



Chocoladelettermallen in het Bakkerijmuseum te Medemblik. Foto: Arch (cc)

## OVER NOBELPRIJZEN EN CHOCOLADE

Wat een mooi begin van de week, een Nobelprijs met een Nederlands tintje. Guido Imbens, econometrist en Erasmus Alumnus, kreeg deze week samen met David Card en Joshua Angrist de Nobelprijs voor de Economie voor het baanbrekende onderzoek dat zij hebben verricht op het gebied van de causale inferentie. Daarmee is Imbens, na Jan Tinbergen en Tjalling Koopmans de derde (geboren) Nederlander die de Nobelprijs voor Economie ontvangt.

### Conclusies uit gevonden data

Imbens' onderzoek heeft betrekking op het aantonen van causale verbanden met data uit natuurlijke experimenten en is daarmee van grote toegevoegde waarde voor onder andere economisch onderzoek. Bijvoorbeeld onderzoek naar de relatie tussen het aantal opleidingsjaren en inkomen. Het is algemeen bekend dat er een positieve correlatie is tussen de duur van het genoten onderwijs en inkomen. De vraag is of langer in de schoolbanken zitten een hoger inkomen oplevert of dat slimme mensen zowel op school als op het werk succesvol zijn. Het probleem om dit te onderzoeken is dat we nooit de situatie van een individu met minder scholing kunnen vergelijken met de situatie waarin hetzelfde individu meer scholing geniet. Dit is het fundamentele probleem van causale inferentie.

Een natuurlijk experiment is een observationele studie waarbij niet gecontroleerd wordt in de traditionele zin van een gerandomiseerd experiment. De data worden als

het ware 'gevonden', en niet gegenereerd met een gecontroleerd experiment. De 'natuur' bepaalt of individuen worden blootgesteld aan de 'behandeling' of tot de controlegroep behoren, buiten de invloed van de onderzoeker. Zo is de invloed van de duur van het genoten onderwijs op inkomen onderzocht door gebruik te maken van een bijzondere eigenschap van het Amerikaanse schoolstelsel. Kinderen kunnen in de VS hun opleiding stoppen op hun zestiende verjaardag. Als gevolg zullen kinderen die eind december geboren zijn, vergeleken met kinderen die een paar weken later zijn geboren (in januari), een heel jaar langer onderwijs hebben genoten als ze beide op hun zestiende stoppen. Of de kinderen stopte op hun zestiende bepaalde niet de onderzoeker, maar de 'natuur'. Onderzoek met de aldus verzamelde data toonde aan dat kinderen geboren in januari die stopte met school en dus een jaar minder onderwijs genoten hadden, meetbaar minder verdienden.

Het proberen uit te zoeken of een patroon (langere scholing → hoger salaris) wordt veroorzaakt door wat we denken dat het veroorzaakt is de kern van het werk van een econometrist. Als de data die voor een dergelijk onderzoek gebruikt worden niet zijn gegenereerd door een gecontroleerd proces vertroebeld dat de analyse. Een probleem dat dankzij het onderzoek van Guido Imbens kan worden aangepakt. Imbens ontwikkelde econometrische methoden om te helpen causale verbanden te vinden in data uit natuurlijke experimenten. Als gevolg is economisch onderzoek steeds meer een empirische weten-



Nobelprijzen vs chocoladeconsumptie

schap geworden, een vakgebied vol empirische bevindingen, waarvan een belangrijk aantal gevonden zijn dankzij het werk van Imbens.

### Kritiek

Toch is de aanpak van Imbens niet zonder kritiek. Het gebruik van data uit natuurlijke experimenten brengt het probleem van reproduceerbaarheid met zich mee. Ook is het proces waarin individuen door de 'natuur' worden toegewezen aan de 'behandeling'- of 'controle'-groepen zelden echt random. Niet waargenomen externe factoren kunnen de selectie beïnvloeden en gecorreleerd zijn met de uitkomst waarnaar onderzoek wordt gedaan. Los van de problemen die samenhangen met het gebruik van data uit natuurlijke experimenten kun je je ook afvragen of data alleen wel voldoende is om een causaal verband aan te tonen. Judea Pearl, schrijver van onder andere *Het boek waarom*, is een fervent tegenstander van de aanpak van Imbens. Het idee dat statistische data voldoende zijn om te bepalen hoe de werkelijkheid in elkaar steekt is in zijn optiek een grote misvatting. Om van een causaal verband te kunnen spreken moet je, aldus Pearl, aantonen dat het verband niet door andere factoren wordt verklaard en dat er een plausibel werkingsmechanisme is. Als voorbeeld noemt hij de correlatie tussen een kraaiende haan en de opkomst van de zon. Om te weten dat de haan kraait omdat de zon opkomt en niet andersom, heb je naast data

een model nodig van de wereld. Dat model vertelt welke variabele (haangekraai) wordt beïnvloed door welke andere variabele (stand van de zon), het ordent de relaties tussen oorzaken en gevolgen. Je kunt zo'n model niet halen uit data volgens Pearl.

### Chocolade

Als econometrist is het mijn dagelijks werk om chocolade te maken van de data van mijn klanten en beantwoord ik vragen over bijvoorbeeld de factoren die de status van een machine beïnvloeden, of wat de belangrijkste drijvers zijn van de CO<sub>2</sub> footprint van een product. Daarin maak ik gebruik van de technieken die ons vakgebied ons biedt, soms die van Imbens, soms die van Pearl. Gelukkig groeit dat instrumentarium en zorgen kruisbestuivingen tussen bijvoorbeeld machine learning en econometrie voor vernieuwing en verbetering. Of het tot nog een Nederlandse Nobelprijs leidt? Daartoe zal de chocoladeconsumptie in ons land flink omhoog moeten, als het onderzoek van Franz Messerli tenminste klopt. In 2012 schreef hij in de *New England Journal of Medicine* dat hoe hoger de chocoladeconsumptie van een land, hoe meer Nobelprijswinnaars het per hoofd van de bevolking voortbrengt. Wie weet, ik sla alvast wat extra chocoladeletters in.

JOHN POPPELAARS, Doing the Math  
E-mail: john@doingthematn.nl