Instituut voor Humane Bewegingsfunctionaliteit (HBF)

Het gangbare wetenschappelijk onderzoeksmodel in de menswetenschappen

Prof. dr A. J. A. Verberk.

1. De kern van iedere onderzoeksvraag: de variabiliteit van X

Een verschijnsel X zal dán aanleiding geven tot onderzoek als het een variabel verschijnsel is.

> Als voorbeeld: Het verschijnsel X = prestatiemotivatie = interne prestatiedruk. Men spreekt van een hoge prestatiemotivatie als iemands wereldbeeld beheerst wordt door eisen van prestatie en concurrentie, van successen en mislukkingen.

Gezocht wordt naar een verklaring voor de variabiliteit van X: waarom bij de ene mens wèl en bij de andere niet; waarom bij de een sterk en bij de ander zwak; waarom vandaag wèl en morgen niet?

Oorzakelijke verklaring hoeft niet altijd het doel te zijn. Het doel kan ook zijn: hoe kan ik X voorspellen. Het verschil in doelstelling zal vanzelf duidelijk worden. Ik ga er voorlopig vanuit, dat het gaat om oorzakelijke verklaring.

Uitgangspunt is, dat de verschillen in X veroorzaakt worden door in invloed van ándere variabelen.

De kernvraag is dus:

(1) Door welke variabelen wordt de variabiliteit van X verklaard?

Bijna altijd is er sprake van een complex van oorzakelijke factoren; A, B, C…..etc. In de menswetenschappen lukt het nooit om alle bepalende factoren in een onderzoek te betrekken.Vaak beperkt een onderzoek zich rechtstreeks op maar één van die bepalende variabelen. Ik noem dat factor A. Alle andere vat ik samen in de term ‘overige’.

Als voorbeeld:

A = externe prestatiedruk; d. w. z. dat er vanuit de omgeving een min of meer sterke druk wordt uitgeoefend om te 'presteren' (b. v. in de opvoeding thuis of op school, op het werk. bij examens, bij een sollicitatieprocedure).

Een eerst benadering van de vraag of de factor a thuishoort in het complex van oorzakelijke factoren, is de vraag:

(2a) Is er een functionele relatie tussen X en A?

De hypothese is: X = f (A, overige)

Met het begrip 'functionele relatie' bedoelt men niet het begrip functie zoals in b.v. orgaanfuncties, maar het wiskundige begrip functie: gaan wijzigen van de variabele A samen met veranderingen in X, b.v. grotere externe druk prestatiedruk → grotere interne prestatiedruk.
Als er een functionele relatie is tussen X en A, wil dat zeggen:

(2b) dat er correlatie is tussen X en A.

Let wel: het bestaan van een functionele relatie (correlatie) is beslist onvoldoend om en causale relatie aan te nemen.
B.v. _de kamertemperatuur is een functie van de stand van de thermometer, maar wordt er niet door veroorzaakt;
_het salaris X van de directeur van een drank bestrijdingsbond is beslist niet een functie van de prijs van een liter jenever A.

Toch is de vraag naar een functionele relatie zinvol, omdat zo'n functionele σ^2 relatie een eerste en noodzakelijke (maar onvoldoende) voorwaarde is, wil er van een oorzakelijke relatie sprake zijn.

2. De variantie σ^2 als maat voor variabiliteit.

Of een gemeten eigenschap (meetprobleem!) bij een groep mensen meer of minder variabel genoemd moet worden, hangt af van het aantal verschillen en grootte van die verschillen tussen alle meetuitkomsten. In de statistiek kent men een maat om die variabiliteit van X in een groep van grootte N in een getal uit te drukken.

Dat is de variantie σ^2 (sigma kwadraat).

Afziende van de gebruikelijke definitie en berekeningsformules voor σ^2 , kan men daarbij het beste denken aan
De som van alle gekwadrateerde onderlinge verschillen.

- Dat zijn er, de verschillen nul natuurlijk meetellend, ½ N (N -1).

Als men die som deelt door N^2 krijgt men σ^2 . De deling door N^2 maakt dat de σ^2 ongevoelig is voor het aantal metingen. [Zie (17)].

- De kwadratering van de verschillen heeft als belangrijkste effect dat grote onderlinge verschillen voor de variabiliteitmeting veel zwaarder wegen dan kleine verschillen. B.v. één verschil van 6 verhoogt de σ^2 evenzeer als vier verschillen 3, omdat $6^2=4x3^2$.

- Bedenk dat men ook het verschil tussen slechts 2 meetuitkomsten als een variantie kan meten. Hun σ^2 is dan het kwadraat van hun verschil gedeeld door (zie b.v. pag. 6).

Daardoor kan men de correlatievraag X = f (A, overige) aldus formuleren:

(3) De grootte van de X-variantie ${}_x\sigma^2$ is gedeeltelijk een functie van veranderingen in A.

Als experimenteel en wiskundig hanteerbaar alternatief gebruikt men de volgende formulering van deze hypothese:

(4) Als de factor A constant gehouden wordt, zal de X-variantie kleiner worden. D.w.z.: een deel van de totale X-variantie ${}_x\sigma^2_{tot}$ is gebonden aan het variëren van de factor A.

Een minstens even belangrijke vraag is natuurlijk, hoe sterk die functionele relatie is tussen X en A, n.l.:
Hoeveel kleiner wordt de X-variantie, als A constant gehouden wordt?

[Duidelijk is, dat als A de enige oorzaak is voor X-verschillen (noodzakelijke en voldoende voorwaarde), dat dan de X-variantie nul moet worden voor X-metingen in een groep, waarin factor A constant is.]

Deze vraag wordt gewoonlijk geformuleerd in de volgende vorm:

(5) Hoe groot is het deel van de X-variantie ${}_x\sigma^2_{tot}$, dat gebonden is aan het variëren van factor A?

Het merkwaardige is, dat deze vraag over de sterkte van de functionele relatie bij het wetenschappelijk onderzoek zelfs, en zeker bij de popularisering van haar uitkomsten, heel vaak achterwege blijft, waardoor de ‘significantiemythe’ (zie verderop) steeds dieper geworteld wordt.

3. Het voorbeeld uitgewerkt.

Het voorbeeld X = f (A, overige) voor X = prestatiemotivatie (interne prestatiedruk) en A = externe prestatiedruk werk ik uit met verzonnen onderzoekgegevens. Die gegevens zijn geconstrueerd om een beeld te geven van een hoogst significante onderzoeksuitkomst. Het is bedoeld als een eerlijk beeld van de gebruikelijke wetenschappelijke basis voor een wetmatigheiduitspraak, niet te ongunstig en niet te gunstig.
X wordt gemeten door het aantal prestatiegerichte interpretaties bij een platentest.
De factor A wordt naar 2 klassen gevarieerd:

A^+ → groep 1 (200 pers.) die onder sterke externe prestatiedruk staan bij de test (b.v. tijdens een sollicitatie),

A- → groep 2 (200 pers.) waarbij geen sprake is van externe prestatiedruk.

Er voor wordt voor gezorgd, m.b.t. de ‘overige’ bepalende factoren, dat-

- ofwel de testsituatie voor beide groepen wat betreft een aantal belangrijk geachte invloeden identiek is, (bij dit onderzoek een rol spelen, met name ook of zij)

- (6) òfwel dat louter toeval bepaalt hoe die invloeden bij dit onderzoek een rol spelen, met name ook of zij in groep 1 een andere rol spelen dan in groep 2 (→ o.a. zorgen voor random-verdeling van de 400 proefpersonen over groep 1 en 2); het belang hiervan zal verderop blijken.

Veronderstel de volgende uitkomsten: zie verder, pagina 5.

Mogelijke X-waarden	Feitelijk waargenomen frequenties (f) bij:		
	constante A		Beide groepen samen
	groep 1 A^-	Groep 2 A^+	
X	f 1	f 2	f_{tot}
50	--	10	10
45	10	40	50
40	80	70	150
35	60	60	120
30	40	20	60
25	10	--	10
N	200	200	400
gem. X	36	39	37.5
var. σ^2	24.00	26.50	27.50

24.00 en 26.50

De X–verschillen binnen de A–groepen, gemiddeld: ½(24.00 + 26.50)=	gebonden aan ándere factoren dan A n.l. de 'overige':
${}_x\sigma^2_{binnen}$ = 25.25 →	${}_x\sigma^2_{overige}$

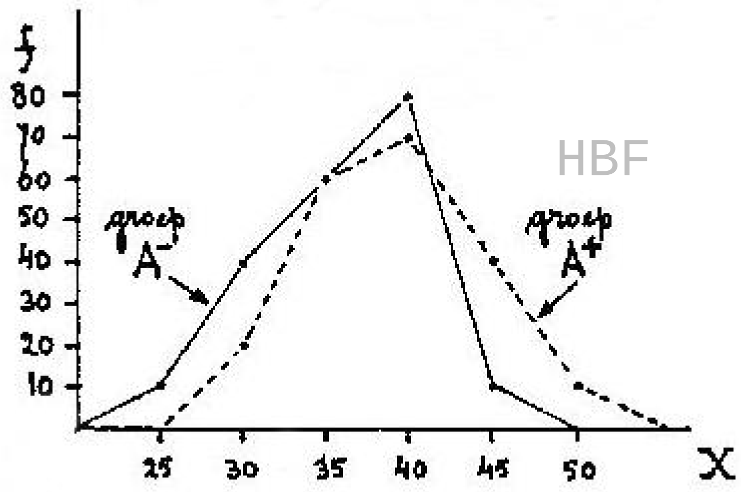
36 en 39

De X–verschillen tussen de groepen d.i. tussen de groepsgemiddelden:	gebonden aan factor A ($A^- \rightarrow A^+$)
${}_{x}\sigma^2_{tussen}$ = 2.25 →	${}_{x}\sigma^2_{A}$

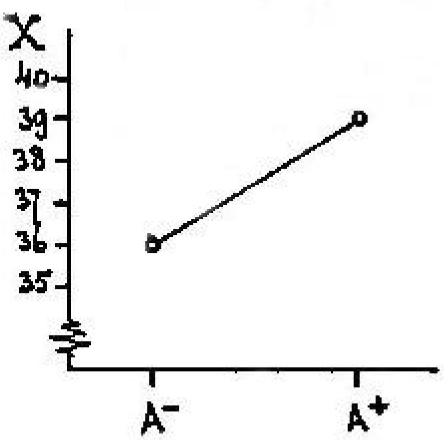
27.50

De X–verschillen in de totale proefgroep	De X-variabiliteit die om verklaring vraagt
${}_{X}\sigma^2_{tot}$ = 27. 50	

Zeldzaam gepubliceerde grafiek



Gewoonlijk gepubliceerde grafiek



Een belangrijke eigenschap van de variantie is de volgende stelling, die altijd geldt:

$$\sigma^2_{tot} = \sigma^2_{tussen} + \sigma^2_{binnen}$$

Dit betekent dat ${}_x\sigma^2_{tussen}$ en ${}_x\sigma^2_{binnen}$ twee complementaire delen zijn van de totale X-variantie en wel twee delen die precies beantwoorden aan de twee termen in de functiehypothese X = f (A, overige):

(7) ${}_x\sigma^2_{tussen} = {}_x\sigma^2_{A}$ (lees: de X-variantie voor zover gebonden aan A),

${}_x\sigma^2_{binnen} = {}_x\sigma^2_{overige}$ (lees: de X-variantie voor zover gebonden aan overige factoren).

Deze twee grootheden geven een direct antwoord op vraag (4).
Voor bovenstaand voorbeeld kunnen we dit uitbeelden in de volgende grafiek van de X-variabiliteit: (8a)

A	OVERIGE	
2.25 +	25.25	= 27.50

Het antwoord op vraag (5) ligt nu ook klaar en is al zichtbaar in de grafiek (8a): neem het percentage van de totale X-variantie (27.50) dat gebonden is aan de factor A (2.25).
Dat percentage noem ik R^2_{XA} .

Dus: R^2_{XA} = 2.25 : 27.50 = 0.0818 ofwel 8.2%

(8b) R^2_{XA} = het aan A gebonden procentuele deel van de X-variantie = ${}_x\sigma^2_{A} : {}_X\sigma^2_{tot}$

N.B. In heel veel onderzoek worden de in (7) aangegeven maten alléén maar heel impliciet berekend in zoverre zij nodig zijn voor de significantie-bepaling (zie hieronder) en worden de bewerkingen (8a) en (8b) niet uitgevoerd.

4. De correlatie tussen X en A.

Het percentage door A gebonden X-variantie heb ik R^2 genoemd, (zie opmerking) omdat de betekenis van deze maat identiek is met die van het kwadraat van een aantal maten, die bekend staan als :

(9) R_{XA} = een correlatiecoëfficiënt voor de functionele samenhang tussen X en A.

In het voorbeeld: $R_{XA} = \sqrt{R^2_{XA}} = \sqrt{0.0818} = 0.286$

Let op! Bijna steeds als men de correlatie tussen twee variabelen X en A weergeeft, vermeldt men de coëfficiënt R_{XA}.

(10) De in (8) gegeven interpretatie geldt alléén voor R^2 en niet voor R. Een gegeven correlatiecoëfficiënt mag men dus nooit interpreteren als 'percentage', maar moet men daartoe eerst kwadrateren. Velen weten dat niet! Let op! Door kwadratering wordt iedere R beduidend kleiner en in die zin geeft iedere R-coëfficiënt een zeer geflatteerd beeld van de sterkte van de samenhang. Veel weten dat niet!

Voor alle duidelijkheid het volgende lijstje: (11)

.10	.20	.30	.40	.50	.60	.70	.80	.90	.100	R
1%	4%	9%	16%	25%	36%	49%	64%	81%	100%	R^2

Opmerking: in meer gecompliceerde onderzoeken zijn de betekenissen niet altijd strikt identiek en kan het kwadraat van de correlatiecoëfficiënt zelfs kleiner zijn dan het percentage gebonden variantie. De strekking van de passages (10) en (11) blijft a fortiori geldig.

5. Gebonden variantie is (nog) geen verklaarde variantie.

Zeer vaak (ook in wetenschappelijke publicaties!) noemt men ${}_X\sigma^2_A$ de 'verklaarde' X-variantie en spreekt men dus bij R^2_{XA} = 0.8 over 8% 'verklaarde' X-variantie.

Dit is hoogst misleidend! Beide maten zeggen op zich alléén iets over de functionele relatie en een functionele relatie 'verklaart' op zich niets, daar 'verklaren' verwijst naar een oorzaak-gevolg relatie.

Er kan bij X = f (A, overige) een zeer hoge R^2_{XA} geconstateerd worden zonder dat X veroorzaakt wordt door A. Zie eerdere voorbeelden:

> 1) X = kamertemperatuur en A = stand thermometer. Een correlatie van 1.00 ! Niet A verklaart X, maar omgekeerd: X verklaart A. Men kan slechts X uit A 'voorspellen'.
>
> 2) X = salarishoogte en A = prijs jenever. Een vrij hoge correlatie, maar het salaris wordt niet groter omdat de jenever duurder wordt. Beide correleren, omdat beide een gevolg zijn van eenzelfde derde variabele: de conjunctuur (die in feite een groep van vele variabele samenvat). Wèl kan men ook hier van (zwakke) voorspellingsmogelijkheden spreken.

(12) Een correlatiecoëfficiënt geeft op zich geen enkele informatie over oorzakelijke samenhang. Wèl informeert de correlatiehoogte over voorspellingsmogelijkheden (zie verderop).

En:

(13) Een hoge correlatie tussen X en A blijkt vaak te berusten op het feit, dat beide (al of niet causaal) samenhangen met een derde (groep van) variabele(n) zonder dat er enige directe causale relatie is tussen X en A.

6. De significantie van een onderzoeksuitkomst; de significantiemythe.

Als er sprake is van ‘een significante onderzoeksuitkomst’, wordt dat steeds weer opnieuw geïnterpreteerd als een ‘belangrijke’ of ‘veelbetekende’ onderzoeksuitkomst. Dat is volkomen fout! Wat betekent ‘significant’ dan wel?

Neem het uitgewerkte voorbeeld: mijn onderzoek is beperkt tot een steekproef van:

1) 400 personen

2) die in één situatie

3) éénmaal gefunctioneerd hebben.

Zou het wellicht zó kunnen zijn, dat in het door mijn hypothese van

1) veel meer personen

2) in veel verschillende situaties

3) op allerlei reactiemomenten

er gemiddeld geen sprake is van enig A-effect d.w.z. ${}_{x}\sigma^{2}_{A} = 0$ en $R^{2}_{XA} = 0$?

(14) Men noemt dit de NULHYPOTHESE betreffende de onderzoeksuitkomst.

Als en in zoverre ik ervoor gezorgd heb [zoals het hoort! zie passage (6)], dat ik genoemde beperkingen van mijn steekproef louter ‘door de dobbelsteen’ heb laten bepalen, dan zou de NULHYPOTHESE slechts op basis van een toevalsgebeuren juist kunnen zijn.

(B.v. door toevalsspeling zijn de 400 proefpersonen nu juist mensen die extra gevoelig zijn voor A; of door toevalsspeling zijn bij de verdeling van die 400 over de 2 groepen, meer ‘hoge – X-mensen’ in groep 2 gekomen dan in groep 1.)

De juistheid van die NULHYPOTHESE kan ik dan onderzoeken met behulp van de wiskundige waarschijnlijkheidsrekening.
Dat gaat als volgt:
Ik bereken de toevalskans om het in deze steekproef geconstateerde A-effect te verkrijgen, als er in feite d.i. in de universumwerkelijkheid geen sprake is van enig A-effect. Met andere woorden dat mijn uitkomst louter een toevalstreffer is?

Als die kans zeer klein is ("geen schijn van kans"), verwerp ik de toevalsverklaring, d.w.z. dan verwerp ik de NULHYPOTHESE en noem ik de uitkomst SIGNIFICANT.
Gebruikelijke normen en termen:

· bij een kans van 5% of minder: significant

· bij een kans van 1% of minder: zeer significant.

Men spreekt daarbij over een ‘overschrijdingskans’ van 5% of 1% etc.
Dus:

> (15) "Een wetenschappelijk vastgestelde (hoogst) significante correlatie tussen X en A" betekent niet méér dan: "De correlatie tussen X en A is in werkelijkheid (hoogst) waarschijnlijk niet nul."
>
> Dat heeft dus NIETS te maken met: "de correlatie tussen X en A is ‘veelbetekenend’ of ‘belangrijk’".

Het moge duidelijk zijn, dat na het aantonen van de significantie pas de wetenschappelijke en praktisch belangrijke vragen aan de orde (dienen te) komen. De vragen, die – hoe ongelofelijk ook – vaak niet meer gesteld worden, zijn:

· Hoe groot is die waarschijnlijk niet – nul – zijnde correlatie?

[Zie (5)].

· Hoe zit het met de oorzakelijkheidstructuur bij deze functionele samenhang?

[Zie (12) en (13)].

· Wat kan men daarmee doen in de dagelijkse praktijk?

Het gebruikelijke tevreden achterover leunen van wetenschappers als het werk van de significantietoetsing achter de rug is; en de gebruikelijke, daarmee samenhangende misverstanden rond de ongelukkige term ‘significant’, construeren

> (16) de significantiemythe zowel in als buiten de wetenschap.

Daar komt nog het volgende bij:
Een belangrijke factor, die in de kansberekeningformules dominant beslist over de significantie van een onderzoeksuitkomst, is het aantal personen in de onderzoeksgroep. Dat betekent:

> (17) Iedere gemeten samenhang tussen variabelen, hoe zwak en onbetekenend ook, is hoogst significant als de steekproef maar groot genoeg is.

De uitkomst van ons voorbeeldonderzoek (R^2_{XA} = 8.2%) blijkt na berekening hoogst significant te zijn: de overschrijdingskans is kleiner dan 0.1%.
Als ik het onderzoek niet met 400 maar met 40 personen had uitgevoerd en dezelfde uitkomsten had gekregen (alle frequenties delen door 10 geeft dezelfde gemiddelden, dezelfde

varianties en dezelfde R^2_{XA}), dan was de uitkomst helemaal níet significant geweest: de overschrijdingskans was dan groter dan 5%.

7. Hoe belangrijk is een gevonden correlatie?

Voor de wetenschappelijke theorievorming is een gevonden correlatie over het algemeen pas echt belangrijk, als door aanvullende argumentatie de overtuiging kan ontstaan,dat de functionele samenhang ook een causale samenhang is, zodat er sprake is van echt ‘verklaarde’ X–variantie [zie (12) en (13)]. Maar ook dán, hangt het belang voor de praktijk geheel af van de grootte van die correlatie.
En als het louter om een voorspellingsfunctie gaat, is de grootte van de correlatie alles bepalend voor het praktisch belang. Algemene regels ter beoordeling zijn er niet. Bedenk het vogende:

1). Bereken bij iedere opgegeven correlatiecoëfficiënt R_{XA} het complement van R^2_{XA}. (Blijf wantrouwend zolang géén correlatiecoëfficiënt gegeven wordt!)

> In het uitgewerkte voorbeeld is er een (hoogst significante) correlatie $R_{XA} \approx 0.29$. $R^2_{XA} = 0.0818$ leert dat factor A ruim 8% van de X-variantie bindt. En dat wil zeggen, dat de bestaande verschillen in X voor 92% gebonden zijn aan ándere factoren dan A. Dat klinkt niet vrolijk. Ook niet voor louter voorspellingsdoeleinden.

2). In ons voorbeeld, afgaande op de frequentieverdelingen, waaruit de R_{XA} berekend is, kan men nagaan, hoe groot de kans is op voorspellingsfouten (c.q. verklaringsfouten) als men op grond van deze hoogst significantie correlatie, als voorspellingsregel zou hanteren $A^+ \rightarrow X^+$ en $A^- \rightarrow X^-$ (waarbij X^+ en X^- betekenen: boven of beneden het algemene X-gemiddelde).
In 42.5% van alle ‘gevallen’ zou ik dan een fout maken. Dat is aardig veel fouten, als men bedenkt, dat de foutenkans maar 47.5% zou zijn, (de hele correlatie met A verwaarlozend, dus ongeacht de geldende A-waarde) zonder meer voor iedereen X^+ voorspel; en maar 50% als ik besluit tot $X^+$of X^- op grond van mijn kruis – of – munt – gooien.

> (Erger is overigens, dat men vaak geneigd is zo’n functionele relatie in omgekeerde vorm te gebruiken, d.w.z. niet X uit A voorspellen (c.q. verklaren), maar uit X tot A besluiten. B.v. "Deze persoon vertoont chronisch een hoge X-waarde; dat dient verklaard te worden door (of minstens dat duidt op) een even chronische A+-situatie".
>
> De omkering van een functionele relatie is slechts geldig als R^2_{XA} = 1.00, zoals het geval is bij gedetermineerde wiskundige functies.)

3). Zolang er geen sprake is van een R^2 = 1.00. blijft het als ‘wet’ gebruik maken van vaststaande correlaties voor deze of die persoon (b.v. voor diagnose) een pure gok. Correlaties kleiner dan 1.00 hebben per definitie slechts kracht van wet zolang het gaat over massaverschijnselen of op den lange duur, b.v. voor de gezondheidszorg in zijn geheel.

(18) Steeds als het belang van afzonderlijke personen prevaleert boven het belang van ‘de massa’, mag iedere z.g. significante wetmatigheid niet ánders gebruikt worden dan als "dat zou hier ook wel eens zo kunnen zijn".

Dus geen ‘wet’, maar een ‘heuristiek’, zolang $R^2 < 1.00$.

Als men er niet aan ontkómt om te gokken (of op A of op B of ….etc.) dan geeft de hoogte van de respectieve R^2) en een indicatie van wat (op den lange duur!) de beste gok is, maar gokken blijft het.

De laatste zin in kader (18) kan óók een wat positievere formulering krijgen.
Berusten immers niet bijna alle beslissingen in ons dagelijks handelen op kiezen voor die gedragsvorm die de meeste kans (lijken te) bieden op vervulling van een beoogd doel, dat zelf weer gekozen is, omdat het de meeste kans (lijkt te) bieden op vervulling van een wat fundamenteler doel?
Maar wegens de manier waarop de uitdrukking "het is wetenschappelijk aangetoond dat …."in het wetenschappelijk en maatschappelijk verkeer als geloofsartikel functioneert, lijkt mij de gekozen formulering beter. Dit klemt des te meer, als men zich realiseert dat praktisch alle wetenschappelijk aangetoonde wetmatigheden in de menswetenschappen gaan over een R^2 die ver beneden de 50% ligt.
In het dagelijks leven kiezen we gewoonlijk pas, als er sprake is van een heel wat betere gokkans, om de eenvoudige reden, dat we geleerd hebben om z.g. intuïtief of z.g. instinctmatig een ontelbaar aantal variabelen uit de concrete context van ons handelen te onderkennen als samenhangend met de te kiezen alternatieven en ze in onze keuze te verdisconteren, al is de één daarin bekwamer dan de ander.