

Niet-lezers en boekenwormen. Een toepassing van het *zero-inflated*-regressiemodel in onderzoek naar leespraktijken

Mart Willekens*¹ & John Lievens¹

Inleiding

Een op het eerste gezicht vrij basale bevinding in onderzoek naar cultuurconsumptie is dat een groot deel van de bevolking niet of nauwelijks interesse vertoont in culturele activiteiten (Van Steen & Lievens, 2011). Dit geldt zowel voor uithuizige participatie (toneel, opera, concerten...) als cultuurconsumptie binnenshuis (lezen van boeken, bekijken van films...). Toch is deze scheidingslijn tussen participanten en niet-participanten een belangrijk onderzoeksobject binnen de cultuursociologie, aangezien de drempel om te participeren sterk samenhangt met een aantal belangrijke stratificerende kenmerken zoals onderwijsniveau en inkomen (de Graaf & Kalmijn, 2001). Daarnaast zijn er binnen de groep participanten ook vaak grote verschillen naar intensiteit van participatie, gaande van occasionele participatie bij toevallige passanten tot zeer frequente participatie bij cultuurliefhebbers, en ook deze groepen verschillen van elkaar (Vlegels & Lievens, 2011). In deze onderzoeksnota bespreken we het *zero-inflated* (ZI)-regressiemodel als een geschikte techniek om dit type van consumptiepraktijken te onderzoeken. Deze variant op normale regressieanalyse laat toe om simultaan (1) het verschil tussen participanten en niet-participanten en (2) de frequentie van participatie binnen de groep participanten te analyseren. Om de mogelijkheden van dit model te demonstreren, bespreken we een toepassing hiervan op het lezen van boeken.

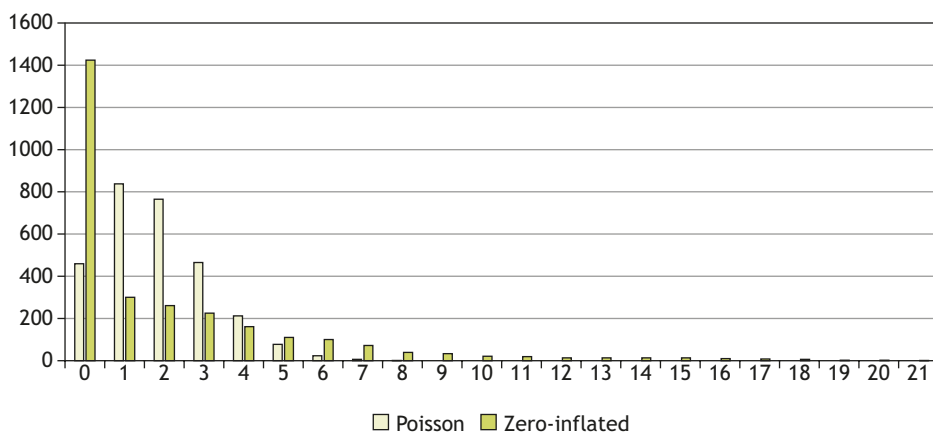
ZI-regressiemodellen

ZI-regressiemodellen kunnen gebruikt worden wanneer de onafhankelijke variabele een telvariabele is met een disproportioneel aantal nulwaarden (Cameron & Trivedi, 2013). Deze situatie doet zich relatief vaak voor bij onderzoek naar culturele consumptiepraktijken, aangezien er vaak een aanzienlijke groep is die niet participeert aan dit type activiteiten (en dus een nulwaarde heeft op het aantal activiteiten dat zij gedaan heeft) en een kleine groep die zeer vaak participeert (en dus extreme telwaarden heeft op de onafhankelijke variabele). Hierdoor vertoont de verdeling van de afhankelijke variabele een disproportioneel aantal nulwaarden en lange uitlopers naar extreme waar-

* Mart.Willekens@ugent.be

¹ Vakgroep Sociologie, Universiteit Gent

den. Een Poisson-regressie, die vaak gebruikt wordt om telvariabelen te analyseren, is hierdoor niet meer bruikbaar, aangezien deze extreme waarden ervoor zorgen dat de variantie groter is dan het gemiddelde (overdispersie: zie figuur 1). Om dit probleem op te lossen, kan een Negatief Binomiaal-model (NB-model) gebruikt worden, waarbij de logaritme wordt genomen van de verwachte waarden van het model, waardoor de overdispersie grotendeels geneutraliseerd wordt (zie bv. Christin, 2012). Verder wordt er een extra dispersieparameter geschat die de extra variantie in rekening brengt. Het grote aantal nulwaarden kan echter nog wel een probleem vormen, aangezien dit extreme overdispersie kan veroorzaken, wat opnieuw tot onbetrouwbare resultaten leidt en in sommige gevallen zal het model dan niet meer convergeren (Atkins & Gallop, 2007).



Figuur 1. Verdeling van een onafhankelijke variabele volgens de Poisson-verdeling en volgens een *zero-inflated*-verdeling met overdispersie.

Daarnaast is deze methodologie niet altijd gepast voor onderzoeksvragen over cultuurparticipatie en consumptiepraktijken. Binnen het NB-model wordt het grote aantal nulwaarden uitsluitend gezien als een afwijking die dispersie veroorzaakt. De assumptie blijft echter dat zowel de nulwaarden als de hogere positieve waarden door dezelfde factoren verklaard worden. Voor onderzoek naar consumptiepraktijken is dit echter vaak een onrealistische assumptie. Zo blijkt voor uithuizige cultuurparticipatie bijvoorbeeld dat een tekort aan cultureel, economisch en sociaal kapitaal vooral een belangrijke drempel vormt om al dan niet te participeren, terwijl de participatiefrequentie ook bepaald wordt door het tijdsbudget (Willekens & Lievens, 2016). Hier kan het dus nuttig zijn om eerst na te gaan wie er tot de groep van participanten behoort en pas in een tweede stap na te gaan welke factoren de frequentie van participatie beïnvloeden.

Om dit soort onderzoeksvragen op te lossen, zijn *hurdle*- en *zero-inflated* (ZI)-modellen beter geschikt (Cameron & Trivedi, 2013). *Hurdle*-modellen gaan ervan uit dat de processen die nulwaarden en hogere positieve waarden genereren, volledig gescheiden zijn. Voor onderzoek naar leespraktijken zou de assumptie dan zijn dat er een vol-

ledig afgescheiden groep niet-lezers en lezers bestaat. In kwantitatief onderzoek wordt echter vaak gevraagd naar het aantal boeken dat men gelezen heeft in een bepaalde tijdsperiode (bv. zes maanden). Hierdoor is de groep van absolute niet-lezers niet volledig gescheiden van de groep die wel af en toe leest. Het is bijvoorbeeld goed mogelijk dat iemand wel af en toe een boek leest, maar toevallig de laatste zes maanden geen boek heeft uitgelezen (door bv. tijdsgebrek bij de geboorte van een kind). Voor dit soort onderzoeksvragen zijn ZI-modellen beter geschikt. Hierbij is de assumptie dat enkel het *disproportioneel* aantal nulwaarden een ander verklaringsmodel nodig heeft en de rest van de telwaarden een reguliere Poisson- of NB-verdeling volgen. ZI-modellen bestaan dus uit twee componenten. De eerste is een binair logistische component die de odds bepaalt of je al dan niet tot de nulgroep behoort (in ons voorbeeld: de niet-lezers). Op basis van deze analyse worden gewichten berekend die corrigeren voor het disproportioneel aantal nulwaarden. In een tweede stap kan dan een Poisson- of NB-model geschat worden om frequenties te verklaren, op basis van de gewogen data. Dus, de logistische component analyseert de odds om tot de nulgroep te behoren en de Poisson- of de NB-regressie analyseert de telwaarden van de afhankelijke variabelen. Indien er verschillende hypothesen zijn over deze twee processen, kunnen verschillende variabelen gebruikt worden om de twee componenten van dit model op te bouwen.

ZI-modellen kennen reeds een brede toepassing binnen bijvoorbeeld de gezondheidssociologie, waar ze gebruikt worden om o.a. doktersbezoeken te analyseren (Elhai, Calhoun, & Ford, 2008). Ook in marketing en econometrisch onderzoek kent deze techniek reeds een brede toepassing om consumentengedrag te verklaren (Ateca-Amestoy & Prieto-Rodriguez, 2013). Binnen sociologisch onderzoek naar participatievormen en consumptie zijn er echter slechts zeer weinig toepassingen van dit model. Nochtans biedt het veel mogelijkheden om een aantal belangrijke onderzoeksvragen binnen dit gebied scherper te stellen en minder vaak voorkomende praktijken op bevolkingsniveau te onderzoeken. In de paper *Who participates and how much? Explaining non-attendance and the frequency of attending arts and heritage activities* passen Willekens en Lievens (2016) dit model toe op het participeren aan kunst- en erfgoedactiviteiten. In deze onderzoeksnota presenteren we een parallelle analyse voor het lezen van boeken als voorbeeld om de bruikbaarheid van dit model te demonstreren.

Niet-lezers en boekenwormen

Het lezen van boeken wordt binnen de cultuur- en onderwijssociologie gezien als een van de belangrijkste indicators van cultureel kapitaal (Bourdieu, 1984). Daarnaast zijn er een aantal specifieke onderzoeksvragen verbonden aan deze praktijk. Zo wijst de literatuur uit dat vrouwen vaker lezen dan mannen (De Laet, Glorieux, & van Tienoven, 2011). Hiervoor zijn verschillende verklaringen te vinden gaande van verschillende socialisatiepatronen, rolpatronen binnen het gezin en verschillen in werksituatie (Mickelson, 1989; Tepper, 2000). Daarnaast zien we ook dat een groot deel van de bevolking niet leest en dit vooral uit desinteresse (Van Steen & Lievens, 2011).

Tabel 1. Zero-inflated negatief binomiaal model voor het wel of niet-lezen (MLZI) en de frequentie van het lezen (ML) van romans en dichtbundels in Vlaanderen (n=2.551).

	N	Maximum Likelihood Zero- Inflation (MLZI) Parameter Estimates		Maximum Likelihood (ML) Parameter Estimates	
		Model 1	Model 2	Model 1	Model 2
Intercept		14,55 ***	15,34 ***	2,40 ***	2,62 ***
Gender	1283				
man					
vrouw	1268	0,38 ***	0,30 ***	1,27 **	1,08
leeftijd		0,99	0,99	1,01 *	1,01 *
lezen moeder	1331				
nee					
ja	1220	0,44 ***	0,44 ***	1,34 **	1,34 **
lezen vader	1487				
nee					
ja	1064	0,63 **	0,64 **	1,15	1,16
onderwijs (in jaren)		0,75 ***	0,75 ***	1,02	1,01
sociaal/culturele richting	2037				
nee					
ja	514	0,50 **	0,52 **	1,25 *	1,28 **
sociaal netwerk	1300				
meer dan 7					
3 tot 6	864	1,04	1,04	0,89	0,87
0 tot 2	387	1,57 *	1,58 *	1,20	1,20
Inkomen	1594				
makkelijk					
Neutraal	506	1,51 *	1,54 *	0,92	0,93
moeilijk	451	1,93 ***	1,94 ***	1,11	1,12
partner	488				
nee					
ja	2063	0,98	0,99	0,92	0,90
Kinderen (jonger dan 12)	1976				
nee					
ja	584	1,32	1,35	0,83	0,87
werksituatie	781				
full time					
part time	754	0,95	0,90	0,94	0,67 *
Vrouw*part time					
Vrouw*niet werkend	1016	0,97	0,82	1,35 *	1,35
dispersieparameter	474		1,18		1,67 *
	571		1,54	1,65 ***	1,05
					1,64 ***

* p < 0,05; ** p < 0,01; *** p < 0,001.

In deze onderzoeksnota gebruiken we data uit de participatiesurvey 2009 (Lievens & Waage, 2011). Dit is een grootschalig, representatief bevolkingsonderzoek naar vrijetijdsparticipatie van Vlamingen. In totaal werden 3.144 respondenten bevraagd aan de hand van *computer-assisted personal interviewing* (CAPI) en voor deze analyse gebruiken we een dataset van 2.511 cases (enkel de cases met volledige informatie op alle items werden opgenomen). We operationaliseren het lezen van boeken als een telvariabele die aangeeft hoeveel romans of dichtbundels de respondent in de laatste zes maanden heeft gelezen. 58% van de respondenten geeft aan geen romans of dichtbundels te hebben gelezen en het maximum aantal is 97.¹ Om de verschillende mechanismen te ontrafelen die bepalen of iemand al dan niet leest en hoeveel een lezer leest, presenteren we in tabel 1 een ZINB-model met een logistische component (met odds ratio's om tot de nulgroep van niet-lezers te behoren) en een NB-component (incidence ratio's voor het aantal gelezen boeken). Om het meest geschikte model te selecteren, kan het BIC van het ZI-Poisson-model en het ZINB-model vergeleken worden. In deze analyse is het BIC voor het ZINB-model -9.102,93 en voor het ZI-Poisson-model -14.616,79, wat duidelijk aangeeft dat het ZINB-model een betere fit heeft. Daarnaast zien we in het ZINB-model dat de dispersieparameter significant is (1.64***: zie tabel 1), wat aangeeft dat er sterke overdispersie is ten opzichte van het Poisson-model. Om na te gaan of het nodig is om een ZINB-model te gebruiken in plaats van het reguliere NB-model, kan het BIC niet vergeleken worden, aangezien de modellen niet genest zijn. De Vuong-test biedt hier een alternatief. Deze test geeft aan of een bepaald geschat model dichter bij het 'echte' (door data gegenereerd) model ligt. De Z-waarde bedraagt hier 13,30***, wat aangeeft dat de nulhypothese die stelt dat beide modellen even goed presteren, moet worden verworpen. Het ZINB-model wordt aangegeven als het geprefereerde model.

In de twee delen van het model nemen we dezelfde variabelen op om na te gaan of het al dan niet lezen van boeken door andere factoren wordt bepaald dan het aantal boeken dat men heeft gelezen. De ZIML-parameters geven de odds weer om tot de (disproportionele) groep van niet-lezers te behoren (de nulgroep) en de gewone ML-parameters geven de rate ratio's weer voor de frequentie van lezen. We nemen in elk deel ook een interactie-effect op tussen gender en werksituatie om na te gaan of het genderverschil in boeken lezen door verschillen in werksituatie verklaard kan worden.² Ten eerste nemen we diverse indicatoren op voor cultureel, economisch en sociaal kapitaal. Als indicator voor cultureel kapitaal nemen we voorbeeldpraktijken binnen het gezin waar men opgroeide (las de vader/moeder wanneer de respondent tussen de 12 en de 14 jaar oud was?), onderwijsniveau (jaren in het onderwijs) en studierichting (volgen van een sociaal-culturele richting). Wanneer de moeder (,44) of de vader (,63) las, zijn de odds kleiner dat men tot de nulgroep van niet-lezers behoort. Ook het aantal jaren in het onderwijs (,75) en het volgen van een sociaal-culturele richting (,50) verkleinen de odds dat men tot deze nulgroep behoort. Daarnaast blijkt een lager niveau van economisch en sociaal kapitaal een drempel voor lezen te vormen. Wanneer men een sociaal netwerk heeft van 0 tot 2 personen (,57) of moeilijk kan rondkomen met het gezinsinkomen (,93), zijn de odds groter om tot de nulgroep van niet-lezers

te behoren. Binnen de NB-component van het model, waar we kijken naar het aantal boeken dat gelezen wordt, heeft enkel het lezen van de moeder (,34) en het volgen van een sociaal-culturele richting een positieve invloed op het aantal boeken dat men leest. Inkomen of sociaal netwerk hebben geen effect. Het hebben van voldoende cultureel, economisch en sociaal kapitaal is dus vooral bepalend voor het al dan niet lezen en enkel cultureel kapitaal heeft dan nog (in mindere mate) een invloed op het aantal boeken dat men leest.

Leeftijd blijkt geen effect te hebben op het al dan niet lezen van boeken, maar het aantal boeken dat gelezen wordt binnen de groep die leest, is hoger op oudere leeftijd. Geslacht heeft zoals verwacht een belangrijk effect in beide componenten van het ZINB-model. De odds om tot de groep van niet-lezers te behoren, zijn veel lager voor vrouwen (,38) en de *incidence rate* voor het aantal gelezen boeken is ook hoger voor vrouwen (1,27) dan voor mannen. Gezinssituatie (het hebben van een partner of kinderen) blijkt in geen van beide modellen een effect te hebben. Werksituatie heeft wel een effect op het aantal boeken dat men leest. Respondenten die niet werken (waaronder ook studenten en gepensioneerden), lezen meer boeken dan respondenten die fulltime werken. In beide modellen voegen we ook een interactie-effect toe tussen geslacht en werksituatie. Voor het al dan niet lezen van boeken blijkt dit geen effect te hebben en blijft het effect van gender aanwezig. Voor het aantal boeken dat men leest, zien we wel een interactie-effect. Vooral vrouwen die deeltijds werken, blijken meer boeken te lezen dan mannen die deeltijds werken. Het hoofdeffect van gender valt weg, wat aangeeft dat er voor voltijdswerkende mannen en vrouwen geen verschil is in het aantal gelezen boeken. Dit is een opvallend resultaat, aangezien dit patroon afwijkt van bevindingen voor uithuizige cultuurparticipatie, waar het gelijkaardige genderverschil in participatiefrequentie vooral te verklaren is door de hogere participatie van voltijds werkende vrouwen ten opzichte van voltijdswerkende mannen (Willekens & Lievens, 2016). Voltijdswerkende vrouwen zullen dus vaker aan uithuizige cultuur deelnemen, terwijl vrouwen die deeltijds werken (waarbij zorg voor het gezin en huishouden vaak een belangrijke motivatie vormt, zie: Hakim, 2002) vaker kiezen voor culturele participatie die binnenshuis kan plaatsvinden.

Besluit

Met deze onderzoeksnota tonen we aan dat ZI-modellen belangrijke toepassingen kunnen hebben binnen cultuursociologisch onderzoek naar participatie en consumptiepatronen. Zo kunnen we simultaan nagaan wat de drempels zijn om tot een bepaalde praktijk over te gaan en wat de verklarende factoren zijn voor de frequentie waarmee deze praktijk voorkomt zodra deze drempel overschreden is. Willekens & Lievens (2016) toonden reeds aan dat dit model toepasbaar is op het bezoeken van kunst- en erfgoedactiviteiten en in deze onderzoeksnota presenteren we een parallelle analyse voor het lezen van boeken. Het model kan echter toegepast worden op een brede waaier van vrijetijdsparticipatie (deelname aan het verenigingsleven, reizen, uitoeffe-

nen van hobby's...) en consumptiepatronen (bekijken van films, gebruik van bepaalde media...), waarbij het al dan niet participeren/consumeren en de frequentie van participatie/consumptie door verschillende factoren verklaard kunnen worden.

Hierbij is een belangrijke assumptie dat beide processen niet volledig van elkaar gescheiden zijn en dat enkel het disproportionele aantal nulwaarden verklaard moet worden. Het model heeft ook beperkingen wanneer de steekproefgrootte beperkt is en het aantal nulwaarden te extreem wordt. In dit geval krijgt de kleine groep participanten een te groot gewicht, wat weer tot onbetrouwbare resultaten kan leiden. Zo kan het model bijvoorbeeld moeilijk apart toegepast worden op het lezen van dichtbundels in de bovenstaande analyse, aangezien 95 procent van de steekproef hier tot de groep van niet-lezers behoort.

Het voordeel van dit model voor participatieonderzoek is dat er niet exclusief gefocust wordt op het al dan niet participeren (waardoor de differentiatie binnen de groep participanten onderbelicht blijft) of op de participatiefrequentie (waardoor er geen aparte verklaringen gezocht kunnen worden voor het niet-participeren). ZI-regressiemodellen vormen hier een geschikte techniek om zowel participanten als niet-participanten als volwaardige onderzoeksobjecten te beschouwen.

Noten

1. Respondenten die meer dan 100 boeken hebben gelezen, werden aangezien als outliers en verwijderd uit de analyse.
2. We nemen het interactie-effect tussen gender en werksituatie als illustratie om aan te tonen dat dit model toelaat om interactie-effecten toe te voegen die makkelijk interpreteerbaar zijn (in tegenstelling tot multinomiaal logistische modellen met gehercodeerde variabelen). In het oorspronkelijke model werden ook interactietermen opgenomen om genderverschillen in socialisatiepatronen en rolpatronen binnen het gezin na te gaan, maar deze bleken niet significant te zijn.

Bibliografie

- Ateca-Amestoy, V. & Prieto-Rodriguez, J. (2013). Forecasting accuracy of behavioural models for participation in the arts. *European Journal of Operational Research*, 229(1), 124-31.
- Atkins, D. C. & Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models. *Journal of Family Psychology*, 21(4), 726-35.
- Bourdieu, P. (1984). *Distinction: a social critique of the judgement of taste* (R. Nice, Trans.). Cambridge: Harvard University Press.
- Cameron, C. & Trivedi, P. K. (2013). *Regression Analysis of Count Data. Second Edition*. New York: Cambridge University Press.
- Christin, A. (2012). Gender and highbrow cultural participation in the United States. *Poetics*, 40(5), 423-43.

- de Graaf, P. M. & Kalmijn, M. (2001). Trends in the intergenerational transmission of cultural and economic status. *Acta Sociologica*, 44(1), 51-66.
- De Laet, P., Glorieux, I. & van Tienoven, T. P. (2011). Van kaft tot kaft: leesgedrag en bezoek aan literaire evenementen. In L. John & H. Waege (Eds.), *Participatie in Vlaanderen 2. Eerste analyses van de Participatiesurvey 2009* (pp. 197-217). Leuven: Acco Academic.
- Elhai, J. D., Calhoun, P. S. & Ford, J. D. (2008). Statistical procedures for analyzing mental health services data. *Psychiatry Research*, 160(2), 129-36.
- Hakim, C. (2002). Lifestyle preferences as determinants of women's differentiated labor market careers. *Work and Occupations*, 29(4), 428-59.
- Lievens, J. & Waege, H. (2011). *Participatie in Vlaanderen: basisgegevens van de participatiesurvey 2009*. Leuven: Acco.
- Mickelson, R. A. (1989). Why Does Jane Read and Write so Well? The Anomaly of Women's Achievement. *Sociology of Education*, 62(1), 47-63.
- Tepper, S. J. (2000). Fiction reading in America: Explaining the gender gap. *Poetics*, 27(4), 255-275.
- Van Steen, A. & Lievens, J. (2011). Geen goesting: over drempels en percepties van cultuurparticipatie. In J. Lievens & H. Waege (Eds.), *Participatie in Vlaanderen. Eerste analyses van de Participatiesurvey 2009*. Leuven: Acco.
- Vlegels, J. & Lievens, J. (2011). Louter een kwestie van voorkeur en goesting? Over kunsten- en erfgoedparticipatie, bekeken door een cultuursociologische bril. In J. Lievens & H. Waege (Eds.), *Participatie in Vlaanderen 2: eerste analyses van de participatiesurvey 2009* (pp. 235-72). Leuven: Acco.
- Willekens, M. & Lievens, J. (2016). Who participates and how much? Explaining non-attendance and the frequency of attending arts and heritage activities. *Poetics*, 56, 50-63.