HET GEWEST UITGETEST II
HET GEBRUIK VAN MODELLEN EN STATISTISCHE TECHNIEKEN IN DE REGIONALE GESCHIEDENIS

eric harbers

Drs. E. Harbers is als wetenschappelijk medewerker verbonden aan de vakgroep Economische- en Sociale Geschiedenis van de Rijksuniversiteit te Groningen.

In deze bijdrage zal nader worden ingegaan op de kwantitatieve aspecten van de regionale geschiedenis: de modelmatige benadering en en de toepassing van de statistiek. De primaire doelstelling is het verduidelijken van deze technieken voor degenen die bezig zijn met (regionale)geschiedenis en wellicht overwegen iets op kwantitatief gebied te ondernemen. Daarom doe ik op plaatsen waar de gebruikte techniek mijns inziens voor verbetering vatbaar is, enige voorstellen aangaande alternatieven.

Dit stuk bestaat uit twee paragrafen. In de eerste wordt een voorbeeld gegeven van een model dat werd toegepast in regionaal-historisch onderzoek en in de tweede paragraaf komen verscheidene statistische technieken aan de orde, die door Nederlandse historici op het terrein dat dit thema-nummer van Groniek bestrijkt worden gebruikt. Om voor de tweede paragraaf de beschikking te hebben over voldoende voorbeelden, heb ik niet vastgehouden aan het strikte onderscheid dat in de vorige bijdrage werd gemaakt tussen locale en regionale geschiedenis, zodat daar zowel stadshistorici als zuiver regionale historici voor het voetlicht verschijnen.

Enige kennis van de macro-economische theorie en de statistiek is overigens wel vereist.(1)

het model van klep en daems

Zoals reeds vermeld in de bijdrage van Pim Kooij, neemt het gebruik van modellen ook in de regionale geschiedenis toe. De werking van één van die modellen zullen we op deze plaats onder de loupe nemen. In het proefschrift van Prof. Dr. Paul M.M. Klep wordt een economisch model gebruikt, ontwikkeld in samenwerking met Prof. Dr. Herman Daems.(2) Daarmee wordt getracht de verhouding tussen de industriële en de agrarische vraag naar arbeid in de provincies Antwerpen en Brabant tussen 1709 en 1910 te verklaren.(3) Voluit, is het een comparatief-statisch macro-economisch vraagmodel met twee sectoren, namelijk landbouw en industrie. Ondanks het feit dat deze betiteling iets heel ingewikkelds doet vermoeden, betreft het hier een relatief simpel model met vrij veel vereenvoudigde vooronderstellingen.

Laten we eerst eens naar de figuur op de volgende pagina kijken om een globale indruk van het model te krijgen. De economie is onderverdeeld in vijf blokken. Twee daarvan hebben betrekking op de produktie, namelijk de landbouw en de industrie. Twee andere blokken hebben een dubbele functie: die van consument en leverancier van factor-diensten. Dit betreft het "arbeiders"- en "kapitalisten"-blok. Deze laatste twee blokken zullen elkaar in de praktijk gedeeltelijk overlappen, omdat men immers tegelijkertijd kan arbeiden en inkomen uit vermogen ontvangen. Het laatste blok heeft enkel de functie van consument: de buitenregio (vergelijkbaar met het buitenland).

De pijlen stellen geldstromen voor. De betekenis der symbolen is te vinden na de wiskundige specificatie van het model.

Door de consumentenblokken wordt vraag naar produktie uitgeoefend. Hierbij wordt aangenomen, dat de arbeiders produkten willen kopen, zowel van de landbouw als van de industrie, maar dat de kapitalisten zich alleen industriële produkten aanschaffen. Ook de buitenregio koopt enkel industriële produkten: dit is dus de export van de regio. De buitenregio treedt in dit model niet op als producent, met andere woorden de regio importeert niet. Deze vereenvoudigende vooronderstellingen zorgen er enerzijds voor dat het model overzichtelijk en wiskundig hanteerbaar blijft, maar anderzijds wordt het model er wel minder realistisch door. Voor het vervaardigen van de gevraagde goederen wordt een inkomen uitgekeerd door de agrarische en de industriële sector aan arbeiders en kapitalisten. Dit inkomen kan weer aangewend worden om produkten te kopen. En daarmee is de kringloop rond. Op de pijl met het vraagteken kom ik later terug.

In wiskundige notatie ziet het model er als volgt uit:

$$(1)\quad Y = A+I+X$$
$$(2)\quad Y = (Y-R)+R$$
$$(3)\quad A = sP+c_1(Y-R)$$
$$(4)\quad I = R+c_2(Y-R)$$
$$(5)\quad P = \overline{P}$$
$$(6)\quad X = \overline{X}$$
$$(7)\quad L = (1/k)A$$
$$(8)\quad E_1 = (1/a)L$$
$$(9)\quad R_L = rL$$
$$(10)\quad R_O = vR_L$$
$$(11)\quad R = R_L+R_O$$
$$(12)\quad E_2 = (1/b)(I+X)$$
$$(13)\quad S = E_2/E_1$$

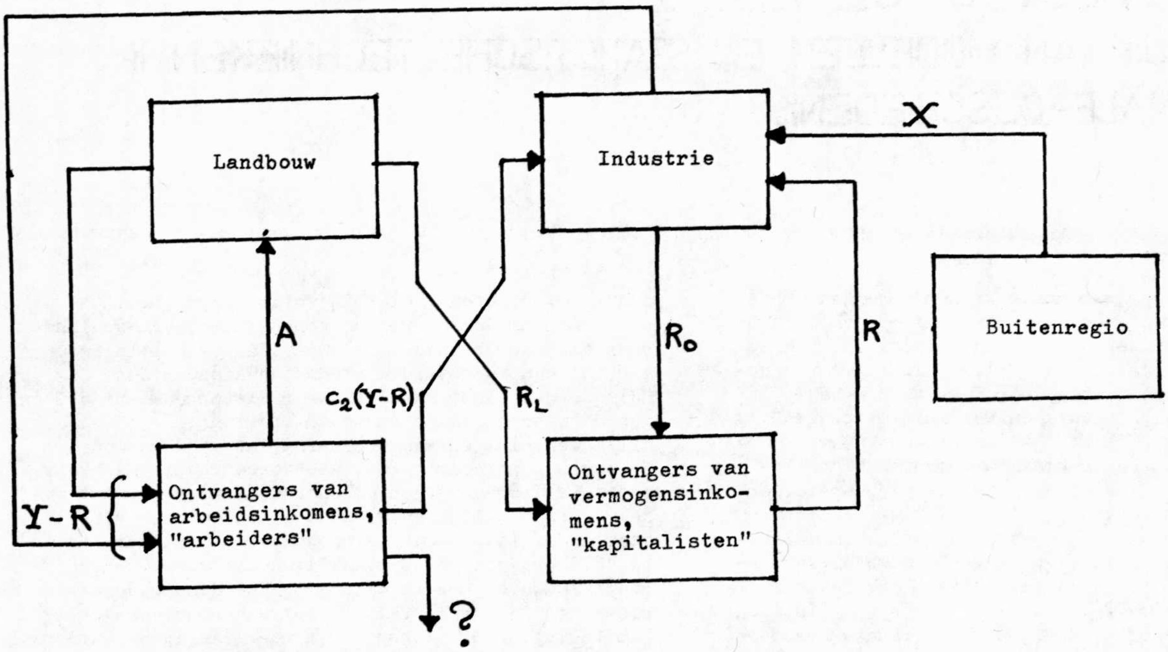
De symbolen hebben de volgende betekenis:

variabelen:

- A : de binnenregionale vraag naar agrarische produkten
- E_1 : het totaal aantal agrarische actieven (het aantal mensen werkzaam in de landbouw)
- E_2 : het totaal aantal niet-agrarische actieven (het aantal mensen werkzaam in de industrie)
- I : de binnenregionale vraag naar niet-agrarische produkten
- L : het toataal aantal eenheden land (ha)
- P : de totale bevolking
- R : het inkomen uit vermogen
- R_L : het inkomen uit land
- R_O : het inkomen uit overig vermogen
- S : de structuur van de tewerkstelling
- X : de exportvraag
- Y : de totale vraag naar de economische produktie én het totale inkomen

parameters:

- a : het aantal eenheden land (ha), dat één agra-



Schematische weergave van het model van Klep en Daems

risch werker bewerkt

b : de gemiddelde arbeidsproduktiviteit in de niet-agrarische sector, dat wil zeggen de hoeveelheid eindprodukten die één niet-agrarisch werker maakt

c_1 : de geneigdheid extra agrarische produkten te consumeren uit het arbeidsinkomen

c_2 : de consumptiegeneigdheid van niet-agrarische produkten en diensten van het arbeidsinkomen (sic)

k : de gemiddelde produktiviteit van het land, dat wil zeggen het aantal eenheden produkt dat wordt voortgebracht door één eenheid land

r : de grondrente

s : het per capita overlevingsminimum

v : de vaste relatie tussen R_L en R_O

De eerste vergelijking is een definitie-vergelijking. De vraag, die wordt omgezet in produkt en daardoor in inkomen, bestaat uit de binnenregionale vraag naar agrarische produkten, de binnenregionale vraag naar niet-agrarische produkten en de exportvraag. In de vergelijkingen (3), (4) en (6) worden deze componenten van de vraag nader gespecificeerd. Vergelijking (2) is, wiskundig gezien, volstrekt overbodig: er volgt immers direct, dat Y gelijk is aan Y. De economische betekenis is, dat het totale inkomen uiteenvalt in arbeidsinkomen (Y-R) en inkomen uit vermogen (R). Bij de oplossing van het model speelt deze vergelijking geen rol.

In vergelijking (3) wordt de binnenregionale vraag naar agrarische produkten bepaald geacht door het bedrag, per hoofd van de bevolking, dat nodig is om te kunnen overleven vermenigvuldigd met de omvang van de totale bevolking, zodat een bedrag ontstaat dat de totale bevolking nodig heeft om in leven te blijven, en door nog een extra gedeelte (c_1) van het arbeidsinkomen. In de vierde vergelijking blijkt de binnenregionale vraag naar niet-agrarische produkten uiteen te vallen in het inkomen uit vermogen en een deel (c_2) van het arbeidsinkomen.

Samen zijn c_1 en c_2 kleiner dan 1, want anders zou er niets voor sP overblijven.

In de vergelijkingen (5) en (6) staat, dat respectievelijk de totale bevolking en de exportvraag exogeen zijn, dat wil zeggen dat ze niet door het model bepaald worden. Men vult de werkelijk opgetreden waarden in. Vergelijking (7) vertelt hoeveel hectares land in totaal zullen worden bewerkt. Dit is afhankelijk van de binnenregionale vraag naar agrarische produkten (A) en de hoeveelheid land die nodig is om één eenheid produkt te maken. Dit laatste wordt weergegeven door 1/k en kan de technische coëfficiënt van het land genoemd worden. Vergelijking (7) is gemakkelijk te herschrijven tot:

$$(14) \quad A = kL$$

en men herkent hierin misschien een gedeelte van de produktiefunctie met vaste technische coëfficiënten. (4) De parameter k is de gemiddelde produktiviteit van het land, dus het aantal eenheden produkt dat door één eenheid land (=één hectare) wordt voortgebracht. De gemiddelde produktiviteit van het land is de inverse van de technische coëfficiënt van het land. Immers, gesteld dat één eenheid land 1000 eenheden produkt oplevert, dan geldt tevens dat voor de vervaardiging van één eenheid produkt 1/1000 eenheid land nodig is.

Vergelijking (8) geeft het verband weer tussen de totale hoeveelheid land die bewerkt wordt en het hiervoor benodigde aantal arbeiders. Dit aantal wordt bepaald door 1/a: het aantal agrarische werkers dat ingezet moet worden bij de bewerking van één eenheid (hectare) land. We kunnen, bij wijze van voorbeeld, stellen dat a gelijk is aan 0.5, dat wil zeggen: één agrarisch werker bewerkt een halve eenheid land. Hieruit volgt, dat 1/a gelijk is aan 2, hetgeen betekent dat één eenheid land twee agrarische werkers nodig heeft.

Uit de vergelijkingen (7) en (8) is op eenvoudige wijze een ander deel van de produktiefunctie met vaste technische coëfficiënten te destilleren:

$$(15) \quad E_1 = (1/ka)A \rightarrow A = kaE_1$$

waar ka de gemiddelde produktiviteit van de arbeid in de agrarische sector voorstelt. Om naar ons cijfervoorbeeld terug t e keren: één agrarisch werker brengt 1000 x 0.5 = 500 eenheden produkt voort. Hierbij lâát Klep het echter ten aanzien van de produktiefunctie behorend bij de agrarische sector. In de economische theorie is het evenwel gebruikelijk daar nog wat aan toe te voegen. We zagen, dat de produktiefunctie bestond uit de beide volgende vergelijkingen:

$$(14) \quad A = kL$$

$$(15) \quad A = kaE_1$$

Deze vergelijkingen zonder meer kunnen echter leiden tot niet realistische gevolgen. Indien de beschikbare hoeveelheid land of arbeid zijn maximum heeft bereikt, zijn deze vergelijkingen slechts bij één bepaalde combinatie van L en E_1 geldig. Deze combinatie volgt direct uit (14) en (15):

$$(16) \quad L = (ka/k)E_1$$

Teruggrijpend op ons cijfervoorbeeld kunnen we stellen, dat (ka/k) gelijk is aan (1000 x 0.5)/1000 = 0.5. De hoeveelheid land land (in hectaren) moet dus twee maal zo groot zijn als het aantal agrarische werkers. Stel nu eens dat er 2500 hectare land aanwezig is en 6000 agrarische werkers. In dat geval is de potentiële agrarische produktie 1000 x 2500 = 2,500,000 eenheden als we kijken naar het maximale aantal eenheden land dat ingezet kan worden, en 1000 x 0.5 x 6000 = 3,000,000 eenheden als we kijken naar het maximale aantal agrarische arbeiders dat tewerkgesteld kan worden. Het mag duidelijk zijn, dat in dit voorbeeld het land de beperkende factor vormt, met andere woorden: het is fysiek niet mogelijk dat er 3,000,000 eenheden agrarisch produkt worden vervaardigd, omdat de aanwezige hoeveelheid land daartoe tekort schiet.

Dus, als een produktiefactor volledig bezet raakt gaan beide vergelijkingen uitsluitend tegelijkertijd op, als de produktiefactoren in een bepaalde verhouding tot elkaar staan. Als dat laatste niet zo is, dan geldt slechts dat deel van de functie waarvan de produktiefactor het meest schaars is. Daarom is het in de economische theorie gebruikelijk om de produktiefunctie met vaste technische coëfficiënten, in de symbolen van het onderhavige model, als volgt te definiëren:

$$(17) \quad \begin{aligned} A &= kL && \text{voor } E_1 \leq E_1\text{max} \\ A &= kaE_1 && \text{voor } L \leq L\text{max} \end{aligned}$$

waar E_1max en Lmax de maximale waarden van E_1 en L zijn. (5) Dit komt er op neer dat zodra hetzij E_1, hetzij L zijn maximale waarde bereikt, de reele produktie niet verder kan oplopen. Er ontstaat dan respectievelijk volledige werkgelegenheid in de agrarische sector gepaard gaande met met landovervloed, of landtekort gepaard gaande met werkloosheid, tenzij arbeid en land in een unieke verhouding tot elkaar staan.

Terug naar ons getallenvoorbeeld herschrijven we (17) als volgt:

$$(18) \quad \begin{aligned} L &= (1/k)A && \text{voor } E_1 \leq E_1\text{max} \\ E_1 &= (1/ka)A && \text{voor } L \leq L\text{max} \end{aligned}$$

en vullen we de gekozen waarden in

$$L = (1/1000)A$$

$$E_1 = (1/500)A$$

en stellen we, zoals voorheen, bovendien E_1max op 6000 en Lmax op 2500, dan zal bij een vraag naar 2,000,000 eenheden produkt de benodigde arbeid 2,000,000/500 = 4000 personen bedragen en het benodigde land 2,000,000/1000 = 2000 hectaren. Dit kan verwezenlijkt worden, omdat geen van beide produktiefactoren zijn plafond heeft bereikt. Indien echter de totale agrarische vraag 2,800,000 eenheden bedraagt dan zou de benodigde arbeid 2,800,000/500 = 5600 personen zijn en en het benodigde land 2,800,000/1000 = 2800 hectares. Nu is de arbeid wel in voldoende mate aanwezig, maar er is niet genoeg land, want de maximale L van 2500 is overschreden. De vraag is dus te hoog geworden voor de fysieke mogelijkheden van het land. De maximale hoeveelheid land is 2500 hectaren en daarbij hoort een maximaal te bevredigen vraag van 1000 x 2500 = 2,500,000 eenheden, hetgeen leidt tot een maximum aantal tewerkstellingen van (1/500) x 2,500,000 = 5000.

Duidelijk moet zijn, dat dit betoog slechts slaat op de door Klep gebruikte en, in de historische context, wel realistische produktiefunctie, waarbij arbeid niet door (bijvoorbeeld vruchtbaarder) land kan worden vervangen en vice versa. De voornaamste conclusie uit de verhandeling rond de vergelijkingen (7) en (8) moet luiden, dat de economische produktie aan bepaalde grenzen gebonden is, ongeacht de gebruikte produktiefunktie. Volledige substitutie tussen land en arbeid is onmogelijk.

Naar mijn mening dient Klep zich zeer wel af te vragen of deze grenzen niet overschreden kunnen worden door zijn model en of daarom de bedoelde plafonds niet moeten worden ingebouwd. Eerlijkheidshalve dient vermeld te worden, dat er in dit opzicht wel een veiligheidsklep is gemonteerd: tussen 1750 en 1850 vermindert het aantal eenheden land dat één agrarisch werker bewerkt (a) tot de helft van zijn oorspronkelijke waarde. (6) Dit betekent, dat het aantal arbeiders in de agrarische sector aanzienlijk kan toenemen als het plafond overschreden dreigt te worden. Aan de hand van de in Klep's proefschrift verstrekte gegevens kan echter niet worden uitgemaakt of bedoelde veiligheidsklep toereikend is.

In de negende tot en met de elfde vergelijking wordt het inkomen uit vermogen (R) bepaald. Dit inkomen wordt onderverdeeld in het (kapitaal-) inkomen uit land (R_L) en het inkomen uit overig vermogen (R_O). Het inkomen uit land wordt bepaald geacht door de, helaas niet door het model verklaarde, grondrente vermenigvuldigd met de hoeveelheid land die benut wordt. (7) Het inkomen uit het overig vermogen (R_O) is gekoppeld aan het inkomen uit het land (R_L) door middel van de parameter v. Deze parameter varieert steeds en is daarom wel aanvaardbaar, maar in principe had men R_O net zo goed exogeen kunnen maken.

In vergelijking (12) wordt het totaal aantal niet-agrarische actieven afhankelijk gesteld van de vraag naar niet-agrarische produkten (binnen- en buitenregionaal) en de technische coëfficiënt van de arbeid in de industriële sector (1/b). Opvallend is dat er geen vergelijking is waarin de industriële kapitaalgoederen gekoppeld worden aan de vraag in de industriële sector. Daarover later nog een opmerking. Wederom doet zich het probleem voor van de produktieplafonds: zijn er genoeg industriële arbeiders om het gevraagde produkt te vervaardigen en, immanent, is de vereiste hoeveelheid kapitaalgoederen wel aanwezig? Ook hier lijkt een ingebouwd plafond nuttig. In feite moet men zich in dit sectorenmodel dus afvragen of de totale vraag naar arbeid (E_1+E_2) niet uitstijgt boven het gedeelte van de totale bevolking (P) dat zich aanbiedt op de arbeidsmarkt. Substitutie tussen agrarische en industriele arbeid is immers relatief eenvoudig.

Vergelijking (13) tenslotte is de vergelijking waar het Klep om gaat: hier wordt de verhouding tussen industrieel tewerkgestelden en agrarisch tewerkgestelden bepaald. De oplossing die Klep geeft voor dit model is dan ook geheel gericht op deze vergelijking. Dat is stellig mede gebeurd omdat deze oplossing min of meer te schatten is voor Brabant en Antwerpen en het model als geheel (nog?) niet. Ik geef de oplossing zonder afleiding. Enthousiaste cijferaars kunnen op een interessante zondagmiddag rekenen!

$$(19) \quad S = (1+v)(a/b)r+(ka/b)\ \frac{c_2s+(1-c_1)x}{(1-c_2)s+c_1x}$$

waar $x = (X/P)$

De verhouding tussen industrieel en agrarisch tewerkgestelden is hier afhankelijk gemaakt van parameters en exogenen. De grondrente figureert in de eerste term van de rechterzijde van de vergelijking en de exporten en de bevolkingsomvang verschijnen in de tweede term.

Aan het begin van deze paragraaf gaf ik een korte omschrijving van de aard van het model. De meeste termen hieruit zijn, voorzover zij dat nog niet waren, waarschijnlijk wel duidelijk geworden in de loop van dit betoog, met uitzondering wellicht van de term comperatief-statisch. Een statisch model is een model dat bij onveranderde exogenen en parameters slechts één oplossing kent, als het lineair is tenminste. De tegenhanger van het begrip statisch is het begrip dynamisch. Een dynamisch model genereert een reeks verschillende oplossingen voor achtereenvolgende periodes, ook al blijven exogenen en parameters constant. Comparatief statisch nu houdt in, dat men een statisch model gebruikt voor een aantal periodes, waarbij exogenen en parameters steeds aangepast worden. De periodes hoeven daarbij niet opeenvolgend te zijn.

Klep gebruikt het model voor de jaren 1709, 1755, 1846, 1880, 1895 en 1910; de bijbehorende werkelijke (dus niet de door het model voorspelde) S-waarden werden geschat op respectievelijk 1, 1, 1, 1.4, 2 en 3.3. Op ingenieuze wijze wordt vervolgens de eerste term van de rechterzijde van vergelijking (19) geschat voor elk van de zes jaren. En nu blijkt de slimheid van dit model en de noodzaak het model juist op deze manier te specificeren: de benodigde parameters kunnen zonder de hiertoe gebruikelijke statistische techniek regressie-analyse, waarvoor vrij lange cijferreeksen nodig zijn, net uit het weinige materiaal dat er is, gewrongen worden. Nadat de eerste term geschat is, berekent Klep, op basis van de werkelijke S en die eerste term, het percentage dat de tweede term van S uitmaakt. Dit leidt tot de volgende percentages, respectievelijk: 30% (ik kom hier zelf trouwens op 40%), 30%, 10%, 30%, 40%, 50%.

Op deze plaats, waar het immers gaat om de gebruikte technieken, wil ik niet ingaan op de conclusies die Klep aan dit resultaat verbindt. Wel moet worden gezegd dat Klep, en mijns insziens terecht, hierbij de nodige voorzichtigheid betracht.

Tot slot wil ik nog enige kritische geluiden laten horen ten aanzien van het model. Klep heeft zonder meer een aardig staaltje van modelbouw laten zien, maar klopt het eigenlijk wel? Het antwoord op deze vraag moet helaas luiden: zeer waarschijnlijk niet, want Klep gaat er bij de berekening van zijn percentages van uit, dat zijn model tot op grote hoogte correct is, dat wil zeggen dat, indien alle parameters en exogenen worden ingevuld in vergelijking (19), er ook bij benadering de werkelijke waarden van S uit zullen komen. En dit nu laat hij, waarschijnlijk door de gebrekkigheid van zijn cijfermateriaal, achterwege.

Het bepalen van alle parameters van het model, het berekenen van de waarheden die het model toekent aan de variabelen en het vergelijken van de modeluitkomsten met de werkelijkheid móét de eerstvolgende stap zijn in een dergelijk onderzoek. Ik geloof, dat de betrouwbaarheid van het model te wensen overlaat omdat het tè simpel is geworden. Om een realistischer model te krijgen moet het importeren mogelijk worden gemaakt, hetgeen Klep overigens zelf ook erkent, en zou het vermogensinkomen ook in de agrarische sector besteed moeten kunnen worden. Bovendien verdient het aanbeveling om in een model, dat een periode van economische bloei beslaat, de kapitaalvorming expliciet op te nemen. Dit laatste houdt in, dat de investeringen meer zichtbaar moeten worden gemaakt en dat er een relatie opgenomen zou moeten worden betreffende de vraag naar industriële arbeid en de omvang van de kapitaalgoederenvoorraad. De investeringen zitten onder andere verborgen in de bestedingen die de kapitalisten doen bij de industriële sector. Er is geen reden om aan te nemen, dat voor het vermogensinkomen geen kapitaalgoederen worden gekocht. Een dergelijke aankoop is dan tevens een besparing, want het is een niet-consumptieve besteding. Nu ontkent Klep de mogelijkheid van het optreden van besparingen in zijn model. (8) Dat is, ook al zou er geen enkel kapitaalgoed worden vervaardigd, onjuist zolang er geëxporteerd wordt. Dit is eenvoudig in te zien, als we kijken naar wat het arbeidersblok en het kapitalistenblok ontvangt en weer uitgeeft. Men raadplege hiertoe, desgewenst, de figuur.

De ontvangers van vermogensinkomens krijgen $R_L+R_0 = R$ en geven ook weer R uit. Daar wordt dus niet bespaard, behalve in de zojuist genoemde zin. De arbeiders echter ontvangen Y-R en geven uit: $A+c_2(Y-R)$. Nu geldt:

$$Y-R = A+I+X-R = A+R+c_2(Y-R)+X-R = A+c_2(Y-R)+X$$

waaruit blijkt, dat

$$Y-R \geqq A+c_2(Y-R)$$

en wel juist met het bedrag der exporten. Er wordt dus voor een bedrag van X bespaard. En dat valt ook zonder formules wel te begrijpen, want eenmaal geëxporteerde goederen komen niet meer in aanmerking om binnen de regio te worden afgezet. Het inkomen dat voor de produktie ervan ontvangen is, kan dus niet worden besteed en wordt daarom gedwongen bespaard. Bovendien vindt deze gedwongen besparing plaats in het arbeidersblok... Een goede aanwending voor deze besparingen zou gelegen kunnen zijn in importgoederen. Een reden te meer om importeren mogelijk te maken in het model! Vandaar het vraagteken in de figuur.

Bovenstaand betoog is in overeenstemming met de, in de elementaire economische theorie voorkomende, ex post gelijkheid tussen spaaroverschot en overschot op de lopende rekening van de betalingsbalans. (9) Als we aannemen, dat er niet geïnvesteerd wordt, houdt dat in, dat de besparingen gelijk zijn aan de exporten (er wordt ook niet geïmporteerd!). Nemen we aan, dat er wel geïnvesteerd wordt, dan geldt dat het besparingsoverschot gelijk is aan de exporten.

Mocht in het voorgaande de indruk zijn gewekt, dat het model van Klep en Daems louter kritiek oproept, dan is deze indruk onjuist. Hun model is noodgedwongen eenvoudig van aard, maar het is wel een der eerste modellen die in de Nederlandse geschiedschrijving werd toegepast en dat is een prijzenswaardig feit. Modelgebruik leidt immers tot explicitering en toetsing van theorieën op een wijze die met behulp van een kwalitatieve methode veelal niet haalbaar is. Het moge zijn, dat dit nog geen ideaal model is (als dat al bestaat), maar het is een ingenieuze eerste stap op de, naar mijn mening, goede weg.

statistische technieken

Statistische technieken worden, evenals modellen, meer en meer toegepast in de regionale geschiedenis. Soms zijn dat zeer eenvoudige, beschrijvende technieken, maar ook ingewikkelder technieken worden niet geschuwd. En kijken we bijvoorbeeld naar de inventarisatie door O.W.A. Boonstra van lopend historisch onderzoek waarbij gebruik wordt gemaakt van de computer, dan blijkt dat ons nog veel meer te wachten staat. (10) In deze paragraaf komt aan de orde wat men met een aantal van de door Nederlandse regionale historici gebruikte technieken kan doen, maar niet hoe men dat doet, want daar is een grotere kennis van de statistiek voor nodig dan in dit artikel bij de lezer wordt verondersteld. Dus geen formules, maar wel verwijzingen. Mocht men zelf met statistiek willen gaan werken, dan verdient het aanbeveling om allereerst een goede inleiding te lezen, zoals bijvoorbeeld de pakweg eerste 500 pagina's van het, bij het bijvak "statistiek voor historici" gebruikte, werk van T.H. Wonnacott en R.J. Wonnacott. (11) Het "blind" invullen van formules kan gemakkelijk leiden tot nonsens. Bij G. Ohlin bijvoorbeeld kan men lezen welke voorzichtigheid betracht dient te worden bij kwantitatieve analyse en wat de gevolgen kunnen zijn indien men die voorzichtigheid uit het oog verliest. (12) Als men zelf nu niet direct de behoefte gevoelt om practicerend kwantitatief historicus te worden, maar als men wel enig passief inzicht in de materie wil verkrijgen, dan kan een keuze worden gemaakt uit de werken van onder meer Joliffe, Floud, Willemsen en Dollar en Jensen. (13)

De eerste stap, na bepaling van onderwerp en vraagstelling, in kwantitatief (en natuurlijk niet uitsluitend kwantitatief) historisch onderzoek is doorgaans het verzamelen van gegevens (in het vakjargon "data" genoemd). Direct kunnen zich dan al twee problemen voordoen. Het kan zijn, dat er zoveel data zijn, dat er een steekproef moet worden getrokken. Het eerste probleem is dan: hoe wordt die steekproef getrokken? Het tweede probleem wordt gevormd door het niveau waarop de data worden gemeten.

Een belangrijke eis waaraan een goede steekproef moet voldoen is, dat elk element in de populatie waaruit de steekproef wordt getrokken, bij aanvang, in ieder geval een bekende kans heeft om te worden opgenomen in de steekproef. Veelal tracht men bedoelde kans voor elk element gelijk te laten zijn, maar strikt noodzakelijk is dat niet. De meest eenvoudige vorm van een steekproef is de aselecte steekproef (simple random sample). Hierbij worden de elementen volstrekt door het toeval aangewezen. Men kan bijvoorbeeld elk element in de populatie een nummer geven, alle nummers in een hoge hoed stoppen en vervolgens een aantal nummers trekken. (14) In de praktijk is dit echter een heel karwei. Daarom werkt men vaak met de zogenaamde systematische steekproef (systematic sample). Hierbij worden alle elementen van de populatie achter elkaar gezet, het eerste element wordt bij toeval getrokken en vervolgens wordt er, terwijl men de populatie van voor naar achter doorwerkt, met vaste tussenafstanden een element uitgelicht. Als men bijvoorbeeld een systematische steekproef wil trekken die qua grootte een tiende van de populatie vormt, dan trekt men eerst een toevalsgetal tussen 1 en 10 (weer uit de hoge hoed bijvoorbeeld). Laten we veronderstellen dat dit getal 4 is. Het eerste element in de steekproef is dan het populatie-element nummer 4, het tweede steekproef-element is dan populatie-element 14, het derde nummer 24 enzovoort. Doorgaans hebben systematische steekproeven eenzelfde betrouwbaarheid als aselecte steekproeven. Er kan echter iets grondig misgaan als er systematische fluctuaties in de populatie aanwezig zijn. Veronderstel bijvoorbeeld, dat het eerste, tiende, twintigste enzovoort element telkens een unieke eigenschap heeft. Als men werkt met tussenafstanden van tien en het aanvangsgetal is niet 1, dan komen deze elementen nooit in de steekproef terecht. In de praktijkis dit overigens geen groot probleem, hoewel men het wel in de gaten moet houden. Van de historici die werken op het regionale vlak, gebruikt onder meer J.A. de Jonge de systematische steekproef. (15) Zowel bij de aselecte als de systematische steekproef is, bij aanvang, de trekkingskans per element gelijk.

Er zijn overigens vele methoden om een steekproef te trekken. Een belangrijke andere steekproef is de gelede of gestratificeerde steekproef (stratified sample). Hierbij wordt de populatie vooraf ingedeeld in groepen (strata) op basis van een bepaalde eigenschap van de groep. Uit elk stratum wordt dan een aselecte steekproef getrokken. Een voorbeeld, ontleend aan Wonnacott en Wonnacott, moge dit verduidelijken. (16) Veronderstel, dat een bevolking voor 60% uit stedelingen en voor 40% uit plattelandsbewoners bestaat en dat aan een statisticus wordt gevraagd om het gemiddelde inkomen van die bevolking te schatten. Indien deze statisticus nu de indruk heeft, dat de inkomens in de stad en op het platteland aanzienlijk verschillen, dan kan hij de bevolking op basis van de eigenschap stadsbewoner of plattelandsbewoner indelen in strata. Vervolgens wordt met behulp van twee aselecte steekproeven het gemiddelde per stratum bepaald, waarna uit deze twee gemiddelden het totaalgemiddelde wordt berekend. Het ligt voor de hand en valt wiskundig ook te bewijzen, dat dergelijke steekproeven betrouwbaarder kunnen zijn dan de volkomen aselecte steekproef. (17) Indien de steekproefgroottes zich net zo verhouden als de groottes van de strata, blijft de trekkingskans voor ieder element gelijk. In alle andere gevallen gaan de trekkingskansen uiteenlopen, maar dat vormt op zich geen bezwaar. Ingeburgerd als deze methode is (bijvoorbeeld bij verkiezingsonderzoek), kan het enige bevreemding wekken dat ze zo weinig in de regionale geschiedenis wordt toegepast. Een voor de hand liggende verklaring voor dit feit is, dat men, om de gelede steekproef te kunnen toepassen, enige kennis moet hebben omtrent de opbouw van de populatie en die kennis ontbreekt vaak. Ook moet worden opgemerkt, dat het kwantitatief historisch onderzoek met speciale problemen te kampen heeft: veelal heeft de historicus nauwelijks vrijheid bij het trekken van een steekproef omdat deze als het ware "door de geschiedenis" getrokken kan zijn. Met andere woorden, het bronnenmateriaal kan door de loop der tijd zijn uitgeselecteerd en men kan dan alleen maar hopen dat dit min of meer "at random" is gebeurd.

Het tweede probleem dat zich reeds bij de aanvang van kwantitatief historisch onderzoek voordoet, wordt gevormd door het meetniveau der data. In het algemeen worden drie niveau's onderscheiden: nominaal, ordinaal en interval/ratio. Nominale en ook ordinale data worden veelal "kwalitatief" genoemd, interval/ratio variabelen "kwantitatief". Nominale data worden aangeduid met namen en er is volstrekt geen rangorde in aan te brengen. De meest eenvoudige nominale variabele bestaat uit twee categorieën, bijvoorbeeld: mannelijke en vrouwelijke of westers en niet-westers. Dit wordt een dummy-variabele, 0-1 variabele of dichotomie genoemd. Bij een dummy zijn er dus twee mogelijkheden: een bepaalde eigenschap is wel (1) aanwezig of niet (0). Ook indien er meer dan twee categorieën bestaan spreken we van een nominale variabele, bijvoorbeeld: geloof of geboorteplaats. Tot nog niet zo heel lang geleden was men van mening, dat nominale variabelen zich slecht zouden lenen voor statistische analyse.

Ordinale data worden aangeduid met namen of getallen en er is een rangorde in aan te brengen: bijvoorbeeld sociale klassen of waarderingscijfers, zoals cijfers op een schoolrapport. Men kan deze variabelen op een schaal afzetten, waarbij iedere categorie een vaste plaats heeft ten opzichte van de andere categorieën. Nominale categorieën hebben deze eigenschap niet. De afstand tussen de ordinale waarden op de schaal kan echter niet nauwkeurig worden aangegeven.

Dit laatste kan wel bij interval/ratio variabelen. Deze variabelen bestaan uit getalsmatige waarden, zoals bijvoorbeeld: bevolkingsomvang of inkomen. Het verschil tussen interval en ratio niveau, dat overigens doorgaans niet belangrijk is, wordt gevormd door het feit, dat bij de ratio-schaal het nulpunt eenduidig vastligt (bijvoorbeeld lengte) en bij de interval-schaal niet (bijvoorbeeld temperatuur gemeten in graden celsius of fahrenheit). Het belangrijke punt bij het niveau der variabelen is dat de te gebruiken statistische techniek er afhankelijk van is. Dus bij nominale variabelen worden andere technieken toegepast dan bij interval/ratio variabelen.

Samenvattend: reeds in een vroeg stadium van kwantitatief onderzoek moet er een aantal ingrijpende beslissingen worden genomen. Wordt er een steekproef getrokken en, zo ja, wat voor soort steekproef? Tevens komt vast te liggen, door het niveau van de data waar men mee gaat werken, welke statistische technieken gebruikt kunnen worden.

De volgende stap zal vrijwel altijd de uitwerking zijn van de, in theorie reeds vooraf bepaalde, vragen die men aan het materiaal wil stellen. Twee groepen vragen kunnen hierbij onderscheiden worden: a. men wil iets weten over elke variabele afzonderlijk en b. men wil iets weten over de relatie tussen twee of meer variabelen. In eerst instantie kan men de data ordenen en samenvattend beschrijven. Veelal worden dan de frequentie-verdelingen van de variabelen samengesteld, dat wil zeggen: er wordt geteld hoe vaak een bepaalde waarde van een variabele voorkomt. Frequentieverdelingen kunnen samengesteld worden voor alle dataniveau's en hebben betrekking op de bovengenoemde categorie a. Tevens kan men een maatstaf van centrale tendentie bepalen. Hieronder vallen modus, mediaan en gemiddelde, die respectievelijk op nominaal, ordinaal en interval/ratio niveau toegepast kunnen worden en ook steeds op hogere niveau's, dat wil zeggen dat bijvoorbeeld de modus enige zin heeft op ordinaal niveau, maar de mediaan ongeschikt is voor nominaal niveau. De modus is de waarde die het meest voorkomt bij een variabele, met andere woorden: de waarde met de hoogste frequentie. De term modaal inkomen is hier min of meer van afgeleid. De mediaan is de waarde van de variabele behorend bij de middelste waarneming van een frequentie-verdeling en het (rekenkundige) gemiddelde is de som van alle waarden die een variabele aan kan nemen, steeds vermenigvuldigd met de bijbehorende frequentie gedeeld door de totale frequentie. De zogenaamde spreidingsmaatstaven geven, globaal gesproken, aan, hoever de afzonderlijke waarnemingen gemiddeld verwijderd zijn van het gemiddelde. Een spreidingsmaatstaf heeft alleen zin op interval/ratio niveau. De meest voorkomende zijn de variantie en de standaartdeviatie. (18) De variatiecoëfficiënt is een relatieve spreidingsmaatstaf, die bestaat uit de standaartdeviatie gedeeld door het gemiddelde. (19) Hiermee kan men verschillende frequentie-verdelingen gemakkelijk vergelijken. Genoemde maatstaven vallen eveneens in de categorie a. De kruistabel valt duidelijk in categorie b. In een kruistabel worden twee (of soms ook meer) frequentie-verdelingen tegenover elkaar afgezet, zodat de frequentie behorend bij elke waarde van de ene variabele wordt verdeeld over de waarden van van de andere variabele. (20) Op deze wijze kan men enige indruk krijgen van het eventuele verband tussen de afgezette variabelen.

Bovengenoemde ordenende en beschrijvende technieken worden door vele regionale historici gebruikt. (21)

Na de beschrijvende en samenvattende werkzaamheden, komen we toe aan de echte statistiek. Er is een steekproef getrokken uit een populatie en er moet op basis van die steekproef iets gezegd worden over die populatie. Vaak berekend men de steekproefgemiddelden of -proporties van respectievelijk ratio/interval-variabelen en nominale of ordinale variabelen. Een proportie is een percentage van een aantal waarnemingen met een bepaalde eigenschap, ten opzichte van het totaal aantal waarnemingen. In feite is de proportie het (reknkundige) gemiddelde van een dummy-variabele. Stel bijvoorbeeld dat we werken met een variabele "beroep" en we willen de proportie timmerlieden in een populatie van mensen bepalen. Van elk individu kan vastgesteld worden of het wel of niet een timmerman is; als dat wel het geval is, dan kennen we het bewuste individu een 1 toe en indien het niet zo is een 0 (nul). Er ontstaat aldus een reeks van nullen en enen. Dit is nu de dummy-variabele. Het gemiddelde van die nullen en enen vormt dan de proportie timmerlieden in de populatie.

Men wil nu weten in hoeverre het steekproefgemiddelde of de steekproefproportie overeenkomt met het populatiegemiddelde respectievelijk de populatieproportie, bijvoorbeeld: hoe groot is het aantal timmerlieden in de populatie? Het doel hiervan is vaak het analyseren van ontwikkelingen in de tijd, bijvoorbeeld: neemt het werkelijke aantal timmerlieden toe of af? In het nu volgende beperk ik me tot de aselecte en de systematische steekproef, waarvoor men de zelfde formules kan gebruiken.

Wil men op basis van <u>één</u> steekproefgemiddelde of -proportie een uitspraak doen omtrent de omvang van het populatiegemiddelde of de -proportie, dan kan men het beste een zogenaamd betrouwbaarheidsinterval construeren (confidence interval). Hiermee kan men aangeven tussen welke grenzen het populatiegemiddelde of de -proportie met welke waarschijnlijkheid ligt; een typisch voorbeeld: de kans is 95%, dat het aantal timmerlui in de populatie ligt tussen de 1% en de 9%. De soort formule die gebruikt moet worden om tot een dergelijke uitspraak te komen, is mede afhankelijk van het aantal elementen in de steekproef. De kritische grens ligt daarbij op ongeveer 100 steekproefelementen. Tevens dient vermeld, dat het bij proporties iets ingewikkelder in elkaar zit dan bij gemiddelden. Met name bij proporties blijkt het benaderende karakter van de meeste formules, want als de steekproefomvang erg gering is (zeg kleiner dan 25) en ook als de proportie erg klein of zeer groot is (pakweg kleiner dan 5% of groter dan 95%), gelden de doorgaans gegeven formules niet meer. In dat geval moet men meer gecompliceerde formules gebruiken, die gebaseerd zijn op de zogenaamde binomiale verdeling, een basis-verdeling voor proporties, of verdelingen die daar zeer sterk op op lijken. (22)

Wil men op basis van <u>twee</u> steekproefgemiddelden of -proporties een uitspraak doen omtrent het verschil tussen twee populatiegemiddelden of -proporties, dan zou dat kunnen door met de voorgaande methode te werken. Men kan bijvoorbeeld het volgend resultaat hebben gekregen: in jaar t loopt het interval voor timmerlieden van 1 naar 9% en in jaat t+1 vinden we een interval van 8% tot 15%. De statistische bewijslast is nu onvoldoende om te beweren, dat de populatieproportie van timmerlui is veranderd. Immers, de intervallen overlappen elkaar en daarmee is het mogelijk, dat de populatieproportie van timmerlieden constant 8% is gebleven. Gelukkig is er een techniek die deze vraag exacter kan beantwoorden: de test op het verschil in gemiddelden of proporties (difference of/in means test, difference of/in proportions test). Door de steekproefgegevens efficiënter te ordenen, werkt deze test meer discriminerend dan de voorgaande, dat wil zeggen dat verschillen in populatiegrootheden eerder zijn vast te stellen. De test voor de verschillen in proporties is overigens alleen betrouwbaar voor grote steekproeven (steekproefomvang groter dan 100). (23) Zojuist vermelde toetsen worden onder meer gebruikt door J.A. de Jonge en L. Blok en J.M.M. de Meere. (24)

Natuurlijk komt het in kwantitatief historisch onderzoek nog al eens voor, dat men wil weten of méér dan twee gemiddelden in de populatie van elkaar verschillen. Deze vraag kan beantwoord worden met behulp van de zogenaamde enkelvoudige variantie-analyse (one-factor analysis of variance). Wat betreft proporties, die zoals we zagen een dummy-verdeling hebben, kan wel gebruik worden gemaakt van de formules van variantie-analyse, maar men zondigt dan tegen op zijn minst twee van de vooronderstellingen ervan (die, voor de meer ingewijden, luiden, dat de populaties een normale - klokvormige - verdeling moeten hebben en bovendien gelijke varianties; aan geen van beide eisen wordt door dummy-variabelen voldaan). De eventuele resultaten worden daardoor zeer onbetrouwbaar!

H. van Dijk gebruikt een aanverwante toets: de Kruskall-Wallis toets. (25) Deze toets is de zogenaamde nonparametrische tegenhanger van de variantie-analyse. Nonparametrische, ook wel verdelingsvrij genoemde statistiek stelt minder eisen aan de populatie dan de, tot nog toe behandelde, parametrische technieken, maar geeft vaak ook wat mindere resultaten. (26) Nonparametrische toetsen worden gebruikt als men het vermoeden heeft, dat de populaties niet aan de eisen van de "normale" parametrische statistiek voldoen en natuurlijk ook als de nonparametrische toetsen beter (efficiënter in vaktaal) zouden zijn dan de parametrische. Beter (efficiënter) wil in dit verband zeggen, dat de betrouwbaarheidsintervallen smaller zijn. De efficiëntie van de Kruskall-Wallis toets ten opzichte van variantie-analyse is ongeveer 95%, hetgeen betekent dat de Kruskall-Wallis toets iets minder goed is. Bovendien kunnen nonparametrische toetsen veelal variabelen van een lager meetniveau aan. Zo heeft de enkelvoudige variantie-analyse een ratio/interval variabele nodig om betrouwbare resultaten te verkrijgen en de Kruskall-Wallis toets slechts een ordinale variabele. Vrijwel elke parametrische toets heeft wel een nonparametrische tegenhanger. (27)

Wordt bij de variantie-analyse met gemiddelden gewerkt, de Kruskall-Wallis toets kijkt naar medianen en kan de volgende vraag (met een bepaalde waarschijnlijkheid) oplossen: als we drie of meer steekproefmedianen hebben, komen die dan alle uit populaties met de zelfde mediaan of niet? Dus let wel: als de toets nee zegt, dan is het mogelijk, dat twee populatie-medianen gelijk zijn en alleen de derde afwijkt. (28) De Kruskall-Wallis toets op zich is niet zo interessant, want ze kan melden dat er verschil is, maar er wordt niet bij verteld hoe groot het verschil is. Als de steekproef maar groot genoeg is, dan is er vrijwel altijd wel een verschil aan te wijzen, hoe miniem ook. Eigenlijk zouden we, net als bij de één- en twee-steekproef toetsen betrouwbaarheidsintervallen moeten kennen. De formules hiervoor zijn inderdaad ontwikkeld zowel voor de Kruskall-Wallis toets als voor de variantie-analyse. (29)

Alle bovenstaande toetsen behoren tot de categorie a. We komen nu toe aan de categorie b: de relatie tussen twee of meer variabelen. Ten aanzien van die relatie kan men zich afvragen: 1. óf er verband is en, zo ja, 2. hoe de omvang van dat verband is en 3. hoe de aard van het verband is. Ook hier kunnen weer uitspraken worden gedaan over de populatie. De te gebruiken technieken verschillen weer met het meetniveau van de variabelen. Op nominaal niveau worden de chi-kwadraat toets (chi-square test) en de contingentie-coëfficiënt (contingency coefficient) wel gebruikt om een antwoord te geven op respectievelijk de vragen 1. en 2. Vrij nieuw in de statistiek is de zogenaamde log-lineaire analyse (voluit: log-linear models and latent-structure analysis), die op nominaal niveau (maar ook in combinatie met andere niveua's) uitspraken kan doen over de aard van verbanden. De chi-kwadraat toets en de contingentie-coëfficiënt worden tot de gepaarde waartrische statistiek gerekend. Log-lineaire analyse zit tussen parametrische en nonparametrische statistiek in. Op ordinaal niveau kunnen de, nonparametrische, rangcorrelatie-coëfficiënten (rank correlation coefficients) een antwoord geven op vraag 2. Bekende rang-correlatie-coëfficiënten zijn de r_s van Spearman en de tau van Kendall. Op ratio/interval niveau zijn zeer veel technieken beschikbaar waaronder de correlatie-coëfficiënt van Pearson (de "gewone" correlatiecoëfficiënt) en de regressie-analyse, die respectievelijk betrekking hebben op de vragen 2. en 3. Regressieanalyse kan met behulp van dummy-variabelen uitgebouwd worden voor nominaal en ordinaal niveau.

De chi-kwadraat toets test of de frequenties in een kruistabel zoveel afwijken van wat men op grond van de (nul)hypothese "geen relatie" zou mogen verwachten, dat er nauwelijks meer van toeval sprake kan zijn. Als de afwijking dermate groot is, dan is er volgens deze toets sprake van verband. (30) Een voorbeeld, ontleend aan S. Siegel, moge dit verduidelijken: (31)
Men kan veronderstellen, dat de startpositie bij een paarderace op een cirkelvormig parcours van invloed is op de uitslag: het paard dat het dichtst bij de binnenrailing start zou bevoordeeld kunnen zijn. Men kan dit onderzoeken door een aantal koersen (144 in het voorbeeld) te volgen en steeds winnaar en bijbehorende startpositie te noteren. Siegel geeft de volgende tabel, waarbij startposietie nummer 1 het dichtst bij de binnenrailing gelegen is, startpositie nummer 2 daarnaast ligt, enzovoort:

Startpositie	Aantal overwinningen
1	29
2	19
3	18
4	25
5	17
6	10
7	15
8	11
Totaal	144

Indien er nu geen sprake is van enige invloed van de startpositie op het aantal overwinningen, dan mag men verwachten, dat het aantal overwinningen per startpositie steeds 18 bedraagt. Met de chi-kwadraat toets wordt nu bekeken of de afwijkingen van het werkelijke aantal gescoorde overwinningen en het onder de nulhypothese te verwachten aantal overwinningen (18) dermate groot zijn, dat niet meer van toeval kan worden gesproken. Als we nu veronderstellen dat er in de populatie geen verband bestaat tussen startpositie en aantal overwinningen, dan blijkt de kans van het toevallig voorkomen van de waarden in bovenstaande tabel tussen de 2% en de 5% te liggen. Indien we nu van te voren zouden hebben gezegd, zoals Siegel dat deed, dat we pas zouden geloven in enig verband als die kans 1% of minder zou zijn geweest, dan kunnen we op basis van deze uitkomst de nulhypothese niet verwerpen. Met andere woorden: we moeten dan concluderen, dat we geen verband aan kunnen tonen.

Wonnacott en Wonnacott uiten zware kritiek op de chi-kwadraat toets: volgens hen zou bij deze toets de verkeerde vraag worden gesteld. (32) Hun redenering is als volgt: er wordt een vraag gesteld waarvan het antwoord in feite al bekend is; ook in de totale populatie van paarderaces, die, laten we zeggen, bestaat uit alle races die ooit op het bewuste parcours zijn gehouden en, zo U wilt, nog zullen worden gehouden, zal, als er geen verband bestaat tussen startpositie en winnaar, het aantal overwinningen per startpositie nooit exact het totaal aantal overwinningen gedeeld door 8 zijn. Als de steekproef nu maar groot genoeg gemaakt wordt, dan zullen deze toevallige verschillen in de populatie tot uiting komen in de in de chi-kwadraat toets, met als resultaat dat de nulhypothese altijd verworpen wordt en er dus altijd een verband te ontdekken valt. Er werd dus gevraagd of er verschillen in resultaat tussen de startposities waar te nemen zijn, terwijl we al wisten dat dat zo was. In hun behandeling van de chi-kwadraat toets geven Wonnacott en Wonnacott een aantal alternatieve toetsen.

Het probleem dat zich hier voordoet zagen we reeds bij de Kruskall-Wallis toets. In het algemeen kan gesteld worden, dat betrouwbaarheidsintervallen de voorkeur verdienen boven de zogenaamde hypothesetoetsen zoals Kruskall-Wallis in zijn normale vorm en de chi-kwadraat toets. Bij hypothese-toetsen wordt slechts één hypothese aan een onderzoek onderworpen, terwijl men betrouwbaarheidsintervallen kan zien als alle (dit is: de verzameling van) aanvaardbare hypotheses. (33) Bijvoorbeeld: een betrouwbaarheidsinterval voor een steekproefproportie kan gezien worden als alle hypotheses omtrent de populatieproportie die aanvaardbaar zijn. Bij hun alternatieven voor de chi-kwadraat toets werken Wonnacott en Wonnacott dan ook met betrouwbaarheidsintervallen. Het gegeven voorbeeld is overigens een goede illustratie van de wijze van redeneren van de statisticus. De chi-kwadraat toets wordt onder meer gebruikt door Van Dijk. (34)

De contingentie-coëfficiënt is een maatstaf voor de mate van verband tussen twee nominale variabelen en is nauw gerelateerd aan de chi-kwadraat toets. (35) Het is één getal, waarvan de bovengrens afhankelijk is van het aantal categorieën in de kruistabel (8 in het bovenstaand voorbeeld). Dit brengt met zich mee, dat contingentie-coëfficiënten van kruistabellen met een verschillend aantal categorieën niet vergelijkbaar zijn. Als de contingentie-coëfficiënt de waarde 0 (nul) aanneemt, is er sprake van een volkomen afwezigheid van verband tussen beide variabelen. Men kan op basis van de contingentie-coëfficiënt een uitspraak doen omtrent de aanwezigheid van verband in de populatie, maar deze uitspraak lijdt aan het zelfde euvel als de chi-kwadraat toets en leidt bovendien tot het zelfde resultaat. Ook de contingentie-coëfficiënt wordt onder meer gebruikt door Van Dijk. (36)

De rangcorrelatie-coëfficiënten van Spearman en Kendall geven de mate van verband aan tussen twee variabelen, die beide tenminste van ordinaal niveau moeten zijn. (37) Beide lopen van -1 naar +1. Een coëfficiënt van -1 geeft aan, dat er een volledig invers (tegen elkaar inwerkend) verband bestaat, 0 (nul), dat er geen verband is en +1, dat er een volledig positief verband bestaat. Spearman's r_s is in feite de, aanstonds te behandelen, correlatie-coëfficiënt van Pearson (r) vertaald voor ordinale variabelen. Omdat er verschillende formules worden gebruikt bij de berekening van r_s en tau zijn deze grootheden niet met elkaar vergelijkbaar, maar wel zijn r_s'en onderling en tau's onderling te vergelijken. Dát is dus een verbetering ten opzichte van de contingentie-coëfficiënt. Beide kunnen een uitspraak doen omtrent het verband in de populatie en zijn daarbij ook even goed, hoewel wat minder dan de r van Pearson. Zoals al eerder vermeld, verdient het aanbeveling om te werken met betrouwbaarheidsintervallen. Het interval voor Spearman's r_s kon ik helaas niet op het spoor komen, maar het bestaat ongetwijfeld. Het interval voor Kendall's tau is te vinden bij M. Hollander en D.A. Wolfe. (38) Rangcorrelatie wordt onder meer gebruikt door Van Dijk, De Meere en Blok en De Meere. (39)

Regressie- en correlatie-analyse (waarbij de correlatie-coëfficiënt van Pearson wordt gebruikt) zijn nauw samenhangende technieken. (40) Een korte behandeling: in enkelvoudige lineaire regressie-analyse worden twee, doorgaans interval/ratio variabelen tegen elkaar afgezet in een vlak, waarbij met punten de gepaarde waarnemingen worden aangegeven. De gepaarde waarnemingen bestaan uit de bij de waarde van de ene variabele behorende waarde van de andere variabele. Als we bijvoorbeeld het verband tussen de consumptie en het nationale inkomen in de loop der tijd zouden bestuderen, dan bestaan de gepaarde waarnemingen uit de consumptie en het nationale inkomen in jaar t, de consumptie en het nationale inkomen in jaar t+1, enzovoort. Het aldus tegen elkaar afzetten van de variabelen levert een "puntenwolk" op. Door deze puntenwolk wordt een rechte lijn getrokken (de zogenaamde regressie-lijn), die zo nauwkeurig mogelijk aansluit bij alle punten. Op basis van deze lijn kunnen we, uitgaande van een waarde van de ene variabele, een schatting maken van de waarde die de andere variabele daarbij aanneemt, met andere woorden: de lijn geeft de aard van het verband weer.

De correlatie-coëfficiënt van Pearson (r) geeft aan in hoeverre de regressie-lijn aansluiting vindt bij de puntenwolk. De correlatie-coëfficiënt neemt de waarde -1 aan indien alle punten op de regressie-lijn liggen en de helling er van neerwaarts (negatief) is. De waarde +1 wordt bereikt als er sprake is van een stijgende (positieve) helling en alle punten op de regressielijn liggen. In beide gevallen is de relatie, binnen de steekproef althans, perfect: de schatting van een waarde van de ene variabele op basis van een, in de steekproef voorkomende waarde van de andere variabele met behulp van de regressie-lijn is altijd correct. Volkomen positieve en volkomen negatieve correlatie komt in de historische praktijk niet voor. De correlatie-coëfficiënt neemt de waarde 0 (nul) aan, indien de regressielijn geen helling vertoont (evenwijdig loopt aan de horizontale as), ongeacht de positie van de punten. In dit geval immers zal de schatting van de op de verticale as afgezette variabele met behulp van de op de horizontale as afgezette variabele altijd de zelfde zijn; dit betekent, dat er totaal geen lineair verband bestaat tussen beide variabelen. Ook een correlatie-coëfficiënt met de waarde 0 (nul) komt in de praktijk niet voor. De conclusie is, dat hoe meer de absolute waarde van de correlatie-coëfficiënt naar 1 nadert, het verband tussen de twee variabelen des te groter is. Uitspraken over de populatie-regressie-lijn en de populatie-correlatie-coëfficiënt kunnen worden gedaan, inclusief betrouwbaarheidsintervallen. Regressie- en correlatie-analyse kan worden uitgebreid tot het opnemen van meer dan twee variabelen en tot niet-lineaire verbanden.

De uitleg van regressie en correlatie was, vanwege de beperkte ruimte en de noodzaak om het relatief eenvoudig te houden, betrekkelijk naïef, zodat het misschien niet duidelijk is geworden, dat regressie de meest krachtige techniek van de statistiek is. Met behulp van regressie-analyse kunnen ook uitspraken gedaan worden als de "verklarende" variabelen van nominaal of ordinaal niveau zijn (in het bovenstaande voorbeeld wordt de consumptie "verklaard" uit het nationale inkomen): men kan in dat geval de nominale of ordinale variabelen omzetten in "dummy"variabelen. Indien echter de "te verklaren" variabele nominaal of ordinaal is, moet men zich tot een andere techniek wenden. In dit geval kan gedacht worden aan log-lineaire-analyse, die een verzoening tussen "correlators" en "crosstabbers" (de twee hoofdstromingen in de toegepaste statistiek) tracht te bewerkstelligen op het nominale meetniveau en daarmee waarschijnlijk de grootste sensatie vormt in de ontwikkeling van de theoretische statistiek in het laatste decennium. Ook als de "te verklaren" variabele een interval/ratio schaal heeft, kan men over-

wegen log-lineaire analyse te gebruiken. (41)
Pearson's correlatie-coëfficiënt wordt gebruikt bij Diederiks c.s., Klep en De Meere (42), regressie-analyse wordt momenteel toegepast door Schuurman (43) en log-lineaire analyse tot nog toe door geen enkele regionale historicus. In deze paragraaf werden een aantal statistische technieken die door Nederlandse regionale historici gebruikt werden en worden aan de orde gesteld, alsmede enige aanverwante technieken.
In een aantal gevallen werden suggesties gedaan om de gebruikte techniek enigszins aan te passen. In de meeste gevallen betrof dit het feit, dat betrouwbaarheidsintervallen de voorkeur verdienen boven hypothesetoetsingen.

noten

1. Zie voor de macro-economische theorie het bijzonder goed bij het model van Klep en Daems aansluitende: S. Korteweg, F.A.G. Keesing, H. de Haan en J.K.T. Postma, Het moderne geldwezen. Deel I: Macro-economische uitgangspunten (Amsterdam, 1978).
Zie voor de aanbevolen statistische werken de tweede paragraaf.
2. Paul M.M. Klep, Bevolking en Arbeid in transformatie: Brabant 1700-1900.Een analyse van ongelijktijdige ontwikkelingen in een maatschappij op weg naar moderne economische groei (Leuven, 1978). Van dit, door mij gebruikte, proefschrift verscheen een handelseditie onder de titel: Paul M.M. Klep, Bevolking en Arbeid in transformatie, een onderzoek in Brabant 1700-1900 (Nijmegen, 1981).
3. Zie voor de exacte afbakening van het door Klep bestudeerde gebied: P.M.M. Klep, Bevolking en Arbeid, 1978.
4. Zie voor de produktiefunctie met vaste technische coëfficiënten bijvoorbeeld: S. Korteweg e.a., Het moderne geldwezen,159-164.
5. Idem,162.
6. P.M.M. Klep, Bevolking en Arbeid, 226.
7. Men kan bij de verklaring van de grondrente echter niet gebruik maken van het klassieke marginalistische verdelingskriterium als men met een lineaire produktiefunctie werkt.
8. P.M.M. Klep, Bevolking en Arbeid, 214.
9. Dus ex post geldt (zonder overheid): S - I = Exp - Imp, waarin S de besparingen voorstelt, I de investeringen, Exp de exporten en Imp de importen. Zie hiervoor: S. Korteweg e.a., Het moderne geldwezen, 67.
10. O.W.A. Boonstra, Gecomputeriseerd Historisch Onderzoek. Stand van zaken; Perspectief (Nijmegen, 1981). Uitgave van de sectie geschiedenis van de Katholieke Universiteit Nijmegen en, voor de belangstellenden, te bestellen door overmaking van ƒ10,- op postgironummer 2491145 t.n.v. O.W.A. Boonstra te Nijmegen onder vermelding van "Rapport G.H.O."
11. Thomas H. Wonnacott en Ronald J. Wonnacott, Introductory Statistics for Business and Economics (Santa Barbera, 1977). Een verkorte versie is verkrijgbaar onder de titel: Introductory Statistics.
12. G. Ohlin, "No safety in Numbers: Some pitfalls of Historical Statistics", in: H. Rosovsky ed., Industrialisation in two systems: Essays in honor of Alexander Gerschenkron (New York, 1966). Herdrukt in: Roderick Floud ed., Essays in Quantitative Economic History (Oxford, 1974)
13. F.R. Jolliffe, Commonsense Statistics for Economists and others (London, 1974).
R. Floud, An Introduction to Quantitative Methods for Historians (London, 1973).
E.W. Willemsen, Understanding Statistical Reasoning (San Francisco, 1974).
C.M. Dollar en R.J. Jensen, Historian's Guide to Statistics. Quantitative Analysis and Historical Research (New York, 1971).
14. Zie voor aselecte steekproeven bijvoorbeeld: T.H. Wonnacott en R.J. Wonnacott, Introductory Statistics, 143 e.v. Als het aantal elementen in de populatie klein is, moet met teruglegging (na trekking van elk element) worden gewerkt om de steekproef 'Simple' te houden. Men kan echter ook achteraf correcties uitvoeren.
15. J.A. de Jonge, "Delft in de negentiende eeuw, van stille nette plaats tot centrum van industrie", in: Economisch- en Sociaal-historisch Jaarboek, 37 (Den Haag, 1974)
16. T.H. Wonnacott en R.J. Wonnacott, Introductory Statistics, 673.
17. Zie voor de verschillende vormen van steekproeftrekking bijvoorbeeld: T.H. Wonnacott en R.J. Wonnacott, Intodùctory Statistics, 673-684, of W.G.Cochran, Sampling Techniques (New York, 1977).
18. Zie voor frequentie-verdelingen, de maatstaven van centrale tendentie en de spreidingsmaatstaven bijvoorbeeld: T.H. Wonnacott en R.J. Wonnacott, Introductory Statistics, 12 e.v.
19. Zie voor de variatie-coëfficiënt bijvoorbeeld: Taro Yamane, Statistics, an introductory analysis (New York, 1973) 77 e.v.
20. Zie voor de kruistabel bijvoorbeeld: R. Floud, An Introduction to Quantitative methods, 48 e.v.
21. L. Blok en J.M.M. de Meere, "Welstand, Ongelijkheid in Welstand en Censuskiesrecht in Nederland omstreeks het midden van de Negentiende Eeuw", in: Economisch- en Sociaal-Historisch Jaarboek, 41 (Den Haag, 1978), H.A. Diederiks, C.A. Davids, D.J. Noordam en H.D. Tjalsma, Een stad in achteruitgang, Sociaal-historische studies over Leiden in de achttiende eeuw. (Leiden, 1978), H. van Dijk, Rotterdam 1810 - 1880, Aspecten van een stedelijke samenleving (Rotterdam, 1976), P.M.M. Klep, Bevolking en Arbeid.
22. Zie voor de bedoelde betrouwbaarheidsintervallen bijvoorbeeld: T.H. Wonnacott, Introductory Statistics, 199 - 211 en 223 - 227, of T. Yamane, Statistics, 669 - 672 en 718 - 722.
Zie voor de binomiale verdeling bijvoorbeeld: T.H. Wonnacott, Introductory Statistics, 82 - 86, of T. Yamane, Statistics, 677 - 733, waarbij op de pagina's 722 - 724 meer over het verband tussen proporties en de binomiale verdeling.
23. Zie voor toetsen op verschillen in twee gemiddelden bijvoorbeeld T.H. Wonnacott, Introductory Statistics, 212 - 220 en 227, of T. Yamane, Statistics, 672 - 675.
24. L. Blok, Ongelijkheid in welstand, J.A. de Jonge, Delft in de negentiende eeuw, 211.
25. H. van Dijk, Rotterdam 1810-1880, 263 en 427 - 428.
26. De termen nonparametrische en verdelingsvrije statistiek dekken niet exact de zelfde lading; zie hiervoor bijvoorbeeld: J.V. Bradley, Distribution-Free Statistical Tests (Englewood Cliffs, New Jersey, 1968).
27. Zie voor een eenvoudige inleiding in de nonparametrische statistiek bijvoorbeeld: T.H. Wonnacott, Introductory Statistics, 471 - 500. Het klassieke handboek voor nonparametrische statistiek is S. Siegel, Nonparametric Statistics for the Behavioral Sciences (Tokyo, 1956).
28. Zie voor de Kruskall-Wallis toets bijvoorbeeld: S. Siegel, Nonparametric Statistics, 184 - 194.
29. Zie voor betrouwbaarheidsintervallen bij de Kruskall-Wallis toets bijvoorbeeld: C.H. Kraft en C. van Eeden, A Nonparametric Introduction to Statistics (New York, 1968) 149 - 153.
Zie voor betrouwbaarheidsintervallen bij de variantie-analyse bijvoorbeeld: T.H. Wonnacott, Introductory Statistics, 290 - 297.
30. Zie voor een diepgaander behandeling van de chikwadraat toets bijvoorbeeld: T.H. Wonnacott, Introductory Statistics, 501 - 522, of S. Siegel, Nonparametric Statistics, 42 - 47.
31. S. Siegel, Nonparametric Statistics, 44 - 46.

vervolg op pagina 20