

1984-155
CBM
R
7626
1984
155



Bestemming <i>ECO</i>	TIJDSCHRIFTENBUREAU BIBLIOTHEEK KATHOLIEKE HOGESCHOOL TILBURG	Nr. <i>R41</i>
--------------------------	---------------------------------------------------------------------------	-------------------

subfaculteit der econometrie

RESEARCH MEMORANDUM



TILBURG UNIVERSITY
DEPARTMENT OF ECONOMICS
Postbus 90153 - 5000 LE Tilburg
Netherlands



Over de kosten per ziekenhuisopname
en levensduurmodellen¹⁾

P.F.P.M. Nederstigt²⁾

1) Dit betreft een deelonderzoek van het project "Kostenverbijzondering naar homogene patiëntengroepen" dat wordt uitgevoerd in samenwerking tussen de Subfaculteit Econometrie van de Katholieke Hogeschool Tilburg en het Nationaal Ziekenhuis Instituut.

2) Dank is verschuldigd aan prof.dr. R. Doornbos en prof.dr. B.B. van der Genugten voor hun waardevolle suggesties en commentaar op een eerdere versie van dit rapport.

Inhoudsopgave

	<u>pag</u>
1. Inleiding	1
2. Gegevensbestand	3
3. Uitbijtertoetsen	5
4. Levensduurmodellen	9
5. 'Maximum Product of Spacings'-schattingsprocedure	14
6. Schattingsresultaten	17
7. Verschillen-analyse	25
8. Budgetbepaling	29
8.1. Amerikaanse oplossingen	29
8.2. Alternatieven	31
9. Verrichtingen- en kostenprofielen	36
10. Samenvatting	39

Over de kosten per ziekenhuisopname en levensduurmodellen

1. Inleiding

Het is mogelijk om op basis van medische patiëntengegevens, zoals vastgelegd door de Stichting Medische Registratie (SMR),¹⁾ homogene patiëntengroepen te definiëren, Diagnosis Related Groups (DRG's) genaamd; zie Fetter e.a. (1980) en Nederstigt (1982). Deze DRG's zijn homogeen in tweërlei opzicht:

- medisch. DRG's zijn herkenbaar voor medici; d.w.z. dat een medicus een mogelijke behandeling zal kunnen aangeven als een DRG aan hem wordt voorgelegd;
- economisch. Patiënten uit dezelfde DRG doen in min of meer dezelfde mate een beroep op de ziekenhuisvoorzieningen; d.w.z. zij veroorzaken min of meer dezelfde kosten.

Door integratie van de SMR-gegevens met de factuurgegevens van elke patiënt is het mogelijk om voor elke DRG een verrichtingenprofiel te bepalen dat het gemiddelde beslag op de ziekenhuisvoorzieningen voor een patiënt uit die DRG aangeeft; zie Nederstigt (1984a).

M.b.v. de gegevens uit de boekhouding is het mogelijk om de kostprijs per verrichting te bepalen. 'Herziening' van de patiëntenfactuur door toepassing van de kostprijs per verrichting i.p.v. het tarief, levert de "kosten per patiënt" op, i.p.v. de "opbrengsten per patiënt" (zijnde de kosten voor de patiënt!). Hieruit vallen vervolgens de kostenprofielen per DRG af te leiden, die de gemiddelde kosten per patiënt uit die DRG specificeren. Deze kostenprofielen zijn vanzelfsprekend ook rechtstreeks uit de verrichtingenprofielen af te leiden door elke verrichting te wegen met zijn kostprijs.

De verrichtingen- en kostenprofielen zijn op velerlei wijze toe te passen. De belangrijkste mogelijkheden zijn:

- Ziekenhuisvergelijking.

Vergelijken van de gemiddelde kosten per opname of per verpleegdag stuit op problemen. De patiëntenpopulaties van de ziekenhuizen zijn immers anders van samenstelling: in een academisch ziekenhuis

zullen andere patiënten behandeld worden dan in een basisziekenhuis (case-mix verschil). Door evenwel de vergelijking op DRG-niveau uit te voeren is dit probleem te ondervangen. Per DRG gaat het immers om patiënten die vanuit medisch oogpunt vergelijkbaar zijn. Verschillen in profielen vinden derhalve hun oorzaak in verschillen in behandeling, zich uitend in een ander beslag op de ziekenhuisvoorzieningen, en in kostprijsverschillen.

- Interne budgettering.

Bij het vaststellen van het budget voor een specialist gaat men veelal uit van het aantal verrichtingen dat hij het voorgaande jaar heeft aangevraagd. Verschillen tussen budget en realisatie zijn daardoor moeilijk te analyseren. DRG's geven echter de mogelijkheid tot een verregaande verschillenanalyse door het onderscheiden van volumever verschillen (aantal patiënten), case-mix verschillen, behandelverschillen en, voor zover het budget in guldens luidt, prijsverschillen. Alleen het behandelverschil mag aan de specialist worden toegeschreven.

Elke patiënt blijft een afzonderlijk geval. Niet alle patiënten uit dezelfde DRG zullen derhalve dezelfde kosten oproepen. Het verrichtingen- en kostenprofiel dat voor een DRG gevonden wordt, staat derhalve onder invloed van het toeval:

- in een DRG kunnen dit jaar iets meer patiënten ingedeeld zijn waarbij zich complicaties hebben voorgedaan;
- er kunnen dit jaar relatief meer oudere patiënten zijn opgenomen, waarbij sprake is van een uitgebreider pre-operatief poliklinisch vooronderzoek;
- de plaatsingsmogelijkheden in verpleeghuizen en de opvangmogelijkheden thuis waren minder groot dan andere jaren.

Weliswaar zal de invloed van deze factoren op de kosten niet groot zijn, omdat in dat geval de DRG-definities daarop ingespeeld zouden hebben door onderscheid te maken naar leeftijdscategorieën of naar aan- resp. afwezigheid van nevenaandoeningen en complicaties, of de wijze van ontslag uit het ziekenhuis; het zal echter duidelijk zijn dat voor een willekeurige patiënt uit een bepaalde DRG de kosten van de ziekenhuisopname beschouwd moeten worden als een stochastische grootheid met een kansverdeling met DRG-specifieke parameters.

De vraagstelling die in het kader van ziekenhuisvergelijking of verschillenanalyse na afloop van de budgetperiode opkomt, dient derhalve in te spelen op dit stochastisch karakter van de kosten per opname en zal luiden:

- zijn de verschillen tussen de frequentieverdelingen van de kosten per opname statistisch significant of, minder verregaand,
- zijn de verschillen van de gemiddelde kosten per opname stochastisch significant.

In het navolgende zal achtereenvolgens beschreven worden:

- in hoeverre het mogelijk is m.b.v. uitbijtertoetsen na te gaan of de verschillen slechts veroorzaakt worden door één of meerdere extreme waarnemingen,
- met welke kansverdelingen men, vanuit theoretisch standpunt bezien, de empirische frequentieverdelingen van de kosten per opname van een DRG kan trachten te beschrijven,
- welke schattingsprocedure bij het aanpassen van deze frequentieverdelingen gebruikt is,
- de schattingsresultaten,
- hoe de hierboven geformuleerde vragen beantwoord kunnen worden.

Allereerst zullen we evenwel kort ingaan op het gegevensbestand waarop de resultaten gebaseerd zijn.

2. Gegevensbestand

Van drie algemene ziekenhuizen hebben we de beschikking gekregen over gegevens m.b.t. de patiënten die in 1980 uit het ziekenhuis zijn ontslagen. Deze gegevens betreffen:

- SMR-gegevens per patiënt. Aan de hand van deze gegevens kan een patiënt in een DRG ingedeeld worden,
- factuurgegevens per patiënt. Aan de hand van onderzoek naar de kostprijs per verrichting was het mogelijk deze factuurgegevens om te zetten in kosten per patiënt.

De in dit rapport gepresenteerde resultaten hebben geen betrekking op het totale patiëntenbestand. Ten eerste hebben we ons beperkt tot patiënten met keel-, neus- en ooraandoeningen. Ten tweede hebben we ons beperkt tot de patiënten die na hun ontslag uit het ziekenhuis naar

huis zijn gegaan; patiënten die zijn overgeplaatst naar een ander ziekenhuis of verpleeghuis, die het ziekenhuis tegen medisch advies in hebben verlaten, of die zijn overleden zijn buiten beschouwing gelaten omdat bij hen duidelijk sprake is van een kenmerk dat zich zal doen gelden in een van de overige patiënten afwijkende kostenpatroon.

Voor een aantal DRG's staat in tabel 1 aangegeven hoeveel patiënten in deze DRG zijn ingedeeld (n), hoe hoog de gemiddelde kosten per opname (m) en de standaardafwijking (s) zijn en hoe groot het excess is.²⁾ Voor een gedetailleerde beschrijving van alle DRG-definities m.b.t. patiënten met keel-, neus- en ooraandoeningen en bij elke DRG behorende verrichtingen- en kostenprofielen zij verwezen naar Nederstigt (1984b).

Tabel 1. Statistics van de kosten per opname voor een aantal DRG's m.b.t. het specialisme KNO

	n	m	s	excess
<u>geopereerd</u>				
tympanoplastiek	108	3349	833	-1,28
mastoidectomie	62	3541	1243	3,39
stapedectomie	24	3223	410	1,40
neusplastiek	128	2006	526	1,28
neusbijholten, > 15 jaar	89	2285	651	0,40
oorschelp	120	2058	508	4,00
overige neusoperaties	211	1755	340	1,31
tonsillectomie, > 15 jaar	195	1672	311	0,03
keeloperaties	145	1176	770	8,03
trommelvliesbuisje, > 15 jaar	39	1147	701	2,96
trommelvliesbuisje, 0-14 jaar	59	845	432	6,14
<u>niet-geopereerd</u>				
otitis media	33	3025	3074	0,70
laryngitis, faryngitis e.d.	58	2703	3074	4,92
sinusitis	24	1743	712	1,41
neustrauma	49	1637	728	1,20
neusbloeding	34	1310	1060	1,65
overige KNO-aandoeningen	67	2046	2163	2,29

3. Uitbijtertoetsen

Bij de beantwoording van de in de inleiding gestelde vragen zal men in eerste instantie willen nagaan in hoeverre de verschillen veroorzaakt worden door enkele extreme waarnemingen. Het identificeren van dergelijke uitbijters is des te meer van belang als men zich realiseert dat er fouten in de patiëntengegevens kunnen zitten:

- Fouten in de medische patiëntengegevens, zoals het helemaal niet of foutief coderen van een (neven)diagnose of operatieve ingreep, kunnen tot gevolg hebben dat een patiënt in de verkeerde DRG wordt ingedeeld.
- Er kunnen fouten zitten in de factuurgegevens, bijv. persoonsverwisseling, die leiden tot verkeerde facturen.
- In de SMR-gegevens wordt een ander patiënten-identificatienummer gehanteerd dan in de factuurgegevens. Het is mogelijk dat bij de koppeling van beide identificatienummers fouten begaan zijn.
- Bij sommige patiënten wordt een (gedeelte van de) factuur meerdere keren toegestuurd aan de ziekenfondsen voordat de facturen worden geaccepteerd. De creditnota's worden evenwel niet op naam van de patiënt gesteld maar op een verzamelingsrekening voor het ziekenhuis geboekt. Daardoor komt bij een aantal patiënten een dubbele factuur voor. Daarop is bij het screenen van de ontvangen gegevens zo goed mogelijk ingespeeld, maar bij een aantal patiënten zullen te veel of te weinig factuuronderdelen als 'dubbel' zijn aangemerkt.

Het is in dit kader dat Andrews e.a. de volgende uitspraak doen (1973, pag. 244): "A good statistician (or research worker) never uses the mean, even though he may honestly claim to do so. In reality, he will look at the data and set a side outliers and suspected outliers in order to treat them separately; then he will take the mean of the rest. But this amounts to using the mean with a (possible very vague and subjective) rejection procedure, not 'the mean' in its strict mathematical sense".

Andrews e.a. stellen vervolgens een aantal robuuste schattingsmethodieken voor het gemiddelde aan de orde. Deze methoden zijn voor ons evenwel niet bruikbaar om dat ze betrekking hebben op symmetrische kansverdelingen; zie ook par. 4 en 7.1.

Een andere mogelijkheid is het toepassen van 'uitbijtertoetsen', die expliciet zijn geconstrueerd met het oog op het toetsen op de aanwezigheid van extreme waarnemingen. De meeste bekende uitbijtertoetsen komen hieronder ter sprake.

Deze uitbijterstoetsen worden veelal impliciet geformuleerd als een zogenaamde 'slippage alternative'-model. De meest algemene omschrijving van het 'normal slippage'-model wordt gegeven door Ferguson (1962). Voor een 'shift in location' luidt dit model: " x_1, x_2, \dots, x_n arise independently from normal distributions with common variance σ^2 . (Under H_0 they have common mean μ .) There are known constants a_1, a_2, \dots, a_n (most of which will be zero), an unknown parameter Δ and an unknown permutation (v_1, v_2, \dots, v_n) of $(1, 2, \dots, n)$ such that the normal distributions from which the x_i arise have means

$$\mu_i = \mu + \sigma \Delta a_{v_i} \quad (i = 1, 2, \dots, n)$$

$H_1 : \Delta \neq 0$ (or one sided analogues, e.g. $\Delta > 0$ the a_i have the same sign)." Op analoge manier kan een 'slippage alternative' geformuleerd worden van een 'shift in dispersion'.

Dixon (1962) geeft voor normaal verdeelde waarnemingen een aantal uitbijtertoetsen waarin getoetst wordt op de aanwezigheid van één (of twee) uitbijters, bijv. $a_1 \neq 0$ en $a_i = 0$ ($i > 1$). Als men wil toetsen op de aanwezigheid van een extreem hoge waarneming, m.a.w. het alternatief luidt dat $\text{sign}(a_1) = \text{sign}(\Delta)$, zijn onderstaande toetsingsgrootheden voorhanden, uitgaande van een oplopend gerangschikte steekproef x_1, x_2, \dots, x_n :

- studentized extreme deviate

$$B = \frac{x_n - \bar{x}}{S}$$

- studentized range

$$C = \frac{x_n - x_1}{S}$$

- gemodificeerde F-toets

$$D_1 = \frac{S_n^2}{S^2} \quad \text{voor uitbijter } x_n$$

of

$$D_2 = \frac{S_{n,n-1}^2}{S^2} \quad \text{voor uitbijters } x_n \text{ en } x_{n-1}$$

met

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\bar{x}_n = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i$$

$$S_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2$$

$$\bar{x}_{n-1} = \frac{1}{n-2} \sum_{i=1}^{n-2} x_i$$

$$S_{n-2}^2 = \frac{1}{n-2} \sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1})^2$$

- ratio voor range en subranges

$$r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}$$

$$r_{20} = \frac{x_n - x_{n-2}}{x_n - x_1}$$

$$r_{22} = \frac{x_n - x_{n-2}}{x_n - x_3}$$

Voor deze en andere toetsingsgrootheden heeft Dixon (1962) tabellen met kritieke waarden verzameld en bepaald.

Voor ons probleem zijn deze toetsen niet bruikbaar:

- deze toetsen zijn ontwikkeld voor kleine steekproeven en toetsen slechts op de aanwezigheid van één of twee uitbijters. Meerdere malen achtereen dezelfde toets toepassen levert een betrouwbaarheid op die

- kleiner is dan de betrouwbaarheid van iedere toets afzonderlijk en houdt het risico in van 'masking' omdat aanwezigheid van meerdere uitbijters kan betekenen dat ze niet als zodanig ontdekt worden;
- ze gaan uit van normaal verdeelde waarnemingen. De kosten per opname zijn echter niet normaal verdeeld; zelfs niet lognormaal onsymmetrisch zoals we in par. 6 zullen constateren.

Walsh (1950) heeft een verdelingsvrije uitbijtertoets ontwikkeld. De nulhypothese is daarbij dat alle waarnemingen uit populaties met een gemeenschappelijke mediaan afkomstig zijn, de alternatieve hypothese dat r waarnemingen afkomstig zijn uit een populatie met een grotere mediaan; r is daarbij niet afhankelijk van n , de steekproefomvang.

Deze toets is evenmin bruikbaar voor ons probleem

- r is vooraf gespecificeerd; dit betekent wederom dat de toets meerdere malen achtereen toegepast moet worden;
- deze toets gaat uit van een symmetrische populatieverdeling.

Deze laatste vooronderstelling heeft Walsh in latere voorstellen voor uitbijtertoetsen laten vallen; zie Walsh (1959). Daaraan kleven nog steeds een aantal praktische bezwaren, zoals het vooraf aangeven van het verwachte aantal outliers en het lage onderscheidingsvermogen. Maar ook theoretisch lijkt een verdelingsvrije toets weinig zinvol: het is weinig zinvol om extreem hoge of lage waarnemingen als uitbijter te bestempelen als men vooraf niets kan zeggen over de (vorm van de) kansverdeling. Afhankelijk van de veronderstelde kansverdeling zal men immers op voorhand reeds meer of minder extreme waarnemingen verwachten.

Darling (1952) heeft een algemenere methode ter benadering van de uitbijterproblematiek afgeleid. Hij heeft een integraalvorm bepaald voor de karakteristieke functie van de verdeling van

$$z_n = \frac{\sum_{i=1}^n x_i}{x_n}$$

met x_1, x_2, \dots, x_n de geordende waarnemingen van een steekproef uit een populatie met een volledig gespecificeerde verdeling, dus eventueel ook een asymmetrische. In bepaalde gevallen is deze karakteristieke functie inverteerbaar of kan de limietverdeling van z_n bepaald worden.

Een op z_n gebaseerde toets brengt echter wederom het bezwaar met zich mee dat hij meerdere malen moet worden uitgevoerd.

In het verlengde van deze uitbijtertoetsen van Darling kunnen een aantal meer specifieke toetsen genoemd worden die zijn verzameld door Barnett en Lewis (1978). Zij beschrijven diverse uitbijtertoetsen voor gespecificeerde verdelingen, waaronder de gamma-verdeling en de (log)normale verdeling, en in mindere mate de weibull-verdeling.

Sommige daarvan zijn gebaseerd op de eerder besproken toetsingsgrootheden van Dixon. Tevens blijkt uit het overzicht een aantal beperkingen die voor ons probleem van doorslaggevend belang zijn:

- voor gamma, weibull en lognormale verdeling moet de locatieparameter a bekend zijn;
- voor de weibull-verdeling moet ook de vormparameter b bekend zijn;
- voor de gamma-verdeling moet de vormparameter b bij de meeste toetsen bekend zijn;
- het aantal outliers op wier aanwezigheid getoetst wordt moet vooraf gespecificeerd zijn.

Omdat niet op voorhand aangegeven kan worden welke kansverdeling de kosten per opname hebben, laat staan welke waarden de parameters van die verdeling aannemen, kan ook van deze uitbijtertoetsen geen gebruik gemaakt worden.

Samenvattend kan geconcludeerd worden dat geen gebruik gemaakt kan worden van uitbijtertoetsen, gezien de eisen die gesteld worden aan de kennis m.b.t. kansverdeling en parameterwaarden en gezien het feit dat de toets meerdere malen achtereenvolgens uitgevoerd moet worden omdat niet vooraf aangegeven kan worden hoeveel uitbijters verwacht worden.

Voor de beantwoording van de in de inleiding gestelde vragen zal dus een andere weg bewandeld moeten worden.

4. Levensduurmodellen

De kosten per opname voor patiënten uit een bepaalde DRG zullen scheef naar rechts verdeeld zijn; immers:

- elke DRG vormt een homogene patiëntengroep. De kansverdeling zal derhalve unimodaal zijn, omdat anders meerdere DRG's gedefinieerd zouden

- zijn teneinde in te spelen op de multimodaliteit;
- er zal voor elke DRG sprake zijn van een minimum pakket aan verrichtingen³⁾ en bijbehorende kosten;
 - een soortgelijke begrenzing van de kosten naar boven toe bestaat niet. Er kan immers bij elke patiënt sprake zijn van kostenverhogende factoren, zoals een of meer bijkomende aandoeningen of complicaties of van minder adequate opvangmogelijkheden buiten het ziekenhuis.

Er zijn meerdere verdelingen die rechts scheef zijn. M.b.v. begrippen die gehanteerd worden m.b.t. levensduurmodellen is deze klasse van kansverdelingen verder in te perken; zie ook Lawless (1982).

Zij $f(t)$ de dichtheidsfunctie en $F(t)$ de cumulatieve verdelingsfunctie van de kosten per opname. Dan is

$$r(t) := f(t)/(1-F(t))$$

de 'failure rate' of uitvalkans. De grootheid $r(t) \cdot \Delta t$ wordt normaliter geïnterpreteerd als 'de kans dat een produkt dat reeds tot tijdstip t is meegegaan niet langer meegaat dan $t+\Delta t$ '. In onze situatie valt dit te herformuleren tot 'de kans dat een patiënt die minstens voor t aan kosten heeft opgeroepen niet meer kosten oproept dan $t+\Delta t$ (c.q. dat hij ontslagen wordt uit het ziekenhuis en zijn behandeling beëindigd is)'.
 Het ligt voor de hand om slechts die kansverdelingen in de beschouwing te betrekken, waarvoor $r(t)$ toeneemt; m.a.w. waarvoor geldt dat, nadat reeds kosten ter grootte van het minimum pakket zijn veroorzaakt, voor elke volgende gulden geldt dat de kans groter wordt dat het de laatste gulden is.

Beschouwen we onderstaande, meest toegepaste kansverdelingen in levensduurmodellen:

- lognormaal

$$f_L = \frac{1}{c\sqrt{2\pi}} \cdot \frac{1}{t-a} \cdot \exp\left\{-\frac{1}{2}\left(\frac{\ln(t-a)-b}{c}\right)^2\right\}$$

- weibull

$$f_w = \frac{b}{c} \cdot \left(\frac{t-a}{c}\right)^{b-1} \exp\left\{-\left(\frac{t-a}{c}\right)^b\right\}$$

- gamma

$$f_g = \frac{1}{c \cdot \Gamma(b)} \cdot \left(\frac{t-a}{c}\right)^{b-1} \exp\left\{-\frac{t-a}{c}\right\}$$

voor $t > a$, met a het minimum bedrag aan te maken kosten.

Voor de weibull-verdeling vallen de eigenschappen van de uitval-functie $r(t)$ eenvoudig af te leiden omdat de cumulatieve verdelingsfunctie valt te schrijven als

$$F(t) = 1 - \exp\left\{-\left(\frac{t-a}{c}\right)^b\right\}$$

zodat

$$r(t) = \frac{b}{c} \left(\frac{t-a}{c}\right)^{b-1}$$

Hieruit volgt

$$r(t) \text{ toenemend} \quad \Leftrightarrow \quad b > 1$$

$$r(t) = 1/c \quad \Leftrightarrow \quad b = 1$$

$$r(t) \text{ afnemend} \quad \Leftrightarrow \quad b < 1$$

Voor de lognormaal-verdeling en de gamma-verdeling kunnen we terugvallen op Glaser (1980). Deze heeft beschreven hoe het verloop van de uitval-kans eenvoudig nagetrokken kan worden voor kansdichtheden van de vorm

$$f(t, \theta) = C(\theta) \cdot \exp\left\{\sum_{i=1}^k U_i(t, \theta)\right\}$$

waarin θ de vector met parameters van de dichtheid aangeeft, en U_i twee keer partieel differentieerbaar is naar t op het interval (a, ∞) . Zij

$$g(t) = - \sum_{i=1}^k U_i'(t, \theta)$$

$$g'(t) = - \sum_{i=1}^k U_i''(t, \theta)$$

$$\lim_{t \rightarrow a} f(t) = e$$

Hij onderscheidt twee situaties

1. Er bestaat een $t_0 > a$ met

$$g'(t) < 0 \quad t \in (a, t_0)$$

$$g'(t_0) = 0$$

$$g'(t) > 0 \quad t \in (t_0, \infty)$$

Als $e = \infty$ is r afnemend tot een bedrag $t_1 < t_0$ en toenemend voor $t > t_1$.

Als $e = 0$ is r toenemend.

2. Er bestaat een $t_0 > a$ met

$$g'(t) > 0 \quad t \in (a, t_0)$$

$$g'(t_0) = 0$$

$$g'(t) < 0 \quad t \in (t_0, \infty)$$

Als $e = 0$ dan is r toenemend tot een bedrag $t_1 < t_0$ en afnemend voor bedragen $> t_1$.

Voor $e = \infty$ is de uitvalkans afnemend.

De bevindingen van Glaser leiden voor de lognormaal- en gamma-verdeling tot de volgende conclusies m.b.t. de uitvalkans

- lognormaal

eerst toenemend, daarna afnemend, met

$$t_0 = a + \exp(b-c^2+1)$$

- gamma

toenemend als $b > 1$

constant (= $1/c$) als $b = 1$ (negatief-exponentiële verdeling)

afnemend als $b < 1$

Voor de gamma-verdeling valt t.a.v. de uitvalkans voorts nog op te merken:

$$r(t) = \frac{\frac{1}{c} \left(\frac{t-a}{c}\right)^{b-1} \exp\left(-\frac{t-a}{c}\right)}{\Gamma(b) - \int_0^{(t-a)/c} u^{b-1} \exp(-u) du}$$

waaruit m.b.v. de stelling van l'Hôpital volgt:

$$\lim_{t \rightarrow \infty} r(t) = 1/c$$

M.a.w. voor $b > 1$ is de uitvalkans weliswaar toenemend, maar met bovengrens $1/c$; voor $b < 1$ is de uitvalkans afnemend, met ondergrens $1/c$.

Op grond van bovenstaande bevindingen m.b.t. de uitvalkans kunnen we concluderen dat de lognormale verdeling niet in aanmerking komt voor het beschrijven van de verdeling van de kosten per opname omdat de uitvalkans vanaf een zeker bedrag afnemend is. Voor de gamma-verdeling is voor $b > 1$ weliswaar sprake van een toenemende uitvalkans, maar blijkens bovenstaande opmerking is voor grote bedragen sprake van een bovengrens $1/c$, zodat dan sprake is van 'lack of memory', juist zoals bij de negatief exponentiële verdeling; m.a.w. de hoogte van de reeds gemaakte kosten geven geen extra informatie meer over de uitvalkans. De vanuit theoretisch oogpunt meest in aanmerking komende kansverdeling ter be-

schrijving van de kosten per opname is derhalve de weibull-verdeling met $b > 1$.

5. 'Maximum Product of Spacings'-schattingsprocedure

Zoals in de literatuur algemeen bekend is, leveren Maximum Likelihood schattingsprocedures voor 3-parameter-verdelingen, waarin slechts sprake is van positieve kansdichtheden rechts van een verschoven onbekende oorsprong a , niet altijd bevredigende resultaten op. Voorbeelden daarvan zijn de hierboven beschreven 3-parameter gamma- en weibull-verdeling:

voor $b < 1$ leidt de ML-procedure niet tot een oplossing. De loglikelihood-vergelijkingen zijn namelijk niet oplosbaar; voor de gamma-verdeling luidt één van de loglikelihood-vergelijkingen immers

$$\frac{n}{c(b-1)} = \sum_{i=1}^n (x_i - a)^{-1}$$

zodat geen oplossing \tilde{a} mogelijk is waarvoor $x_i > \tilde{a}$ voor alle i . Derhalve wordt voor $b < 1$ veelal $\tilde{a} = \min\{x_i\}$ als 'super-efficiënte' schatter gehanteerd; waardoor de schatters voor b en c evenwel inconsistent worden.

Voor $b > 1$ kan de ML-procedure nog steeds problemen veroorzaken. Weliswaar levert de ML-procedure consistente schatters op, voor steekproeven van gegeven omvang kunnen zich situaties voordoen waarbij (zelfs) geen sprake is van een (locaal) maximum. De kans dat dergelijke situaties zich voordoen is groter naarmate b meer in de buurt ligt van 1. Voor Johnson & Kotz (1970) is dit aanleiding om de ML-procedure voor de 3-parameter gamma-verdeling slechts aan te bevelen als $b > 2.5$.

Bovendien geldt voor $1 < b < 2$ dat de ML-schatter van de 3-parameter weibull- en gamma-verdeling niet asymptotisch normaal verdeeld is omdat niet aan de regulariteitsvoorwaarden voldaan is.

Voor de lognormale verdeling leidt de ML-procedure weliswaar steeds tot asymptotische normaal verdeelde schatters, ook hier kan de ML-procedure voor sommige steekproefresultaten problemen opleveren doordat de loglikelihood-schattingen resulteren in

$$\tilde{a} = \min_i \{x_i\}, \quad \tilde{b} = -\infty \text{ en } \tilde{c} = \infty$$

Cheng en Amin (1983) hebben voor rechts scheve verdelingen met onbekende verschoven oorsprong een alternatieve schattingsmethode voorgesteld, die consistente schatters oplevert met dezelfde asymptotische efficiëncy als de ML-schatter, voor zover deze bestaat, maar bovendien consistente, asymptotisch efficiënte schatters oplevert als de ML-procedure niet toegepast kan worden.

Daartoe gaan zij uit van een kansdichtheid f die slechts positieve waarden aanneemt op (p, q) ,⁴⁾ en een verdelingsfunctie F ; laat $x_1 \leq x_2 \leq \dots \leq x_n$ de geordende steekproefresultaten zijn. Zij $x_0 := p$, $x_{n+1} := q$. Vervolgens definiëren zij:

$$y_i = F(x_i, \theta) \quad i = 0, 1, \dots, n, n+1$$

en

$$D_i = y_i - y_{i-1} = \int_{x_{i-1}}^{x_i} f(x, \theta) dx \quad i = 1, 2, \dots, n, n+1$$

Vanzelfsprekend geldt $\sum_{i=1}^{n+1} D_i = 1$.

De MPS-methode komt neer op het maximaliseren van

$$G = \left\{ \prod_{i=1}^{n+1} D_i \right\}^{1/(n+1)}$$

of, equivalent daarmee, het maximaliseren van

$$H = \log G = \frac{1}{n+1} \sum_{i=1}^{n+1} \log D_i \quad 5)$$

Ter vergelijking, de ML-functie kan als volgt geschreven worden

$$L = \frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta)$$

Daar in par. 6 ook schattingsresultaten voor gegroepede waarnemingen worden besproken, geven we hier ter volledigheid ook de bijbehorende ML-functie. Als we m klassen onderscheiden met n_1, n_2, \dots, n_m waarnemingen en met klassegrenzen k_1, k_2, \dots, k_{m-1} , dan valt de ML-functie te schrijven als

$$L_g = n_1 \log F(k_1) + \sum_{j=2}^{m-1} n_j \log(F(k_j) - F(k_{j-1})) + n_m \log(1 - F(k_{m-1}))$$

De consistentie en asymptotische efficiëncy-eigenschappen van de MPS en ML-schatter worden door Cheng en Amin als volgt naast elkaar gezet:

MPS

Voor de 3-parameter weibull- en gamma-verdeling volgt de MPS-schatter $\hat{\theta} = (\hat{a}, \hat{b}, \hat{c})'$ uit

$$\frac{\partial H}{\partial \theta} = 0$$

i) als $b > 2$ dan is $\hat{\theta}$ consistent en asymptotisch normaal verdeeld:

$$\sqrt{n} (\hat{\theta} - \theta) \xrightarrow{L} N(0, - (E(\frac{\partial^2 \log f}{\partial \theta^2}))^{-1})$$

ii) als $0 < b < 2$ dan is er 'in probability' een oplossing $\hat{\theta}$ met

$$\hat{a} - a = O_p(n^{-1/b})$$

en met $\hat{\phi} := (\hat{b}, \hat{c})'$ asymptotisch normaal verdeeld:

$$\sqrt{n} (\hat{\phi} - \phi) \xrightarrow{L} N(0, - (E(\frac{\partial^2 \log f}{\partial \phi^2}))^{-1})$$

ML

Beschouw m.b.t. de 3-parameter weibull- en gamma-verdeling de ML-vergelijking

$$\frac{\partial L}{\partial \theta} = 0$$

- i) als $b > 2$ dan is $\tilde{\theta}$ consistent en asymptotisch normaal verdeeld met dezelfde limietverdeling als de MPS-schatter $\hat{\theta}$.
- ii) als $1 < b < 2$ dan is 'in probability' een oplossing $\tilde{\theta}$ met

$$\tilde{a} - a = O_p(n^{-1/b})$$

en met $\tilde{\phi} = (\tilde{b}, \tilde{c})'$ asymptotisch normaal verdeeld met dezelfde limietverdeling als $\hat{\phi}$.

- iii) als $b < 1$ bestaat er geen consistente oplossing voor $\frac{\partial L}{\partial \theta} = 0$.

Voor de lognormale verdeling geldt zowel voor de MPS als voor de ML-schatter dat ze asymptotisch normaal verdeeld zijn

$$\sqrt{n}(\hat{\theta} - \theta) \text{ en } \sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{L} N(0, - (E(\frac{\partial^2 \log f}{\partial \theta^2}))^{-1})$$

De hierboven aangegeven asymptotische variantie-covariantiematrices van de 3-parameter weibull-, lognormale- en gamma-verdeling staan expliciet uitgewerkt in Lemon (1975), Hill (1963) resp. Johnson en Kotz (1970). Voor de asymptotische variantie-covariantiematrices van de ML-schatters voor gegroepeerde waarnemingen zij verwezen naar Kendall en Stuart (1973, vgl. 30.22 en 30.30).

Samengevat:

- waar de ML-schatter geen bruikbare schatter is voor de weibull- of gamma-verdeling als $b < 1$, levert de MPS-schatter een superefficiënte schatter voor a , met een variantie van een orde kleiner dan $1/n$, terwijl de asymptotische verdeling van \hat{b} en \hat{c} dezelfde is als ware a bekend;
- van de weibull- en gamma-verdeling met $b > 1$ en voor de lognormale verdeling is de asymptotische efficiëncy van de MPS- en ML-schatter hetzelfde.

6. Schattingsresultaten

Voor de in tabel 1 besproken DRG's m.b.t. het specialisme KNO is

nagegaan of de frequentieverdelingen van de kosten beschreven kunnen worden m.b.v. de weibull-verdeling. We hebben daarbij de in par. 5 besproken MPS-schattingsprocedure gehanteerd.

Als aanpassingstoets is de χ^2 -toets toegepast. Het aantal te onderscheiden klassen k is zoveel mogelijk gebaseerd op de door Kendall & Stuart (1973) gegeven formule (30.72) die is afgeleid ter maximalisatie van het onderscheidingsvermogen:

$$k = b \cdot \left\{ \frac{(n-1)/2}{\lambda_{\alpha} + \phi^{-1}(P_0)} \right\}^{0.4}$$

met n het aantal waarnemingen, α het niveau van de toets, λ_{α} het $(1-\alpha)$ kwantiel van de standaardnormale verdeling, ϕ^{-1} de inverse verdelingsfunctie van de standaardnormale verdeling, P_0 de waarde van de powerfunctie die gemaximaliseerd wordt en b een constante, idealiter te kiezen tussen 2 en 4. Kendall en Stuart hebben geadviseerd k kleiner te kiezen naarmate men meer geïnteresseerd is in het gebied waar het onderscheidingsvermogen hoger is. Wij hebben $\alpha = 0,05$, $P_0 = 0,75$ en $b = 2,5$ genomen.

Daar de schatting van de parameters van de weibull-verdeling niet op gegroepede waarnemingen is gebaseerd maar op individuele waarnemingen, is de limietverdeling van X^2 echter niet langer een χ^2 -verdeling met $k-3-1$ vrijheidsgraden. "However, the distribution of X^2 is bounded between a χ_{k-1}^2 and a χ_{k-3-1}^2 variable, and as k becomes large these are so close together that the difference can be ignored... But for k small, the effect of using the χ_{k-3-1}^2 distribution for test purposes may lead to serious error; for the probability of exceeding any given value will be greater than we suppose ... but it is as well to be sure, when ordinary (ongegroeped P.N.) ML-estimation is being used, that the critical values of χ_{k-3-1}^2 and χ_{k-1}^2 are both exceeded by X^2 ", aldus Kendall en Stuart (1973).

De $k-1$ klassegrenzen zijn zodanig vastgesteld dat het geschatte verwachte aantal waarnemingen voor alle klassen gelijk is ($= n/k$). De toetsingsgrootte valt daardoor als volgt te herschrijven

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_{0i}(\hat{\theta}))^2}{np_{0i}}$$

$$= \frac{k}{n} \sum_{i=1}^k n_i^2 - n$$

met n_i het aantal waarnemingen in klasse i en $p_{0i}(\hat{\theta})$ de geschatte kans dat een waarneming in klasse i valt onder de nulhypothese dat de kosten per opname van de betreffende DRG een weibull-verdeling hebben.

In tabel 2 zijn de schattingsresultaten voor de in tabel 1 aangegeven patiëntengroepen weergegeven. Ter volledigheid zijn daarin ook weergegeven de MPS-schattingsresultaten voor de 3-paramater lognormale verdeling en de 2 parameterlognormale verdeling met $a = 0^6$). Schattingsresultaten m.b.t. de 3-parameter gamma-verdeling blijven achterwege i.v.m. numerieke problemen; de onvolledige gamma-functie kan immers slechts numeriek bepaald worden, hetgeen leidt tot (te) veel computertijd vergende schattingsprocedures.

Van de weibull-verdeling wordt steeds een waarde voor de parameter b gevonden die groter is dan 1 zodat inderdaad voldaan lijkt te worden aan de hypothese van een toenemende hazard-functie. De aanpassing van de 3-parameter weibull-verdeling is echter slecht. Op het 5% niveau leiden bij $k-4$ vrijheidsgraden 16 van de 17 toetsen tot verwerping van de nulhypothese dat de kosten per opname een weibull-verdeling volgen; bij $k-1$ vrijheidsgraden worden nog altijd 13 van de 17 nulhypothesen verworpen.

Ook de aanpassing van de lognormale verdeling is slecht; zowel voor de 3-parameter versie als voor de 2-parameter versie worden volgens de χ^2 -toets de meeste nulhypothesen verworpen.

Voor enkele kansverdelingen, waaronder de normale verdeling, is de kansverdeling bekend van de Kolmogorov-Smirnov toetsingsgroottheid; zie ook Kendall & Stuart (1973) en Pearson & Hartley (1972):

$$D_n = \sup_t \{F_n(t) - F(t)\}$$

Tabel 2. Schattingsresultaten op basis van individuele waarnemingen.¹⁾

n	k-4	weibull (MPS)				lognormaal						D _n	D ²⁾	
		\hat{a}	\hat{b}	\hat{c}	χ^2	MPS	\hat{a}	\hat{b}	\hat{c}	χ^2	\tilde{b}			ML (met a = 0)
108	8	0	4,84	3638	22,67	0	8,06	0,39	86,22	8,07	0,37	82,22	0,21	2,25
62	3	1433	1,70	2399	43,23	529	7,95	0,36	22,68	8,13	0,28	18,84	0,14	1,13
24	1	2750	1,08	517	5,58	2672	6,07	0,79	1,00	8,07	0,12	6,42	0,20	1,01
128	10	1108	1,72	1016	65,38	716	7,09	0,41	55,09	7,57	0,25	39,78	0,11	1,22
89	7	0	3,51	2528	19,39	0	7,69	0,35	27,80	7,69	0,33	22,11	0,13	1,27
120	9	802	2,28	1403	196,33	0	7,60	0,23	92,98	7,61	0,22	102,73	0,17	1,86
211	13	923	2,44	936	81,38	340	7,23	0,24	82,03	7,45	0,18	76,39	0,10	1,39
195	13	74	4,62	1719	247,78	0	7,40	0,24	226,69	7,40	0,23	204,89	0,25	3,54
145	11	421	1,32	838	93,24	249	6,70	0,47	51,45	6,99	0,35	43,59	0,10	1,26
39	1	360	1,23	882	24,21	211	6,67	0,61	21,90	6,93	0,44	3,95	0,16	1,02
59	3	188	1,66	743	98,44	0	6,68	0,33	60,00	6,68	0,31	60,00	0,25	1,98
33	1	0	1,64	3442	1,39	0	7,81	0,76	1,39	7,82	0,69	0,79	0,11	0,65
58	3	218	1,04	2626	14,41	0	7,60	0,79	5,48	7,60	0,73	3,79	0,10	0,74
24	1	956	1,07	857	5,58	853	6,52	0,81	3,08	7,40	0,35	6,83	0,21	1,08
49	3	129	1,99	1713	22,00	0	7,29	0,54	34,00	7,29	0,50	25,14	0,17	1,21
34	1	246	0,99	1120	6,59	144	6,72	0,90	1,29	6,92	0,69	4,53	0,13	0,78
67	4	260	0,88	1748	16,94	173	7,04	1,04	9,30	7,24	0,84	7,87	0,11	0,93

1) Varianties op de schattingen zijn bij de auteur verkrijgbaar. Omdat de algemene aanpassing echter zo slecht is, zijn ze voor verdere conclusies niet van belang.

2) Voor $\alpha = 0.01, 0.025, 0.05, 0.10$ is de grens van het kritieke gebied, resp. 1.035, 0.955, 0.895, 0.819.

Hierin is F_n de empirische verdelingsfunctie en F de verdelingsfunctie van de normale verdeling met geschatte MVU-parameters b en c^2 . M.b.v. deze toetsingsgrootheid is een normaliteitstoets uitvoerbaar en derhalve ook een lognormaliteitstoets, met $a = 0$. Omdat deze toets van Lilliefors een hoger onderscheidingsvermogen heeft dan de χ^2 -toets hebben we in tabel 2 ook de hierbij behorende toetsingsgrootheden D_n en $D_n^{(7)}$ opgenomen. Werden op een toetsingsniveau van $\alpha = 0.05$ bij een χ^2 -toets met $k-3$ vrijheidsgraden 12 van de 17 nulhypotheses verworpen en bij een χ^2 -toets met $k-1$ vrijheidsgraden 10 van de 17, de toets van Lilliefors verwerpt zelfs 14 van de 17 nulhypotheses.

Teneinde na te kunnen gaan in hoeverre deze tegenvallende schattingsresultaten veroorzaakt worden door een aantal extreme waarnemingen hebben we ook schattingsresultaten bepaald aan de hand van gegroepeerde waarnemingen. Deze schattingsprocedures zijn immers robust m.b.t. eventuele uitschieters. We kunnen nu evenwel niet meer terugvallen op de MPS-schattingsprocedure, zodat de resultaten in tabel 3 gebaseerd zijn op de ML-procedure voor gegroepeerde waarnemingen; zie ook par. 5. Het aantal onderscheiden klassen is hetzelfde als hiervoor; het aantal vrijheidsgraden bedraagt derhalve $k-4$. De klassegrenzen kunnen nu vanzelfsprekend niet zodanig vastgesteld worden dat het geschatte verwachte aantal waarnemingen voor alle klassen gelijk is. Ze worden nu vastgesteld op basis van gegevens m.b.t. de mediaan en de range van de waarnemingen, zodanig dat de klassebreedte en het verwachte aantal waarnemingen per klasse enigszins aansluiten bij de vorm van de kansverdeling.

In tabel 3 staan de schattingsresultaten; niet alleen voor de weibull-verdeling, maar ook voor de lognormale en de gamma-verdeling. Het aantal evaluaties van de onvolledige gamma-functie blijft nu immers beperkt tot een veelvoud van het aantal onderscheiden klassen i.p.v. een veelvoud van het aantal waarnemingen, zodat de rekentijd van de computer binnen de perken blijft.

De conclusies zijn niet veel anders dan m.b.t. de ongegroepeerde waarnemingen:

Tabel 3. Schattingsresultaten op basis van gegroepeerde waarnemingen; ML-procedure

n	k-4	Weibull				Lognormaal				Gamma			
		a	b	c	χ^2	a	b	c	χ^2	a	b	c	χ^2
108	8	48	5,91	3655	43,39	0	8,16	0,22	51,34	318	17,46	185	50,06
62	3	1569	4,19	1938	5,11	529	7,94	0,18	3,78	1569	12,60	144	3,72
24	1	2003	4,38	1212	2,37	2341	6,65	0,39	2,62	2523	4,08	158	2,64
128	10	1155	1,05	1065	16,32	716	7,09	0,71	16,73	743	2,26	608	16,43
89	7	0	5,70	2381	24,95	0	7,71	0,24	28,91	33	18,28	123	28,30
120	9	0	5,10	2265	50,02	0	7,65	0,24	57,86	0	18,38	116	56,40
211	13	891	3,41	904	85,46	340	7,22	0,21	85,60	927	7,40	110	85,80
195	13	0	11,91	1748	455,09	0	7,44	0,13	608,04	(275	40,62	35)
145	11	0	4,95	1185	108,13	0	7,00	0,23	115,33	0	18,55	60	113,84
39	1	396	3,35	667	2,43	396	6,38	0,34	1,77	396	8,89	69	1,89
59	3	0	7,93	838	81,61	0	6,67	0,15	101,69	(118	31,76	21)
33	1	0	2,01	3218	11,32	0	7,89	0,61	12,07	356	2,39	1123	11,81
58	3	268	1,70	2024	3,28	268	7,38	0,66	2,88	268	2,52	740	3,07
24	1	998	1,79	582	0,54	853	6,42	0,50	0,51	998	1,84	319	0,95
49	3	162	2,36	1747	7,30	0	7,41	0,46	7,84	0	5,20	340	7,60
34	1	227	1,41	784	1,40	144	6,60	0,67	1,25	277	1,45	536	1,63
67	4	0	0,90	1831	0,67	0	7,10	1,29	0,68	140	0,70	2649	0,68

() resultaten laatste iteratieslag van wegens numerieke problemen afgebroken schattingsprocedures.

- de hazard-functie van zowel de weibull-verdeling als de gamma-verdeling lijkt toenemend te zijn ($b > 1$), op één uitzondering na;
- de aanpassing van de kansverdelingen is slecht. Voor de weibull-verdeling worden op het 5% toetsingsniveau 8 van de 17 nulhypotheses verworpen. Voor de lognormale en gamma-verdeling liggen deze aantallen op 9 van de 17 resp. 8 van de 17. De acceptatiegraad van de nulhypothese ligt weliswaar hoger i.v.m. de robustheid van een op gegroepeerde waarnemingen gebaseerde schattingsprocedure, de algemene conclusie blijft dat deze levensduurmodellen de kosten per opname slecht beschrijven.

Samenvattend kunnen we concluderen dat de meest toegepaste levensduurmodellen een slechte beschrijving geven van de kosten per opname voor patiënten uit dezelfde DRG. Robuste schattingsmethoden doen aan deze conclusie niets af.

Een van de mogelijke oorzaken van deze tegenvallende resultaten zou kunnen liggen in het feit dat de schattingen gebaseerd zijn op gepoolde gegevens van drie ziekenhuizen. Als echter verschillen bestaan in de wijze waarop patiënten uit dezelfde DRG in elk van deze ziekenhuizen behandeld worden, zullen ook de kansverdelingen van de kosten per opname voor deze DRG in elk ziekenhuis anders zijn. Derhalve is nagegaan in hoeverre per ziekenhuis een betere aanpassing van de kosten per opname aan een kansverdeling wordt gevonden. Deze nadere analyse doet de conclusies echter niet veranderen. In tabel 4 zijn de resultaten gegeven voor de Lilliefors aanpassingstoets m.b.t. de lognormale verdeling met $a = 0$. Werden m.b.t. de gepoolde gegevens op het 5% niveau 14 van de 17 nulhypotheses verworpen, voor de drie ziekenhuizen afzonderlijk liggen deze cijfers op 4 van de 9, 4 van de 5 en 13 van de 16, zodat we ook nu niet mogen concluderen dat de kosten per opname lognormaal verdeeld zijn.

Voor de 3-parameter weibull-, gamma- en lognormale verdelingen hebben we analoge resultaten gevonden, zodat ook invoering van de drempelwaarde a de aanpassing niet verbetert.

Tabel 4. Lilliefors D-toetsingsgrootheid bij lognormale verdeling (met $\alpha = 0$).

	pool		zhs 1		zhs 2		zhs 3	
tympanoplastiek	2.25	(108)	1.28	(19)	1.39	(13)	2.10	(76)
mastoïdectomie	1.13	(62)	-		-		1.10	(50)
stapedectomie	1.01	(24)	-		-		0.99	(15)
neusplastiek	1.22	(128)	0.77	(79)	-		1.79	(48)
neusbijholten ≥ 15 jaar	1.27	(89)	0.86	(28)	-	(7)	1.24	(54)
oorschelp	1.86	(120)	-		-		1.86	(120)
overige neusoperaties	1.39	(211)	1.01	(52)	1.43	(28)	1.77	(131)
tonsillectomie ≥ 15 jaar	3.54	(195)	2.16	(78)	1.11	(18)	3.01	(99)
keeloperaties	1.26	(145)	0.99	(52)	-		1.87	(91)
trommelvliesbuisjes ≥ 15 jaar	1.02	(39)	-		-		1.06	(26)
trommelvliesbuisjes 0-14 jaar	1.98	(59)	-		1.69	(37)	1.72	(20)
otitis media	0.65	(33)	0.68	(12)	-		0.68	(20)
laryngitis, faryngitis	0.74	(58)	0.54	(26)	-		0.48	(28)
sinusitis	1.08	(24)	-		0.49	(18)	-	
neustrauma	1.21	(49)	-		-		1.14	(42)
neusbloeding	0.78	(34)	-		-		0.84	(21)
overige KNO-aandoeningen	0.39	(67)	0.69	(15)	-		0.93	(46)

() aantal patiënten in deze DRG.

- minder dan 10 patiënten in deze DRG.

120 PATIENTEN, WAARVAN 119 IN DEZE TABEL; KLASSEBREEDTE = 100

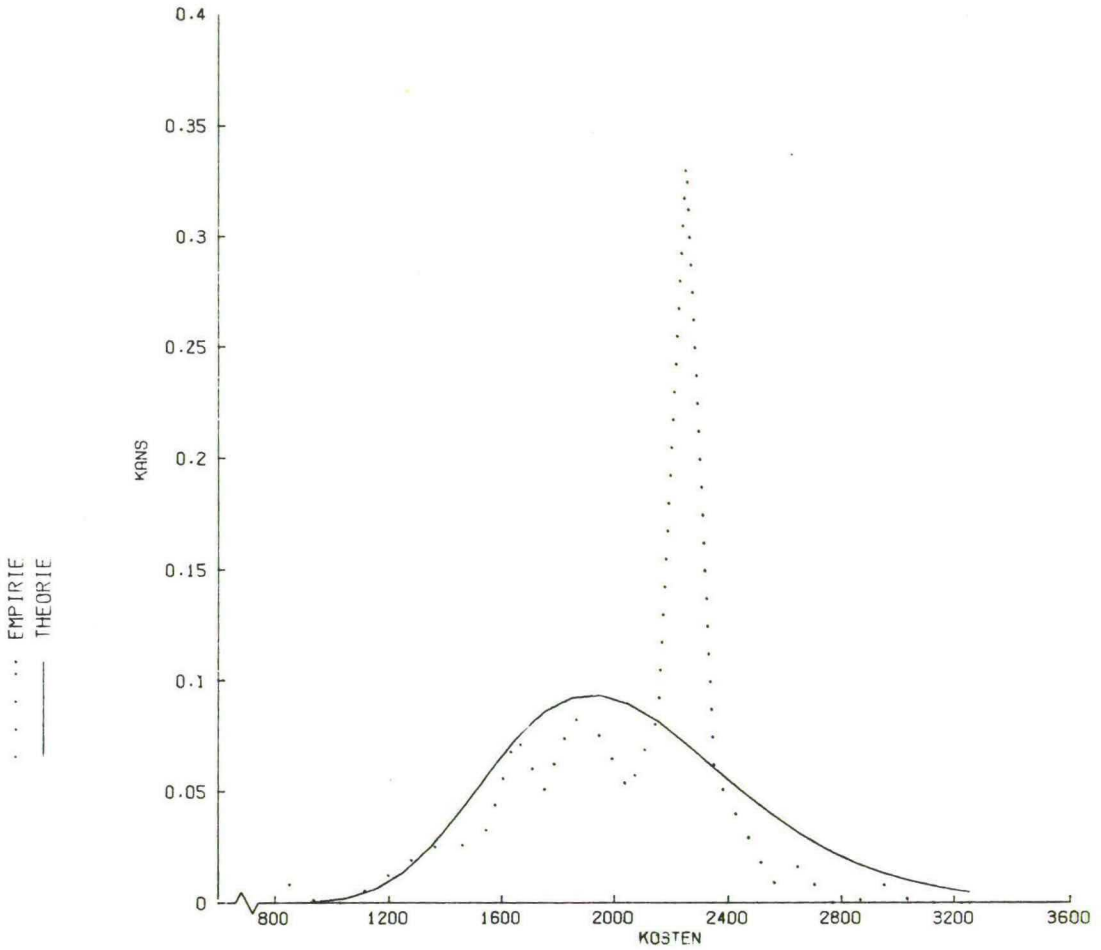


fig. 1 Frequentieverdeling van de kosten per opname voor patiënten met "oorschelpoperaties".

211 PATIENTEN, WAARVAN

210 IN DEZE TABEL; KLASSEBREEDTE =

100

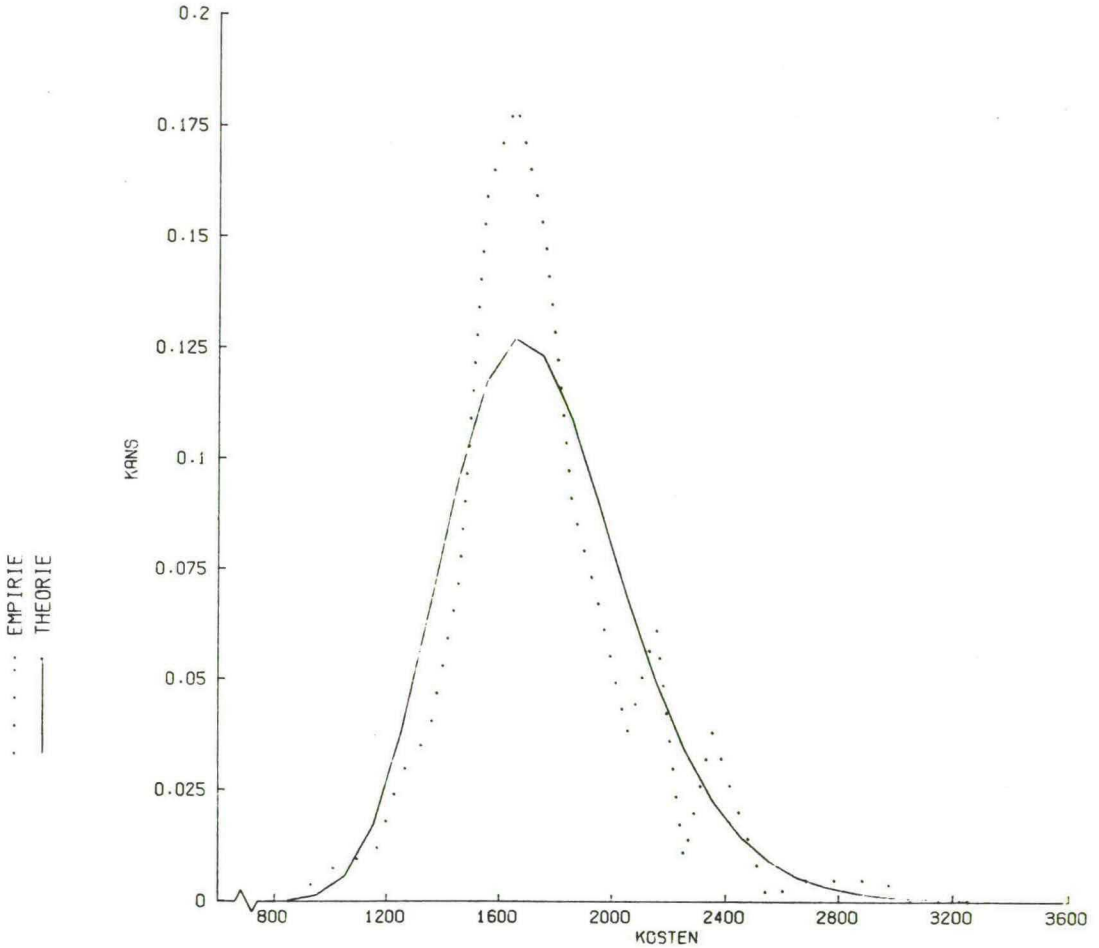


fig. 2 Frequentieverdeling van de kosten per opname voor patiënten met "overige neusoperaties".

Deze resultaten overziende moet men zich wel realiseren dat we de nulhypothese van lognormale, weibull of gamma-verdelingen verwerpen op statistische gronden. Het is daarbij een bekend verschijnsel dat naarmate men over meer waarnemingen beschikt men de nulhypothese vaker moet verwerpen. In hoeverre dit ook een voor de praktijk 'significant verschil' oplevert, is nog maar de vraag. In fig. 1 en 2 zijn de theoretische en empirische frequentieverdelingen van de DRG's 'oorschelpoperaties' en 'overige neusoperaties' getekend, behorend bij de in tabel 2 opgenomen ML-schattingsresultaten voor de lognormale verdeling met $a = 0$. Daaruit blijkt dat men tot verwerping van de nulhypothese overgaat in situaties dat de aanpassing op het oog redelijk voldoet.

7. Verschillen-analyse

Daar de schattingsresultaten niet wijzen op een weibull-, lognormale of gamma-verdeling voor de kosten per opname van een DRG, zal de vraagstelling uit de inleiding hoe nagegaan kan worden of twee frequentieverdelingen dan wel de gemiddelden van twee kansverdelingen statistisch significant verschillen, niet expliciet gebruik maken van op deze kansverdelingen gebaseerde toetsen.

Verschillen tussen twee empirische frequentieverdelingen zijn echter ook te toetsen m.b.v. de twee steekproeven-toets van Kolmogorov. In het licht van ziekenhuisvergelijking laat deze toets zich als volgt beschrijven:

beschouw de patiëntenpopulaties die in ziekenhuis A resp. B worden behandeld en in een bepaalde DRG zijn ingedeeld.⁸⁾ Laat de opnamekosten voor deze DRG de onbekende verdelingsfunctie F resp. G hebben. Beschouw de n patiënten uit deze DRG die in een bepaalde periode in ziekenhuis A zijn opgenomen en opnamekosten x_1, x_2, \dots, x_n hebben en de m patiënten uit deze DRG die in ziekenhuis B zijn opgenomen en opnamekosten y_1, y_2, \dots, y_m hebben. Dan laten de empirische verdelingsfuncties van de kosten per opname zich als volgt definiëren:

$$F_n(t) = (\text{aantal der } x_i \leq t)/n$$

en

$$G_m(t) = (\text{aantal der } y_i < t)/m$$

Voor een toets met als nulhypothese $H_0 : F = G$, dat de verdelingsfuncties van de kosten voor beide ziekenhuizen identiek zijn tegen het alternatief $H_1 : F \neq G$, luidt de toetsingsgrootheid

$$D_{nm} = \max_t \{ F_n(t) - G_m(t) \}$$

Kim & Jennrich (1970) hebben een tabel voor de verdelingsfunctie van D_{nm} onder de nulhypothese berekend. Zie ook Pearson & Hartley (1972).

Als de vraag, of er sprake is van gelijkheid van gemiddelde kosten per opname, wordt opgevat als het toetsingsprobleem

$$H_0: P(\underline{x} > \underline{y}) = \frac{1}{2}$$

$$H_1: P(\underline{x} > \underline{y}) \neq \frac{1}{2}$$

waarin \underline{x} de kosten per opname voor een willekeurige patiënt uit ziekenhuis A en \underline{y} de kosten per opname voor een willekeurige patiënt uit ziekenhuis B, beide m.b.t. dezelfde DRG, dan kunnen we gebruik maken van de twee-steekproeventoets van Wilcoxon; ook bekend als Mann-Whitney-toets of rangsomtoets. Hierbij worden de waarnemingen x_1, x_2, \dots, x_n en y_1, y_2, \dots, y_m geordend. Laat $(z_1, \dots, z_n, z_{n+1}, \dots, z_{n+m})$ de rangnummers zijn van $(x_1, \dots, x_n, y_1, \dots, y_m)$. De toetsingsgrootheid luidt in dit geval:

$$U = \sum_{i=1}^n r_i$$

de rangsom van de kosten van opname van patiënten uit ziekenhuis A in de gecombineerde steekproef. Onder de nulhypothese is deze toetsingsgrootheid symmetrisch verdeeld rond $\frac{1}{2}n(n+m+1)$. Pearson en Hartley (1972) hebben tabellen opgenomen met een aantal kwantielen van deze toetsingsgrootheid.

Het exacte antwoord op de vraag naar de gelijkheid van de gemiddelde kosten per opname is verwant aan het Behrens-Fisher-probleem. Voor twee steekproeven uit de normale verdeling met parameters (μ_1, σ_1^2) resp. (μ_2, σ_2^2) kan ingeval $\sigma_1^2 = \sigma_2^2$ de toets $H_0: \mu_1 = \mu_2$ tegen $H_1: \mu_1 \neq \mu_2$ uitgevoerd worden m.b.v. de t-toets; de toetsingsgrootheid luidt

$$t = \frac{\bar{x} - \bar{y}}{\left\{ S^2 \left(\frac{1}{n} + \frac{1}{m} \right) \right\}^{\frac{1}{2}}}$$

met

$$S^2 = \left(\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^m (y_j - \bar{y})^2 \right) / (n+m-2)$$

De toetsingsgrootheid t heeft onder de nulhypothese een t-verdeling met $n+m-2$ vrijheidsgraden.

Als $\sigma_1^2 \neq \sigma_2^2$ is sprake van het Behrens-Fisher-probleem. De t-toets mag nu niet langer toegepast worden. Een van de mogelijke toetsen voor deze situatie is de Aspin-Welch-toets, met als toetsingsgrootheid

$$Z = \frac{\bar{x} - \bar{y}}{\left\{ \frac{s_x^2}{n-1} + \frac{s_y^2}{m-1} \right\}^{\frac{1}{2}}}$$

met

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

en

$$s_y^2 = \frac{1}{m} \sum_{j=1}^m (y_j - \bar{y})^2$$

Aspin & Welch en Tricket e.a. hebben enkele kwantielen van deze toetsingsgrootheid bepaald onder H_0 , als functie van n , m en σ_x^2/σ_y^2 ; zie ook Pearson & Hartley (1972).

Of er sprake is van $\sigma_1^2 = \sigma_2^2$ kan m.b.v. een F-toets getoetst worden. In geval van gelijkheid van varianties kan men vervolgens de t-toets hanteren ter vergelijking van de gemiddelde kosten per opname van

de DRG, in het andere geval moet men terugvallen op de Aspin-Welch-toets. Waarschijnlijk zal deze gelijkheid van varianties zich eerder voordoen bij vergelijking in een ziekenhuis van de gemiddelde kosten per DRG over twee opeenvolgende jaren, dan in geval van ziekenhuisvergelijking.

Overigens ligt aan beide toetsen de normaliteitsveronderstelling ten grondslag. Weliswaar geeft Scheffé (1970) aan dat zowel de toets als de Aspin-Welch-toets in zekere mate robuust zijn t.a.v. afwijkingen van normaliteit, het is nog maar de vraag in hoeverre dat ook nog opgaat voor erg scheve verdelingen; zie ook tabel 1. Wellicht dat transformatie van de opnamekosten hier uitkomst biedt. M.b.v. een log-transformatie zou aldus beter aan de normaliteitsveronderstelling voldaan kunnen worden. Toch wordt daarmee ons toetsingsprobleem niet volledig opgelost. Toetsen op $b_1 = b_2$ en $c_1^2 = c_2^2$ is immers niet equivalent met toetsen op de gelijkheid van de gemiddelde kosten $m_1 = m_2$, met $m_i = \exp\{b_i + \frac{1}{2} c_i^2\}$, hetgeen equivalent is met $b_1 + \frac{1}{2} c_1^2 = b_2 + \frac{1}{2} c_2^2$. "... and this equivalence shows that the means m_1 and m_2 may be equal even through the parent populations differ in respect both of b and c^2 . Unfortunately, there is no test of the null hypothesis for the means m , since there is as yet no theory of joint confidence intervals for b and c^2 for normal populations: so that the statistician must confine himself to separate statements in regard to these two parameters", aldus Aitchinson en Brown (1957). Daar de mediaan van de log-normale verdeling $\exp\{b\}$ bedraagt, is echter wel een toets voor de gelijkheid van de medianen van de verdelingen van de kosten per opname te bepalen.

Voor grote n en m kan overigens een beroep gedaan worden op de centrale limietstelling, immers

$$\sqrt{n}(\bar{x}_n - \mu_x) \xrightarrow{L} N(0, \sigma_x^2)$$

met

$$E(\underline{x}_t) = \mu_x, \quad V(\underline{x}_t) = \sigma_x^2$$

zodat de normaliteitsveronderstelling die aan bovengenoemde toetsen ten grondslag ligt overbodig is.

Daarnaast geldt voor grote waarden van n en m , als vuistregel hanteert men veelal $n, m > 30$, dat de kwantielen van de toetsingsgrootheden van de t -toets en Aspin-Welch-toets goed benaderd kunnen worden door de kwantielen van de standaard-normale verdeling; een conclusie die ook rechtstreeks uit het eerder vermelde resultaat van de centrale limietstelling getrokken kan worden, als men zich realiseert dat

$$\frac{1}{n-1} \sum (x_i - \bar{x})^2 \xrightarrow{P} \sigma_x^2$$

8. Prospectief

Bij het opstellen van een budget is het verleidelijk uit te gaan van de realisatie van het afgelopen jaar. Het budget wordt in deze situatie gefixeerd op het produkt van het verwachte aantal patiënten in een DRG en de gemiddelde kosten per opname van een patiënt uit die DRG.

Men dient zich evenwel te realiseren dat dit gemiddelde onder invloed van het toeval tot stand is gekomen. Hoe daar op in te spelen?

8.1. Amerikaanse oplossingen

Fetter (1982) realiseert zich dat de gemiddelde kosten per patiënt erg beïnvloed worden door een paar patiënten met extreem hoge kosten. Hij stelt derhalve voor deze patiënten buiten beschouwing te laten bij het bepalen van de gemiddelde kosten en tevens buiten het budget van de specialist te houden.

Andrews e.a. (1972) hebben een uitgebreide studie verricht m.b.t. 'robust estimates of location'. Zij beschreven de belangrijkste families van robuuste schatters: trimmed means, lineaire combinaties van order-statistics, M -estimates (waaronder Huber-estimates), single and iterative skipping procedures en folded means. Voor 68 robuuste schatters zijn m.b.v. Monte Carlo technieken onder meer bias en mean square error bepaald. Daarbij bleek dat met name enkele M -estimates en, in mindere mate, combinaties van skipping procedures en trimming procedures aantrekkelijke robuuste schatters zijn. Echter, de studie van Andrews e.a. heeft zich volledig geconcentreerd op robuuste schatters voor locatie in situaties met symmetrische verdelingen.

Desondanks heeft Fetter (1982) zich op Andrews' resultaten gebaseerd en heeft hij gebruik gemaakt van de volgende skipping procedure: laat k_p het $p\%$ percentiel zijn van de empirische verdeling

$$c_1 = k_{25} - a(k_{75} - k_{25})$$

$$c_2 = k_{75} + a(k_{75} - k_{25})$$

b is het aantal waarnemingen buiten het interval (c_1, c_2) . Vervolgens worden de b hoogste en b laagste (als $b = 0$ de hoogste en laagste waarneming) waarnemingen die binnen het interval (c_1, c_2) vallen buiten beschouwing gelaten bij het berekenen van de gemiddelde kosten per opname \bar{x} en de standaardafwijking s . De waarnemingen binnen het interval $(\bar{x} - d \cdot s, \bar{x} + d \cdot s)$ zijn de waarnemingen waarop Fetter zich uiteindelijk baseert bij het bepalen van de gemiddelde kosten per DRG. Hij beveelt daarbij de parameterwaarden $a = 1,5$ en $d = 3$ aan.

Medicare, de Amerikaanse ziekenfondsverzekering voor bejaarden, heeft beter ingespeeld op de asymmetrische verdeling van de kosten per opname binnen een DRG; zie Schweiker (1982). Er van uitgaande dat de kosten lognormaal verdeeld zijn, worden de ML-schatters \tilde{b} en \tilde{c}^2 bepaald:

$$\tilde{b} = \frac{1}{n} \sum_{i=1}^n \ln(x_i)$$

$$\tilde{c}^2 = \frac{1}{n} \sum_{i=1}^n (\ln(x_i) - \tilde{b})^2$$

Medicare gebruikt deze resultaten bij het vaststellen van één tarief voor de behandeling van alle patiënten uit een bepaalde DRG. Daarvan wil men echter de patiënten met de hoogste en de laagste kosten uitzonderen. Daartoe gaat men uit van het rekenkundig gemiddelde van de kosten per opname van de patiënten wier kosten binnen het 95% betrouwbaarheidsinterval van de gemiddelde kosten⁹⁾ liggen:

$$(e^{\tilde{b} - \tilde{c} \cdot u_{0.025}}, e^{\tilde{b} + \tilde{c} \cdot u_{0.025}})$$

met $u_{0.025}$ het 97.5% percentiel van de standaardnormale verdeling.

Ook deze oplossing is echter niet correct

- als überhaupt al sprake is van een log-normale verdeling, dan komt de 3-parameterverdeling meer in aanmerking;
- volgens de resultaten van par. 6 zijn de kosten per opname niet log-normaal verdeeld;
- men schat de parameters \tilde{b} en \tilde{c}^2 op basis van gepoolde gegevens. Als er evenwel sprake is van verschillen in behandelwijze tussen de verschillende ziekenhuizen zal er sprake zijn van kansverdelingen met verschillende parameters, zie ook par. 6.

8.2. Alternatieven

Het vaststellen van een budget, of dat nu een totaalbudget of een budget voor alle patiënten uit een bepaalde DRG is, komt neer op het bepalen van de verwachte kosten. Veelal zal men niet alleen geïnteresseerd zijn in dit budget, maar ook in een voorspellingsinterval voor dit budget, teneinde aldus ook rekening te houden met het stochastisch karakter van de kosten van een opname.

Als bekend zou zijn welke kansverdeling de kosten per opname in een bepaalde DRG hebben, zouden m.b.v. convolutietheorie dit budget en voorspellingsinterval bepaald kunnen worden.

Daar we in par. 6 aan de hand van de schattingsresultaten geconcludeerd hebben dat de kosten per opname voor een DRG niet goed te beschrijven zijn met verdelingen die veelvuldig in de theorie van levensduurmodellen worden toegepast, kunnen we de bepaling van budget en voorspellingsinterval niet op deze wijze benaderen.

Als het aantal patiënten per DRG echter voldoende groot is, kunnen we terugvallen op de centrale limietstelling. Immers als x_1, \dots, x_n de kosten voorstellen voor n patiënten uit een bepaalde DRG en als we verwachten m patiënten uit die DRG op te nemen met kosten y_1, \dots, y_m , geldt volgens de centrale limietstelling:

$$\sqrt{n} (\bar{x} - \mu) \xrightarrow{L} N(0, \sigma^2)$$

$$\sqrt{m} (\bar{y} - \mu) \xrightarrow{L} N(0, \sigma^2)$$

zodat, vanwege de onafhankelijkheid der x_i en y_j ,

$$\sum_{i=1}^m y_i - \frac{m}{n} \sum_{i=1}^n x_i \rightarrow N\left(0, \left(m + \frac{m^2}{n}\right) \sigma^2\right)$$

Met

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \rightarrow \sigma^2$$

volgt derhalve

$$\frac{\sum_{i=1}^m y_i - m\bar{x}}{\left\{ \left(m + \frac{m^2}{n}\right) S_x^2 \right\}^{\frac{1}{2}}} \rightarrow N(0, 1)$$

Het rechtseenzijdig voorspellingsinterval wordt derhalve

$$\left(0, m\bar{x} + u_\alpha * S_x * \sqrt{m \left(\frac{n+m}{n}\right)}\right)$$

Ter vergelijking, het rechtseenzijdig betrouwbaarheidsinterval voor $m\mu$ luidt

$$\left(0, m\bar{x} + u_\alpha * S_x * m\sqrt{n}\right)$$

Dat we bij het voorspellingsinterval te maken krijgen met $\sqrt{(m(n+m))/n}$ i.p.v. $m\sqrt{n}$ wordt veroorzaakt door de onvolledige kennis omtrent de werkelijke kansverdeling van de kosten per opname voor patiënten uit een DRG. We beschikken immers slechts over de waarnemingen x_1, x_2, \dots, x_n en willen aan de hand daarvan een uitspraak doen over \bar{y} . Het voorspellingsinterval dat hierboven is aangegeven houdt derhalve zowel rekening met het feit dat \bar{x} van μ kan verschillen als dat \bar{y} van μ kan verschillen.

Als we de mogelijkheid dat \bar{y} van μ kan verschillen nog niet in-calculeren bij de budgetopstelling, maar daar pas op inspelen bij de verschillenanalyse na afloop van de budgetperiode, kunnen we het budget baseren op een schatting van μ en hoeven we aan de hand van de waarne-

mingen x_1, x_2, \dots, x_n slechts een betrouwbaarheidsinterval voor μ te specificeren. Een mogelijkheid daartoe is het hierboven gegeven interval voor $m\mu$.

Kleijnen e.a. (1984) hebben m.b.v. Monte-Carlo technieken aangetoond dat bij kleinere steekproeven uit asymmetrische verdelingen de t-toets voor $\mu = \mu_0$ een percentage fouten van de le soort oplevert dat hoger is dan de gespecificeerde betrouwbaarheid α . M.a.w., als men m.b.v.

$$(0, \bar{x} + t_{n-1; \alpha} S/\sqrt{n})$$

een rechtsezijdig betrouwbaarheidsinterval voor μ bepaalt is dit te klein. Johnson (1978) stelt een modificatie van de t-toets voor, die expliciet corrigeert voor enkele ongewenste invloeden op de toetsingsgrootte t die uitgaan van een asymmetrische verdeling. De voorgestelde toetsingsgrootte luidt:

$$t_1 = [(\bar{x} - \mu) + \frac{\mu_3}{6\sigma^2 n} + \frac{\mu_3}{3\sigma^4} (\bar{x} - \mu)^2] / (S/\sqrt{n}) \quad 10, 11)$$

De coverage van de gemodificeerde t-toets is volgens de studie van Kleijnen e.a. hoger dan voor de t-toets. M.a.w. als sprake is van asymmetrische verdelingen doet men er beter aan om betrouwbaarheidsintervallen voor μ te bepalen aan de hand van de toetsingsgrootte van de gemodificeerde t-toets dan aan de hand van de t-toets.

Omdat de gemodificeerde t-toets met name geconstrueerd is t.b.v. asymmetrische verdelingen en de minimum steekproefomvang door Johnson slechts op 13 is gesteld, hebben wij deze methode ter bepaling van betrouwbaarheidsintervallen voor μ verkozen boven de op de centrale limietstelling gebaseerde intervallen

$$(0, \bar{x} + u_{\alpha} * S/\sqrt{n})$$

of de op normaliteit gebaseerde intervallen

$$(0, \bar{x} + t_{n-1; \alpha} * S/\sqrt{n})$$

Uit de toetsingsgrootheid voor de gemodificeerde t-toets valt onderstaande rechtseenzijdig betrouwbaarheidsinterval voor μ af te leiden:¹²⁾

$$(0, \bar{x} + (1 - [1 - 4 \frac{\text{excess}}{3} (\frac{\text{excess}}{6n} + \frac{t_{n-1; \alpha}}{\sqrt{n}})]^{\frac{1}{2}} / 2 \frac{\text{excess}}{3S})$$

Waar de gemodificeerde t-toets expliciet rekening houdt met de asymmetrische verdeling van de kosten per opname voor een bepaalde DRG, is het ook mogelijk om m.b.v. een transformatie van de opnamekosten te proberen beter aan de normaliteitsveronderstelling te voldoen. Een veel toegepaste techniek is de log-transformatie. Omdat dit echter betekent dat wordt uitgegaan van een log-normale verdeling is het ook mogelijk zich rechtstreeks op deze verdeling te baseren bij het bepalen van budget en budgetinterval. De MVU-schatter voor de gemiddelde kosten per opname m in geval van een log-normale verdeling met $a = 0$ luidt (Aitchinson en Brown, 1957):

$$\hat{m} = \exp\{\tilde{b}\} \psi_n\left(\frac{1}{2} \tilde{c}^2\right)$$

met \tilde{b} en \tilde{c}^2 de eerder gedefinieerde ML-schatters voor b en c^2 en

$$\psi_n(t) = 1 + \frac{n-1}{n} t + \sum_{j=2}^{\infty} \frac{(n-1)^{2j}}{n^j (n-1)(n+1)\dots(n-3+2j)} \frac{t^j}{j!}$$

M.b.t. betrouwbaarheidsintervallen voor de gemiddelde kosten per opname merken Aitchinson en Brown vervolgens op: "Theory provides no means of obtaining exact confidence intervals for m ; all that can be said is that \hat{m} may be treated as asymptotically normal with mean m and variance $V\{\hat{m}\}$ ", waarin

$$V\{\hat{m}\} = m^2 (c^2 + \frac{c^4}{2})/n$$

Een rechtseenzijdig betrouwbaarheidsinterval luidt derhalve

$$(0, \hat{m} + u_{\alpha} * \hat{V}\{\hat{m}\}^{\frac{1}{2}})$$

Tabel 5. Rechtseenzijdige 95%-betrouwbaarheidsintervallen voor de gemiddelde kosten per opname per DRG.

	n	$m_1^{1)}$	$m_2^{1)}$	t-toets	Johnson	MVU ¹⁾
<u>geopereerd</u>						
tympanoplastiek	108	3349	3413	3483	3472	3623
mastoïdectomie	62	3541	3524	3807	4071	3732
stapedectomie	24	3223	3222	3370	3411	3354
neusplastiek	128	2006	2005	2084	2090	2077
neusbijholten, ≥ 15 jaar	89	2285	2304	2400	2403	2442
oorschelp	120	2058	2056	2135	2171	2124
overige neusoperaties	211	1755	1754	1794	1796	1791
tonsillectomie, ≥ 15 jaar	195	1672	1680	1709	1709	1727
keeloperaties	145	1176	1153	1282	- ²⁾	1210
trommelvliesbuisje, ≥ 15 jaar	39	1147	1126	1339	-	1263
trommelvliesbuisje, 0-14 jaar	59	845	835	939	-	892
<u>niet-geopereerd</u>						
otitis media	33	3025	3139	3544	3584	3840
laryngitis, faryngitis, e.d.	58	2703	2600	3384	-	3065
sinusitis	24	1743	1732	1998	2070	1944
neustrauma	49	1637	1662	1813	1836	1869
neusbloeding	34	1310	1283	1622	1708	1566
overige KNO-aandoeningen	67	2046	1967	2490	2613	2354

1) m_1 : gemiddelde kosten per DRG (momentenschatter)

m_2 : MVU-schatter voor de gemiddelde kosten per opname, uitgaande van de log-normale verdeling

MVU: het betrouwbaarheidsinterval, gebaseerd op de MVU-schatter, uitgaande van een log-normale verdeling.

2) De grootte onder het wortelteken van het in par. 6 gegeven betrouwbaarheidsinterval is negatief, zodat de bovengrens van het betrouwbaarheidsinterval onbepaald is.

In tabel 5 staan voor elke DRG achtereenvolgens aangegeven het aantal patiënten, de gemiddelde kosten per opname (momentenschatter), de MVU-schatter voor de gemiddelde kosten per opname (uitgaande van een log-normale verdeling met $a = 0$) en de grenzen van de rechtsezijdige betrouwbaarheidsintervallen voor de gemiddelde kosten gebaseerd op de t-toets, gemodificeerde t-toets van Johnson (die voor $\mu_3 > 0$ inderdaad tot een grotere bovengrens leidt dan de t-toets) en de MVU-schatter (uitgaande van een lognormale verdeling met $a = 0$).

Uitgaande van dergelijke betrouwbaarheidsintervallen is de fase van de budgetvaststelling af te ronden. Bij de verschillenanalyse na afloop van de budgetperiode zal men alsnog rekening dienen te houden met het feit dat niet alleen de gegevens waarop het budget is gebaseerd, maar ook de realisaties gedurende de budgetperiode stochastisch van aard zijn.

9. Verrichtingen- en kostenprofielen per DRG

Bij de presentatie van verrichtingen- en kostenprofielen van een DRG worden gemiddelde en met name standaardafwijking erg beïnvloed door de aanwezigheid van één of meer patiënten met extreem hoge kosten; ongeacht of deze hoge waarden als uitbijters beschouwd dienen te worden of bij de veronderstelde rechts scheve verdeling verwacht kunnen worden. Teneinde de invloed van deze extreme waarnemingen op de profielen te mitigeren, is het zinvol deze waarnemingen niet mee te nemen bij de profielbepaling. De daardoor te verkrijgen profielen zullen meer overeenstemmen met het profiel van de 'modale' patiënt en daardoor meer geschikt zijn om met de medici door te spreken. Omdat aan de aldus verkregen profielen geen budgetconsequenties vastzitten, die problematiek is immers hierboven reeds besproken, kunnen we ter oplossing van dit probleem een minder geavanceerde benadering kiezen.

Bij het analyseren van de schattingsresultaten hebben we kunnen constateren dat de aanpassing van de weibull, gamma en lognormale verdeling vergelijkbaar was. Weliswaar is op grond van eigenschappen van de hazard-functie de weibull-verdeling met $b > 1$ vanuit theoretisch oogpunt te prefereren, vanuit praktisch oogpunt is de lognormale verdeling te verkiezen omdat de parameters van de lognormale verdeling het eenvoudigst te schatten zijn. Het voorstel m.b.t. de buiten beschouwing te

laten waarnemingen luidt nu: bepaal de parameters b en c van de lognormale verdeling waarin $a = 0$ en laat patiënten wier kosten groter zijn dan $\exp\{u_{.05} * c + b\}$ buiten beschouwing, met $u_{.05}$ het 95% kwantiel van de standaardnormale verdeling.¹³⁾ De motieven achter dit voorstel zijn:

- de aanpassing van de lognormale verdeling met $a = 0$ is weliswaar niet overtuigend maar toch redelijk te noemen; zeker als we de aanpassing voor elk ziekenhuis afzonderlijk uitvoeren; zie tabel 4. Fig. 1 en 2 versterken deze conclusie omdat de aanpassing op het oog redelijk voldoet, al moet de nulhypothese statistisch gezien verworpen worden;
- de buiten beschouwing gelaten patiënten tellen wel mee bij vaststellen van budgetten; de nu te verkrijgen profielen dienen slechts om een algemeen beeld van het middelenbeslag van de 'modale' patiënt te geven t.b.v. de medici;
- de invloed van de gekozen drempelwaarde a op het 95% kwantiel is niet groot, voor zover deze drempelwaarde kleiner is dan de werkelijke drempelwaarde a . Johnson & Kotz (1970) geven dit m.b.t. een gestandaardiseerde verdeling als volgt aan: "There can be considerable variation in 'a' with little effect on the percentiles and little effect on the cumulative distribution function for fixed values of x_1 . Insensitivity is most marked for large negative values of 'a'".

Toepassen van deze methode leidt tot de in tabel 6 vermelde gemiddelde kosten per DRG, standaardafwijking, aantal patiënten excl. extreme waarnemingen en afkapgrens. Vergelijking met de in tabel 1 weergegeven cijfers toont dat gemiddelde en met name de standaardafwijking van de kosten nu lager uitvallen; zoals verwacht. Ook kunnen we constateren dat het percentage extreme waarnemingen meer rond 5% ligt naarmate de aanpassing middels de 2-parameter lognormale verdeling beter is; zie ook tabel 2. Dat in een aantal DRG's het percentage outliers lager uitvalt dan 5% is niet bezwaarlijk: zoals uit de standaardafwijkingen uit tabel 1 blijkt is in die DRG's sprake van een kleine standaardafwijking en derhalve grote homogeniteit, zodat er weinig aanleiding is om veel patiënten buiten beschouwing te laten.¹⁴⁾

Tabel 6. Extreme waarnemingen.

	m	s	n	afkap- grens	% extreme waarne- mingen
tympanoplastiek	3349	833	108	5883	0
mastoidectomie	3311	601	59	5338	4,84
stapedectomie	3129	275	22	3888	8,33
neusplastiek	1953	430	124	2913	3,13
neusbijholten, \geq 15 jaar	2203	543	85	3770	4,49
oorschelp	2015	336	118	2865	1,67
overige neusoperaties	1695	251	197	2332	6,64
tonsillectomie \geq 15 jaar	1652	265	192	2399	1,54
keeloperaties	1087	279	141	1928	2,76
trommelvliesbuisjes $>$ 15 jaar	1002	320	37	2099	5,13
trommelvliesbuisjes 0-14 jaar	791	149	58	1325	1,69
otitis media	2866	1502	32	7703	3,03
laryngitis, faryngitis	2350	1536	57	6621	1,72
sinusitis	1451	285	20	2887	16,67
neustrauma	1532	530	47	3323	4,08
neusbloeding	1046	648	31	3151	8,82
overige KNO-aandoeningen	1539	1175	62	5497	7,46

10. Samenvatting

De kosten per opname voor patiënten uit dezelfde DRG zijn stochastisch van aard. Dit betekent dat conclusies m.b.t. verschillen in gemiddelde kosten per DRG, budgetbepaling en verschillenanalyses van budget en realisatie daarop in moeten spelen. Elk probleem vereist een eigen benadering.

- De waarnemingen per DRG kunnen één of meer uitbijters bevatten. Dit probleem kan evenwel niet m.b.v. uitbijtertoetsen benaderd worden, omdat we over te weinig kennis beschikken m.b.t. de kansverdeling van de kosten per opname en evenmin vooraf kunnen aangeven hoeveel uitbijters verwacht worden; zie par. 3.
 - In het kader van de budgetbepaling is het dienstig te weten welke kansverdeling de kosten per opname hebben. Schattingsresultaten hebben uitgewezen dat de meest voor de hand liggende rechts-scheve verdelingen daarvoor niet in aanmerking komen; noch als we uitgaan van een 2-parameter verdeling ($a = 0$), noch als we uitgaan van een 3-parameter verdeling ($a > 0$). Zie par. 4, 5 en 6.
 - Nagaan of de kosten per opname voor een bepaalde DRG in twee ziekenhuizen hetzelfde verdeeld zijn is terug te voeren tot de twee-steekproeuntoets van Kolmogorov. De toets is ook bruikbaar om na te gaan of de verdeling van de kosten per opname in twee opeenvolgende jaren hetzelfde is gebleven. Zie par. 7.
 - Als men slechts wil nagaan of de gemiddelde kosten per opname voor een bepaalde DRG in twee opeenvolgende jaren hetzelfde zijn gebleven is de t-toets bruikbaar, voor zover de veronderstelling van gelijke varianties juist is. Omdat deze veronderstelling voor twee ziekenhuizen minder waarschijnlijk is, is de t-toets niet langer bruikbaar voor het toetsen op de gelijkheid van de gemiddelde kosten per DRG, en moet men terugvallen op de Aspin-Welch-toets. Overigens gaan beide toetsen uit van de normaliteitsveronderstelling.
- Als we de vraagstelling een weinig aanpassen, is dit probleem ook te benaderen m.b.v. de twee-steekproeuntoets van Wilcoxon (Mann-Whitney-toets, rangsomtoets). Zie par. 7.
- In het kader van de budgetvaststelling kunnen we geen gebruik maken van convoluties van kansverdelingen, zodat we voor de vaststelling van

voorspellingsintervallen moeten terugvallen op de centrale limietstelling. Betrouwbaarheidsintervallen voor de gemiddelde kosten per opname zijn af te leiden m.b.v. de t-toets. Een modificatie van deze t-toets die expliciet rekening houdt met de asymmetrie van de kansverdeling is daar evenwel beter voor geschikt. Zie par. 8.2.

- Bij de presentatie van verrichtingen- en kostenprofielen per DRG aan de media worden de cijfers erg beïnvloed door enkele extreme waarnemingen. Omdat dit probleem evenwel van ondergeschikt belang is, hebben we een eenvoudig te berekenen afkapgrens voorgesteld, zodanig dat patiënten wier kosten boven deze grens uitkomen buiten beschouwing gelaten worden bij het vaststellen van de DRG-profielen. De aldus te verkrijgen profielen zijn derhalve niet 'vervuild' door enkele extreme waarnemingen, zodat ze meer overeen zullen komen met de profielen van de 'modale' patiënt en daardoor beter herkenbaar zijn voor de specialist; zie par. 9.

Voetnoten

- 1) Hoofddiagnose, nevendiaognoses, operatieve ingrepen, leeftijd, geslacht, wijze van ontslag, al dan niet spoedopname e.d.
2. In verband met afspraken die met directie en medische staf van de drie betrokken ziekenhuizen zijn gemaakt m.b.t. de beschikbaar gestelde gegevens, is het niet mogelijk om een gedetailleerd overzicht van de kosten per patiënt te geven. Nadere informatie is bij de auteur verkrijgbaar.
- 3) Dit minimum pakket kan per ziekenhuis of per specialist verschillen.
- 4) p en q kunnen $-\infty$ resp. $+\infty$ zijn; p en q mogen onbekend zijn en derhalve deel uitmaken van de te schatten parametervector θ .
- 5) Als $x_i = x_{i-1}$ geldt $D_i = 0$. Dit probleem wordt ondervangen door D_i te interpreteren als

$$\lim_{x_i \rightarrow x_{i-1}} \int_{x_{i-1}}^{x_i} f(x, \theta) dx = \lim_{x_i \rightarrow x_{i-1}} f(x_i, \theta) * (x_i - x_{i-1})$$

Daar $x_i - x_{i-1}$ niet van θ afhangt wordt $\log D_i$ derhalve vervangen door $\log f(x_i, \theta)$.

- 6) Het aantal vrijheidsgraden bedraagt nu $k-2-1$ i.p.v. $k-3-1$.
- 7) Teneinde het toetsen te vereenvoudigen, is het mogelijk de volgende transformatie toe te passen:

$$D = D_n * (\sqrt{n} - 0.01 + 0.85/\sqrt{n})$$

Pearson en Hartley (1972) geven tabellen voor D.

- 8) We gaan er aldus bij deze toets impliciet vanuit dat de behandelwijze in een ziekenhuis in de loop der tijd niet verandert.
- 9) Het gemiddelde van een lognormaal verdeelde grootte bedraagt $\exp\{b + \frac{1}{2} c^2\}$; Medicare bepaalt derhalve in werkelijkheid een 95% betrouwbaarheidsinterval voor de mediaan $\exp\{b\}$.
- 10) Johnson suggereert om bij praktische toepassingen μ_3 en σ^2 te vervangen door de steekproefmomenten.
- 11) Omdat in t_1 de kwadratische term $(\bar{x} - \mu)^2$ voorkomt zal bij de bepaling van het acceptatiegebied $\{\bar{x} | t_1 \leq t_{n-1, \alpha}\}$ een rechtseenzijdig én een linkseenzijdig interval gevonden worden. Op statistische gronden kan men het linkseenzijdige interval negeren. De toetsingsuitspraak mag derhalve slechts op het rechtseenzijdige interval gebaseerd worden.
- 12) Het is evident dat:

$$(\bar{x} + (1 + [1 - 4 \frac{\text{excess}}{3} (\frac{\text{excess}}{6n} + \frac{t_{n-1; \alpha}}{\sqrt{n}})]^{\frac{1}{2}}) / \frac{2 \text{ excess}}{3\hat{\sigma}}, \infty)$$

geen zinnig betrouwbaarheidsinterval voor μ is.

- 13) Deze aanpak verschilt van de Medicare aanpak zoals beschreven in par. 8.1. Medicare laat de extreme waarnemingen buiten beschouwing bij de tariefbepaling per DRG. Wij laten ze alleen buiten beschouwing bij de bepaling van de DRG profielen. De in par. 8.2 beschreven budgetbepaling houdt expliciet rekening met het voorkomen van extreme waarnemingen.
- 14) Ook hier hebben we te maken met de vertekende invloed van het 'poolen' van gegevens. Afkapgrenzen bepalen voor elk ziekenhuis afzonderlijk, levert percentages extreme waarnemingen op die dichter in de buurt van 5% liggen dan op basis van de gepoolde gegevens. Dit ligt ook in de lijn der verwachting omdat de aanpassing d.m.v. de 2-parameter lognormale verdeling per ziekenhuis beter uitvalt dan voor de gepoolde gegevens, zie tabel 4.

Literatuur

- Aitchinson, J. en J.A.C. Brown,
"The lognormal distribution",
Cambridge University Press, Cambridge, 1957.
- Andrews, D.F., P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers, J.W. Tukey,
"Robust estimates of location",
Princeton University Press, Princeton, New Jersey, 1972.
- Barnett, Vic. and Toby Lewis,
"Outliers in statistical data",
John Wiley & Sons, 1978.
- Cheng, R.C.H. and N.A.K. Amin,
"Estimating parameters in continuous univariate distributions with a shifted origin",
Journal of the Royal Statistical Society B, 45, 1983, p. 394-403.
- Darling, D.A.,
"On a test for homogeneity and extreme values",
Annals of Mathematical Statistics, 23, 1952, p. 450-456.
- Dixon, J.W.,
"Rejection of observations",
in: Contributions to order statistics,
A.E. Sarhan, B.G. Greenberg (editors),
John Wiley & Sons, 1962.
- Ferguson, T.S.,
"On the rejection of outliers",
in: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability,
1962, Vol. I, pp. 253-287.

Fetter, Robert B., Youngsoo Shin, Jean L. Freeman, Richard F. Averill,
John D. Thompson,

"Case mix definition by Diagnosis Related Groups",
Medical Care, supplement, 18, feb. 1980.

Fetter, Robert B., e.a.,

"The new ICD-9-CM Diagnosis Related Groups Classification
Scheme", Section IV,
Yale University, Health Systems Management Groups, may 1982.

Glaser, Ronald E.,

"Bathtub and related failure rate characterizations",
Journal of the American Statistical Association, 75, 1980, p.
667-672.

Hill, Bruce M.,

"The three-parameter lognormal distribution and baysian analysis
of a point source epidemic",
Journal of the American Statistical Association, 58, 1963, p.
72-84.

Johnson, M.L. & S. Kotz,

"Continuous univariate distribution, I",
John Wiley & Sons, 1970.

Johnson, Norman J.,

"Modified t-test and confidence intervals for asymmetrical popu-
lations",
Journal of the American Statistical Association, 73, 1978, p.
536-544.

Kendall, Maurice G and Alan Stuart,

"The advanced theory of statistics, II, Inference and relation-
ship".
Charles Griffin & Company Ltd, London, 1973.

- Kim, P.J. & R.I. Jennrich,
"Selected tables in mathematical statistics I",
H.L. Harter and D.B. Owen, editors,
Markham Publishing Co for the Institute of Mathematical Statistics, Chicago, 1970.
- Kleijnen, J.P.C., G.J. Cuypers, J.W.M. van Rooijen,
"Testing the mean of a asymmetric population: four procedures evaluated",
Research Memorandum 138, Katholieke Hogeschool Tilburg, 1984.
- Lawless, J.F.,
"Statistical models and methods for life time data",
John Wiley & Sons, 1982.
- Lemon, Glen H.,
"Maximum Likelihood Estimation for the three parameter weibull distribution based on censored samples",
Technometrics, 17, may 1975, pp. 247-254.
- Lilliefors, Hubert W.,
"On the Kolmogorov-Smirnov test for normality with mean and variance unknown",
Journal of the American Statistical Association, 62, 1967,
pp. 399-402.
- Nederstigt, P.F.P.M.,
"Diagnosis Related Groups als bron van informatie voor interne budgettering",
Tijdschrift voor Sociale Geneeskunde, 60, 1982, pp. 476-480.
- Nederstigt, P.F.P.M.,
"Diagnosis Related Groups en verrichtingenprofielen",
Tijdschrift voor Sociale Gezondheidszorg, 62, 1984a, pp. 502-504

Nederstigt, P.F.P.M.,

"DRG definitives m.b.t. keel-, neus- en oor-aandoeningen",
Utrecht, NZI, 1984b.

Pearson, E.S. and H.O. Hartley,

"Biometrika tables for statisticians",
Cambridge University Press, London, 1972.

Scheffé, H.,

"Practical solutions of the Behrens Fisher problem",
Journal of the American Statistical Association, 65, 1970, p.
1501-1508.

Schweiker, Richard S.,

"Report to congress hospital prospective payment for Medicare",
Department of Health and Human Services, dec. 1982.

Walsh, J.E.,

"Some non-parametric tests of whether the largest observations
of a set are too large or too small",
Annals of Mathematical Statistics, 21, 1950, p. 583-592,
"Correction
Annals of Mathematical Statistics, 24, 1953, p. 134-135.

Walsh, J.E.,

"Large sample non-parametric rejection of outlying obser-
vations",
Annals of the Institute of Statistical Mathematics, Tokyo, 1959,
pp. 223-232.

IN 1983 REEDS VERSCHENEN

- 126 H.H. Tigelaar
Identification of noisy linear systems with multiple arma inputs.
- 127 J.P.C. Kleijnen
Statistical Analysis of Steady-State Simulations: Survey of Recent Progress.
- 128 A.J. de Zeeuw
Two notes on Nash and Information.
- 129 H.L. Theuns en A.M.L. Passier-Grootjans
Toeristische ontwikkeling - voorwaarden en systematiek; een selectief literatuuroverzicht.
- 130 J. Plasmans en V. Somers
A Maximum Likelihood Estimation Method of a Three Market Disequilibrium Model.
- 131 R. van Montfort, R. Schippers, R. Heuts
Johnson S_U -transformations for parameter estimation in arma-models when data are non-gaussian.
- 132 J. Glombowski en M. Krüger
On the Rôle of Distribution in Different Theories of Cyclical Growth.
- 133 J.W.A. Vingerhoets en H.J.A. Coppens
Internationale Grondstoffenovereenkomsten.
Effecten, kosten en oligopolisten.
- 134 W.J. Oomens
The economic interpretation of the advertising effect of Lydia Pinkham.
- 135 J.P.C. Kleijnen
Regression analysis: assumptions, alternatives, applications.
- 136 J.P.C. Kleijnen
On the interpretation of variables.
- 137 G. van der Laan en A.J.J. Talman
Simplicial approximation of solutions to the nonlinear complementarity problem with lower and upper bounds.

IN 1984 REEDS VERSCHENEN

- 138 G.J. Cuypers, J.P.C. Kleijnen en J.W.M. van Rooyen
Testing the Mean of an Asymmetric Population:
Four Procedures Evaluated
- 139 T. Wansbeek en A. Kapteyn
Estimation in a linear model with serially correlated errors when
observations are missing
- 140 A. Kapteyn, S. van de Geer, H. van de Stadt, T. Wansbeek
Interdependent preferences: an econometric analysis
- 141 W.J.H. van Groenendaal
Discrete and continuous univariate modelling
- 142 J.P.C. Kleijnen, P. Cremers, F. van Belle
The power of weighted and ordinary least squares with estimated
unequal variances in experimental design
- 143 J.P.C. Kleijnen
Superefficient estimation of power functions in simulation
experiments
- 144 P.A. Bekker, D.S.G. Pollock
Identification of linear stochastic models with covariance
restrictions.
- 145 Max D. Merbis, Aart J. de Zeeuw
From structural form to state-space form
- 146 T.M. Doup and A.J.J. Talman
A new variable dimension simplicial algorithm to find equilibria on
the product space of unit simplices.
- 147 G. van der Laan, A.J.J. Talman and L. Van der Heyden
Variable dimension algorithms for unproper labellings.
- 148 G.J.C.Th. van Schijndel
Dynamic firm behaviour and financial leverage clienteles
- 149 M. Plattel, J. Peil
The ethico-political and theoretical reconstruction of contemporary
economic doctrines
- 150 F.J.A.M. Hoes, C.W. Vroom
Japanese Business Policy: The Cash Flow Triangle
an exercise in sociological demystification
- 151 T.M. Doup, G. van der Laan and A.J.J. Talman
The $(2^{n+1}-2)$ -ray algorithm: a new simplicial algorithm to compute
economic equilibria

IN 1984 REEDS VERSCHENEN (vervolg)

- 152 A.L. Hempenius, P.G.H. Mulder
Total Mortality Analysis of the Rotterdam Sample of the Kaunas-
Rotterdam Intervention Study (KRIS)
- 153 A. Kapteyn, P. Kooreman
A disaggregated analysis of the allocation of time within the
household.
- 154 T. Wansbeek, A. Kapteyn
Statistically and Computationally Efficient Estimation of the
Gravity Model.

Bibliotheek K. U. Brabant



17 000 01059788 9