeling of een t-verdeling) te bepalen. *Toegepast op sensorisch onderzoek*, komt de nulhypothese er meestal op neer dat er geen smaakverschil is, dat er geen verschil in voorkeur is, dat product A even scherp (of zout of hard of…) is als product B, dat de gemiddelde score op een JAR-schaal gelijk is aan 50 enzovoort.

Tegenover de nulhypothese staat de *alternatieve hypothese*, $H_1$. Deze staat bij de toetsingsprocedure hierboven verwoord bij stap 1. In de alternatieve hypothese komt doorgaans tot uitdrukking wat de onderzoeker verwacht of probeert aan te tonen. $H_1$ heeft de vorm 'als de nulhypothese niet opgaat, dan moet er meer aan de hand zijn dan alleen toeval, en moet dus het volgende waar zijn: ....'

$H_0$ en $H_1$ hebben *altijd betrekking op de populatie*. Vandaar dat er doorgaans Griekse symbolen in worden gebruikt.

De nulhypothese en de alternatieve hypothese *moeten elkaar volledig aanvullen en uitsluiten*. Dus als bijvoorbeeld in $H_0$ het is-gelijk-teken (=) staat, bevat $H_1$ het is-ongelijk-teken ($\neq$). Als in de nulhypothese $\geq$ staat, moet de alternatieve hypothese $<$ bevatten. Enzovoort.

Nulhypothesen worden verworpen of niet, alternatieve hypothesen worden geaccepteerd of niet. Dit taalgebruik maakt duidelijk dat nulhypothesen niet kunnen worden bewezen. Nulhypothesen krijgen het voordeel van de twijfel: zij worden gehandhaafd bij gebrek aan bewijs van het tegendeel.

### *Een- en tweezijdig toetsen*

De vraag of een toets een- of tweezijdig moet worden uitgevoerd, hangt af van de onderzoeksvraag. Als men wil weten óf er een verschil met het toeval bestaat, is de toets tweezijdig. Wie wil weten of er een verschil *in een bepaalde richting* (meer *of* minder kans dan door toeval) bestaat, toetst eenzijdig. Bij een eenzijdige toets kijkt men naar de linker óf naar de rechter overschrijdingskans. Men noemt de toets dan ook respectievelijk links- of rechtseenzijdig. Je herkent een eenzijdige toets aan de 'ongelijk'-tekens in de hypothesen. De hypothesen van een tweezijdige toets bevatten daarentegen 'is-gelijk'- en 'is-ongelijk'-tekens.

Wie tweezijdig toetst, moet de tweezijdige overschrijdingskans bepalen. Dat is echter meestal lastig of onmogelijk. In de praktijk werkt men meestal omgekeerd en *vergelijkt men bij een tweezijdige toets de eenzijdige overschrijdingskans met ½α*.

### *Toetsingsgrootheden en kritieke gebieden*

Voor het uitvoeren van een toets moet de overschrijdingskans van een *toetsingsgrootheid* worden bepaald (zoals k of z voor de binomiaaltoets, $\chi^2$ voor de $\chi^2$-toets of F voor een variantieanalyse). Voor z en k kan dat met behulp van de tabellen B of C uit het boek. Zulke tabellen zijn echter gauw erg omslachtig en omvangrijk. Daarom bestaan er ook tabellen met alleen de waarden van waaraf de toetsingsgrootheid significant is. Men spreekt van *kritieke waarden*. De tabellen D tot en met J zijn volgens dat principe uitgevoerd. Als een toetsingsgrootheid *op of voorbij* de kritieke waarde komt, luidt de toetsuitslag 'significant' en wordt de nulhypothese dus verworpen. Zodra een in de steekproef gevonden toetsingsgrootheid de kritieke waarde evenaart, is de kans daarop kleiner dan α. Men zegt dan dat de toetsingsgrootheid in het *kritieke gebied* ligt.

Voor het uitvoeren van een toets kunnen dus twee wegen worden bewandeld, die overigens op hetzelfde neerkomen:

1 het berekenen van een overschrijdings*kans* p, die wordt vergeleken met de gekozen *kans* α;
2 het vaststellen van de kritieke *waarde* van de toetsingsgrootheid die hoort bij de gekozen α; vervolgens nagaan of de *toetsingsgrootheid* (berekend) uit de steekproef deze kritieke waarde evenaart.

De toetsuitslag is voor beide wegen uiteraard hetzelfde. Toch blijkt het in de praktijk verwarring te wekken. De eerste methode dwingt je er altijd toe de nulhypothese te verwerpen als de gevonden *kans* kléiner is dan α. De tweede leidt er vaak toe, dat je de nulhypothese verwerpt als de *toetsingsgrootheid* gróter is dan de kritieke waarde. Computerprogramma's als SPSS werken door overschrijdingskansen te geven met de eerstgenoemde methode.

## 2. De binomiaaltoets

- Wordt *toegepast* in situaties waarin herhaalde *gelijke* kansprocessen elk slechts twee mogelijke uitkomsten hebben (zoals: kruis/munt, goed/fout, monster A/monster B).
- Toepassing in hier behandeld sensorisch onderzoek: paarsgewijze vergelijking, driehoekstest, 3-AFC-test, tetrad-test, duo-triotest en twee-uit-vijftest.
- Hierbij geeft n het aantal proefpersonen/tests aan en π de kans per keer op een bepaalde uitkomst (meestal de gokkans bij het niet waarnemen van een verschil). De toetsingsgrootheid is k, dat is het aantal keren dat een bepaalde uitkomst is opgetreden.
- Tabel D en tabel E tonen de kritieke waarden van k voor respectievelijk π = ½ en π = ⅓ voor diverse waarden van n.
- In situaties waarin de tabellen D en E niet voorzien, kunnen voor enkele waarden van π en n in tabel C eenzijdige linker overschrijdingskansen worden opgezocht. (Bij die tabel staat aangegeven hoe daaruit rechter overschrijdingskansen kunnen worden afgeleid. Voor tweezijdige toetsing wordt α gehalveerd.)
- In situaties waarin de tabellen C, D en E niet voorzien, kunnen eenzijdige overschrijdingskansen via de *normale verdeling* worden benaderd (mits zowel nπ als n(1–π) minstens 10 bedraagt). Daarvoor wordt z berekend met de formule:

$$z = \frac{k \pm 1/2 - n\pi}{\sqrt{n\pi(1-\pi)}}$$

  De zogenoemde continuïteitscorrectie van ½ in de teller is negatief voor de rechter overschrijdingskans en positief voor de linker. De rechter overschrijdingskans van de berekende z kan worden opgezocht in tabel B, de linker kan daarvan worden afgeleid. Voor tweezijdige toetsing wordt α gehalveerd.

***Voorbeelden:***

- *Driehoekstest (dus $\pi = 1/3$) met ($n =$) 40 panelleden en $k = 22$ (goede antwoorden).* Toetsing met $\alpha = 5\%$ (langs elk van de drie genoemde wegen):
  - Tabel E laat zien dat $k = 19$ genoeg is om de nulhypothese te verwerpen. Met $k = 22$ is er dus een significant verschil.
  - Uit tabel C (bij $n = 40$ en $\pi = 1/3$) valt af te leiden dat de rechter overschrijdingskans 100% – 99,61% = 0,39% bedraagt. Deze is kleiner dan $\alpha$, dus wordt de nulhypothese verworpen.
  - De normale benadering mag zo nodig worden toegepast. Invullen van de formule levert op: $z = (22–1/2–40/3)/\sqrt{(40\times1/3\times2/3)} = +2{,}74$. Tabel B laat zien dat hier een overschrijdingskans bij hoort van 0,31%. Deze is kleiner dan $\alpha$, dus wordt de nulhypothese verworpen. (Merk het verschil op van deze benadering met de exacte werkwijze volgens tabel C.)
- *Paarsgewijze vergelijking (dus $\pi = 1/2$) met ($n =$) 76 panelleden en $k = 45$.* Tweezijdige toetsing met $\alpha = 1\%$.
  Tabel D noch tabel C voorziet in $n = 76$. De normale benadering kan en mag echter worden toegepast. $z = (45–1/2–76/2)/\sqrt{(76\times1/2\times1/2)} = +1{,}49$. Tabel B laat zien dat hier een rechter overschrijdingskans bij hoort van 6,81%. Deze is groter dan ½$\alpha$, dus de nulhypothese wordt niet verworpen. Er kan geen verschil worden aangetoond.

## 3. De $\chi^2$-toets (chi-kwadraat-toets)

- Wordt – als het om sensorisch onderzoek gaat – *toegepast* in situaties waarin gegevens bestaan uit *absolute frequenties* die zijn ondergebracht in een ‘kruistabel’. Het gaat dan om het verband tussen twee variabelen die beide op nominaal niveau zijn gemeten.
- Toepassing in hier behandeld sensorisch onderzoek: A/niet-A-test en eenvoudige verschiltest. Hiervan worden de uitkomsten in 2-bij-2-tabellen ondergebracht.
- De gegevens staan in een kruistabel met werkelijk gevonden frequenties W. Daarnaast wordt een even grote tabel met verwachte frequenties V opgezet. Deze V’s zijn de (niet op hele getallen afgeronde) frequenties die de ‘eerlijkste’ verdeling van de variabelen over elkaar weergeven, ervan uitgaande dat de nulhypothese waar is en er dus geen enkel verband tussen de variabelen bestaat.
- De V’s worden berekend aan de hand van de *randtotalen* van de tabel met W’s: de V van elke cel bereken je door het bijbehorende kolomtotaal te vermenigvuldigen met het bijbehorende rijtotaal, en dit product te delen door het algemene totaal:

| **Werkelijke frequenties $W_1$ t/m $W_4$:** | | |
|---|---|---|
| | | |
| $W_1$ | $W_2$ | $A = W_1 + W_2$ |
| $W_3$ | $W_4$ | $B = W_3 + W_4$ |
| | | |
| $C = W_1 + W_3$ | $D = W_2 + W_4$ | $n = A + B\ (= C + D)$ |

| **Verwachte frequenties $V_1$ t/m $V_4$:** | | |
|---|---|---|
| | | |
| $V_1 = (A\times C)/n$ | $V_2 = (A\times D)/n$ | A |
| $V_3 = (B\times C)/n$ | $V_4 = (B\times D)/n$ | B |
| | | |
| C | D | n |

- De omvang van een kruistabel wordt getypeerd door het aantal zogeheten *vrijheidsgraden (degrees of freedom), afgekort tot df*. Df is het aantal kolommen min 1 vermenigvuldigd met het aantal rijen min 1. Voor een 2-bij-2-tabel geldt dus df = 1.
- Voor df = 1 is voor de berekening van $\chi^2$ een *continuïteitscorrectie* nodig. Hiervoor wordt elke W met ½ in de richting van de bijbehorende V gebracht. Als W groter is dan V wordt W dus met ½ verminderd, als W kleiner is wordt er ½ bij opgeteld.
- De toetsingsgrootheid $\chi^2$ wordt berekend met de formule:

$$\chi^2 = \sum \frac{(W-V)^2}{V}$$

- Tabel F toont de kritieke waarden van $\chi^2$ voor verschillende waarden van df en α voor zowel een- als tweezijdige toetsing.
- $\chi^2$ kan nooit negatief zijn. Voor een eenzijdige toets moet daarom nog goed naar de kruistabel worden gekeken om na te gaan of een significante waarde van $\chi^2$ wel overeenkomt met de in $H_1$ geformuleerde richting van het verband.
- Voorbeeld: de A/niet-A-test van paragraaf 6.10.

| **Werkelijke frequenties W:** | | | |
|---|---|---|---|
| **Gegeven:** | **Antwoord panellid:** | | |
| | **'M'** | **'Niet M'** | **Totaal** |
| merk M | 15 | 15 | 30 |
| merk P | 9 | 21 | 30 |
| Totaal | 24 | 36 | 60 |

| **Verwachte frequenties V:** | | |
|---|---|---|
| **Antwoord panellid:** | | |
| **'M'** | **'Niet M'** | **Totaal** |
| 12* | 18 | 30 |
| 12 | 18 | 30 |
| 24 | 36 | 60 |

| **Gecorrigeerde W's:** | | |
|---|---|---|
| **Antwoord panellid:** | | |
| **'M'** | **'Niet M'** | **Totaal** |
| 14,5 | 15,5 | 30 |
| 9,5 | 20,5 | 30 |
| 24 | 36 | 60 |

* = 30×24/60 enzovoort

$\chi^2 = (14{,}5-12)^2/12 + (15{,}5-18)^2/18 + (9{,}5-12)^2/12 + (20{,}5-18)^2/18 = 1{,}74.$

Tabel F toont voor eenzijdige toetsing (df = 1) voor α = 1% en α = 5% achtereenvolgens 5,41 en 2,71 als kritieke waarden voor $\chi^2$. De gevonden $\chi^2$ is kleiner, de resultaten zijn dus niet significant (afwijkend van het toeval), de nulhypothese blijft staan, er kan niet worden aangetoond dat liefhebbers van pilsmerk M pils van hun merk kunnen onderscheiden van dat van merk P.

## 4. *Rangordetoets*

- Wordt – als het om sensorisch onderzoek gaat – *toegepast* in situaties waarin gegevens bestaan uit de volgordes van door panelleden op basis van een of ander criterium gerangordende monsters of producten. Het gaat daarbij dus om drie of meer producten.
- Het is een toets die rechtstreeks op de getabelleerde gegevens wordt toegepast zonder berekening van een enkele toetsingsgrootheid. In paragraaf 6.12 worden uitgebreid de procedure en het gebruik van tabel G beschreven aan de hand van een voorbeeld.

## 5. *De t-toets*

- Wordt *toegepast* in situaties waarin moet worden bepaald of *gemiddelden* verschillen.
- Toepassing bij sensorisch onderzoek: daar waar gegevens het berekenen van gemiddelden en standaarddeviaties toelaten. Met name mogelijk na gebruik van lijnschalen en soms mogelijk geacht bij de verschil-met-referentietest.
- Er zijn drie varianten:
  1. een variant waarmee wordt nagegaan of het gemiddelde in één steekproef afwijkt van een bepaald gegeven getal;
  2. een variant waarmee wordt nagegaan of het gemiddelde in een steekproef afwijkt van dat van een daarvan *onafhankelijke* andere steekproef;
  3. een variant waarmee wordt nagegaan of het gemiddelde in een steekproef afwijkt van dat van een daarvan *afhankelijke* andere steekproef (waarbij dus sprake is van gepaarde waarnemingen).
- De toetsingsgrootheid is t. Deze kan worden berekend met de volgende formules:
  - voor variant 1: $t = \dfrac{\bar{x} - a}{s/\sqrt{n}}$

    Hierbij is $\bar{x}$ het gemiddelde over n gegevens, s is er de standaarddeviatie van en a is het gegeven getal.

  - voor variant 2: $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{(n_1-1)s_1^{\,2} + (n_2-1)s_2^{\,2}}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

    De indices 1 en 2 geven de betreffende steekproef aan.

  - voor variant 3: $t = \dfrac{\bar{d}}{s_d/\sqrt{n}}$

    Hierbij is d het verschil tussen de meetwaarden van *elk paar*. Er zijn dus evenveel d's als gegevensparen. Het gemiddelde van al die d's is $\bar{d}$, terwijl $s_d$ de standaarddeviatie van de d's is.

- Van belang is verder ook hier het aantal vrijheidsgraden, df.
  Hiervoor geldt:
  voor variant 1: $df = n - 1$
  voor variant 2: $df = n_1 + n_2 - 2$
  voor variant 3: $df = n - 1$ (waarbij n het aantal *paren* van waarnemingen is!)

- Tabel H toont de kritieke waarden van t voor diverse waarden van α en df. Merk op dat t zowel negatief als positief kan zijn. Om $H_0$ te verwerpen moet bij een eenzijdige toetsing de gevonden t soms juist positief, soms juist negatief zijn, afhankelijk van de hypothesen.
- Toepassing van de toets is gebonden aan bepaalde voorwaarden, die vooral de normaliteit van de verdelingen en de variantie van de verdelingen betreffen. Zie daarvoor meer statistisch gespecialiseerde literatuur.
- Voorbeelden van de toepassing:
  - *Variant 1.* Een onderzoeker heeft een panel van 81 consumenten gevraagd de textuur van een product (hedonisch) te beoordelen op een bipolaire lijnschaal (veel te zacht/veel te hard) van 100 mm. ‘Precies goed’ zit in het midden, op 50 mm dus. Hij vindt een gemiddelde van 55 mm, wat er op duidt dat het product wat te hard zou worden bevonden. De standaarddeviatie van de scores bedraagt 18 mm. De vraag is of 55 significant (met α = 5%) boven het ‘precies goed’-punt van 50 ligt. Dus:
    $H_0: \mu \leq 50$
    $H_1: \mu > 50$

    De gevonden t bedraagt $(55–50)/(18/\sqrt{81}) = +2{,}50$. Df = 81–1 = 80. De kritieke waarde van t ligt dan tussen 1,658 en 1,671 en wordt dus overschreden. $H_0$ wordt daarom verworpen, $H_1$ geaccepteerd.

  - *Variant 2.* Een sensoricus wil (met α = 5%) weten of mannen en vrouwen verschillen in hun oordeel over een product. Een panel van 40 mannelijke en 60 vrouwelijke consumenten geeft een totaaloordeel in de vorm van een rapportcijfer: mannen gemiddeld een 7,1 (s = 1,2), vrouwen een 7,7 (s = 1,4).
    $H_0: \mu_{mannen} = \mu_{vrouwen}$
    $H_1: \mu_{mannen} \neq \mu_{vrouwen}$

    De gevonden t bedraagt:

    $$\frac{7{,}1 - 7{,}7}{\sqrt{\frac{(40-1)1{,}2^2 + (60-1)1{,}4^2}{60+40-2}\left(\frac{1}{60}+\frac{1}{40}\right)}} = -2{,}220$$

    (Let op de volgorde van de bewerkingen.) Df = 60 + 40 – 2 = 98. De kritieke waarde van t ligt dan tussen –1,980 en –2,000. $H_0$ wordt dus verworpen, $H_1$ geaccepteerd.

- *Variant 3.* Een kwaliteitsonderzoeker wil met een analytisch panel van 16 personen met behulp van lijnschalen van 100 mm (0 = helemaal niet zout, 100 = heel sterk zout) nagaan of product A zouter smaakt dan B ($\alpha$ = 5%). Hij trekt voor elk panellid diens score van product B af van diens score van product A. Aldus ontstaat een serie van 16 verschilscores, met een gemiddelde van *min* 1,8 mm en een standaarddeviatie van 3,6 mm. Dus:

  $H_0$: $\mu_A \leq \mu_B$
  $H_1$: $\mu_A > \mu_B$

  De gevonden t bedraagt $-1{,}8/(3{,}6/\sqrt{16}) = -2{,}000$. Df = 16–1 = 15. De kritieke waarde van t is hier **plus** 1,753. $H_0$ wordt dus niet verworpen, $H_1$ wordt niet geaccepteerd.

## 6. Variantieanalyse

- Een ingewikkelde techniek, waar we hier maar globaal op ingaan.
- Eigenlijk gaat het om een uitbreiding van de t-toets, die wordt *toegepast* in situaties waarin moet worden bepaald of *meer dan twee gemiddelden* verschillen.
- Wordt in sensorisch onderzoek toegepast als bij de t-toets, maar dan als er meer dan twee gemiddelden in het geding zijn. De toets heet ook wel F-toets.
- De toetsingsgrootheid F (of F-ratio) wordt als volgt berekend:

$$\text{F-ratio} = \frac{\text{TKS}/(r-1)}{\text{BKS}/(n-r)}$$

  Waarbij r het betrokken aantal producten is, n het totaal aantal waarnemingen/scores is in het onderzoek (dus over alle r producten *samen*), BKS de zogeheten binnenkwadratensom (de som van de gekwadrateerde afwijkingen van elke score ten opzichte van zijn eigen productgemiddelde) en TKS de tussenkwadratensom (de som van de gekwadrateerde afwijkingen van de productgemiddelden ten opzichte van het algemene gemiddelde, nadat elk van die kwadraten vermenigvuldigd is met het bijbehorend aantal waarnemingen).
- Er zijn bij elke F twee soorten vrijheidsgraden in het geding: een voor de teller en een voor de noemer. Df bedraagt voor de teller r–1, voor de noemer n–r.
- Tabel I bevat kritieke waarden van F voor verschillende combinaties van vrijheidsgraden.
- Variantieanalyse is, door de aard van de hypotheses, een tweezijdige toets. Dat wil zeggen dat er is-gelijktekens (=) in de nulhypothese staan.

# Bijlage 3

## De relatie tussen de omvang van de R-index en d′

(Overgenomen van: Patricia Elliott, in: J.A. Swets: Signal Detection and Recognition by Human Observers. John Wiley, New York, 1964, page 683.)

| R | d' |
|---|---|
| 50% | 0,00 |
| 51% | 0,04 |
| 52% | 0,07 |
| 53% | 0,11 |
| 54% | 0,14 |
| 55% | 0,18 |
| 56% | 0,21 |
| 57% | 0,25 |
| 58% | 0,28 |
| 59% | 0,32 |
| 60% | 0,36 |
| 61% | 0,40 |
| 62% | 0,43 |
| 63% | 0,47 |
| 64% | 0,51 |
| 65% | 0,54 |
| 66% | 0,60 |
| 67% | 0,62 |
| 68% | 0,66 |
| 69% | 0,71 |
| 70% | 0,74 |
| 71% | 0,78 |
| 72% | 0,82 |
| 73% | 0,86 |
| 74% | 0,90 |

| R | d' |
|---|---|
| 75% | 0,95 |
| 76% | 1,00 |
| 77% | 1,05 |
| 78% | 1,09 |
| 79% | 1,14 |
| 80% | 1,19 |
| 81% | 1,24 |
| 82% | 1,29 |
| 83% | 1,34 |
| 84% | 1,40 |
| 85% | 1,47 |
| 86% | 1,53 |
| 87% | 1,60 |
| 88% | 1,66 |
| 89% | 1,74 |
| 90% | 1,81 |
| 91% | 1,90 |
| 92% | 1,98 |
| 93% | 2,08 |
| 94% | 2,19 |
| 95% | 2,32 |
| 96% | 2,48 |
| 97% | 2,66 |
| 98% | 2,90 |
| 99% | 3,28 |

# Bijlage 4

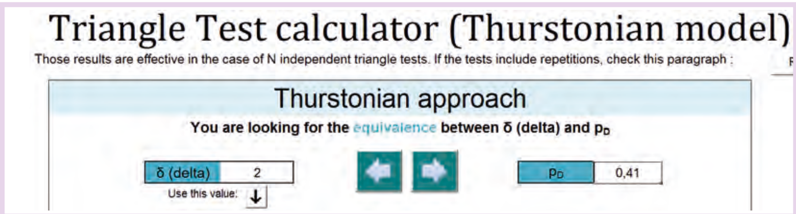
## Werken met V-Power

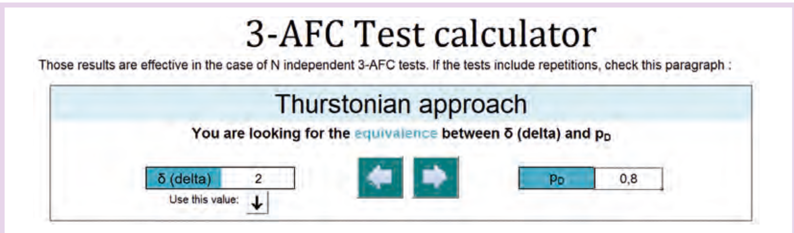
Via de website senstools.com kun je een gratis Excelbestand, ‘V-Power’ genaamd, downloaden. Je moet even ervaren hoe het werkt, en bijvoorbeeld ontdekken dat een ingevoerd getal pas wordt opgenomen als de entertoets is gebruikt. V-Power maakt geen onderscheid tussen $\delta$ en $d'$, en gebruikt steeds de Griekse letter. Hier en daar zitten er wat onvolkomenheden in het programma, maar met wat gewenning en de kennis die in hoofdstuk 6 is geboden kom je daar wel uit. We hopen dat V-Power nog lang beschikbaar is op deze wijze, want het is een veelzijdig en handig hulpmiddel.

### *De relatie tussen $\delta$ en $p_d$ (en $p_c$) nagaan*

Met V-Power kun je voor een driehoekstest, 3-AFC-test, duo-triotest, paarsgewijze vergelijking (door V-Power 2-AFC-test genoemd) en tetradtest nagaan hoeveel ‘discriminatoren’ (= ‘gecorrigeerde’ proportie juiste detecties, $p_d$) je gemiddeld kunt verwachten, uitgaande van een bepaalde omvang van $\delta$. Kies daarvoor bij de gekozen testmethode voor ‘Thurstonian model’. In het werkblad kun je een waarde voor $\delta$ invoeren, waarna je de betreffende $p_d$ terugkrijgt. Omgekeerd kun je een waarde voor $p_d$ invoeren om het bijbehorende productverschil af te lezen.

De volgende figuur toont voor zowel de driehoekstest als de 3-AFC-test het betreffende schermfragment. In beide gevallen is voor $\delta$ het getal 2 ingevuld. Deze waarden komen overeen met die van de verticale rode lijnen in figuur 6.15.




Misschien wekt het verbazing dat de in deze schermen teruggegeven waarden 0,8 en 0,41 niet overeenkomen met wat er in figuur 6.15 is af te lezen. Dat komt doordat V-Power de $p_d$ geeft, wat dus het aantal discriminatoren betreft. Figuur 6.15 bevat echter de kans op een correct antwoord, wat neerkomt op het te verwachten aandeel van correcte antwoorden, $p_c$. Met de geregeld in hoofdstuk 6 voorkomende reken- en redeneerwijze kun je $p_d = 0{,}80$ (= 80%) omrekenen tot $p_c = 86{,}7\%$ (afgerond). Immers: als 80%

het proeft, moet 20% gokken, waardoor nog eens 20/3 = 6,7% extra het goede antwoord geeft.

### *Het betrouwbaarheidsinterval rond een gevonden d' bepalen*

Voor de resultaten van een driehoekstest, 3-AFC-test, duo-triotest, paarsgewijze vergelijking en tetradtest kun je op een werkblad (te bereiken via 'guessing model') nagaan hoe groot $p_d$ is. Tegelijk kun je voor elk gewenst betrouwbaarheidsniveau aflezen wat het betrouwbaarheidsinterval rond $p_d$ is. Met de onder het vorige kopje besproken optie vertaal je vervolgens zowel $p_d$ als de grenzen van het betrouwbaarheidsinterval rond $p_d$ naar de bijbehorende waarden van d′.

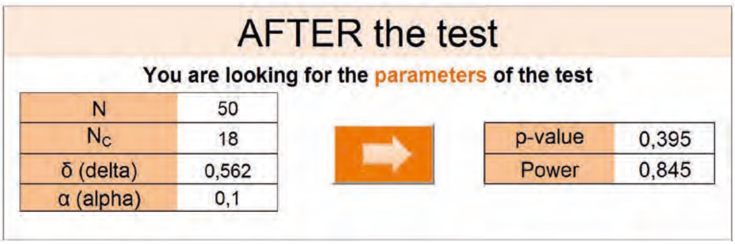
### *Na een test de overschrijdingskans en de power bepalen*

(Dit geldt voor de driehoekstest, 3-AFC-test, duo-triotest, paarsgewijze vergelijking en tetradtest. Voor de A/niet-A-test en de eenvoudige verschiltest gaat onderstaande op met soms wat andere maten voor de grootte van het verschil.)

Je kunt in V-Power testuitslagen invoeren, waarna je de overschrijdingskans terugkrijgt. Hiermee kun je dus, zonder formules in te vullen en zonder de tabellen achter in dit boek te raadplegen, nagaan of een verschil significant is.

Daarnaast kun je vaststellen hoe hoog de power van de test is. Als er geen significant verschil is, mag je, als de power maar hoog genoeg is en er dus wel goed genoeg naar een verschil is gezocht, concluderen dat er geen (groot) verschil is tussen de geteste producten.

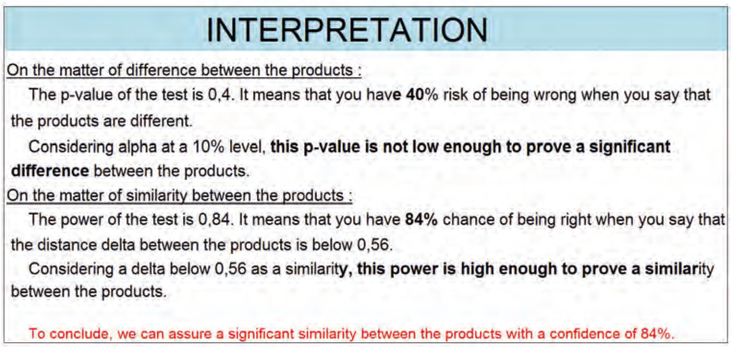
Om deze functies te illustreren, toont de volgende figuur twee fragmenten van een V-Power-werkblad, zoals dat is gebruikt ten behoeve van een 3-AFC-test.


AFTER the test

You are looking for the parameters of the test

| | |
|---|---|
| N | 50 |
| $N_C$ | 18 |
| δ (delta) | 0,562 |
| α (alpha) | 0,1 |

| | |
|---|---|
| p-value | 0,395 |
| Power | 0,845 |


INTERPRETATION

On the matter of difference between the products :

The p-value of the test is 0,4. It means that you have **40%** risk of being wrong when you say that the products are different.

Considering alpha at a 10% level, **this p-value is not low enough to prove a significant difference** between the products.

On the matter of similarity between the products :

The power of the test is 0,84. It means that you have **84%** chance of being right when you say that the distance delta between the products is below 0,56.

Considering a delta below 0,56 as a similarity, **this power is high enough to prove a similarity** between the products.

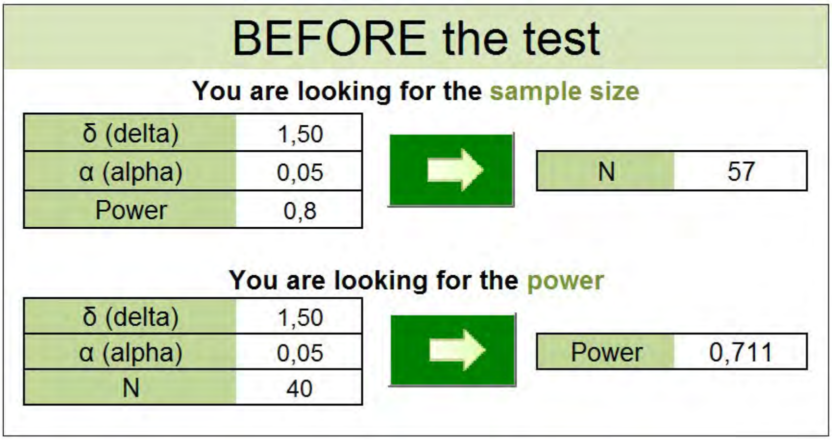
To conclude, we can assure a significant similarity between the products with a confidence of 84%.

In het bovenste fragment ('After the test') zijn links de panelomvang (hier 50), het aantal correcte detecties (hier 18) en de gekozen α (hier 10% oftewel 0,1) ingevoerd. De gekozen δ is door het programma overgenomen na eerdere invoer door de gebruiker van $p_d = 0{,}25$. Rechts van de oranje drukknop met de pijl staat de uitkomst van de toets. De overschrijdingskans p blijkt 0,395 oftewel 39,5% te zijn. Het aantal juiste detecties is dus verre van significant: een verschil is niet aantoonbaar. Is er dan goed genoeg naar een verschil gezocht? De power bedraagt 0,845, oftewel 84,5%. Als je ervan uitgaat dat een power van 80% voldoende is, is er dus goed genoeg gezocht en is de conclusie dat de producten niet (sterk) verschillen. Deze interpretatie van de uitkomst staat verwoord in het onderste fragment, 'Interpretation'.

### *Voorafgaand aan een test op gelijkheid bepalen hoeveel panelleden je nodig hebt of hoe groot de power is*

Met V-Power kun je nagaan hoe groot in een test op gelijkheid je panel moet zijn, uitgaande van het maximaal aanvaardbare verschil (uitgedrukt in $p_d$ of δ), de te hanteren α en de gewenste power. Omgekeerd kun je nagaan hoe groot de power is, uitgaande van het maximaal aanvaardbare verschil (zoals uitgedrukt in $p_d$), de te hanteren α en de (beoogde) panelomvang.

Het bovenste deel van de volgende figuur, die betrekking heeft op een driehoekstest, laat zien dat je bij een voorgenomen α van 5%, een gewenste power van 80% en een maximaal aanvaardbaar productverschil van δ = 1,5 een panel van (minstens) 57 personen nodig hebt.




Het onderste deel maakt duidelijk dat je, als je een panel hebt van 40 personen, bij dezelfde δ en α een power bereikt van 71,1%.

Deze werkwijze geldt voor de driehoekstest, 3-AFC-test, duo-triotest, paarsgewijze vergelijking en tetradtest. Voor de A/niet-A-test en de eenvoudige verschiltest is iets soortgelijks mogelijk, maar in plaats van δ wordt φ (phi) als effectmaat gehanteerd. Deze wordt in het boek niet behandeld, maar V-Power geeft globaal aan wat φ zegt over de omvang van een productverschil.