

Gaat de AD Haringtest om meer dan de haring?

Ben Vollaard, Universiteit Tilburg, juli 2017¹

Samenvatting. *Verkooppunten in Rotterdam en omstreken krijgen een hogere score in de AD haringtest dan uit de kwaliteit van de geserveerde haring is te verklaren. Dit blijkt uit een analyse van de kenmerken van de haring waar het AD team haar oordeel op baseert. Het voordeel bedraagt bijna een halve punt, op een schaal van 0 tot 10. Dit komt neer op een voordeel van ongeveer zes plaatsen in de ranking van verkooppunten. Als verkooppunten uit de regio Rotterdam dit voordeel niet zouden genieten, dan komen slechts drie zaken uit deze regio in de top 10 van de haringtest – in plaats van zeven, zoals nu het geval is. Dit aandeel van drie van de 10 is gelijk aan het aandeel van de 148 dit jaar bezochte verkooppunten dat in de regio Rotterdam ligt. Kortom, uitgaande van de kwaliteitskenmerken van de haring die het team meeneemt in haar oordeel heeft een zaak buiten de regio Rotterdam evenveel kans om in de top 10 te komen als een zaak binnen deze regio.*

Aanleiding

Algemeen Dagblad (AD) test elk jaar de kwaliteit van Hollandse Nieuwe in een aantal verkooppunten verspreid over Nederland. Deze ‘haringtest’ kan nuttig zijn voor consumenten. Als de test meer is dan alleen een momentopname, dan weten consumenten immers waar zij goede haring kunnen verwachten. Ook brengt de test wettelijke eisen voor de verkoop van haring onder de aandacht, waaronder een verkooptemperatuur van niet meer dan 7 graden celcius.

Tegelijk brengt de haringtest voor AD verantwoordelijkheid met zich mee. Zowel lezers als verkopers mogen verwachten dat de test eerlijk is. De beoordeling trekt veel aandacht, anders zou die immers ook niet interessant zijn voor de krant, en heeft naar verluidt grote invloed op het aantal klanten dat een verkooppunt bezoekt. Consumenten baseren er dus hun oordeel op en ondernemers kunnen er flink voordeel van hebben of nadeel van ondervinden.

De haringtest is niet onomstreden. De minst erge klacht is dat de beoordeling niet zoveel zegt, omdat smaken nu eenmaal verschillen. Het testteam zou vooral heel koude haring willen.² Een ernstiger klacht is

¹ Ben Vollaard is als hoofddocent verbonden aan de vakgroep economie van de Universiteit Tilburg. De voor dit onderzoek gebruikte gegevens zijn afkomstig van de website van Algemeen Dagblad (<http://www.ad.nl/haringtest>) en omgezet in een formaat die analyse toelaat. De STATA code voor de analyse is opvraagbaar bij de auteur (vollaard@uvt.nl). Martijn de Leeuw heeft in het kader van een Masterscriptie aan het onderzoek bijgedragen.

² Zo serveerde de top drie van 2017 haring van 3 graden celcius of minder. Zie verder ook een recent blog op het webforum Jongebazen.nl <http://www.jongebazen.nl/trage-stromen/binnenkort-ad-haringtest-is-wel-betrouwbaar>.

dat het team niet onafhankelijk is.³ Zakelijke belangen van het testteam zouden hun oordeel kleuren.⁴ Derde klacht is dat het team vishandels die kritiek hebben op de haringtest met opzet slechte cijfers zou geven als een soort straf.⁵ Tenslotte zou de test niet eerlijk zijn, omdat de wijze van testen dat in de weg staat. Beoordelingen zijn vaak vertekend als experts de te beoordelen personen (in dit geval de verkopers) kunnen zien, laat staan hen persoonlijk kennen, zo is bekend uit wetenschappelijk onderzoek.⁶ Zo zouden vishandels in de regio Rotterdam/Den Haag (inclusief Scheveningen) een streepje voor hebben, omdat het team uit die regio komt of althans daar werkzaam is.⁷ Verrassend genoeg staan in de top 10 van dit jaar zeven verkooppunten die binnen een straal van 30 kilometer van Rotterdam liggen, terwijl maar één op de drie van alle 148 bezochte punten binnen deze regio ligt (vorig jaar ging het om acht van de top 10).

Dit onderzoek richt zich exclusief op het laatstgenoemde punt. Centraal staat de volgende vraag: *hebben verkooppunten in de regio Rotterdam/Den Haag een streepje voor vergeleken met verkooppunten buiten deze regio?*⁸ Dit is voor zover bekend de eerste keer dat de haringtest zelf wordt onderzocht. Een kwantitatief onderzoek is mogelijk, omdat alle details achter de testresultaten op de website van Algemeen Dagblad zijn terug te vinden. Alvorens op de aanpak in te gaan, geven we hieronder eerst kort aan hoe de haringtest volgens de makers ervan tot stand komt.

Hoe werkt de 'haringtest'?

Het team van de AD haringtest bestaat uit drie personen: twee mensen uit de viswereld, de 'keurmeesters', en een redacteur die de test aan elkaar schrijft. Het team bezoekt een kleine 150 verkooppunten in de start van het seizoen van de verkoop van eerste, jonge haring. Het kan gaan om een klassieke vishandel, maar ook om een viskraam of een supermarkt. Dit kost hen zo ruim twee weken. Rond de 90 procent van de verkooppunten bezoeken zij niet.

Bij elk verkooppunt worden vier haringen gekocht. Van twee van deze haringen meten zij direct na aankoop de temperatuur. Ook beoordeelt het team deze twee haringen op de kwaliteit van het schoonmaken, geur, uiterlijk buitenzijde, uiterlijk binnenzijde, zoutgehalte, mate van rijping en smaak. Tezamen resulteert dit in een voorlopig cijfer. De twee andere haringen worden in een steriele zak verpakt, gewogen en ingevroren, en doorgestuurd naar het laboratorium. Daar wordt het vetgehalte gemeten en de microbiologische gesteldheid van de haringen bepaald. Het cijfer wordt hierop aangepast. De verkooppunten die in de top

³ Zie onder meer een artikel in NRC Handelsblad <https://www.nrc.nl/nieuws/2015/07/21/jongens-ik-gooi-ze-weg-een-nul-1517301-a584524> een uitzending op Den Haag FM <http://denhaagfm.nl/2017/07/24/sprake-van-belangenverstremgeling-bij-ad-haringtest/> en een veelgelezen blog op Jongebazen.nl <http://www.jongebazen.nl/lifestyle-kunst/haringtest-ad-daar-zit-luchtje-aan>.

⁴ Het hierboven aangehaalde artikel uit NRC Handelsblad stelt dat het team ook viszaken beoordeelt waar één van de teamleden cursussen geeft. Ook zou AD vishandels die zijn getest benaderen of ze een advertentie willen plaatsen.

⁵ Zie het blog aangehaald in voetnoot 3.

⁶ Claudia Goldin & Cecilia Rouse, 2000, Orchestrating impartiality: the impact of 'blind' auditions on female musicians, *American Economic Review*, 90 (4), 715-741; Danielle Li, 2017, Expertise versus bias in evaluation: evidence from the NIH, *American Economic Journal: Applied Economics*, 9 (2), 60-92.

⁷ Voor een aanpak waarbij de keurmeesters niet weten waar de haring vandaan komt, zie de Leidsch Dagblad Haringsmaakttest: <http://sleutelstad.nl/2017/06/24/stevenshof-visspecialist-wint-leidsch-dagblad-haringsmaakttest/>

⁸ Bij gebrek aan gegevens onderzoeken we niet de mogelijke invloed van belangenverstremgeling.

10 eindigen worden een tweede maal bezocht. Ook hier wordt het cijfer op aangepast. De uiteindelijke ranking is bepaald op basis van het eindcijfer, waarin al deze zaken zijn verwerkt.

Aanpak

We splitsen Nederland op in twee gebieden: alles in een straal van 30 kilometer rond Rotterdam – dit noemen we in het vervolg de ‘regio Rotterdam’ – en de rest van het land. De regio Rotterdam omvat naast Rotterdam onder meer de plaatsen Scheveningen, Leiden en Zwijndrecht.

Wie zich afvraagt of het cijfer van de haringtest vertekend is in het voordeel van zaken binnen de regio Rotterdam, kan niet simpelweg het verkregen testcijfer relateren aan de vestigingsplaats. *Het is immers mogelijk dat de in de regio Rotterdam bezochte verkooppunten gemiddeld genomen betere haring serveren dan de verkooppunten buiten deze regio.* Dat kan twee redenen hebben: (1) zaken in de regio Rotterdam serveren betere haring dan zaken in de rest van het land; (2) het AD team selecteert in de regio Rotterdam relatief veel kwalitatief goede zaken vergeleken met de rest van het land. Dat in 2016 acht van de zaken in de top 10 binnen deze regio vallen en in 2017 zeven is op zichzelf dus geen bewijs voor een vertekening in het oordeel. Om te bepalen of deze vertekening daadwerkelijk bestaat, moeten we dus rekening houden met hoe de gemeten kenmerken van de haring zoals rijping en kwaliteit van schoonmaken het cijfer beïnvloeden.

Het verklaringsmodel

Om rekening te houden met de kwaliteit van de haring maken we een model dat het cijfer in de haringtest verklaart uit de gemeten kenmerken van de haring. We gebruiken hiervoor gegevens van de test van 2016 en die van 2017. Het verklaringsmodel ziet er als volgt uit:

eindcijfer = a. gewicht + b. temperatuur + c. vet + d. prijs + e. vers + f. microbiologische gesteldheid + g. rijping + h. schoonmaken + j. indicator 2017 + constante + restterm

In dit model staat *gewicht* voor het gewicht van de haring in grammen; *temperatuur* voor de temperatuur in graden celcius van de haring (beneden 7 graden/tussen 7 en 10 graden/boven 10 graden); *vet* voor het door het laboratorium vastgestelde vetpercentage van de haring (beneden 10%/tussen 10 en 14%/boven 14%); *prijs* voor de prijs in euro's per 100 gram haring (beneden €2,84/tussen €2,84 en €3,48/boven €3,48); *vers* voor of de haring vers bereid is of niet; *microbiologische gesteldheid* voor de door het laboratorium vastgestelde microbiologische gesteldheid van de haring ((zeer) goed/voldoende/slecht/waarschuwingfase/afgekeurd); *rijping* voor de mate van rijping van de haring (licht/gemiddeld/sterk/bedorven); *schoonmaken* voor de kwaliteit van schoonmaken van de haring (zeer goed/goed/matig/slecht); *indicator 2017* voor in 2017 bezochte zaken (de rest is in 2016 bezocht). Het model heeft verder een constante term en tenslotte een restterm; dat wat niet verklaard wordt uit voorgaande variabelen. De termen a, b, c, etc. geven aan hoe deze kenmerken het testcijfer beïnvloeden.

De eerste kolom van de tabel in de appendix geeft de resultaten weer van de schatting van het verklaringsmodel. Samengevat scoort de volgende haring het *allerhoogst* in de test:

... haring van 80 gram met een temperatuur beneden de 7 graden celcius, een vetpercentage boven de 14 procent, een prijs van rond de €2,50, vers van het mes, een goede microbiologische gesteldheid, licht gerijpt, die zeer goed schoongemaakt is.

Het model verklaart 83 procent van de variantie in het eindcijfer. Veel, maar niet alles. Wat is die overige 17 procent? Allereerst zou de voorspelling van het model op basis van de kenmerken als rijping en kwaliteit van schoonmaken imperfect kunnen zijn. Zo gaat het model ervan uit dat elk kenmerk een afzonderlijke invloed heeft op het eindcijfer, terwijl het ook zou kunnen zijn dat het puntenverlies voor een relatief warme haring die ook nog slecht is schoongemaakt groter is dan het puntenverlies van de hoge temperatuur en het slechte schoonmaken afzonderlijk. Dit voorbehoud moeten we maken. Daarnaast kan het zo zijn dat het eindcijfer beïnvloedt wordt door *andere* zaken dan de kenmerken van de geteste haring. Denk bijvoorbeeld aan de invloed van de volgorde waarin verkooppunten worden bezocht op het oordeel.⁹ Daarnaast staat bij de achtergrondgegevens van de haringtest voor elke beoordeelde zaak ook een vrij invulveld met opmerkingen over schoonmaak, geur, smaak en presentatie (bijvoorbeeld: ‘frisheid verloren, karamel verkleurd’, enz.). Deze opmerkingen komen goeddeels terug in de gemeten kenmerken, maar wellicht niet voor de volle 100 procent. Kortom, het model voorspelt het gegeven cijfer goed, maar niet perfect. Hier komen we in de conclusie op terug.

Tabel 1. *Alternatieve top 10 haringtest 2017 op basis van het geschatte model*

Plaats in haringtest 2017 (daadwerkelijk)		Plaats in haringtest 2017 (op basis van model)	
1	Vishandel Ruud den Haan (Rotterdam)	1	De Graaf's Visspecialiteiten (Leerdam)
2	Visgilde De Zeeland (Bergen op Zoom)	2	Lindsey's Visspecialzaak (Wijk bij Duurstede)
3	Vishandel Simonis (Den Haag)	3	Viscircus Gebr. Koning (Almere)
4	De Graaf's Visspecialiteiten (Leerdam)	4	Volendammer Vispaleis Tol (Haarlem)
5	Visspecialist Corné Struik (Krimpen a/d Lek)	5	Vishandel Ruud den Haan (Rotterdam)
6	Schmidt Zeevis (Rotterdam)	6	Volendammer Vishandel Snoek (Amsterdam)
7	Viswinkel De Lange (Den Haag)	7	Fijnvishandel Santpoort aan Zee (Santpoort)
8	Volendammer Vishandel Koning (Rijswijk)	8	Visspecialist Martin den Heijer (Den Haag)
9	M.J. de Graaf Fishmasters (Zwijndrecht)	9	Volendammer Vishandel Sanny Bont (Nieuwegein)
10	Vistaria Schaap (Almelo)	10	Schmidt Zeevis (Rotterdam)
<i>Aandeel verkooppunten in straal van 30 kilometer rond Rotterdam:</i>			
7 van de 10		3 van de 10	

Laten we dit model het cijfer voor elk in 2017 bezocht verkooppunt voorspellen, dan krijgen we een heel andere top 10. Dit laat tabel 1 zien. Opvallend is dat de alternatieve top 10 een veel bredere spreiding van hoog scorende verkooppunten over het land laat zien. In de alternatieve top 10 komen *drie* zaken uit de regio Rotterdam in de top 10 voor – tegen zeven zaken in de daadwerkelijke top 10. Dit aandeel van drie van de 10 is gelijk aan het aandeel van de 148 dit jaar bezochte verkooppunten dat in de regio Rotterdam

⁹ Zie Victor A. Ginsburgh en Jan C. van Ours, 2003, Expert opinion and compensation: evidence from a musical competition, *American Economic Review*, 93 (1), 289-296.

ligt. Kortom, *uitgaande van het model heeft een zaak buiten de regio Rotterdam even veel kans om in de top 10 te komen als een zaak binnen deze regio.* Maar binnen de haringtest lijken andere regels te gelden.

Het is belangrijk op te merken dat de alternatieve top 10 niet noodzakelijk de 'juiste' top 10 is. De Graaf's Visspecialiteiten uit Leerdam serveert niet noodzakelijkerwijs de beste haring van alle 148 bezochte zaken. Zoals eerder gezegd kan het gebruikte model imperfecties kennen. Punt is dat *het voordeel dat zaken uit de regio Rotterdam nu in de haringtest hebben verdwijnt zodra we het cijfer alleen laten afhangen van de door het team gemeten kwaliteitskenmerken van de haring.*

Om te toetsen hoe groot deze vertekening is breiden we het model uit met één extra kenmerk: een indicator voor gevestigd zijn in de regio Rotterdam. De resultaten van dit tweede model staan in de tweede kolom van de tabel in de appendix. Daaruit blijkt dat zaken in de regio Rotterdam gemiddeld genomen 0,4 punt extra krijgen op het cijfer van de haringtest (op een schaal van 0 tot en met 10). Dit is met 95 procent zekerheid geen toeval. Het voordeel komt neer op ongeveer zes plaatsen hoger scoren in de ranking van verkooppunten dan op basis van de kwaliteitskenmerken van de geserveerde haring is te verwachten.¹⁰

Conclusie

Alles wijst er op dat verkooppunten van haring in Rotterdam en omstreken een hogere score in de AD haringtest krijgen dan uit de kwaliteit van de geserveerde haring is te verklaren. Als we het testcijfer voorspellen aan de hand van de door het AD team gegeven kwaliteitskenmerken van de haring dan genieten zaken binnen de regio Rotterdam niet langer dit voordeel en zijn zij ook niet langer oververtegenwoordigd in de top 10 van de test.

De conclusie dat de resultaten van de haringtest vertekend zijn zou alleen anders uitvallen als we geloven dat met een ander voorspelmodel wél de bestaande ranking – met de sterke oververtegenwoordiging van zaken uit de regio Rotterdam – is te reproduceren. Dat is niet waarschijnlijk, omdat we ons volledig baseren op de kwaliteitskenmerken op basis waarvan het AD team tot haar oordeel zegt te komen. Daarnaast voorspelt het gebruikte model goed. Dit suggereert dat de weging van de verschillende kenmerken zoals kwaliteit van schoonmaken en temperatuur van de haring overeenkomt met de wijze waarop het daadwerkelijke cijfer tot stand komt. Kortom, er lijkt iets buiten deze kenmerken en de manier waarop ze een rol spelen in het oordeel te zijn dat resulteert in een voordeel voor zaken in de regio Rotterdam. Over de oorzaak van deze vertekening kunnen we niet anders dan speculeren – en dat laten we aan anderen over.

¹⁰ Het toevoegen van deze indicator voor vestigingsplaats verandert weinig aan de andere geschatte coëfficiënten van het model. Dit is een andere aanwijzing dat de factor vestigingsplaats volledig losstaat van de opgenomen kwaliteitskenmerken van de haring.

Appendix. Schattingsresultaten

Verklaring van het eindcijfer in de haringtest 2016 en 2017

Te verklaren: eindcijfer haringtest (0 tot 10)	Model 1: uitsluitend kenmerken haring	Model 2: inclusief invloed vestigingsplaats
Gewicht (gr)	0,04 (0,01) ^{***}	0,05 (0,01) ^{***}
Temperatuur		
beneden 7°C	referentiecategorie	referentiecategorie
tussen 7 en 10 °C	-0,68 (0,19) ^{***}	-0,65 (0,19) ^{***}
boven 10 °C	-1,80 (0,22) ^{***}	-1,76 (0,22) ^{***}
Vetpercentage		
beneden 10%	referentiecategorie	referentiecategorie
tussen 10 en 14%	0,17 (0,20)	0,17 (0,19)
boven 14%	0,57 (0,25) ^{**}	0,64 (0,25) ^{***}
Prijs (per 100 gr)		
beneden €2,84	referentiecategorie	referentiecategorie
tussen €2,84 en €3,48	0,44 (0,22) ^{**}	0,46 (0,21) ^{**}
boven €3,48	0,17 (0,27)	0,27 (0,27)
Vers van het mes		
niet	referentiecategorie	referentiecategorie
wel	1,77 (0,20) ^{***}	1,76 (0,20) ^{***}
Microbiologische gesteldheid		
(zeer) goed	referentiecategorie	referentiecategorie
voldoende	-0,16 (0,32)	-0,20 (0,31)
slecht	-0,61 (0,45)	-0,53 (0,44)
waarschuwingsfase	-0,19 (0,29)	-0,19 (0,29)
afgekeurd	-2,37 (0,68) ^{***}	-2,59 (0,68) ^{***}
Rijping		
licht	referentiecategorie	referentiecategorie
gemiddeld	-0,43 (0,34)	-0,39 (0,33)
sterk	-1,95 (0,39) ^{***}	-1,88 (0,39) ^{***}
bedorven	-4,45 (0,51) ^{***}	-4,43 (0,51) ^{***}
Schoonmaken		
zeer goed	referentiecategorie	referentiecategorie
goed	-0,99 (0,21) ^{***}	-0,93 (0,21) ^{***}
Matig	-1,71 (0,22) ^{***}	-1,65 (0,22) ^{***}
Slecht	-2,83 (0,44) ^{***}	-2,75 (0,44) ^{***}
Regio		
binnen straal van 30 km rond R'dam		referentiecategorie
meer dan 30 km van R'dam		-0,40 (0,17) ^{**}
Indicator 2017		
haringtest 2016	referentiecategorie	referentiecategorie
haringtest 2017	0,18 (0,17)	0,17 (0,17)
Constante	3,60 (0,94) ^{***}	2,84 (0,99) ^{***}
Aantal waarnemingen	292	292
Verklaarde variantie (R ²)	0,83	0,83

Noot: Tussen haakjes staan de standaardfouten rond de geschatte coëfficiënten. ^{***} statistisch significant op het 1 procent-niveau; ^{**} op het 5 procent-niveau en ^{*} op het 10-procent niveau.