

Selectie op school naar gedrag

Een selectie naar het gedrag in schoolklassen valt uiteen in twee componenten. Allereerst zal men willen weten hoeveel leerlingen in een klas voldoen aan een gegeven norm. In de tweede plaats zal men willen weten welke leerlingen dat precies zijn. Het is dus zaak de hele klas te karakteriseren en verder binnen die klas een rangorde van gedrag vast te stellen. Door combinatie van deze twee gegevens valt in een klas een selectie op gedrag te maken.

Hoeveel leerlingen voldoen aan een norm?

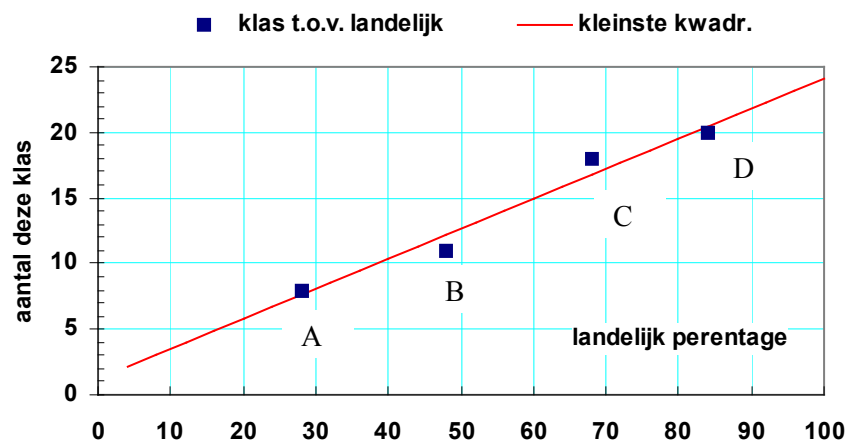
Hoe maak je een goede ijking van een hele klas, qua gedrag? Je wilt een aantal vinden dat aan een zekere norm voldoet. Je begint met een aantal tests:

- De groep rekt, bij een duidelijk aanwezige maar verboden afleiding. Hoeveel kinderen rekenen nog na een kwartier?
- De kinderen krijgen voor de maaltijd een koek. Die moeten ze bewaren tot hun brood op is. Hoeveel kinderen bewaren de koek?
- De kinderen doen een balspel, de ene helft met een rood petje, de andere helft met een groen. De petjes mogen ze houden, maar vóór het spel wordt gezegd dat ze de petjes niet in de klas mogen dragen. Hoeveel leerlingen dragen een petje in de klas, na afloop van het spel?
- enz.

Elke test levert een aantal op dat niet aan de gestelde eis voldoet. Geef elke test een kenmerk: A, B, C, enz.

Een gecommiteerde bezoekt scholen en test klassen. Na een aantal scholen te hebben bezocht, maakt de gecommiteerde een tabel. Van alle leerlingen mist 28% test A, 48% test B, 68% test C en 84% test D. Test D is dus de meest strenge.

Hij maakt een grafiek voor een speciale klas. Op de x-as staat het percentage van alle kinderen dat een test mist. Op de y-as staat het aantal in die speciale klas dat dezelfde test mist. Hij trekt een curve door de punten, en krijgt voor de klas een ijking tegen de populatie van allen.



De rode curve is een aanpassing tussen de landelijke percentages en de aantallen in de speciale klas. Nu eisen we bijvoorbeeld voor toelating dat leerlingen hoger scoren dan de laagste 70%. In de grafiek zien we dat in deze klas de landelijke norm van 70% wordt gehaald door 8 leerlingen, allen boven nr. 17. Met een beperkt aantal proeven weet de gecommiteerde dat hij van deze klas 8 leerlingen kan toelaten.

Hier passen enkele punten van kritiek.

De proeven van de gecommiteerde moeten zo veel mogelijk individueel gedrag toetsen. Gedrag dat vooral berust op groepsconformisme, geeft een onzeker resultaat. Daarbij is nooit duidelijk wat men precies meet.

Het valt niet geheel uit te sluiten dat de persoonlijke stijl van de gecommiteerde invloed heeft op de uitkomst. Na een groot aantal geteste klassen, die goed gespreid zijn over allerlei sociale achtergrond, kan men de persoonlijke invloed van iedere gecommiteerde schatten en daarvoor corrigeren.

Welke leerlingen in de klas halen de norm?

Om een vertaling te maken van een aantal naar de concrete leerlingen, maak je gebruik van een relatieve lijst met gedragsscores. Deze lijst wordt door de docent opgemaakt, eventueel na een onafhankelijke “second opinion” door een collega. Hij kent daarvoor een aantal punten toe aan elk aspect van gedrag dat in de klas relevant is.

- Hoe ongedurig is een kind, ofwel, in hoeverre kan het kind langere tijd met een taak bezig zijn?
- Hoe luidruchtig is een kind? Is het een luidkeels schreeuwend element of kan het zich rustig houden?
- Heeft het een grote mond telkens als de docent een opdracht geeft, of volgt het de instructie zonder morren op?
- enz.

Hoe maakt men een relatieve beoordeling? Laat een panel van 10 docenten een algemene indruk geven van hun leerlingen, waarbij elke leerling een algemeen cijfer A krijgt. Daarnaast geeft elke docent cijfers $c_1, c_2, c_3 \dots$ naar een aantal specifieke criteria. Neem alle beoordelingen van het panel samen en maak een aanpassing $A \approx a_1 c_1 + a_2 c_2 + a_3 c_3 + \dots$ zodanig dat in het rechterlid de best mogelijke aanpassing ontstaat aan het algemene cijfer A. Hierbij kan men denken aan een aanpassing met kleinste kwadraten, ook genaamd lineaire regressie. Elke docent wiskunde kan die maken. Met een rijtje zakelijke criteria voorkomt men de invloed van willekeur. Dan wordt een score minder beïnvloed door zaken als een lief snoetje of de kleur van de ogen.

De relatieve lijst geeft een rangorde naar puntentotaal. Als we de 8 besten van de klas toelaten, kunnen we met enig recht stellen dat we een selectie hebben op een grens van landelijk 70%. Daarmee hebben we een landelijk genormeerde selectie gerealiseerd op opvoeding en gedrag.

Verwerking van enkele commentaren.

De tests waarmee je een hele klas karakteriseert, moeten valide zijn. De kwaliteit waarop je discrimineert, moet datgene zijn wat je weten wilt. Concreet: zijn de kinderen die hun koek voortijdig opeten ook echt de minderen qua gedrag? Er moet een goede correlatie bestaan tussen de uitkomsten van een test en de eigen beoordeling door de docent.

De tests moeten statistisch betrouwbaar zijn. Als je dezelfde test na enige tijd herhaalt in dezelfde klas, moet je min of meer hetzelfde aantal tellen. Ook kun je dezelfde test zo vaak herhalen dat de gemiddelde uitkomst weinig gevoelig meer is voor toevalstreffers.

De invloed van degene die de test afneemt, moet niet te groot zijn. Neem eens aan dat een docent zijn eigen klas test op het einde van de basisschool. Collega's elders doen dat ook. Vervolgens komen de leerlingen van verschillende scholen bij elkaar in een brugklas. Is er dan geen verschil waar te nemen in de scores bij de verschillende docenten? Als allen tot ongeveer dezelfde norm komen, is de methode niet gevoelig voor de persoon die haar hanteert.

Een gedragstest voor jonge kinderen kan niet zinvol zijn vóór de leeftijd van ca. 7 jaar. De kleutertijd is de fase waarin het kind zijn gedragspatronen leert en ermee experimenteert. Daarna pas kan het resultaat betrouwbaar worden gemeten.

J.M.M.J. Vogels. mei, uitgebreid juli 2013

statistische notitie (toevoeging juni 2015)

In de grafiek zagen we voor de tests het aantal in de klas uitgezet tegen het landelijk percentage. Het is mogelijk dat de relatieve scores van de docent niet egaal gespreid zijn over de klas. Dan kun je de weergave van de meting verbeteren. Je stelt vast welke docentscore het best past bij het landelijk percentage van een test. Je kunt nu een grafiek maken met horizontaal de docentscores en verticaal de landelijke percentages. Daaruit is met betere nauwkeurigheid te bepalen welke docentscore correspondeert met de selectiedrempel.

Theoretisch kan men zelfs als volgt redeneren. Als die ene docent alle kinderen in het land zou beoordelen, zou het resultaat een Gaussverdeling zijn (normaalverdeling). Zet verticaal de landelijke percentages en horizontaal de docentscores op waarschijnlijkheidspapier, dan kun je een rechte lijn verwachten. Die geeft voor elke leerling een landelijk percentage voor het gedrag.

Voor percentages echter tussen 30% en 70% is waarschijnlijkheidspapier niet nodig. De grafiek van percentage tegen score is dan vrijwel een rechte lijn. Als de ijking en de selectiedrempel binnen die percentages vallen, heb je aan lineair grafiekpapier genoeg.