Antwoorden

bij (de vijfde druk van) Proeven van succes

Joep Brinkman

vijfde, herziene druk

Antwoorden bij (de vijfde druk van) Proeven van succes

1.1
a. hedonisch;
b. analytisch;
c. analytisch;
d. analytisch (want het gaat er niet om wat men lekker vindt, maar wat – of hoe groot – de afwijkingen zijn);
e. het eerste deel is analytisch, het tweede hedonisch.

1.2
b. beschrijvend;
c. discriminatief;
d. discriminatief en/of beschrijvend;
e. het eerste deel beschrijvend; het tweede deel discriminatief.

2.1
Detectiedrempel, herkenningsdrempel, bovendrempel. (De verschildrempel is relatief, en kan dus niet met de andere drempels worden vergeleken. Die zijn absoluut.)

2.2
Volgens de wet van Weber is niet het absolute maar het relatieve verschil van belang. Het relatieve verschil is in het eerste geval 100%, in het tweede slechts 10%.

2.3
a. 0,006 g/100 ml.
b. 0,008 g/100 ml.
c. 26% en 22%.
d. Het absolute verschil in concentratie is in het tweede paar weliswaar groter, maar volgens de wet van Weber gaat het om het relatieve verschil.
e. 0,036 + 0,26×0,036 = 0,045 g/100 ml.

2.4
Tussen 10,00 mg/l en 14,40 mg/l. (De Weberfractie is kennelijk 20%. 12 is 120% van 10; 120% van 12 is 14,40.)

2.5
Zie tekstkader ‘Sensorische luchtverontreiniging’ in §2.2.4.

2.6

Overeenkomsten:

- De prikkels komen binnen via chemoreceptoren.
- Er zijn ‘opgeloste’ moleculen nodig.

Verschillen:

- Het reukzintuig is gevoelig voor bijzonder veel kwaliteiten en het smaakzintuig voor slechts (vier of) vijf.
- Het reukzintuig werkt met in gas opgeloste of met gas gemengde moleculen, het smaakzintuig met in vloeistof opgeloste.
- Het reukzintuig kan waarnemingen mengen.
- Het reukzintuig werkt direct en veel sterker in op het ‘emotionele deel’ van de hersenen.

2.7

Bijvoorbeeld: karamels kennen een grote cohesie, maar kunnen zeer zacht zijn.

2.8

Geur. Dat komt doordat er zoveel geuren zijn, ze goed kunnen mengen en het reukzintuig uiterst gevoelig is terwijl de werking ervan amper bekend is.

3.1

Voorbeeldantwoorden voor alleen de gevallen a, b en c:

a. Consumentenpanel van minstens 80 à 100 leden. Het moeten in ieder geval tieners zijn. Er valt over te twisten of deze tieners al gebruikers van chocolademelk moeten zijn, of dat het (ook) om potentiële gebruikers mag gaan. (Is het doel bijvoorbeeld gebruikers van andere merken over te halen of om huidige niet-gebruikers aan te trekken?) Geen relevante allergie of dieet.

b. Een geselecteerd en min of meer getraind panel, van minimaal 10 à 12 personen. Ze moeten worden geselecteerd op in het algemeen goed kunnen proeven, kunnen samenwerken en verbaal vaardig zijn. Ze hoeven niet van kaas(wafels) te houden, maar mogen er ook geen afkeer van hebben. Geen relevante allergie of dieet.

c. Eventueel het panel van vraag b. Voor deze taak zijn overigens geen bijzondere sociale en verbale vaardigheden nodig. Het is ook denkbaar dat hier een consumentenpanel wordt ingezet (van bijvoorbeeld 100 volwassen gebruikers van het product), het gaat er wellicht uiteindelijk om of de consument het verschil kan proeven!

3.2

Opmerking vooraf: de moraal van deze opgave is dat geschiktheid van ‘proevers’ niet zozeer of alleen een persoonskenmerk is, maar vooral of ook afhangt van de paneltaak!

a. Bij dit onderzoek gaat het niet om paneldiscussies enzovoort, maar vooral of alleen om de basissmaak zout. A en E lijken dus geschikt. B eventueel ook.

b. Voor het boven tafel krijgen van termen zijn paneldiscussies nodig en zijn dus verbale en sociale vaardigheden van de panelleden vereist. Wat dat betreft komen dus C, D en E in aanmerking. Verder is het kennen, herkennen en benoemen van aroma’s van belang.
Wat dat betreft doet E het erg goed, en zou ook D (mede vanwege zijn positievere score op geurherkenning) in aanmerking kunnen komen.

c. Nu is vooral het vermogen aroma's (en smaken) te *leren* herkennen van belang. Dan zijn B en C en E dus geschikte kandidaten. C is geschikt, al doet deze het niet zo goed op basissmaken. Als er geen paneldiscussies nodig zijn, is B geschikt (ook al vanwege diens score op geurherkenning). E heeft het voordeel dat hij ook nog goed scoort met de basissmaken.

4.1

Voorbeeldantwoord voor a:
Batchvariatie:
Onder andere: verschil in hoeveelheid vet, vetverdeling, hoeveelheid bot, dikte (met gevolgen voor de garing als het gaat om de bereiding), versheid; het vlees kan afkomstig zijn van verschillende varkens, een net even ander deel of uitsnede van het dier; de varkens kunnen op verschillende momenten zijn geslacht.
Variatie door aanbieding en bereiding:
Onder andere: verschillen in kruiding, hoeveelheid vet waarin wordt gebakken, baktijd, baktemperatuur, temperatuur waarop het product wordt beoordeeld, uitsnede.

4.2

(n! =) 1 × 2 × 3 × 4 × 5 = 120 panelleden.

4.3

a.
32965 93443287 1668074 1600668933504551893250687
BB**EDC** **EBC**CBBE**D** **AD**D**E**AD**C** ADAADDEEBBCACCCAEEBBCADED

Panellid 1 krijgt de producten in volgorde BEDCA.
Panellid 2 krijgt de producten in volgorde EBCDA.
Panellid 3 krijgt de producten in volgorde ADECB.

b. Dat kan van alles zijn, afhankelijk van de koppeling van cijfers aan letters. Let er daarbij in ieder geval wel op dat elke letter maar twee cijfers krijgt! Het gevolg daarvan is dat twee cijfers niet gebruikt worden. Hier gaan we uit van de volgende afspraken:
Bij 0 en 1 hoort monster A.
Bij 2 en 3 hoort monster B.
Bij 4 en 5 hoort monster C.
Bij 6 en 7 hoort monster D.

32965 93443287 1668074 1600668933504551893250687
BB-**DC** -**BC**CBB-**D** **AD**D-AD**C** ADAADDEEBBCACCCAEEBBCADED

Panellid 1 krijgt de producten in volgorde BDCA.
Panellid 2 krijgt de producten in volgorde BCDA.
Panellid 3 krijgt de producten in volgorde ADCB.

c. Daar zit veel systeem in: sommige cijfercombinaties komen heel vaak voor, andere helemaal niet.

4.4

(Dit zijn niet per se de enige goede antwoorden!)

a. zoute haring, filet americain, fricandel speciaal;
b. brood, pudding;
c. saté, patat;
d. in eerst instantie natuurlijk avocado om een dipsaus te maken, de dipsaus zou vervolgens weer tortillachips als drager kunnen hebben.

5.1

- *Onacceptabele bijsmaak*, *geschikt voor bij de borrel* en *nare vetsmaak* betreffen hedonische oordelen (wat betreft *geschikt voor bij de borrel* gaat het misschien niet om een zuiver hedonische uitspraak, maar deze is wel subjectief en normatief).
- *Vetgehalte* betreft geen smaak of ander sensorisch attribuut, maar slaat op de hoeveelheid van een bestanddeel.
- *Oud* is een verwijzing naar een mogelijke oorzaak.
- *Koud* is geen eigenschap die eigen is aan het product maar eerder met de aanbieding ervan te maken heeft. Koud is daarmee bovendien niet onderscheidend.
- *Frissig* is halfslachtig: waarom niet *fris* of *frisheid*? (Merk op dat *ranzig, hartig* en misschien zelfs *romig* geen halfslachtige termen zijn, de bedoelde attributen heten gewoon voluit zo.)
- *Zuur-ranzig* is dubbelhartig.
- *Kleur* is hier een zeer vaag begrip: gaat het om de intensiteit/sterkte van de totale kleurprikkel, om de mate waarin de olijf groen of zwart of donker is of om de aard van de kleur?
- *Romigheid* lijkt niet echt op een relevant en onderscheidend attribuut betrekking te hebben.
- *Hartigheid en zoutheid* zouden wel eens op hetzelfde attribuut betrekking kunnen hebben, en dus verre van orthogonaal zijn.
- ....

5.2

Hier zijn geen eenduidige antwoorden mogelijk. Beschouw de volgende antwoorden als voorbeelden of suggesties.

a.
Hoe sterk is de zoetheid van deze cake?

1. heel zwak
2. tamelijk zwak
3. tamelijk sterk
4. heel sterk

b.
Wat vindt u van de zoetheid van deze cake?

veel te weinig zoet | precies goed | veel te veel zoet

c.
Hoe beoordeelt u het uiterlijk van deze truffels?

heel erg onaantrekkelijk	heel erg aantrekkelijk

d.
Wat vindt u van de geur van deze appelstroop?
1. heel erg vies
2. erg vies
3. vies
4. enigszins vies
5. noch vies, noch lekker
6. enigszins lekker
7. lekker
8. erg lekker
9. heel erg lekker

e.
Hoe sterk is de *rooksmaak* nadat u dit product geheel hebt doorgeslikt?

(1 = heel zwakke rooksmaak, 7 = heel sterke rooksmaak)
1 -------- 2 -------- 3 -------- 4 -------- 5 -------- 6 -------- 7

f.
Proeft u de soep uit alle vier de koppen, en zet ze alle vier op volgorde van het minst romig naar het meest romig. Neem, wanneer u daarmee klaar bent, de codes hieronder over:

Het minst romig is ______.
Dan komt ______.
Het een na romigste is ______.
Het meest romig is ______.

5.3

a Het linkeranker is zuiver analytisch, dat aan de rechterkant behoort bij een JAR-schaal (en houdt tevens een subjectief oordeel in).
b Flauw en pittig zijn niet (automatisch) elkaars tegengestelde. *Flauw* betekent (meestal) *weinig zout* of soms ook *weinig sterke* smaak.
c De eerste drie antwoordmogelijkheden zijn analytisch, de laatste is hedonisch.
d De antwoorden zijn niet in balans: er zijn twee negatieve antwoorden (1 en 2, met *onaangenaam*) en drie positieve (4, 5 en 6, met *aangenaam*).

5.4

Panelleden A en B komen elk voor de producten gemiddeld op 65. Zij blijken dus op hetzelfde *level/niveau* te scoren. Toch hebben A en B een geheel verschillend patroon: A is 'uitbundiger' en geeft veel hoge naast veel lage scores (die lopen van 32 tot en met 98: 66 punten verschil tussen de hoogste en de laagste), terwijl B veel minder reliëf heeft in zijn oordelen (van 49 tot en met 85, 36 punten verschil). Zij verschillen dus in *spreidingsbreedte/range*.

C scoort gemiddeld 45, veel lager dan A en B. Hij heeft dus een ander, lager, *level/niveau* dan zijn collega's. Zijn *spreidingsbreedte/range* ligt tussen die van A en B in.

5.5

Het verschil tussen A en B bedraagt 8 punten. Dat is 2 maal de standaarddeviatie (die immers 4 bedraagt), zodat $\delta_{AB} = 2$. Op vergelijkbare wijze kom je tot $\delta_{AC} = 5$ en $\delta_{BC} = 3$.

5.6

a. (De volgende rijtjes zijn niet uitputtend.)
Toevallige fouten: Het ene monster komt van boven uit de pan, en heeft daardoor misschien een wat andere temperatuur dan monsters die van onder komen, het is ook wat minder vet en minder vochtig, wellicht blijven er meer of minder kruiden aanhangen. Sowieso verschillende monstertemperaturen.
Systematische fouten: Wellicht verschil in hoogte/hitte van de pitten waarop A en B zijn gebakken. A en B zijn wellicht niet even rul gemaakt. B is misschien kouder dan A bij aanbieding en heeft langer gestaan. Volgorde-effecten doordat er niet is gebalanceerd. Het gehakt kan voor beide pannen enigszins verschillend zijn samengesteld.

b. De test zou al een stuk verbeteren in bijvoorbeeld de volgende opzet (waar ongetwijfeld aanvullingen en verbeteringen op mogelijk zijn):

Monsterbereiding
Indien toepasbaar kun je al veel problemen voorkomen door zo lang mogelijk het ongekruide gehakt in één pan, pan A, te bereiden. Dit gehakt moet op een gegeven moment natuurlijk wel in twee gelijke porties worden verdeeld over de gelijke pannen B en C (die op een even hete kookplaat staan). Dat kan niet door zomaar de helft van boven uit A in B over te scheppen en de rest in C. Je kunt wel bijvoorbeeld een aantal keren *ombeurten* een lepel uit A in B en een lepel uit A in C overscheppen. B en C moeten precies dezelfde hoeveelheid gehakt bevatten. De kruiding wordt daarna toegevoegd, uiteraard tegelijk en in gelijke hoeveelheid, en op gelijke wijze gemengd en doorgeroerd.

Monsteraanbieding
De pannen B en C moeten even lang op dezelfde temperatuur worden gebracht en gehouden. Voor de eerste testronde wordt er steeds tegelijkertijd van dezelfde plaats uit pan B en pan C een monster geschept. De monsters uit pan B gaan naar de ene helft van het panel, de monsters uit pan C tegelijkertijd naar de andere helft ervan. In de volgende ronde krijgt de eerste helft van het panel een monster uit pan C, de tweede een monster uit pan B. De monsters worden in dezelfde panellidvolgorde uitgedeeld als in de eerste ronde. Deze procedure wordt in volgende sessies herhaald, net zo lang tot alle panelleden aan de beurt zijn geweest.

6.1

a. Analytisch.
b. Omdat het er uiteindelijk om zal gaan of de consument het alcoholarmere bier als anders of gelijk ervaart. Blijkbaar bevindt het product zich al in een vergevorderde fase van ontwikkeling.
c. Het zou kunnen dat het makkelijker is om B of juist M als afwijkend te herkennen.
d. Er is niet de ideale testopzet gehanteerd: M is tweemaal zo vaak het afwijkende monster als B. Er zijn geen drie, maar zes driehoeken mogelijk. Die moeten dan ook allemaal (even vaak) gebruikt worden.
e. 40 goed en 20 fout.

6.2

a. Ja, want volgens tabel E zijn 30 juiste antwoorden al genoeg voor verwerping van de nulhypothese.
b. 20 panelleden hebben het fout. Deze moeten het dus fout hebben gegokt. Je verwacht tweemaal zoveel fout-gokkers dan goed-gokkers. Daarom gokten, ‘statistisch gezien’, ook nog eens (ongeveer) 10 panelleden goed. Dan hebben ‘gemiddeld’ 30 personen van de 60 moeten gokken, ofwel 50%. De overige 30 personen (hier dus eveneens 50%) zouden het verschil dus hebben kunnen proeven. Dit is het gecorrigeerde aantal.
c. $n_c = 40$, $n_d = 30$, $p_c = 40/60 = 2/3 = 67\%$ en $p_d = 30/60 = 1/2 = 50\%$.

6.3

a. Voor u staan drie monsters bier. Eén daarvan heeft een sterkere alcoholsmaak dan de andere twee. Proef de monsters en geef aan welke van de drie het bier met de sterkste alcoholsmaak is.
b. Er is een significant verschil aangetoond, want volgens tabel E zijn 38 juiste antwoorden al genoeg voor verwerping van de nulhypothese.
c. Blijkbaar is de doelstelling van de ontwikkeling dat het alcoholarme bier door consumenten niet van ‘gewoon’ te onderscheiden is. Het risico bestaat er daarom uit dat het alcoholarme bier wordt ‘goedgekeurd’ (H_0 wordt aangenomen) terwijl in werkelijkheid consumenten het verschil wél kunnen proeven. Om dit risico te verminderen moet dus de kans dat H_1 (‘consumenten proeven verschil’) wordt aangenomen juist groot worden. In deze opzet is die kans groter dan in de situatie van opgave 6.1 door maar liefst drie maatregelen: er is gekozen voor een 3-AFC-test, een groter panel en een grotere α. Merk op dat we hier komen op de problematiek van het testen op gelijkheid, zoals genoemd in de inleiding van dit hoofdstuk.

6.4

a. Nee, want de referentie is hier steeds gelijk. Wanneer de referentie gebalanceerd zou zijn, zou voor elk van de vier tests in (ongeveer) de helft van de gevallen het bewaarde product als referentie moeten fungeren.
b. Niet alleen moet afwisselend de referentie dan wel het te testen koekje het eerst of het laatst worden geproefd (dus links of rechts op het dienblad en de vragenlijst staan); het is eveneens belangrijk dat de vier subtests (de vier bewaartijden) in volgorde variëren.
c. Met een α van 5% zijn 18 juiste detecties genoeg om de nulhypothese te verwerpen (tabel D (eenzijdig!)). Bij een α van 1% ligt dat aantal op 19. Voor een bewaartijd van 20 weken is de uitkomst voor beide alfa's significant. Bij een alfa van 5% is de derde conditie ook significant, maar bij 1% niet. De bewaartijd van 12 weken levert sowieso geen significante uitslag op. Merk op dat de uitkomst bij een bewaartijd van 8 weken niet significant is: er zijn weliswaar 18 gelijke (foute) antwoorden, maar de toets is eenzijdig naar de andere kant!
   Wat betreft de keuze van α: met een lage α wordt de afwijking niet gauw significant. Dan loop je kans op de omgekeerde fout (verderop in boek zal blijken dat die de β-fout wordt genoemd). Dan wordt er een onterecht lange houdbaarheidstermijn op de verpakking gezet. Het risico is dan dus dat consumenten ontevreden worden. Vandaar is er veel te zeggen voor een hoge α, misschien zelfs wel voor 10%.

6.5

40% van 60 = 24 personen kunnen het verschil echt proeven en geven dus sowieso het goed antwoord. De resterende (60 – 24 =) 36 panelleden proeven het niet en moeten dus gokken. (Ongeveer) de helft daarvan zal het juiste monster gokken, wat dus nog eens (36/2 =) 18 goede antwoorden oplevert. In totaal zijn dat (24 + 18 =) 42 goede antwoorden, ofwel 70%.

6.6

a. Er zijn 4 aanbiedingsmogelijkheden. Er zijn 6 sessies nodig met steeds 8 panelleden. Hij moet dus zorgen dat elke volgorde in elke sessie tweemaal voorkomt.
b. Significant, want volgens tabel D zijn 31 juiste antwoorden genoeg voor verwerping van de nulhypothese.
c. 18 (want tegenover de (48 – 33 =) 15 fout*gegokte* antwoorden staan ook circa 15 goed*gegokte* antwoorden).

6.7

a. Analytisch.
b. 11 + 13 = 24.
c. Volgorde B-A komt onterecht veel meer (25 keer) voor dan A-B (15 keer). Het lijkt erop dat bij de volgorde A-B naar verhouding veel vaker het goede antwoord wordt gegeven, wat dus duidt op een mogelijk volgorde-effect.
d. Volgens tabel D (eenzijdig, want je wilt weten of B zoeter wordt gevonden en niet of er überhaupt een verschil in zoetheid wordt waargenomen) zijn er 26 juiste antwoorden nodig om H_0 te verwerpen. De uitkomst is 24 van de 40, dat is niet significant.

6.8

Merk om te beginnen op dat de toets tweezijdig is. Het gaat er namelijk niet om vast te stellen of een bepaalde worst (A óf B) de voorkeur heeft, maar of er überhaupt een verschil in voorkeur bestaat.

a. Nee.

b. Tabel D voorziet niet in kritieke waarden voor deze n. De toets kan wel worden uitgevoerd met de normale benadering van de binomiaaltoets. (Zie zo nodig de statistische bijlage in het boek.) Dan is (met continuïteitscorrectie) $z = (k - ½ - n\pi) / \sqrt{(n\pi(1-\pi))}$ $= (88 - ½ - 150\times½) / \sqrt{(150\times½\times½)} = 2{,}04$. De eenzijdige overschrijdingskans is (volgens tabel B) 2,07%. De tweezijdige overschrijdingskans bedraagt het dubbele ervan, 4,14%. Deze ligt onder α. De uitkomst is significant: A wordt lekkerder gevonden. (NB Het is netter α te halveren (tot 2,5%), en de *een*zijdige overschrijdingskans daarmee te vergelijken. Maar de uitkomst is natuurlijk gelijk.)

6.9

a. Omdat hij niet weet waarin een eventueel verschil precies kan of zal zitten, kan hij geen gerichte test uitvoeren.

b. Voor een beschrijving van de *instructie* en een *formulier*: zie boek.
Aanbiedingsschema (deze zoveel mogelijk spreiden over sessies / tijdstippen):
30 keer N als referentie, aanbiedingsvolgorde N-O.
30 keer N als referentie, aanbiedingsvolgorde O-N.
30 keer O als referentie, aanbiedingsvolgorde N-O.
30 keer O als referentie, aanbiedingsvolgorde O-N.
Denk wat betreft de *monsterbereiding en -aanbieding* onder andere aan uitsnede van de haring, versheid, moment van 'schoonmaken', temperatuur (van bewaren en aanbieden), eventueel gebruik van drager (roggebrood?) of het toevoegen van uitjes. Monsters moeten *gecodeerd* worden.

6.10

a. 5 panelleden krijgen O-O-N-N.
5 panelleden krijgen O-N-O-N.
5 panelleden krijgen O-N-N-O.
5 panelleden krijgen N-N-O-O.
5 panelleden krijgen N-O-N-O.
5 panelleden krijgen N-O-O-N.

b. Voor u staan vier kokosmakronen, met elk een codenummer. De bedoeling is dat u deze monsters indeelt in twee groepjes van twee, zo dat de koeken *binnen* deze groepjes zoveel mogelijk op elkaar lijken wat betreft hun smaak, geur en mondgevoel.
Proeft u de makronen eerst in de volgorde zoals ze van links naar rechts op het dienblad staan. (Daarna mag u ze opnieuw proeven zo vaak en in de volgorde die u zelf wenst.)

c. De toetsing verloopt als bij de driehoekstest. Tabel E laat zien dat (voor $n = 30$ en $\alpha = 5\%$) 15 goede antwoorden nodig zijn om H_1 aan te nemen. Het zijn er in deze test 16, dus is de uitkomst significant: statistisch is er is een merkbaar verschil aangetoond.

d. Tegenover de 14 *fout*gokkers staan 7 *goed*gokkers. Van de 16 panelleden met het goede antwoord zouden er dan 7 gegokt hebben, en 9 het dus hebben geproefd: $p_d = 9$ van de 30 = 30%.

6.11

a. Als bij opgave 6.10, maar dan steeds 4 panelleden per volgorde.

b. ‘Voor u staan vier porties mayonaise, met elk een codenummer. Twee daarvan kunnen zouter smaken dan de andere twee. Welke zijn volgens u die twee *meest zoute* porties? Proeft u de mayonaise ...’

c. De hypothesen luiden:
H_0: $\pi \leq 1/6$
H_1: $\pi > 1/6$
Hier moet een binomiaaltoets met $\pi = 1/6$ worden toegepast. De werkwijze is vergelijkbaar met het gebruik van tabel C in het boek. Die voorziet echter niet in $\pi = 1/6$.
Volgens de bij de opgave opgenomen tabel hoort bij (k =) 8 (!) goede antwoorden een linker overschrijdingskans van 98,82%. Dan is de rechter overschrijdingskans voor (k =) 9 (!) goede antwoorden 100% - 98,82 = 1,18%. Deze p-waarde is groter dan α, zodat de uitkomst niet significant is en H_0 blijft staan. Er kan geen verschil worden aangetoond.
(Omdat $n\pi = 24\times(1/6) = 4$ kleiner is dan 10 mag de binomiale verdeling hier niet worden benaderd door een normale.)

6.12

(Zie zo nodig de statistische bijlage in het boek voor de formule en de berekeningswijze van χ^2.)
De gevonden frequenties (W) staan in de volgende tabel:

Gegeven monster	Antwoord panellid		Totaal
	‘Het is S’	‘Het is niet S’	
S	11	4	15
X	6	9	15
Totaal	**17**	**13**	**30**

De verwachte frequenties (V) zijn dan:

Gegeven monster	Antwoord panellid		Totaal
	‘Het is S’	‘Het is niet S’	
S	(15×17/30 =) 8,5	6,5	15,0
X	8,5	6,5	15,0
Totaal	**17,0**	**13,0**	**30,0**

De voor continuïteit gecorrigeerde werkelijke frequenties zijn:

Gegeven monster	Antwoord panellid		Totaal
	‘Het is S’	‘Het is niet S’	
S	(11 – 0,5 =) 10,5	4,5	15,0
X	6,5	8,5	15,0
Totaal	**17,0**	**13,0**	**30,0**

Voorbeeldberekening voor de bijdrage aan χ^2 van de cel linksboven (uitgaande van de gecorrigeerde W): $(W-V)^2/V = (10{,}5-8{,}5)^2/8{,}5 = 4/8{,}5 = 0{,}47$.

Wanneer je voor elk van de vier cellen de overeenkomstige uitkomsten optelt, kom je op de totale χ^2: 2,17. Deze ligt onder de kritieke waarde van 2,71. Er is dus geen significant verschil.

6.13

a. Er bestaan vier aanbiedingen, die in zo gelijk mogelijke mate moeten plaatsvinden. Dus bijvoorbeeld 13 keer KK, 13 keer CC, 13 keer KC en 14 keer CK. Deze volgordes moeten weer zo veel mogelijk over de sessies worden gespreid.

b. χ^2 (berekend met continuïteitscorrectie) = 4,32, wat significant is.

6.14

Een volgorde-effect zou eruit bestaan dat voor de ene aanbieding de kans op het goede antwoord anders is dan voor de andere aanbieding. Het meest gezoete monster is B. Voor de volgorde A B is ‘2e’ dus het goede antwoord, voor wie de monsters kreeg in volgorde B A is ‘1e’ correct. Het uitgangspunt van de toets is dan ook de volgende tabel. (Merk op dat die dus anders is dan de tabel van opgave 6.7.)

Aangeboden volgorde	Antwoord	
	Fout	Correct
A – B	4	11
B – A	12	13

Merk verder op dat de toets tweezijdig wordt uitgevoerd, omdat er geen verwachting vooraf is uitgesproken welke volgorde tot de meeste correcte antwoorden zou leiden.

Berekening (met continuïteitscorrectie) leidt tot een χ^2 van 1,00. Deze ligt veel lager dan de kritieke χ^2 van 3,84. Een volgorde-effect kan hiermee niet worden aangetoond (wat nog niet betekent dat zo’n effect niet bestaat).

6.15

De kolomtotalen zijn respectievelijk 24, 16, 29 en 31 (ter controle: samen is dit 100, wat inderdaad 10 (panelleden) keer 10 (‘punten’ per persoon) is.)

Volgens tabel G zijn de grenswaarden 17-33 (er bestaat namelijk geen verwachting over de volgorde). Kolomtotalen van 16 of lager en van 34 of hoger wijzen dus op een significante uitkomst. Alleen speculaasje B voldoet hieraan: dit wordt significant lekkerder gevonden.

6.16

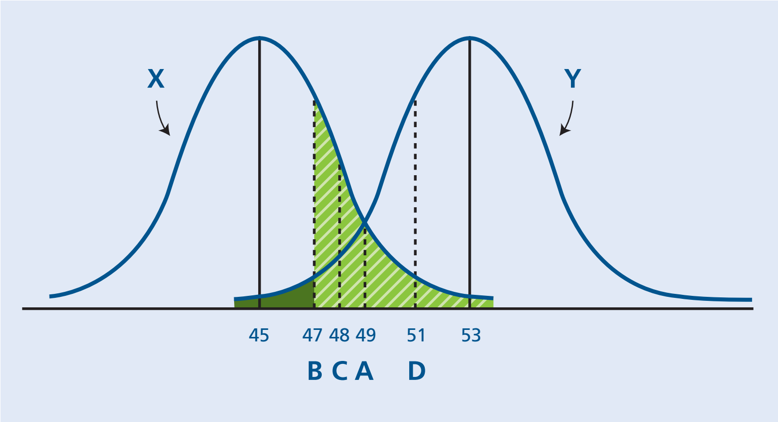
Hier past een t-toets voor onderling afhankelijke (ofwel herhaalde) waarnemingen. Daarvoor worden eerst de verschilscores per panellid berekend (in de vierde kolom).

Panellid	Score		Verschilscore (T–R)	(verschilscore – 2)²
	Testproduct	Referentie		
A	6	3	3	$(3–2)^2 = 1^2 = 1$
B	5	3	2	0
C	6	2	4	4
D	4	2	2	0
E	5	5	0	4
F	6	2	4	4
G	4	3	1	1
H	7	5	2	0
I	4	1	3	1
J	2	3	–1	9
K	4	2	2	0
Totaal			22	24
Gemiddeld			(22/11=) 2	
Variantie				= 24/(n–1) = 24/10 = 2,4
Standaard-deviatie				$\sqrt{2,4}$ = 1,55

Deze verschilscores zijn gemiddeld 2. De variantie van de verschilscores bedraagt 2,4 (zie de rechterkolom, waarin steeds van de verschilscore de gemiddelde verschilscore van 2 is afgetrokken). De standaarddeviatie van de verschilscores, s_d, is de wortel daaruit, en bedraagt 1,55. De formule van t (zie statistische bijlage) is $t = \bar{d}/(s_d/\sqrt{n})$. Dat levert op: $t = 2/(1,55/\sqrt{11}) = 4,280$. Tabel H laat zien dat voor een eenzijdige toets, de gegeven α en df = 11 – 1 = 10 de kritieke waarde voor t op 1,812 ligt. De uitkomst is dus significant: de onderzochte partij smeerkaas wijkt af van de referentie.

6.17

a. (Zie eventueel ook figuur 6.10 in het boek.) Bijgaande tekening laat zien dat A een neutraal criterium hanteert, B en C zijn daarin liberaal (B meer dan C), terwijl D hier streng is.



b. Het gaat om normale verdelingen, zodat de statistische tabel B achter in het boek kan worden geraadpleegd zodra je over z-scores beschikt. Als voorbeeld worden de z-scores van persoon B voor beide verdelingen berekend. B's grens van 47 ligt 2 punten boven het gemiddelde van de verdeling van X. Dat is 0,5 keer de standaarddeviatie boven het gemiddelde, dus bedraagt $z_{\text{van B voor verdeling X}}$ +0,5. Tabel B laat zien dat dit correspondeert met een kans van 30,85%. Voor persoon B geldt kennelijk dat wanneer hij product X krijgt er een kans van 30,85% is dat hij dit ervaart als 47 of sterker, en dus (zoals is gegeven) dat hij het monster benoemt als ongelijk aan de referentie. Zijn antwoord is dan fout. De kans op een vals alarm bedraagt daarom 30,85%. (Zie ook het gearceerde stuk onder de verdeling van X in de tekening.)
De grens van persoon B ligt op 47, dat is 6 punten onder het gemiddelde van de verdeling van Y. Dat is 1,5 keer de standaarddeviatie onder het gemiddelde, dus bedraagt $z_{\text{van B voor verdeling Y}}$ −1,5. Tabel B laat zien dat dit correspondeert met een *linker* overschrijdingskans van 6,68%. Voor persoon B geldt kennelijk dat wanneer hij product Y krijgt er een kans van 6,68% is dat hij dit ervaart als 47 of minder, en dus (zoals is gegeven) dat hij het monster benoemt als de referentie. Dat antwoord is dan fout. De kans op een 'miss' bedraagt daarom 6,68%. (Zie ook het gearceerde stuk onder de verdeling van Y in de tekening.)

Voor personen A, C en D verlopen de berekeningen overeenkomstig. Dat leidt tot de volgende kansen:
A: kans op vals alarm = kans op miss = 15,87%.
C: kans op vals alarm = 22,66%, kans op miss = 10,56%.
D: kans op vals alarm = 6,68%, kans op miss = 30,85%.

6.18

a. Elk product is 15 maal aangeboden. Daarmee komt de n in tabel J op 15 uit. De toets is eenzijdig, omdat er steeds een 'correct' antwoord is. In tabel J wordt dan voor $n = 15$ en $\alpha = 5\%$ de waarde '16,42' afgelezen. Dit betekent dat een R-index van 50 + 16,42 = 66,42% of hoger significant is. Deze kritieke waarde wordt niet overschreden, dus is verschil niet overtuigend aangetoond.
b. Net als bij a is de toets eenzijdig. Je moet steeds uitgaan van het product dat het minst is aangeboden, dus bedraagt n hier 22. Tabel J leidt tot de waarde '18,64'. De kritieke waarde is daarom 68,64%, de uitkomst is significant: een verschil is aangetoond.
c. De toets is ook nu natuurlijk eenzijdig, maar door de berekeningswijze is de R-index blijkbaar onder de 50% uitgevallen. Tabel J levert voor $n = 45$ en $\alpha = 5\%$ de waarde '9,82' op. Volgens de werkwijze van de tabel kan dit getal nu van 50% worden afgetrokken, wat leidt tot een kritieke waarde van 40,18%. Uiteraard moet nu de gevonden R-index *onder* de kritieke waarde liggen om significant te zijn. Dat is het geval: een verschil is hiermee aangetoond.

6.19

a. 19
b. $P(k<19) = P(k\leq 18)$ = (volgens tabel C, $n = 40$ en $\pi = 0{,}5$(!)) 31,79%. Dat is dus de kans om H_0 te laten staan. Dat is dus ook de kans β, want H_0 is niet waar. De power is dan 100% – 31,79% = 68,21%.
c. 12 (= 10%).
d. Toetsen met de normale benadering van de binomiale verdeling levert op:
$z = (k - \frac{1}{2} - n\pi) / \sqrt{(n\pi(1-\pi))} = (48 - \frac{1}{2} - 120\times 1/3) / \sqrt{(120\times 1/3\times 2/3)} = 1{,}45$.
De bijbehorende overschrijdingskans is 7,35%, dus wordt H_1 niet aangenomen.

6.20

40% proeft het. Dus 60% moet gokken, waarvan naar verwachting een derde deel het goed zal gokken. Verwacht wordt dus 40% + 1/3 van 60% = 60% goede antwoorden. Dus $\pi = 0{,}6$. Dan volgt uitwerking als bij 6.19b: $P(k<19) = P(k\leq 18) = 3{,}92\%$ (volgens tabel C, $n = 40$ en $\pi = 0{,}6$(!)). Dat is dus de kans om H_0 te laten staan. De power is dan 96,08%.

6.21

a. toepasbaar;
b. niet toepasbaar (want je weet niet waarop je het panellid zou moeten laten letten);
c. toepasbaar;
d. niet toepasbaar (want je weet niet waarop je zou moeten vergelijken);
e. toepasbaar;
f. niet toepasbaar (want je weet niet waarop je zou moeten groeperen);
g. kan in theorie, maar is niet toepasbaar omdat de test te veel eist van de proevers;
h. toepasbaar (maar omslachtig);
i. toepasbaar (maar omslachtig);
j. rangordening van twee is sowieso mal, maar het kan ook niet omdat je niet weet waarop je het panellid zou moeten laten rangschikken;
k. toepasbaar.

6.22

	Opgave a		Opgave b	Opgave c	Opgave d	
	p_c	n_c	Sign.?	p_d	n_c	Sign.?
Driehoekstest	51%	(0,51×20) = 10	nee	26,5%	20	ja
3-AFC	77%	15	ja	65,5%	31	ja
Duo-trio	67%	13	nee	34,0%	27	ja
Paarsgew.verg.	85%	17	ja	70,0%	34	ja

NB Hier zijn kleine afleesverschillen mogelijk.

6.23

-

6.24

a. 16 panelleden *gokten* blijkbaar fout. Daar staan er (circa) 8 tegenover die goed gokten. n_d is daarom 24 – 8 = 16. 16 is 40% van 40, zodat p_d = 40% = 0,40.
b. 1,98.
c. Dit interval loopt van 17% tot 60%.
d. Het interval loopt van 1,18 tot 2,71.
e. p_d = (50–26)/50 = 48%. d' = 1,96. Het interval rond p_d loopt van 28% tot 68%. Het interval rond d' van 1,36 – 2,67.

NB Hier zijn kleine afrondingsverschillen mogelijk.

6.25

-

6.26

a. 23,3%.
b. 95,4%.
c. 54,5%.

6.27

a. n = 57.
b. n = 65.
c. n = 20.

6.28

Als 1/3 het verschil proeft, moet 2/3 het antwoord gokken en gokt daardoor (1/2 × 2/3 =) 1/3 van het panel het antwoord goed. In dat geval geeft in totaal 2/3 het goede antwoord. Hij moet dus aantonen dat de proportie correcte antwoorden kleiner dan 2/3 is. Dus:

H_0: $\pi > 2/3$
H_1: $\pi \leq 2/3$

Volgens tabel C bedraagt de kans dat 21 of minder panelleden het juiste monster aanwijzen 4,41%. Deze overschrijdingskans is lager dan α = 10%. Het is significant: π ligt onder 2/3, waarmee 'gelijkheid' is aangetoond.

Merk op dat de spelregels niet tijdens het spel mogen worden veranderd, en er dus niet alsnog een α van 5% mag worden gekozen!

6.29

a. De p-waarde is 51,3%, de power bedraagt 55,5%. De uitkomst is dus: niet significant (verschillend). Doordat de power niet hoog genoeg is, is ook gelijkheid hier niet aangetoond.
b. De p-waarde is 48,1%, de power bedraagt 81,7%. De uitkomst is dus: niet significant (verschillend), terwijl de power hoog genoeg is: gelijkheid is hier aangetoond.
c. De p-waarde is 51,3%, de power bedraagt 95,1%. De uitkomst is dus: niet significant (verschillend), terwijl de power hoog genoeg is: gelijkheid is daardoor aangetoond. Merk op dat de gegevens gelijk zijn aan die van opgave a. De p-waarde is dan ook gelijk, maar nu is de power een stuk hoger!
d. De p-waarde is 2,2%, de power bedraagt (nog steeds!) 95,1%. De uitkomst is nu echter significant (verschillend). De power is dus hoog genoeg maar doet er nu niet meer toe, hier is immers juist *on*gelijkheid aangetoond. Anders gezegd: er is goed gezocht naar het verschil, zodat het inderdaad ook gevonden is.

7.1

a. Veel tijdsbesparing doordat er geen lijst met descriptoren hoeft te worden vastgesteld en er geen overeenstemming over de betekenis ervan hoeft te komen.
b. Het attribuut/de descriptor ligt hier al vast; niks free choice dus.

7.2

a. De verschillen lijken vooral kruidnagel, sinaasappel en zoetheid te betreffen.
b. -
c. - (Let erop dat de attributen op volgorde staan naar de omvang van het verschil!)
d. Ja. Gescoord wordt wat er geproefd wordt, niet wat erin zit. Zeker als een product citroen en/of limoen bevat, kan dat ook andere citrussmaken als sinaasappel tot gevolg hebben. Bovendien zul je zelden een gemiddelde van nul aantreffen. Dan moeten namelijk zonder uitzondering álle panelleden nul scoren.

7.3

Met de t-toets voor afhankelijke waarnemingen, maar dan moeten ook de varianties van de verschilscores en het aantal panelleden bekend zijn (of men moet over de ruwe gegevens beschikken, zodat die alsnog berekend kunnen worden).

7.4

a. A ‘wint’ 8 van de 11 keer, dit is niet significant.

b. (De gemiddelde score voor A is 7,0, die voor B 5,0. De varianties (s^2) zijn respectievelijk 4,60 en 3,00.) De betreffende algemene formule (zie bijlage 3 in het boek) luidt:

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

Ingevuld levert dit op:

$$t = \frac{7 - 5}{\sqrt{\left(\frac{(11-1) \times 4{,}60 + (11-1) \times 3{,}00}{11+11-2}\right)\left(\frac{1}{11} + \frac{1}{11}\right)}} = 2{,}406$$

Dus t = 2,406, wat voor een eenzijdige toets bij df = 20 niet significant is voor $\alpha = 1\%$ maar wel voor $\alpha = 5\%$.

c. (Het gemiddelde verschil is 2,0. De variantie (s^2) van de verschillen bedraagt 4,80. De standaarddeviatie (s) is $\sqrt{4{,}8}$.) Dan is t (= $2{,}0/(\sqrt{4{,}8}/\sqrt{11})$) = 3,028, wat voor een eenzijdige toets bij df = 10 significant is.

d. De uitkomst van de binomiaaltoets is helemaal niet significant. Die van de t-toets voor onafhankelijke steekproeven wel voor $\alpha = 5\%$ maar niet voor $\alpha = 1\%$. De t-toets voor afhankelijke steekproeven biedt zelfs voor $\alpha = 1\%$ een significante uitkomst. De kans op de β-fout neemt kennelijk af van a naar c (en de power neemt dus juist toe)!

e. Deplorabel! De scores liggen onacceptabel ver uiteen.

8.1

H_0: $\mu \leq 50$
H_1: $\mu > 50$
t = $(55-50)/(13{,}5/\sqrt{81})$ = 3,33. Kritieke t (eenzijdig, df = 80) = circa 2,37. De worst wordt dus significant te zout gevonden.

8.2

Hier past de t-toets voor onafhankelijke steekproeven. Invullen van $\bar{x}_1 = 55$, $n_1 = 81$, $s_1 = 13{,}5$ en $\bar{x}_2 = 65$, $n_2 = 81$, $s_2 = 13{,}5$ in de formule (zie bijlage) levert bij df = 160 een t op van 4,71. Dit is significant. (De toets is tweezijdig.)

8.3

Hier is sprake van gepaarde waarnemingen, dus past de derde variant van de t-toets. Toetsingsgrootheid t = $(7{,}1-6{,}9)/(0{,}5/\sqrt{144})$ = 4,800 (tweezijdige toets). Dit is significant: A wordt beter gewaardeerd.

9.1

a. Zoetheid, vruchtensmaak en frisheid. Zoetheid en frisheid zijn bijna tegengesteld. Vruchtensmaak correleert niet met zoetheid, wel enigszins met frisheid.
b. Vruchtensmaak hangt er sterk positief mee samen, frisheid enigszins. Er bestaat een zwakke negatieve samenhang met de zoetheid.
c. C.
d. A is een heel klein beetje frisser dan C, heeft minder vruchtensmaak en is minder zoet.

(Dit alles althans binnen dit (tweedimensionale) stelsel van factoren).

9.2

(Op deze vragen is geen laatste antwoord mogelijk. Merk verder op dat een dergelijk plaatje een bepaalde 'fit' moet hebben om de werkelijkheid te kunnen representeren, en dat die werkelijkheid bovendien vaak weerbarstiger is en zich grilliger gedraagt dan je met zo'n model zou willen.)

a. Wat betreft dit soort producten houden deze consumenten van een reep met de volgende kenmerken:
   - (relatief) flink knisperend (nog meer dan reep P_1);
   - een (relatief) nogal zoete smaak (iets zoeter dan reep P_1, maar iets minder zoet dan P_2);
   - een (relatief) weinig sterke vanillesmaak (nog minder dan reep P_1);
   - een (relatief) sterke bessensmaak (nog sterker dan de repen P_1 en P_2);
   - (relatief) erg ongelijkmatig van structuur.
b. P_4: dit ligt in geen enkel cluster, en ligt veraf van de clusters die er wel zijn. Grote kans dat dit mede ligt aan de (relatief) zeer zwakke bessensmaak, terwijl dit niet wordt gecompenseerd door attributen die voor consumenten ook belangrijk zijn.
c. P_3 ligt nu in alles zo'n beetje in het midden, waardoor het een soort allemansvriend is, maar niemands béste vriend. P_3 zou door naar beneden op te schuiven naar het witte gebied het grootste segment beter kunnen bedienen, beter dan welke concurrent ook. Als het wat naar rechts gaat, bijvoorbeeld naar de plaats waar de vanillevector het witte gebied raakt, onderscheidt P_3 zich nog steeds van P_1. De productontwikkelaar zou het product dan wat meer vanillesmaak moeten geven (meer dan P_5, maar minder dan P_4), meer gelijkmatig moeten maken (zoals P_2), en tikje minder bessensmaak moeten geven, minder laten knisperen (tot ongeveer het niveau van P_5). Aan de zoetheid hoeft niets te veranderen.