

Ontwerpen en Analyseren van Computer Experimenten: “Maximizing the Simulation Output: A Competition” Opnieuw Bekeken

Pieter van Casteren*, Freek Huele**, en Ronald Mulder***

Samenvatting

In situaties waar fysisch experimenteren teveel tijd of geld kost of eenvoudigweg onmogelijk is, kunnen experimenteerders hun toevlucht nemen tot het uitvoeren van een zogenaamd computereperiment. De Bedrijfssectie van de VVS organiseerde in januari 1998 een lezingenmiddag over computereperimenten en een “competitie” waarin de output van een computermodel gemaximaliseerd diende te worden. Kleijnen en Pala (*Kwantitatieve Methoden* 59, pp. 5-14) bespreken in hun artikel getiteld “Maximizing the Simulation Output: A Competition” de door hen gevolgde strategie. Hoewel hun heuristiek zeker een aantal interessante aspecten bevat, zijn wij van mening dat hun artikel een nogal eenzijdig beeld geeft van het verloop van deze competitie, aangezien de bijdragen van de overige deelnemers worden genegeerd. In dit artikel zullen wij de aanpak van Kleijnen en Pala daarom in een breder kader plaatsen door onze oplossingen te bespreken en de methoden onderling te vergelijken.

1. Inleiding

In situaties waar fysisch experimenteren teveel tijd of geld kost of eenvoudigweg onmogelijk is, kunnen experimenteerders hun toevlucht nemen tot het uitvoeren van een zogenaamd computereperiment. Computer- experimenten worden bijvoorbeeld gebruikt in de ontwikkelingsfase van geïntegreerde circuits, zie e.g. Welch, Yu, Kang, en Sacks (1990). Bij een computereperiment worden, voor een aantal combinaties van inputparameters, de responsies berekend met behulp van een computermodel. Zo'n computermodel kan bijvoorbeeld een expliciet gedefinieerde functie zijn of een stelsel differentiaalvergelijkingen. Vanwege hun complexiteit zijn computermodellen echter veelal weinig inzichtelijk en vergt het doorrekenen ervan vaak nogal wat rekentijd. Daarom wordt getracht om, op basis van een beperkt aantal runs, het verband tussen de respons en de inputparameters door middel van een eenvoudig empirisch model te beschrijven. Met behulp van zo'n empirisch model kunnen vervolgens de inputparameters worden geoptimaliseerd, voorspellingen worden gedaan, etc.

* Afdeling Kwantitatieve Economie, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, tel: 020-5254232, e-mail: casteren@fee.uva.nl

** CQM B.V., Postbus 414, 5600 AK Eindhoven, tel: 040-2758641, e-mail: Huele@cqm.nl

*** Hoogovens Research & Development, Postbus 10 000, 1970 CA IJmuiden, tel: 0251-497607, e-mail: Ronald.Mulder@hoogovens.com

Hoewel er dus een flink aantal overeenkomsten bestaat tussen fysische experimenten en computerexperimenten, is er ook een wezenlijk verschil. Daar waar men bij fysische experimenten altijd wordt gebeten door ruis in de meetwaarde (denk maar aan de meetfout of omgevingsvariatics) zijn de responsies bij computerexperimenten deterministisch van aard. Dit verschil heeft gevolgen voor zowel de wijze waarop een computerexperiment wordt ontworpen, i.e. de keuze van de waarden van de inputparameters, als de wijze waarop een computerexperiment wordt geanalyseerd. Zo is bijvoorbeeld het uitvoeren van herhaalde metingen zinloos. Immers, iedere herhaling levert altijd dezelfde uitkomst op.

Het vakgebied dat dit alles omvat wordt vaak aangeduid met de term DACE: Design and Analysis of Computer Experiments. Een belangrijk artikel op dit terrein is dat van Sacks, Welch, Mitchell, en Wynn (1989), waarin met name de statistische aspecten van DACE worden belicht. Ondanks het deterministische karakter van DACE, blijkt namelijk dat de statistiek ook hier een belangrijke bijdrage kan leveren. We komen daar in paragraaf 2.2 op terug. Ook vanuit het bedrijfsleven valt een toenemende belangstelling voor DACE te bespeuren. Al met al aanleiding genoeg voor de Bedrijfssectie van de VVS om op 8 januari 1998 een lezingenmiddag te organiseren met als onderwerp "computerexperimenten" (VVS Bulletin 12/97, pp. 162-163). Aan het eind van de middag werden de oplossingen besproken van de deelnemers die mee hadden gedaan aan een casus waarin de output van een bepaald computermodel gemaximaliseerd diende te worden.

Kleijnen en Pala (1998), voortaan aan te duiden door K&P, omschrijven deze casus en bespreken de door hen gevolgde strategie, welke leidde tot de hoogste output. Hoewel de strategie van K&P beslist interessant is, waren de resultaten van deze competitie echter dusdanig, dat er niet van een duidelijke winnaar gesproken kon worden. Immers, een flink aantal deelnemers heeft zich vroegtijdig uit de competitie terug getrokken vanwege het onrealistische gehalte ervan, en bovendien waren K&P niet de enige met een eindscore vlakbij het echte maximum.

In dit artikel zullen wij de aanpak van Kleijnen en Pala daarom in een breder kader plaatsen door onze oplossingen te bespreken (paragraaf 2) en de methoden onderling te vergelijken (paragraaf 3).

2. Drie andere oplossingen

2.1. Oplossing van Van Casteren

Wegens het ontbreken van voorinformatie werd als eerste probleem gezien de globale positie van het optimum te achterhalen, om daarna waarnemingspunten rondom het optimum te kunnen kiezen, want interpoleren werkt beter dan extrapoleren. Vervolgens werd geprobeerd om stapsgewijs de positie van het optimum preciezer te bepalen en de te interpoleren ruimte kleiner te maken. De laatste runs werden gereserveerd om enkele mogelijke locaties van het optimum uit te proberen ("prijschieten").

Hieronder wordt de toegepaste methode in fasen omschreven en vervolgens wordt enig commentaar geleverd. In Tabel 1 in de bijlage staat een overzicht van de toegepaste runs (run 0 is de initiële run).

FASE 1 (run 1-6) en FASE 2 (run 7-12) In fase 1 werd elke factor afzonderlijk verhoogd met ca. 33% t.o.v. de initiële instelling. In fase 2 werd de beste van de zeven beschikbare runs als referentiepunt genomen en werd elke factor van daaruit nog eens veranderd. Was in fase 1 de verhoging gunstig geweest, dan volgde in fase 2 een verdere verhoging met ca. 67%, en anders volgde in fase 2 een verlaging t.o.v. de initiële waarde met ca. 33%. In eerste instantie werd gekozen voor verhoging van elke factor, omdat er geen bovengrenzen bekend waren, terwijl werd aangenomen dat alle factoren positief moesten zijn. Voor de globale plaatsbepaling van het optimum kregen tamelijk grote verhogingen de voorkeur, omdat de optimale waarden ver boven de initiële waarden konden liggen.

FASE 3 (run 13-18) Eerst werd op basis van run 0-12 een kwadratische functie (zonder interactietermen, dus met 13 parameters) aangepast. Maximalisatie hiervan leidde tot het optimum (635, 878, 490, 51, 526, -6), en na vervanging van de ontoelaatbare -6 voor factor F door 10 (namelijk een gewogen gemiddelde van de grenswaarde 0 en de initiële waarde 100 met gewichten 0.9 resp. 0.1), werd een geschat optimum $\hat{z}$ verkregen. Nu lag $\hat{z}$ in een bepaalde positie t.o.v. de beschikbare punten, samengevat door het zwaartepunt $\bar{z}$. Om later te kunnen intrapoleren werden runs 13-18 afgeleid door vanuit $\hat{z}$ verder van $\bar{z}$ te gaan en daar een aantal punten neer te zetten. Om precies te zijn werden vanuit $\hat{z}$ telkens twee factoren $\hat{z}_j$ gelijktijdig vervangen door $\hat{z}_j + 0.3(\hat{z}_j - \bar{z}_j)$.

FASE 4 (run 19-24) en FASE 5 (run 25-27) In zowel fase 4 als fase 5 werd eerst via OLS een kwadratische functie (zonder interactietermen) aangepast. Vervolgens werd op de 15 interactietermen in het volledige tweede orde model (met in totaal 28 parameters) stepwise selection toegepast met een F-to-enter van 4 en met als voorwaarde dat elke kwadratische term een negatieve coëfficiënt moest hebben. Met het geselecteerde model werd het optimum geschat ($\hat{z}$). De runs werden verkregen door telkens vanuit $\hat{z}$ twee factoren $\hat{z}_j$ gelijktijdig te vervangen door $\hat{z}_j \pm \max \{0.3|\hat{z}_j - \bar{z}_j|, |\hat{z}_j - \hat{z}_j^{(-1)}|\}$, waarbij $\hat{z}^{(-1)}$ het geschatte optimum uit de vorige fase voorstelt.

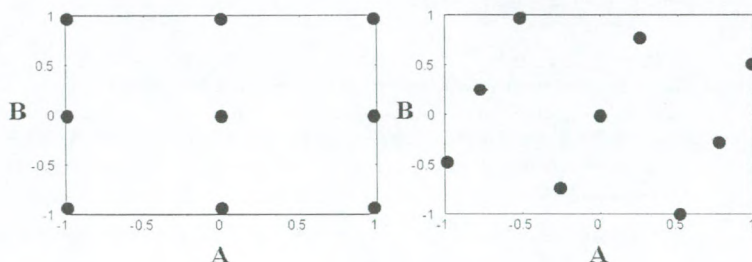
FASE 6 (run 28-32) Fase 6 bestond uit vijf deelfasen, waarin telkens één run werd gebruikt om daarmee een schatting van het optimum uit te proberen. In de eerste deelfase werd op basis van run 0-27 het optimum geschat via een tweede orde benadering met geselecteerde interactietermen (net als in fasen 4 en 5; het hierbij geselecteerde model bevatte overigens geen interacties). Het geschatte optimum vormde run 28 en dit bleek later de beste run te zijn. Run 29-31 werden op soortgelijke wijze gegenereerd, maar in plaats van OLS werd WLS toegepast, waarbij de waarnemingspunten lagere gewichten kregen naarmate zij verder verwijderd waren van het vorige geschatte optimum. Omdat met WLS geen verbetering t.o.v. run 28 werd bereikt, werd run 32 weer gegenereerd via OLS, waarbij nog een soort "lokale foutencorrectie" uitgevoerd werd om de benadering te verbeteren nabij het geschatte optimum.

De essentie van de methode is dus om na een zo klein mogelijk aantal runs de positie van het optimum te schatten, en de nieuwe runs te bepalen door vanuit het geschatte optimum verder van het zwaartepunt te gaan en daar nieuwe punten neer te zetten. Hiermee wordt beoogd om zo vroeg mogelijk te kunnen interpoleren. Vervolgens wordt geprobeerd om nieuwe runs steeds dichter rond het optimum te kiezen ter verkleining van de te interpoleren ruimte.

Voor het schatten van het optimum werd een tweede orde benadering gebruikt, eerst zonder interactietermen en later met interactietermen. Deze opbouw heeft als nadeel dat de interacties pas in tweede instantie aan bod komen. Om die reden kan worden overwogen om al eerder rekening te houden met mogelijke interacties, bijvoorbeeld door uit alle tweede orde effecten de belangrijkste te selecteren. Ook kan bij het kiezen van de runs in de eerste fasen reeds meer nadruk gelegd worden op het identificeren van mogelijke interacties.

2.2. Oplossing van Huele

Zoals in paragraaf 1 reeds is opgemerkt, is het uitvoeren van herhaalde metingen in een computereperiment niet nuttig. Herhalingen kunnen echter ook ontstaan als een bepaalde factor geen invloed heeft op de respons. Denk bijvoorbeeld maar aan een 3^2 experiment met twee factoren A en B ieder op drie niveaus (negen runs dus). Zie de proefopzet links in onderstaande Figuur.



Als factor B er nu niet toe doet, houden we drie herhaalde metingen over bij iedere instelling van factor A, waarmee bij een fysisch experiment de processpreiding geschat kan worden. Dergelijke klassieke schema's (i.e. 2^n , 3^n) en ook het Rechtschaffner ontwerp dat K&P gebruikten (pag. 8 en 14) zijn in eerste instantie minder geschikt voor DACE. In het geval van DACE houden we namelijk als factor B inactief is maar drie runs over voor factor A: de proefopzet is in elkaar geklapt ("collapsed").

Een alternatieve proefopzet zien wij rechts in de Figuur: wanneer factor B inactief is, blijven er toch nog negen runs voor A over. Naast het feit dat DACE designs niet collapsen hebben zij in het algemeen de eigenschap dat in zekere zin de ruimte van inputparameters optimaal overdekt wordt: DACE designs zijn "space-filling" (cf. Morris en Mitchell, 1995).

Als het experiment eenmaal is uitgevoerd, dient het benaderende model bepaald te worden. Gezien de afwezigheid van ruis is het wenselijk dat dit model de data exact fit. Hiervoor wordt in het algemeen gebruik gemaakt van kriging, een bekend model in de geostatistiek (zie e.g. Sacks, Mitchell en Wynn, 1989).

Toch werd de DACE methodiek niet losgelaten op het vraagstuk van de competitie. Daar zijn twee redenen voor. Ten eerste was er totaal geen achtergrondinformatie beschikbaar, over het proces bijvoorbeeld. In de praktijk is de kennis van bijvoorbeeld een chemicus of technoloog onmisbaar als men in een industriële omgeving succesvol wil kunnen experimenteren. Ten tweede is voor DACE bepaalde software nodig die niet ter beschikking stond. Binnen CQM is recentelijk DOT ontwikkeld: het Design Optimization Tool, waarmee DACE binnen handbereik van iedere ontwikkelaar is gekomen.

2.3. Oplossing van Mulder

Over het toelaatbare werkgebied van de ingangsvariabelen was niets bekend maar uit de omschrijving van het probleem kon –achteraf ten onrechte gehinderd door kennis van de chemie– worden afgeleid dat interacties een overheersende rol zouden spelen. De snelheid van chemische reacties hangt immers meestal af van meerdere componenten tegelijk en is volgens Arrhenius niet-lineair van de temperatuur afhankelijk. (Voor bijvoorbeeld de reactie $2A+B \rightarrow A_2B$ is de reactiesnelheid $\sim e^{-E/RT}[A]^2[B]$). Er zou ook sprake kunnen zijn van een katalytische werking waardoor het gevaar bestond dat onder sommige omstandigheden de opbrengst sterk afneemt door ‘vergiftiging’ van de katalysator. Kortom, een ingewikkelde relatie.

EASE 1 (run 0-15) Om een indruk te krijgen van de grootte-orde van de hoofdeffecten en de belangrijkste interacties, is gekozen voor een eerste serie van $2^{6-2}=16$ runs waarbij de zes hoofdeffecten vrij van tweefactor-interacties, en de tweefactor-interacties vrij van driefactor-interacties kunnen worden geschat. Vreemd genoeg bleken de opbrengsten nagenoeg met uitsluitend de hoofdeffecten te kunnen worden verklaard. De variantiebijdrage van het zwakste hoofdeffect was maar liefst 10^5 keer groter dan de bijdrage van de sterkste interactie! Omdat ondanks dit resultaat, de afwezigheid van kruistermen vanuit de chemie ondenkbaar leek, werd voor alle mogelijke modellen bestaande uit combinaties van hoofdeffecten en tweefactor-interacties, de determinatiecoëfficiënt berekend. Daaruit bleek dat ook een model met twee hoofdeffecten (E en F) en drie kruistermen ($A*B$, $B*C$ en $B*D$, zonder de hoofdeffecten zelf) de gegeven opbrengsten al zeer goed benaderde.

EASE 2 (run 16-24) Met dat model als uitgangspunt is in eerste instantie een serie van acht (2^{3-2}) runs opgezet. Uit het model bleek dat alle factoren behalve F verder moesten worden verhoogd. De factoren A t/m E werden ingesteld op 100 en 120% van de waarde van hun hoge niveaus uit de eerste fase, F op 80 en 100% van het lage niveau. Achteraf blijkt deze stap wat klein, maar wel in de goede richting. Met het oog op het opsporen van kwadratische termen is dit keer ook een run op het gemiddelde van de niveaus toegevoegd (run 24). De voorspelde opbrengsten op basis van het model uit fase 1 zijn in Tabel 2 in de bijlage tussen () naast de werkelijke opbrengsten vermeld. De overeenkomst tussen voorspelde en werkelijke waarden was qua rangorde goed en dit gaf vertrouwen in deze aanpak. Bovendien werd vastgesteld dat

een aantal interacties geheel niet aanwezig is (A*C, C*E, A*D, D*E en acht driefactor interacties).

FASE 3-9 (run 25-31) De laatste runs zijn één voor één uitgevoerd. Met de gegevens uit alle voorafgaande runs zijn net zoals in fase 2, alle mogelijke (soms derde orde) kruistermen berekend om daarna als potentiële verklarende variabelen te worden gebruikt. Uit de modellen met geringe restspreiding en klein aantal variabelen werd zonder formeel criterium het eenvoudigste model geselecteerd. Om de optimale stapgrootte te schatten werden A tot en met F op 80, 90, 100, 110 en 120% van hun laatste waarde genomen en zijn vervolgens voor elke van deze 6⁵ combinaties met brute rekenkracht de voorspelde opbrengsten en bijbehorende standaardafwijkingen berekend. De volgende run werd gekozen bij de instelling waarvoor het quotiënt van de voorspelde toename in opbrengst en standaardafwijking maximaal was. Voor run 25, 26 en 27 was deze strategie succesvol: de opbrengst steeg in deze drie stappen van 119 naar 153. Omdat werd aangenomen dat de respons ver van het optimum een ander karakter zou hebben, zijn vanaf run 26 alleen de runs met een opbrengst groter dan 100 voor de modelbouw ingezet. De laatste vier runs hebben nauwelijks nog aan verbetering van de opbrengst bijgedragen. Achteraf kwam dat natuurlijk door de nadruk op interactietermen ten koste van 'eenvoudige' kwadraten.

3. Slotcommentaar

Na afloop van de competitie werd de volgende doelfunctie bekend gemaakt aan de deelnemers (N.B. de doelfunctie op p.12 van K&P is niet correct):

$$y = 160 - (A-420)^2/5000 - (B-870)^2/10000 - (C-480)^2/10000 - (D-40)^2/70 \\ - (E-520)^2/10000 - (F-40)^2/1000 + 30/\{(A-420)(F-40)/1000\}^2 + 5\} - 30/5.$$

De deelnemers waren enigszins onthutst te vernemen dat het een zeer eenvoudige, onrealistische functie was met zes kwadratische termen en slechts één twee-factorinteractie, dus geen complexe functie die een echt probleem vertegenwoordigt. Bij deze doelfunctie hadden degenen die van een standaard kwadratische responsfunctie uitgingen een onverdiend voordeel (o.a. K&P, Van Casteren, Mulder). Zelfs iemand die in één keer al zijn designpunten homogeen random had gekozen in een willekeurig gebied, en vervolgens een tweede orde benadering had aangepast, had tot de winnaars kunnen behoren.

Vele strategieën kunnen in een bepaalde situatie de beste zijn. Een methode "bewijst" zich pas wanneer deze in uiteenlopende situaties goed blijkt te werken. Om die reden moet niet te veel waarde worden gehecht aan de uitkomst van deze competitie, die gebaseerd is op slechts één doelfunctie.

Gezien de "foto finish" en de eenvoudige, onrealistische doelfunctie levert de competitie onvoldoende grond om tot een voorkeur voor een bepaalde methode te komen. De bedoeling van dit artikel was ook slechts om meerdere succesvolle methoden onder het voetlicht te brengen en de aandacht te vestigen op het relevante vakgebied, dat aangeduid wordt als Design and Analysis of Computer Experiments (DACE).

In dit kader is het interessant om de vier besproken oplossingen onderling te vergelijken. De methode van Huele onderscheidt zich van die van de overige drie, omdat deze als enige gebaseerd is op DACE. Vergelijkt men K&P, Van Casteren en Mulder, dan zijn enkele opvallende overeenkomsten te zien. Bij alle drie werd gebruik gemaakt van een tweede orde model (Mulder gebruikte ook nog enkele derde orde termen). De runs werden niet allemaal in één keer gekozen, maar in fasen, waarbij de nieuwe runs steeds afhankelijk van de eerdere uitkomsten bepaald werden. In de laatste runs werd enkele keren een schatting van het optimaal punt uitgetoetst (hierbij kenden Van Casteren en Mulder aan bepaalde runs een groter gewicht toe dan aan andere runs).

Maar er zijn ook belangrijke verschillen. K&P begonnen met een proefopzet van Rechtschaffner voor het identificeren van hoofdeffecten en interacties, legden zich daarna toe op kwadratische effecten en pasten tenslotte een compleet tweede orde model toe. Van Casteren legde een grote nadruk op vroegtijdige interpolatie door het verkrijgen van runs rondom het optimum en koos daarom voor grote stappen en voor snelle schatting van het optimum m.b.v. kwadratische termen. Mulder legde op grond van chemische kennis vanaf het begin sterk de nadruk op interacties. In tegenstelling tot K&P selecteerden Van Casteren en Mulder telkens bepaalde termen die werden opgenomen in het beschrijvende model. Ondanks de verschillen was er in deze competitie nauwelijks een verschil in eindresultaat. Het zou interessant zijn te onderzoeken hoe de beschreven methoden functioneren in andere gevallen.

Literatuurverwijzingen

Kleijnen, J.P., en Pala, Ö. (1998), "Maximizing the Simulation Output: A Competition", *Kwantitatieve Methoden* **59**, 5-14.

Morris, M.D., en Mitchell, T.J. (1995), "Exploratory Designs for Computational Experiments", *Journal of Statistical Planning and Inference* **43**, 381-402.

Sacks, J., Welch, W.J., Mitchell, T.J., en Wynn, H.P. (1989) "Design and Analysis of Computer experiments", *Statistical Science* **4**, 409-435 (met discussie).

Welch, W.J., Yu, T.-K., Kang, S.M., en Sacks, J. (1990) "Computer Experiments for Quality Control by Parameter Design", *Journal of Quality Technology* **22**, 15-22.

Bijlage: Kwantitatieve gegevens van de experimenten beschreven in §2

TABEL 1. Experiment Van Casteren (§2.1)

run	A	B	C	D	E	F	output
0	150	400	250	10.0	300	100	90.90
1	200	400	250	10.0	300	100	95.81
2	150	540	250	10.0	300	100	102.06
3	150	400	350	10.0	300	100	94.46
4	150	400	250	14.0	300	100	94.06
5	150	400	250	10.0	400	100	94.26
6	150	400	250	10.0	300	140	84.38
7	400	540	350	14.0	400	100	131.30
8	200	800	350	14.0	400	100	127.61
9	200	540	500	14.0	400	100	118.86
10	200	540	350	20.0	400	100	121.15
11	200	540	350	14.0	600	100	118.01
12	200	540	350	14.0	400	60	121.47
13	768	878	490	51.0	573	10	127.12
14	768	993	490	51.0	526	10	125.89
15	635	993	490	51.0	526	5	140.76
16	635	878	543	51.0	526	5	141.88
17	635	878	543	63.0	526	10	136.53
18	635	878	490	63.0	573	10	136.64
19	496	790	469	34.0	538	79	152.30
20	635	860	469	34.0	512	148	132.60
21	496	930	441	34.0	512	79	152.47
22	496	860	469	51.0	512	148	139.84
23	496	860	497	34.0	486	79	152.83
24	357	860	469	51.0	512	79	152.65
25	445	849	442	38.8	512	79	157.19
26	496	849	415	38.8	493	45	158.07
27	445	901	442	34.0	493	45	159.00
28	431	849	418	39.9	489	37	159.44
29	425	888	443	39.2	398	31	158.26
30	419	958	387	39.5	476	15	157.53
31	425	846	409	39.9	478	31	159.16
32	429	848	422	40.2	491	28	159.36
*	420	870	480	40	520	40	160.00

TABEL 2. Experiment Mulder (§2.3)

run	A	B	C	D	E	F	output
0	150	400	250	10.0	300	100	90.9
1	150	400	250	12.0	360	120	91.9
2	150	400	300	10.0	360	120	92.3
3	150	400	300	12.0	300	100	94.6
4	150	440	250	10.0	360	100	96.7
5	150	440	250	12.0	300	120	93.3
6	150	440	300	10.0	300	120	93.7
7	150	440	300	12.0	360	100	100.4
8	180	400	250	10.0	300	120	91.1
9	180	400	250	12.0	360	100	97.9
10	180	400	300	10.0	360	100	98.3
11	180	400	300	12.0	300	120	94.8
12	180	440	250	10.0	360	120	97.0
13	180	440	250	12.0	300	100	99.2
14	180	440	300	10.0	300	100	99.6
15	180	440	300	12.0	360	120	100.7
16	180	440	300	14.4	360	80	(109) 107.5
17	180	440	360	12.0	360	80	(109) 107.5
18	180	528	300	12.0	360	100	(112) 110.3
19	180	528	360	14.4	360	100	(118) 114.0
20	216	440	300	14.4	432	80	(116) 112.6
21	216	440	360	12.0	432	80	(116) 112.6
22	216	528	300	12.0	432	100	(120) 115.4
23	216	528	360	14.4	432	100	(126) 119.0
24	198	484	330	13.2	396	90	(116) 112.9
25	300	528	288	20.0	622	80	(136) 128.5
26	360	581	346	22.0	560	64	(153) 142.0
27	400	700	415	27.0	600	50	(168) 153.4
28	480	840	450	28.0	720	49	(185) 152.6
29	530	770	510	31.0	660	51	(186) 151.9
30	440	900	430	26.0	630	70	(168) 154.3
31	530	1080	480	21.0	700	56	(198) 142.2
32	(niet gebruikt)						
*	420	870	480	40	520	40	160.0

Op de onderste regel (zie *) het werkelijke optimum. Bij wisselende arcering start een nieuwe fase. Tussen () de voorspelde output op basis van alleen de metingen en modellen uit voorgaande fasen.

Ontvangen: 10-01-1999

Geaccepteerd: 15-01-1999

ERRATUM

in het artikel van:

Peter G.M. van der Heijden, Ger van Gils, Jan Bouts & Joop Hox

A comparison of randomized response, CASAQ, and direct questioning: eliciting sensitive information in the context of social security fraud

in Kwantitatieve Methoden nr. 59, pagina 15 e.v., is abusievelijk de navolgende tabel weggevalen.

Table 3. Logistic regression. Response variable is income fraud. Covariates are Gender (1 = males, 2 = females); Age (measured in years); Education (1 = low, 8 is high); Born in the Netherlands (1 = yes, 2 = no); Control of Dutch language (scale of four items); Benefit (dummies for health, unemployment, else, versus no benefit anymore); Duration of benefit (measured in years); Attitude towards social security department (scale, seven items); Perception of fraud (scale, four items); Are questions annoying (scale, four items).

Multivariate results, in two blocks

Variable	Forced alternative ¹		Kuk ²		CASIQ ³		Direct ⁴		Total		var test	
	b	z	b	z	b	z	b	z	b	z	chi2	p
Block 1												
constant	3.98	1.43	-.01	-.01	1.77	1.47	.68	1.51				
Gender	-1.07	-1.66	-1.16	-1.36	-1.56	-1.72	-.79	-1.45	-1.01	-3.01	.57	.90
Age	-.06	-1.87	-.03	-1.07	-.08	-1.50	-.02	-.77	-.38	-2.40	1.62	.65
Education	.31	.63	1.07	1.75	-.25	-.42	.11	.28	.25	1.02	2.69	.44
Born in Netherlands.	-.40	-.54	.14	.18	-.44	-.48	-1.62	-2.55	-.68	-1.58	3.48	.32
Control of Dutch language	-.33	-1.05	-.36	-1.10	-.13	-.37	-.07	-.33	-.19	-1.33	.79	.85
Block 2												
constant	11.50	1.46	1.22	.43	-.94	-.36	1.55	1.76				
Benefit												
Health	.36	.29	-2.19	-.72	-.61	-.44	-1.95	-1.93	-1.00	-1.43	2.26	.52
Unemploym.	-1.72	-1.66	-.80	-.84	.31	.30	-1.02	-1.58	-.86	-2.00	2.03	.56
Else	-.06	-.04	-1.94	-1.07	1.58	.86	* ⁵	*	-.13	-.15	1.87	.39
Duration benefit	-.24	-1.29	-.30	-1.39	-.05	-.54	.01	.23	-.04	-.78	3.97	.23
Attitude towards soc. sec. off.	.15	.16	-.48	-.33	-.91	-.85	-2.71	-2.94	-1.07	-1.52	4.98	.17
Perception fraud	-.95	-1.53	-.32	-.50	-.23	-.49	.45	1.12	-.14	-.47	3.96	.26
Questions annoy	-2.90	-1.47	.52	.36	1.63	.66	-.04	-.03	-.17	-.20	2.69	.44

¹n = 96; in analyses including Control of Dutch language n = 95; in analyses including Duration n = 95

²n = 105; in analyses including Duration n = 103

³n = 47; in analyses including Control of Dutch language n = 45; in analyses including Duration n = 44

⁴n = 99; in analyses including Control of Dutch language n = 99; in analyses including Duration n = 92

⁵only one observation. Inclusion leads to unidentified model.