KANSREKENING, STATISTIEK EN ACCOUNTANSCONTROLE

door Prof. Drs. J. Muilwijk

0. Samenvatting

Naar aanleiding van een aantal artikelen van M. Vermaas en J.H. Blokdijk worden enkele methoden uit de kwaliteitszorg besproken. Behandeld worden: de keuringskarakteristiek en de daarmee samenhangende begrippen zoals het criterium bij enkelvoudige steekproefschema's, het controlepunt en de steilheid, afgeknotte en meervoudige schema's; toetsen van hypotheses in de accountantscontrole; schatting van een fractie fouten (exacte en benaderde methoden), betrouwbaarheidsgrenzen.

Vervolgens wordt stelling genomen in de discussie over de toelaatbaarheid van het uitbreiden van de steekproef. Tenslotte wordt aangegeven hoe men in de accountantscontrole tot zo goed mogelijke systemen kan komen.

1. Inleiding

In 1974 werden door het NIVRA in Paterswolde en Eindhoven bijeenkomsten georganiseerd over de mogelijkheden van de steekproef in de accountantscontrole. Een gedegen verslag daarvan gaven Tuitjer en Zuijdervliet (1975). In dit blad werd door M. Vermaas in een aantal artikelen op de daarin verdedigde opvattingen en op het boek van Kriens en Dekkers (1979) kritiek geleverd, hetgeen tot enkele reacties van J.H. Blokdijk leidde. In februari 1980 wendde de redactie zich tot mij met het verzoek om mijn mening over de gerezen meningsverschillen uiteen te zetten. Deze uiteenzetting is nu - tot mijn spijt met een flinke vertraging - gereed.

Het leek mij nuttig te beginnen met een schets van enkele methoden uit de kwaliteitszorg, die model gestaan hebben voor de steekproeftoepassingen in de accountantscontrole. Het betreft enkelvoudige, dubbele, meervoudige en afgeknotte schema's, die allereerst vanuit de kansrekening zijn belicht. Vervolgens komen statistische aspecten (toetsen en schatten) aan de orde (hoofdstuk 2).

In hoofdstuk 3 wordt commentaar geleverd op de discussie Vermaas - Blokdijk; hoofdstuk 4 is gewijd aan de vraag hoe het in de praktijk beter zou kunnen dan via enkele omstreden technieken.

In een vijftal appendices worden o.m. de notatie, uitgewerkte voorbeelden en bewijzen gegeven.

2. Statistische kwaliteitsbeheersing en accountantscontrole

Terwijl de statistische kwaliteitsbeheersing (Quality control) reeds in de twintiger jaren ingang vond, begon de ontwikkeling van administratieve controlemethoden op steekproefbasis enkele tientallen jaren later. Boeken als Vance and Neter (1956) en Arkin (1963) zijn duidelijk geïnspireerd door de kwaliteitszorg.

Ter inleiding van mijn stellingname in de discussie over steekproefonderzoek bespreek ik in dit hoofdstuk enkele begrippen uit de kwaliteitsbeheersing, die voor de accountantspraktijk van belang zijn.

2.1. Keuring van populaties met een tweedeling

In de industrie wordt o.a. „partijcontrole op uitval" toegepast. Daarbij wordt een partij of populatie die uit goede en foute (of „defectieve") elementen bestaat op grond van een steekproef goed- of afgekeurd. Uitgangspunt is dus de aanwezigheid van een *tweedeling* (alternatieve eigenschap).

2.1.1. Enkelvoudig steekproefvoorschrift

Uit een populatie van de omvang N produkten (bij accountantscontrole: posten) wordt een enkelvoudig aselecte steekproef getrokken, waarvan de grootte n tevoren vaststaat. *Gemakshalve* nemen wij aan dat de steekproef slechts een klein deel, bijv. niet meer dan 10%, van de populatie uitmaakt, zodat de trekkingen als *onafhankelijk* kunnen worden beschouwd. Zie voor afhankelijke trekkingen paragraaf 2.2.3.

De onbekende *fractie* defectieven in de populatie noemen wij P (het percentage is dus $100P$) en het aantal foute exemplaren in de steekproef r. Het steekproefvoorschrift luidt nu: keur de partij goed als r ten hoogste gelijk is aan het *criterium* (of goedkeurgrens) c.[1]) De aantallen n en c bepalen dit *enkelvoudige voorschrift* (of *schema*) volledig.

Daar de *goedkeurkans* S zowel van het voorschrift als van de „kwaliteit" P afhangt schrijven wij voor deze kans $S\,(P; n; c)$ of als er geen misverstand kan ontstaan $S\,(P)$. Op grond van de veronderstelling dat $n \leqslant 0,1N$ is kan men de volgende formule van de binomiale verdeling[2]) gebruiken:

$$S\,(P) = (1-P)^n + nP\,(1-P)^{n-1} + \frac{n\,(n-1)}{2}\,P^2\,(1-P)^{n-2} + \ldots +$$

$$\frac{n\,(n-1)\ldots(n-c+1)}{1.2.\ldots.c}\,P^c\,(1-P)^{n-c}$$

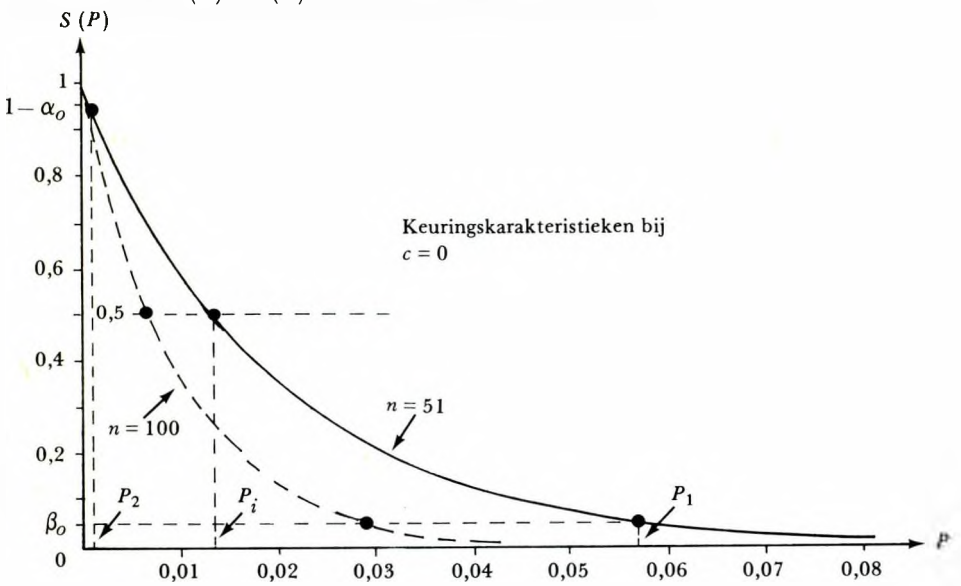
De $(c+1)$ termen in het rechterlid stellen de kansen op resp. $0, 1, \ldots, c$ defectieven voor; als $c = 0$ is alleen de eerste term aanwezig, enz. $S\,(P)$ heet de *keuringskarakteristiek*, zie fig. 2.1.

---

[1]) Bij Kriens en Dekkers vindt men de notatie k_o voor c. In Appendix 1 wordt de notatie uit dit boek met de onze vergeleken.

[2]) Zie bijv. De Wolff (1973) of Kriens en Dekkers (1979).

Wij beschouwen eerst het geval $c = 0$: bij het vinden van één fout wordt de populatie afgekeurd. Hoe groter n, des te strenger is de keuring. De kolommen (1) en (2) in tabel 2.1. illustreren dit.



Figuur 2.1. Enkelvoudige steekproefvoorschriften
Bron: Tabel 2.1.

Tabel 2.1. Goedkeurkansen $S(P)$ bij enkelvoudige schema's*)

	(1)	(2)	(3)	(4)
Fractie fouten in de populatie, P	$n = 51$ $c = 0$	$n = 100$ $c = 0$	$n = 100$ $c = 1$	$n = 160$ $c = 2$
0,01	0,599	0,366	0,736	0,784
0,02	0,357	0,133	0,402	0,377
0,03	0,212	0,048	0,195	0,139
0,04	0,125	0,017	0,087	0,043
0,05	0,073	0,006	0,037	0,012
0,06	0,043	0,002	0,015	0,003
0,07	0,025	0,001	0,006	0,001
0,08	0,014	0,000	0,002	0,000
Goedkeurkansen, $S(P)$	Fractie fouten in de populatie, P			
$1 - \alpha_o = 0,95$	$P_2 = 0,0010$	0,0005	0,0036	0,0051
0,50	$P_i = 0,0135$	0,0069	0,0167	0,0167
$\beta_o = 0,05$	$P_1 = 0,057$	0,0295	0,0466	0,039

*) n = steekproefomvang; c = criterium.
De verklaring van de overige symbolen vindt men in de tekst.
In de literatuur over kwaliteitszorg treft men veelal de volgende symbolen aan: p_a of θ_1 voor P_2; p_t of θ_2 voor P_1; p_o of θ_o voor P_t. Zie ook Appendix 1.

Hoe moet men nu n en c kiezen? Enerzijds moet het praktisch zeker zijn dat een slechte partij wordt afgekeurd, anderzijds moet het zeer waarschijnlijk zijn dat een goede partij wordt goedgekeurd. Zowel de producent als de consument dienen te worden beschermd (met name als het om *niet-ernstige* fouten gaat). In de kwaliteitszorg definieert men de volgende vier begrippen. Laat P_1 de fractie fouten in de „beste slechte partij" voorstellen. Men noemt P_1 de grenskwaliteit voor de consument (het *„grensgeval"*) en eist dat de afnemer slechts een kleine kans, het *consumentenrisico* β_o loopt dat zo'n partij wordt goedgekeurd. Dan geldt dus

$$S(P_1) = \beta_o.$$

Omdat de keuringskarakteristiek een dalende functie van P is volgt hieruit

$$S(P) < \beta_o \quad \text{als} \quad P > P_1.$$

Omgekeerd moet de fabrikant slechts een kleine kans lopen dat een goede partij wordt afgekeurd. Laat P_2 de fractie fouten in de „slechtste goede partij" voorstellen; P_2 heet de grenskwaliteit voor de producent en men eist nu dat de afkeurkans van zo'n partij slechts een eveneens kleine kans α_o, het *producentenrisico*, bedraagt. In formule:

$$1 - S(P_2) = \alpha_o$$

of

$$S(P_2) = 1 - \alpha_o.$$

Hieruit volgt

$$S(P) > 1 - \alpha_o \quad \text{als} \quad P < P_2.$$

Voorbeelden:

1. Stel $\alpha_o = 0{,}05$ dan blijkt uit kolom (1) van tabel 2.1. dat $P_2 = 0{,}001$ bedraagt voor $n = 51$ en $c = 0$. Zij verder $\beta_o = 0{,}05$, dan is voor dit steekproefvoorschrift $P_1 = 0{,}057$. Kiest men deze vier getalwaarden dan volgt daar omgekeerd uit $n = 51$ en $c = 0$.
2. $n = 100$, $c = 0$. Nu is (kolom (2)) $P_2 = 0{,}0005$ en $P_1 = 0{,}0295$.

Wij beschouwen vervolgens $c = 1$, waarbij geldt:

$$S(P) = (1 - P)^n + nP(1 - P)^{n-1}.$$

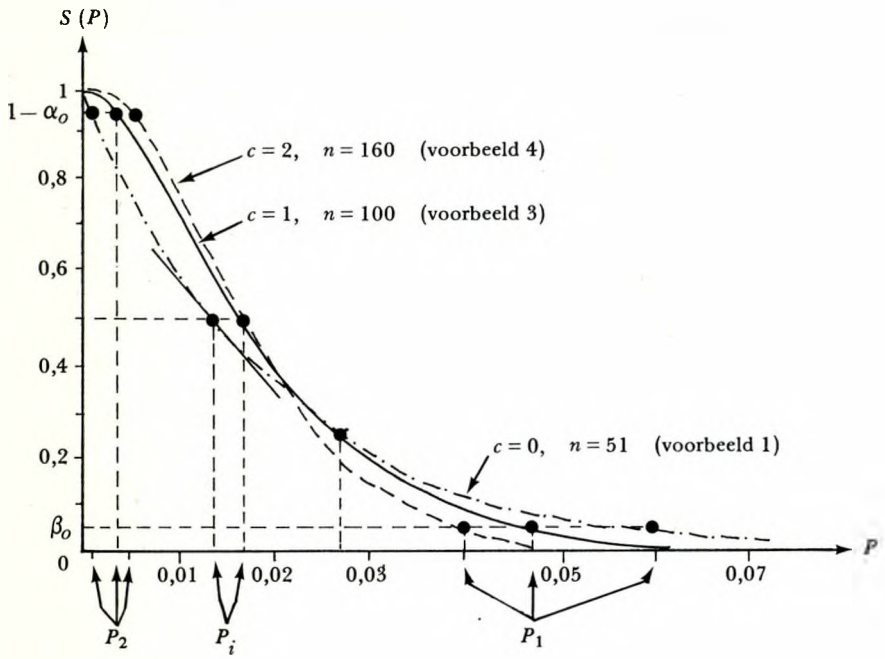
Voor $n = 100$ vindt men in kolom (3) van tabel 2.1. een aantal waarden. Deze goedkeurkansen liggen alle uiteraard hoger dan de overeenkomstige in kolom (2), omdat nu (*voorbeeld* 3) pas bij twee fouten wordt afgekeurd. In de onderste helft van de tabel vindt men P_1 en P_2 (zelfde risico's).

Zij verder $c = 2$; dan geldt:

$$S(P) = (1-P)^n + nP(1-P)^{n-1} + \frac{n(n-1)}{2} P^2 (1-P)^{n-2}.$$

Voor $n = 160$ (*voorbeeld* 4) vindt men $S(P)$ enz. in kolom (4).

Om enkelvoudige steekproefschema's met verschillende omvang en/of criterium te vergelijken letten wij op figuur 2.2., waarin de voorbeelden 1, 3 en 4 zijn afgebeeld.



Figuur 2.2. Enkelvoudige schema's
Bron: Tabel 2.1.

Voor kleine waarden van de fractie foute exemplaren P, nl. kleiner dan 0,0265, geldt

$$S_3(P; 100; 1) \text{ (voorbeeld 3)} > S_1(P; 51; 0) \text{ (voorbeeld 1)}$$

en voor $P > 0{,}0265$

$$S_3(P; 100; 1) < S_1(P; 51; 0).$$

Verder ziet men dat bij voorbeeld 1 de waarden van P_2 en P_1 verder uiteen liggen dan bij voorbeeld 3, zodat de kromme tussen P_2 en P_1 in voorbeeld 3 gemiddeld steiler verloopt. Bij voorbeeld 4 liggen de grenskwaliteiten nog weer dichter bijeen.

Merk op dat voor $c = 1$ en $c = 2$ de grafiek van de keuringskarakteris-

tiek een omgekeerde *S*-vorm bezit: bij toenemende P eerst een geringe daling, daarna een steil gedeelte en tenslotte een bijna horizontaal verloop. Bij alle $c \geqslant 1$ heeft de keuringskarakteristiek één buigpunt tussen $P = 0$ en $P = 1$ (behalve als $c = n - 1$, maar dit komt niet voor: goedkeuren bij één goed exemplaar); dit buigpunt ligt bij $P = c/(n - 1)$, zie Appendix 4.

Samenvattend: bij fabricagecontrole kan de eis worden gesteld dat de keuringskarakteristiek door de punten $(P_2, 1 - \alpha_o)$ en (P_1, β_o) gaat; voor de manier om n en c te berekenen zie men Muilwijk (1977). Een grotere n betekent dat de kromme voor $0 < P < 1$ lager ligt; (alleen) een grotere c dat zij overal hoger ligt. Als men zowel n als c vergroot, terwijl n meer toeneemt dan c - bijv. als men ze evenredig laat toenemen - dan verloopt het middenstuk steiler. Zie voor het bewijs *op. cit.* paragraaf 6.2.1; de krommen 3 en 4 uit fig. 2.2. illustreren dit.

Een andere methode, o.m. bij Philips toegepast, berust op het *controlepunt* (indifference quality level), dat is de fractie fouten waarbij goed- en afkeurkans beide 0,5 bedragen; wij geven deze fractie aan met P_i (i van indifferentie). In de tabellen en grafieken in deze paragraaf vindt men de waarden van het controlepunt bij de voorbeelden 1-4. Uiteraard geldt steeds

$$P_2 < P_i < P_1$$

omdat

$$1 - \alpha_o > \tfrac{1}{2} > \beta_o.$$

Wij komen op P_i in paragraaf 2.1.4. terug; zoals de voorbeelden 3 en 4 leren is het controlepunt alleen niet voldoende om n en c te bepalen.

2.1.2. Afgeknotte schema's (pseudo-enkelvoudig)

Als het doel van de controle *uitsluitend* het goed- of afkeuren van een populatie is, dan zijn er verschillende schema's die bij gelijke grenskwaliteiten en risico's lagere controlekosten met zich meebrengen dan het enkelvoudige. Een voor de hand liggend voorschrift is het *afgeknotte* (curtailed) of „verkorte". Daarbij wordt de controle beëindigd zodra het resultaat vaststaat, ook al zijn er nog geen n waarnemingen verricht. Dit is het geval als er $(c + 1)$ fouten gevonden zijn of $(n - c)$ goede posten; men keurt dan onmiddellijk af resp. goed en laat de verdere waarnemingen achterwege. Gemiddeld kan men dan met minder dan n waarnemingen volstaan, zonder dat de keuringskarakteristiek verandert (het opzoeken van de posten kan duurder worden).

Bij zeer kleine fracties fouten is de besparing gering. Als $c = 0$ dan is de gemiddelde steekproefomvang

$$\frac{1 - (1 - P)^n}{P}$$

(Kriens en Dekkers (1979) blz. 219), hetgeen voor een maximale omvang

$n = 100$ tot de volgende gemiddelde steekproefgrootten leidt (*voorbeeld* 5): 99,5 bij 0,01% fouten in de populatie, 95 bij 0,1% maar 63 bij 1%.

Wanneer $c = 1$, dus één fout in de steekproef wordt nog geaccepteerd, bedraagt de gemiddelde steekproefomvang indien men ophoudt bij 2 fouten dan wel bij $(n - 1)$ correcte posten:

$$\frac{2 - 2\,(1 - P)^{n-1}}{P} - (n - 1)\,(1 - P)^{n-1}$$

(zie Appendix 2 voor een toelichting). *Voorbeeld* 6: maximale grootte $n = 100$. De gemiddelden bedragen 99,0 bij $P = 0{,}01\%$; 98,9 bij 0,1%; 89 bij 1%. Voor andere waarden van n en c zie men de tabel in Appendix 2.

Bij controle op *ernstige* fouten biedt de verkorte methode nauwelijks voordelen. Men kan dan immers aannemen dat P klein is (bijv. $< 0{,}001$) zodat bij niet al te grote steekproeven (bijv. $n_{max} < 500$) het produkt nP als regel een kleine waarde bezit (i.c. $< 0{,}5$). Uit tabel A2.2. blijkt dat de besparingen dan niet groot zijn.

Bij *niet-ernstige* fouten, bijv. bepaalde formele fouten, komt dit systeem wèl in aanmerking, al moet men er rekening mee houden dat het ongeschikt is om schattingen en betrouwbaarheidsgrenzen voor P mee te berekenen. De steekproefomvang hangt immers van het toeval af, zodat de gebruikelijke schattingsformules niet opgaan. Een andere mogelijkheid is het combineren van afknotting met meervoudige systemen, zie de volgende paragraaf.

2.1.3. Dubbele schema's e.d.: uitbreiding van de steekproef

Het enkelvoudige en het afgeknotte schema hebben beide het bezwaar dat één fout teveel onmiddellijk tot afkeuring leidt. Flexibeler is het *dubbele* steekproefvoorschrift, waarbij begonnen wordt met n_1 waarnemingen. Vindt men daarin c_1 of minder fouten, dan wordt de populatie goedgekeurd; treft men $c_1 + 1$, $c_1 + 2$, ..., c_2 fouten aan, dan neemt men een tweede steekproef van de omvang n_2. Blijken er meer dan c_2 fouten in de eerste steekproef aanwezig te zijn, dan wordt de populatie zonder meer afgekeurd. Bij zeer goede partijen is de kans groot dat men in de eerste steekproef c_1 of minder fouten vindt, zodat men dan al een beslissing kan nemen; bij slechte partijen is de kans groot dat men meer dan c_2 fouten in de eerste steekproef aantreft en ook dan behoeft de steekproef niet te worden uitgebreid. Als er wèl een tweede steekproef wordt getrokken let men op het totaal aantal fouten in beide steekproeven samen. Bedraagt dit ten hoogste c_2, dan wordt de populatie alsnog goedgekeurd; is het meer dan c_2, dan volgt afkeuring.

Voorbeeld 7. Men begint met $n_1 = 20$ waarnemingen en accepteert de populatie als daarin $c_1 = 0$ fouten voorkomen. Indien de fractie fouten in de populatie $P \leqslant 0{,}01$ is, dan bedraagt de kans daarop tenminste $0{,}99^{20} = 0{,}82$. De populatie wordt direct afgekeurd als er in de n_1 posten

meer dan $c_2 = 2$ fouten worden ontdekt; bij een slechte partij, bijv. $P \geqslant 0{,}20$, bedraagt de kans daarop tenminste

$$1 - 0{,}8^{20} - 20 \times 0{,}2 \times 0{,}8^{19} - \frac{20 \times 19}{2} \times 0{,}2^2 \times 0{,}8^{18} = 0{,}79.$$

In de overige gevallen, hier dus 1 of 2 fouten in de eerste steekproef, wordt er een tweede steekproef van $n_2 = 40$ exemplaren getrokken. In de totale steekproef van $n_1 + n_2 = 60$ mogen er ten hoogste $c_2 = 2$ fout zijn (dus bij één fout in de eerste steekproef: 0 of 1 in de tweede; bij twee fouten in de eerste geen fouten in de tweede). In tabelvorm:

Tabel 2.2. Dubbel schema (voorbeeld 7) met $c_1 = 0$ en $c_2 = 2$

	Aantallen fouten in:			
	Eerste steekproef ($n_1 = 20$)	Tweede steekproef ($n_2 = 40$)	Totaal ($n_1 + n_2 = 60$)	Beslissing
1	0 (= c_1)	blijft achterwege	–	Goedkeuren
2	1	0	1	Goedkeuren
3	1	1	2 (= c_2)	Goedkeuren
4	1	$\geqslant 2$	$\geqslant 3$	Afkeuren
5	2 (= c_2)	0	2	Goedkeuren
6	2	$\geqslant 1$	$\geqslant 3$	Afkeuren
7	$\geqslant 3$ ($> c_2$)	blijft achterwege	–	Afkeuren

De keuringskarakteristiek vinden wij uit het volgende schema, vergelijk ook Appendix 3:

Schema 2.3. Kansen bij dubbele schema's met $c_1 = 0$ en $c_2 = 2$

	Kansen in de eerste steekproef	Voorwaardelijke kansen in de tweede steekproef	Totale kans op goedkeuren
1	$T_0 = (1-P)^{n_1}$	–	$(1-P)^{n_1}$
2	$T_1 = n_1 P\,(1-P)^{n_1-1}$	$(1-P)^{n_2}$	$n_1 P\,(1-P)^{n_1+n_2-1}$
3	T_1	$n_2 P\,(1-P)^{n_2-1}$	$n_1 n_2 P^2\,(1-P)^{n_1+n_2-2}$
5	$T_2 = \frac{n_1\,(n_1-1)}{2} P^2\,(1-P)^{n_1-2}$	$(1-P)^{n_2}$	$\frac{n_1\,(n_1-1)}{2} P^2\,(1-P)^{n_1+n_2-2}$

In het algemeen gebruiken wij voor een dubbel schema de notatie $S\,(P; n_1; n_2; c_1; c_2)$. Voor $c_1 = 0$ en $c_2 = 2$ luidt de keuringskarakteristiek zodoende

$$S\,(P; n_1; n_2; 0; 2) = (1-P)^{n_1} + n_1 P\,(1-P)^{n_1+n_2-1} +$$

$$\frac{n_1\,(n_1+2n_2-1)}{2}\,P^2\,(1-P)^{n_1+n_2-2}.$$

In voorbeeld 7 komt er

$$S\,(P; 20; 40; 0; 2) = (1-P)^{20} + 20P\,(1-P)^{59} +$$

$$990P^2\,(1-P)^{58}.$$

Wij vergelijken deze kansen met de overeenkomstige bij twee vergelijkbare enkelvoudige schema's, nl. *voorbeeld* 8: $n = 45$, $c = 2$ en *voorbeeld* 9: $n = 50$, $c = 2$.

Tabel 2.4. Vergelijking van de voorbeelden 7 - 9

P	Keuringskarakteristieken (enkel en dubbel)			Verwachte steekproefgrootte in voorbeeld 7
	$S_7\,(P; 20; 40; 0; 2$	$S_8\,(P; 45; 2)$	$S_9\,(P; 50; 2)$	
0,01	0,984	0,990	0,986	27,2
0,02	0,912	0,939	0,922	33,0
0,03	0,796	0,848	0,811	37,4
0,04	0,662	0,732	0,677	40,6
0,05	0,553	0,608	0,541	42,6
0,06	0,420	0,488	0,416	43,8
0,07	0,326	0,382	0,311	44,2
0,08	0,250	0,291	0,226	44,0
0,09	0,192	0,217	0,161	43,3
0,10	0,148	0,159	0,114	42,2
0,15	0,041	0,027	0,014	34,6
0,20	0,012	0,003	0,001	27,8
0,25	0,003	0,001	0,000	23,5
0,30	0,001	0,000	0,000	21,4

Uiteraard is steeds $S_8 > S_9$. Men ziet dat voor $P \leqslant 0{,}04$, dus relatief goede partijen, geldt $S_8 > S_9 > S_7$: de kansen op goedkeuren zijn bij het dubbele schema kleiner dan bij de enkelvoudige.[3]) Voor $0{,}05 \leqslant P \leqslant 0{,}10$ liggen de kansen voor het dubbele schema tussen die voor de beide enkelvoudige in. Voor $P \geqslant 0{,}15$ (slechte partijen) is S_7 iets groter. De drie steekproefvoorschriften ontlopen elkaar niet veel wat de keuringskarakteristieken betreft.

In de laatste kolom zijn de verwachte steekproefgrootten bij het dubbele schema vermeld.[4]) Zij zijn in alle gevallen kleiner dan 45. Voor $0{,}01 \leqslant P \leqslant 0{,}04$ is de besparing in de gemiddelde steekproefomvang 10%

[3]) Het geldt ook voor $0 < P < 0{,}01$.

[4]) Zie voor de berekening Appendix 3.

tot 40% van 45, voor $0{,}05 \leqslant P \leqslant 0{,}10$ 2% tot 6%, voor $0{,}15 \leqslant P \leqslant 0{,}30$ 23% tot 52% (voor nog grotere P wordt 25/45 = 56% benaderd). Uiteraard zijn de besparingen groter als men uitgaat van $n = 50$.

In verband met de discussie tussen Vermaas en Blokdijk[5]) geef ik nog een voorbeeld waarover in Appendix 3 meer details te vinden zijn.
Voorbeeld 10: $n_1 = 75$; $n_2 = 150$; $c_1 = 0$; $c_2 = 2$. Berekening van de keuringskarakteristiek levert $P_i = 0{,}0142$ en voor $\alpha_o = 0{,}05$ een $P_2 = 0{,}0041$ alsmede voor $\beta_o = 0{,}05$ een $P_1 = 0{,}04$.

Een algemenere vergelijking tussen enkelvoudige en dubbele systemen wordt door Schaafsma en Willemze (1978, blz. 276) gemaakt. Zij veronderstellen:

a. Vergelijkbare keuringskarakteristieken, waaronder wordt verstaan dat het controlepunt en de steilheid (zie paragraaf 2.1.4.) gelijk zijn.

b. Efficiënte dubbele schema's, waarvoor zij in navolging van Hamaker de volgende voorwaarden stellen (*op. cit.* blz. 260):

$$2n_1 \leqslant n_2 \leqslant 3n_1$$

$$\frac{c_1}{c_2} < \frac{n_1}{n_1 + n_2}$$

(hieraan voldoen de voorbeelden 7 en 10). De achtergrond is dat de tweede steekproef strenger toetst dan de eerste. Zij vinden:

1. Voor de slechtste goede partij ($P = P_2$): circa 30% besparing.
2. Voor een partij met $P = P_i$: circa 15% besparing.

Opgemerkt zij dat P_i in de voorbeelden 7-9 tussen 0,05 en 0,06 ligt (voorbeeld 7: 0,053) en P_2 tussen 0,01 en 0,02; P_1 is bij de voorbeelden 7, 8 en 9 resp. 0,14; 0,13; 0,12.

Behalve het dubbele steekproefvoorschrift zijn bij de kwaliteitszorg soms ook *meervoudige* schema's in gebruik. Bij een dergelijk voorschrift kan de tweede steekproef behalve tot accepteren of verwerpen ook tot voortzetting van het steekproeftrekken leiden. Wij komen daarop in paragraaf 3.1.1. terug en vermelden thans alleen dat deze tot een gemiddelde besparing van 25% tot 45% van de steekproefomvang kunnen leiden. Voor *sequente* steekproeven verwijzen wij naar Kriens en Dekkers (1979) of Schaafsma en Willemze (1978).

Men combineert soms het dubbele of meervoudige met het *afgeknotte* schema: de eerste steekproef wordt in zijn geheel gecontroleerd (zodat daaruit P geschat kan worden); bij een eventuele tweede of latere steekproef houdt men op zodra het aantal fouten in de steekproeven samen boven het toegelaten maximum komt of als het aantal correcte posten zo groot is geworden dat in ieder geval de beslissing „goedkeuren" zal zijn.

---

[5]) Zie paragraaf 3.1.

2.1.4. Het controlepunt en de steilheid

Door H.C. Hamaker is het controlepunt P_i (waarvoor zoals gezegd de goedkeurkans 50% bedraagt) als eerste grondslag voor de keuze van steekproefschema's aanbevolen. Bij een enkelvoudig voorschrift moet daarnaast nog één andere grootheid worden gespecificeerd teneinde n en c te kunnen bepalen. Meestal wordt hiertoe de *steilheid* h_i van de keuringskarakteristiek in het controlepunt gekozen met als definitie:

$$h_i = -\frac{P_i}{0{,}5} S'(P_i),$$

waarin $S'(P_i)$ de helling van de kromme in $P = P_i$ voorstelt.[6])

Daar de keuringskarakteristiek een dalende functie is, krijgt men voor h_i een positief getal. Naarmate de steilheid groter is, is de goedkeurkans gevoeliger voor kleine (relatieve) afwijkingen ten opzichte van P_i. Vergelijk ook figuur 2.2.

In Appendix 4 wordt uiteengezet hoe men bij gegeven P_i en h_i (afgeronde) waarden voor n en c vindt. Schaafsma en Willemze (1978) behandelen de voordelen van het gebruik van controlepunt en steilheid (blz. 227-228).[7])

2.1.5. Keuren en toetsen

Tot dusver hebben wij steeds over *keuring* van een populatie op grond van een steekproef gesproken. In termen van de statistische theorie betekent dit dat er veelal een *nulhypothese* over een onbekende fractie fouten P wordt *getoetst* tegen een *alternatieve* hypothese. Deze begrippen kunnen vertaald worden in de grootheden die wij bij keuringskarakteristieken hebben leren kennen.

Beschouw eerst het geval dat men zich met *niet-ernstige* fouten bezig houdt en zij P_2 weer de fractie fouten in de slechtste goede partij. Goede partijen dienen zelden afgekeurd te worden en daar $S(P_2) = 1 - \alpha_o$ is het risico om dit wèl te doen ten hoogste α_o. Men noemt $1 - S(P)$ voor $P \leqslant P_2$ de *kans op fouten van de eerste soort*, nl. dat de nulhypothese $P \leqslant P_2$ ten onrechte verworpen wordt. α_o heet de onbetrouwbaarheidsdrempel van de toets. Zij verder P_1 de fractie fouten in de beste slechte partij; wegens $S(P_1) = \beta_o$ worden slechte partijen goedgekeurd met kansen $S(P) \leqslant \beta_o$. Deze goedkeuring geschiedt ten onrechte; de kans daarop wordt de *kans op fouten van de tweede soort* genoemd. Deze is dus $S(P)$ voor $P \geqslant P_1$.

Als $P_2 < P < P_1$ geldt kan men geen van beide soorten fouten maken. Voor $P_2 < P < P_i$ is de goedkeurkans het grootst, voor $P_i < P < P_1$ de af-

---

[6]) De steilheid heeft de vorm van een elasticiteit, want men kan ervoor schrijven

$$\frac{P_i}{S(P_i)}\left(-\frac{dS}{dP}\right)_{P=P_i} \sim -\frac{\Delta S}{S} : \frac{\Delta P}{P}.$$

[7]) Naar mijn mening is (voor de accountantscontrole) een andere combinatie te overwegen. Zie hoofdstuk 4.

keurkans (voor $P = P_i$ zijn beide kansen gelijk). Men noemt $P_1 - P_2$ wel de *tolerantie*.

Vervolgens letten wij op *ernstige* fouten. Kriens en Dekkers (1979) kiezen in dat geval $P_2 = 0$, zodat ook $\alpha_o = 0$; er kunnen geen fouten van de eerste soort worden gemaakt.[8]) Hanteert men nu een *enkelvoudig* schema met $c = 0$ - dus *één fout leidt tot afkeuring* - dan geldt $\beta_o = S(P_1) = (1 - P_1)^n$, vgl. de voorbeelden 1 en 2. Ik prefereer in deze situatie andere waarden van c (dus ook een grotere n)[9]) maar nog beter lijkt het mij een dubbel of meervoudig schema te gebruiken. Ik kom daarop in het vervolg terug.

Alvorens de discussie Vermaas-Blokdijk onder de loep te nemen moeten wij ons evenwel nog met de begrippen schatten en betrouwbaarheidsintervallen bezighouden. Dat geschiedt in paragraaf 2.2.

2.2. Schatting van een fractie fouten bij enkelvoudige steekproeven

Als men een enkelvoudige steekproef van de omvang n trekt uit een populatie met een onbekende fractie fouten P en men vindt r fouten, dan is $p = r/n$ een zuivere schatter voor P. Dit houdt in dat het gemiddelde van alle waarden p in de steekproevenruimte (dat is de verzameling van alle mogelijke aselecte steekproeven van n uit deze populatie), ook wel de *verwachting* van p genoemd, gelijk is aan P. Van belang is (vooral) ook hoe sterk p kan variëren. Hierover kan men zich uit de steekproef mits n groot is - bijv. 50 of meer - en r niet al te klein is - bijv. ten minste 5 - op de bekende manier een oordeel vormen: men past de normale benadering toe. Bij accountantscontrole komen echter vaak lage aantallen fouten in de steekproef voor.

2.2.1. Betrouwbaarheidsgrenzen bij de binomiale verdeling

Wij gaan uit van niet te kleine n en van trekking met teruglegging (of van trekking zonder teruglegging bij zo grote populaties dat bijv. $n \leqslant 0{,}1\,N$ is), zodat de trekkingen als onafhankelijk kunnen worden beschouwd. In paragraaf 2.2.3. bespreken wij de exacte verdeling bij het trekken zonder teruglegging.

Wij voeren nu het begrip eenzijdige *overschrijdingskans* in. Onder de *linker* overschrijdingskans van een aantal fouten r verstaat men de kans op r of minder fouten; in formule, vergelijk paragraaf 2.1.1.:

$$(1-P)^n + nP(1-P)^{n-1} + \ldots + \frac{n(n-1)\ldots(n-r+1)}{1.2.\ldots.r} P^r (1-P)^{n-r} \qquad (r = 0, 1, \ldots)$$

Onder de *rechter* overschrijdingskans van de waarde r verstaat men de

[8]) Bij Kriens en Dekkers wordt een ander tweetal hypothesen tegen elkaar getoetst, nl. $P_2 = 0$ tegen $P_2 > 0$.

[9]) Vermaas (1979b) noemt voorschriften met $c = 0$ onlogisch; dat vind ik niet, maar ik acht ze wèl onbevredigend. Zie paragraaf 3.2.1. en 4.2.3.

kans op r of meer fouten. Deze is gelijk aan één min de kans op $(r-1)$ of minder fouten:

$$1 - \{ (1-P)^n + nP(1-P)^{n-1} + \ldots + \frac{n(n-1)\ldots(n-r+2)}{1.2.\ldots.(r-1)} P^{r-1}(1-P)^{n-r+1} \} \qquad (r = 1, 2, \ldots)$$

(voor $r = 0$ is deze kans uiteraard = 1).

In het voetspoor van J. Neyman en E.S. Pearson definiëren wij nu de linker en rechter *betrouwbaarheidsgrenzen* bij gegeven waarden van de eenzijdige overschrijdingskansen; wij kiezen voor beide 0,05. Dan heeft men de volgende definities:

Onder de *linker* betrouwbaarheidsgrens p_- voor P wordt verstaan die fractie fouten in de populatie waarbij de *rechter* overschrijdingskans van de gevonden waarde r gelijk is aan 0,05. In formule:

$$1 - \{ (1-p_-)^n + np_-(1-p_-)^{n-1} + \ldots + \frac{n\ldots(n-r+2)}{1\ldots(r-1)} p_-^{r-1}(1-p_-)^{n-r+1} \} = 0{,}05 \qquad (r = 1, 2, \ldots)$$

Dit is een fractie die kleiner dan p is, bijv. voor $n = 50, r = 1$: $p_- = 0{,}001 < 0{,}02$.

Onder de *rechter* betrouwbaarheidsgrens p_+ voor P wordt verstaan die fractie fouten in de populatie waarbij de *linker* overschrijdingskans van de waarde r gelijk is aan 0,05. In formule:

$$(1-p_+)^n + np_+(1-p_+)^{n-1} + \ldots + \frac{n\ldots(n-r+1)}{1\ldots r} p_+^r (1-p_+)^{n-r} = 0{,}05 \qquad (r = 0, 1, \ldots)$$

Er geldt $p < p_+$, bijv. voor $n = 50, r = 1$: $p_+ = 0{,}091 > 0{,}02$.

Het bepalen van binomiale betrouwbaarheidsgrenzen is een heel gereken. Wij gaan daarop niet in en vermelden slechts enkele getalwaarden, afgeleid uit Muilwijk en Schouten (1960, tabel f3).

Tabel 2.5. Betrouwbaarheidsgrenzen (eenzijdig; 0,05) bij binomiale verdelingen, $n = 50$

Aantal fouten r	Linker grens p_-	Rechter grens p_+	Overschrijdingskansen van r	
			$P = p_-$ (rechts)	$P = p_+$ (links)
0	–	0,058	–	0,0504
1	0,001	0,091	0,0488	0,0509
2	0,007	0,121	0,0481	0,0492
3	0,017	0,148	0,0533	0,0497
4	0,027	0,174	0,0458	0,0496

Men kan deze overschrijdingskansen zonder moeite narekenen (tengevolge van de afronding van de grenzen in drie decimalen wijken de kansen enigszins af van de gewenste waarde 0,0500).

Wat is de betekenis van deze grenzen voor de steekproefuitkomsten? Er geldt een beroemde stelling die wij niet zullen bewijzen:

a. Bij alle aantallen $r < n$ is voor iedere (vaste) P de ongelijkheid

$$P \leqslant p_+$$

juist, behoudens in 5% van alle steekproeven van de omvang n uit deze populatie (en dus voor alle populaties). 0,05 heet de eenzijdige *onbetrouwbaarheidsdrempel.*

b. Als r tenminste één is geldt voor iedere P de ongelijkheid

$$p_- \leqslant P$$

behoudens in een fractie 0,05 van alle steekproeven van n.

c. Als r tenminste één is geldt voor iedere P:

$$p_- \leqslant P \leqslant p_+$$

behoudens in een fractie van 0,10 van alle steekproeven van n. Men noemt 0,10 de *tweezijdige* onbetrouwbaarheidsdrempel.

Het complement (d.w.z. het verschil met 1) van de genoemde onbetrouwbaarheidsdrempels heet de *betrouwbaarheid* van deze uitspraken. Deze is in de gevallen *a.* en *b.* dus 95% en in geval *c.* 90%.

In de gevallen *a.* en *b.* wordt over een eenzijdig (rechts resp. links) *betrouwbaarheidsinterval* gesproken voor de onbekende fractie P. In geval *c.* heeft men uiteraard een tweezijdig interval.

Om elk misverstand uit te sluiten zij opgemerkt (hoewel men deze opmerking in elk goed statistisch leerboek vindt) dat de uitspraken *a.* - *c.* *geen kansuitspraken* over de onbekende fractie P zijn. Het zijn kansoordelen in steekproevenruimten: bij iedere steekproef behoort een r en dus een p_+; voor alle $r > 0$ is er ook een p_-. De intervallen (p_-, p_+) zijn dus stochastisch, zij variëren over de steekproevenruimte; 90% van deze intervallen bevat P, bij 5% is P kleiner en bij 5% is P groter.

Omdat de begrippen betrouwbaarheid, -sgrens en -sinterval vrij gecom-

pliceerd zijn en in de discussie een rol hebben gespeeld, heb ik ze hier nog eens uiteengezet.

Volledigheidshalve zij opgemerkt dat men analoge definities kan geven voor andere waarden van de onbetrouwbaarheidsdrempels; eenmaal gekozen waarden dient men echter steeds te blijven gebruiken.

2.2.2. De benadering van Poisson

Daar zowel de kansverdelingen bij de binomiale verdeling veel parameters (n, r en P) bezitten als de betrouwbaarheidsgrenzen (n, r en de drempel) is het opstellen en gebruiken van tabellen voor deze verdeling geen sinecure. In gevallen waarin de normale benadering niet bruikbaar is omdat P te klein is, terwijl n wèl groot is kan men de *verdeling* van *Poisson* als benadering van de binomiale hanteren. Stel $nP = M$. Dan is de kans op r fouten (bij benadering) gegeven door

$$e^{-M} \qquad (e = 2{,}718) \qquad (r = 0)$$

$$e^{-M}\ \frac{M^r}{1.\,2.\,.r} \qquad (r = 1, 2, \ldots)$$

Als r/n klein is, bijv. $< 0{,}1$, dan zijn deze uitdrukkingen goede benaderingen van de binomiale kansen.

Analoog met de binomiale verdeling kan men bij de Poissonverdeling overschrijdingskansen vinden. Alles wordt veel overzichtelijker omdat alleen het produkt nP en niet n en P afzonderlijk bekend behoeven te zijn.

Onder de linker *betrouwbaarheidsgrens* m_- voor M verstaat men die waarde waarvoor de rechter overschrijdingskans van het gevonden aantal fouten r gelijk is aan 0,05. De formule hiervoor is gemakkelijk op te schrijven, evenals die voor de rechter betrouwbaarheidsgrens m_+ (die waarde voor M waarvoor r een linker overschrijdingskans 0,05 bezit). Ook nu heeft men betrouwbaarheidsintervallen $M \leqslant m_+$, $m_- \leqslant M$ en $m_- \leqslant M \leqslant m_+$ met drempels resp. 0,05; 0,05; 0,10. Als men de Poissonverdeling als benadering van de binomiale verdeling gebruikt, vindt men voor deze laatste benaderde grenzen $p_- = m_-/n$ en $p_+ = m_+/n$. Hier volgt een beknopte tabel, waarvan de tweede en derde kolom aan Kriens en Dekkers (1979) zijn ontleend.

Tabel 2.6. Betrouwbaarheidsgrenzen voor de Poissonverdeling

Aantal fouten r	Linker grens m_-	Rechter grens m_+	Overschrijdingskansen		Grenzen voor $n = 50$	
			$M = m_-$ (rechts)	$M = m_+$ (links)	p_-	p_+
0	–	3,000	–	0,0498	–	0,060
1	0,051	4,744	0,0488	0,0502	0,001	0,095
2	0,355	6,296	0,0512	0,0498	0,007	0,126
3	0,818	7,754	0,0503	0,0501	0,016	0,155
4	1,366	9,154	0,0504	0,0501	0,027	0,183
5	1,97	10,51			0,039	0,210
6	2,61	11,84			0,052	0,237

Vergelijking met tabel 2.5. leert dat p_- zeer goed klopt met de exacte binomiale grenzen en p_+ vrij goed. Bij grote waarden van n wordt de benadering precieser.

2.2.3. De hypergeometrische verdeling met benaderingen

In de praktijk trekt men uiteraard niet met teruglegging (vaak systematisch). De exacte verdeling van het aantal fouten r in een enkelvoudige aselecte steekproef *zonder* teruglegging is de *hypergeometrische*. Daarbij zijn de achtereenvolgende trekkingen enigermate afhankelijk. Stel dat zich in een populatie van N exemplaren $R = PN$ fouten bevinden. De kans (1) om bij eerste trekking een fout te vinden is ook nu $P\ (= R/N)$ en de kans (2) om geen fout te vinden $1 - P$. Daarna zijn er in geval (1) nog $(R - 1)$ fouten over en de voorwaardelijke kans op een fout bij de tweede trekking bedraagt $(R - 1)/(N - 1)$. In geval (2) zijn alle R fouten in de populatie achtergebleven zodat de voorwaardelijke kans op het vinden van een fout bij de tweede trekking gelijk is aan $R/(N - 1)$. Steeds hangen de kansen op het vinden van een fout van de uitkomsten van eerdere waarnemingen af. Zoals in Kriens en Dekkers (1979, blz. 172) wordt bewezen, is de kans op het vinden van r fouten:

$$\frac{R\,(R-1)\ldots(R-r+1)\,(N-R)\,(N-R-1)\ldots(N-R-n+r+1)\,n\,(n-1)\ldots(n-r+1)}{r\,(r-1)\ldots 2.\,1.\,N\,(N-1)\ldots(N-n+1)}$$

Berekenen, tabelleren en opzoeken is nog omslachtiger dan bij de binomiale verdeling; Cochran (1977, paragraaf 3.6.) noemt enkele tabellen. In de praktijk worden vier benaderingen gebruikt:

1. De *rechtstreekse* benadering met de *normale* verdeling. Men berekent

$$u = (r - nP) \,/\, \sqrt{n\,(N-n)\,P\,(1-P)\,/\,(N-1)}$$

en gebruikt voor de overschrijdingskansen de tabel van de standaardnormale verdeling. Hieruit kunnen omgekeerd betrouwbaarheidsgrenzen worden afgeleid. Cochran (*op. cit.*, table 3.3) geeft aan hoe dit ge-

schiedt en wanneer de uitkomsten nauwkeurig zijn; voor $r \leqslant 0,1n$ dient r minimaal 60 à 80 te zijn. In de accountantscontrole zal dit zelden voorkomen.
2. De benadering met *exacte binomiale* kansen resp. betrouwbaarheidsgrenzen. Deze is goed bruikbaar als $n \leqslant 0,1N$, terwijl r ook kleine waarden mag aannemen. In Muilwijk en Schouten (1960) zijn de genoemde grootheden getabelleerd tot $n = 100$; voor de berekening van betrouwbaarheidsgrenzen is gebruik gemaakt van Fisher and Yates (1957, table VIII1).
3. De benadering met de *normale* verdeling *via* de *binomiale*. Deze vindt men o.a. in Kriens en Dekkers (1979). Ook nu gelden de voorwaarden dat r niet te klein mag zijn, n middelgroot of groot (bijv. $\geqslant 50$) maar $\leqslant 0,1N$.
4. De benadering met de *Poissonverdeling* (paragraaf 2.2.2.). n moet groot zijn maar $\leqslant 0,1N$; $r \leqslant 20$ bijvoorbeeld.

2.2.4. Slotopmerkingen over schattings- en toetsingsprocedures
1. Bij een enkelvoudige aselecte trekking zonder teruglegging is de exacte verdeling, zoals gezegd, de hypergeometrische. Als regel kan daarop een benadering worden toegepast. In uitzonderingsgevallen moeten bij de kwaliteitszorg de correcties van Hamaker worden toegepast, zie Schaafsma en Willemze (1978, blz. 231).
2. Alle in paragraaf 2.2.3. genoemde verdelingen gelden voor enkelvoudige aselecte trekking; bij meervoudige en afgeknotte steekproefvoorschriften gelden deze verdelingen, de daarop gebaseerde betrouwbaarheidsgrenzen en toetsen niet. Bij niet-afgeknotte dubbele en meervoudige schema's gelden samengestelde binomiale (hypergeometrische) verdelingen.
3. Welke keus moet de accountant maken: keuren of schatten, en welke techniek? Zie hiervoor hoofdstuk 4.

3. Commentaar op de discussie tussen Vermaas en Blokdijk

3.1. Uitbreiden van de steekproef

3.1.1. Steekproefplannen volgens de „nieuwe" en de „oude" opvatting
In Vermaas (1979a, blz. 209) oppert de auteur - als men de door hem verworpen „nieuwe" opvatting zou willen volgen - een viervoudig steekproefvoorschrift. Ik geef dat eerst in algemene vorm weer. Men begint met een steekproef van n_1 stuks. De populatie wordt geaccepteerd als er $r_1 = 0$ fouten optreden; als $r_1 = 1$, 2 of 3 wordt de steekproef uitgebreid met n_2 exemplaren. Vindt men echter 4 of meer fouten in de eerste steekproef, dan wordt de populatie afgekeurd. Stel dat in de tweede steekproef r_2 fouten gevonden worden. Indien $r_1 + r_2 = 1$ (dus $r_1 = 1$, $r_2 = 0$) dan wordt de populatie aanvaard; in de gevallen $r_1 + r_2 = 2$ of 3 wordt de steekproef opnieuw uitgebreid met n_3 stuks. Op $r_1 + r_2 = 4$ of

meer volgt afkeuring. Zij r_3 het aantal fouten in de eventuele derde steekproef; $r_1 + r_2 + r_3 = 2$ leidt tot aanvaarding, 3 tot voortzetting en 4 tot afkeuring. Als er een vierde steekproef van n_4 exemplaren getrokken wordt leidt $r_4 = 0$ tot goedkeuren, $r_4 \geqslant 1$ tot verwerpen.

Voorbeeld 11: $n_1 = 96$,[10]) $n_2 = 51$, $n_3 = 35$ en $n_4 = 32$. Zie Appendix 5 voor de keuringskarakteristiek. Wij gaan uit van een grenskwaliteit voor de consument („grensgeval", beste slechte populatie) $P_1 = 0{,}04$ en een consumentenrisico (eenzijdige onbetrouwbaarheidsdrempel) $\beta_o = 0{,}05$. Het volgende schema vat dit steekproefvoorschrift samen.

Schema 3.1. Viervoudig steekproefvoorschrift (voorbeeld 11): goedkeuren

	Aantallen fouten in steekproef i				Totaal aantal fouten Σr_i	Totale steekproefomvang Σn_i
	r_1	r_2	r_3	r_4		
1	0	–	–	–	0	96
2	1	0	–	–	1	147
3	1	1	0	–	2	182
4	1	1	1	0	3	214
5	1	2	0	0	3	214
6	2	0	0	–	2	182
7	2	1	0	0	3	214
8	2	0	1	0	3	214
9	3	0	0	0	3	214

Op blz. 207 (*op. cit.*) geeft de heer Vermaas een ander meervoudig schema bij $P_1 = 0{,}04$:

Voorbeeld 12: [11]) $n_1 = 75$, $n_2 = 44$, $n_3 = 39$, $n_4 = 36$, $n_5 = 35$, $n_6 = 34$, $n_7 = 33$. Wij duiden de gecumuleerde aantallen fouten met R_i aan ($i = 1, \ldots, 7$): $R_1 = r_1$; $R_2 = R_1 + r_2$; $R_3 = R_2 + r_3$ enz. Goedkeuring vindt plaats als $R_1 = 0$, $R_2 = 1, \ldots, R_7 = R_6 + r_7 = 6$, uitbreiding als R_i ($i = 1, \ldots, 6$) groter dan de goedkeurwaarde ($i - 1$) maar kleiner dan 7 is, afkeuring als $R_i \geqslant 7$.

Hoe komt de auteur aan deze aantalllen n_i? Hij gaat uit van de rechter betrouwbaarheidsgrenzen m_+ bij de Poissonverdeling (tabel 2.6.) en stelt dan:

$n_1 P_1 = m_+ \ (r = 0)$; $n_1 = 3{,}000/0{,}04 = 75$

$(n_1 + n_2) P_1 = m_+ \ (r = 1)$; $n_1 + n_2 = 4{,}744/0{,}04 = 119$

$(n_1 + n_2 + n_3) P_1 = m_+ \ (r = 2)$; $n_1 + n_2 + n_3 = 6{,}296/0{,}04 = 157{,}4$

(afgerond 158) enz., waaruit n_2, n_3 enz. volgen.

[10]) Vermaas (1979a, blz. 209) geeft 98; de overige aantallen nam ik over (hoewel ik niet begrijp hoe hij n_2 enz. gevonden heeft). Bij $n_1 = 96$ is $S(P_1) = 0{,}04946$, bij 98: 0,04633 (binomiaal). Vermaas gaat uit van $\beta_o = 0{,}02$ voor de eerste steekproef.

[11]) Wij hebben zijn voorschrift aangevuld, zodat de beslissing in alle gevallen vaststaat.

Vermaas beroept zich nu op de „oude" opvatting om het bewijs te leveren dat de betrouwbaarheid van deze procedure 95% is. Daarbij interpreteert hij de betrouwbaarheidsgrenzen als volgt: als er r fouten in een steekproef van $(n_1 + \ldots + n_{r+1})$ gevonden worden, is er 5% kans dat $P > p_+ = 0{,}04$ is. Deze interpretatie wordt in Blokdijk (1979) en Blokdijk (1980) bestreden, waarna diens opponent in Vermaas (1980b) stelt dat het verschil tussen beide auteurs slechts op een woordenspel berust. Ik merkte in paragraaf 2.2.1. op dat de kansinterpretatie van betrouwbaarheidsgrenzen onjuist is. De heer Vermaas verdedigt zich (*op. cit.*) weliswaar door te zeggen dat zijn opvatting beter bij de gewone spreektaal aansluit, maar dat is geen reden om deze foutieve opvatting te gebruiken. In de beide artikelen van Blokdijk wordt naar mijn mening een afdoende argumentatie gegeven.

Als wij het - gecompliceerde - begrip betrouwbaarheid even terzijde laten, blijkt uit de *kansrekening* dat Vermaas een tweede fout maakt, die veel ernstiger is omdat hij daardoor tot onjuiste aanbevelingen voor de praktijk komt. In voorbeeld 11 bleek dat $S(P_1)$ voor dit viervoudige schema (vrijwel) de juiste waarde 0,05 aanneemt. In voorbeeld 12 is $S(P_1)$ aanzienlijk groter, zie Appendix 5.

Vermaas beroept zich verder op de samenhang tussen toetsen en schatten. Maar hij leidt uit de rechter betrouwbaarheidsgrens voor *enkelvoudige* steekproeven een toetsingsprocedure af voor *meervoudige*; dat kan natuurlijk niet, waarmee verklaard is hoe hij tot ten enen male onjuiste waarden voor de betrouwbaarheid komt. Ook hierop is in Blokdijk (1979) gewezen („iedere uitbreiding is in feite een herkansing").

In Vermaas (1979a, blz. 208) wordt voorbeeld 12 nog langs andere weg behandeld. Hij leidt in feite de goedkeurkans bij $n = 194$ en $c = 3$ af, en deze is 0,046 voor $P_1 = 0{,}04$. Zoals eerder gezegd is $\Sigma\, r_i / \Sigma\, n_i$ geen zuivere schatter van P.

Nogmaals: bij toetsen of keuren zijn dubbele, meervoudige e.d. systemen toelaatbaar mits men de juiste keuringskarakteristiek hanteert. Bij schatten is een meervoudig systeem toelaatbaar als de eerste steekproef niet wordt afgeknot, en P alleen uit die eerste steekproef wordt berekend.[12])

Een derde bezwaar tegen de dubbele en meervoudige steekproefvoorschriften van Vermaas is, dat bij hem $c_1 = 0$, $c_2 = 1$, $c_i = i - 1$ gecombineerd wordt met $n_1 > n_2 > n_3 > \ldots$ In Vermaas (1979b) wordt een dergelijk voorschrift aanbevolen:
Voorbeeld 13: $n_1 = 738$, $n_2 = 376$, $n_3 = 331$, $n_4 = 308$ enz. De bedoeling van de auteur is dat $P_1 = 0{,}005$ en $\beta_o = 0{,}025$ wordt, maar dat is ook hier niet het geval. In ieder geval is dit schema inefficiënt.

Opgemerkt zij nog dat de voorwaardelijke kans om in een volgende steekproef volgens de systemen van Vermaas nul fouten te vinden steeds toeneemt:

---

[12]) Omgekeerd zijn er ook trekkingsmethoden waarmee men - bij gebruik van de juiste formules - zuiver kan schatten, zoals getrapte steekproeven, maar waarvoor de toetsing als die mogelijk is zeker niet via de gebruikelijke methoden kan geschieden.

$$(1-P)^{n_1} < (1-P)^{n_2} < (1-P)^{n_3} < \ldots ; n_1 > n_2 > n_3 > \ldots$$

zodat men steeds gemakkelijker tot goedkeuren besluit. Dit is in strijd met een verantwoorde herkansing. Immers, als door een ongelukkig toeval in de eerste steekproef fouten gevonden zijn gaat men de norm voor goedkeuring verlagen!

3.1.2. Grensgeval en procesfunctie

In de toetsingstheorie en de daarmee equivalente keuringsprocedures wil men dat de kans op fouten van de tweede soort ten hoogste β_o (bijv. 0,05) bedraagt. Als men slechte populaties definieert als zulke waarvoor $P \geqslant P_1$ is, dan is aan deze eis voldaan: de kans op goedkeuren is daarvoor dan $\leqslant \beta_o$ (paragraaf 2.1.1.). Vermaas (1979a) laat eerst $S(P_1) = \beta_o$ gelden maar vindt deze eis als hij tot verwerping zou leiden te scherp en gaat dan de steekproef uitbreiden: niet alleen het *grensgeval* $P = P_1$. Dit is in strijd met de theorie van Neyman en Pearson; de door de heer Vermaas te hulp geroepen statistici - De Jonge en Wielinga (1973) en De Wolff (1973) - gaan van deze theorie uit, evenals Kriens en Dekkers (1979), Blokdijk (1979, 1980) en ik. Daaruit vloeien de begrippen betrouwbaarheidsinterval enz. voort.

In plaats van de door hem bestreden gedachtengang beschouwt Vermaas (*op. cit.* blz. 210) een a priori-verdeling van 1 mln te controleren populaties. Voor fracties fouten van 1% tot en met 7% berekent hij de goedkeurkansen (hoe hij dit doet is mij niet duidelijk geworden) en weegt deze met de frequenties. De foutenfracties en hun frequenties zijn in tabel 3.2. af te lezen.

Tabel 3.2. A priori-verdeling van 1 000 000 populaties volgens Vermaas (1979a)

Percentage fouten in de populatie	1	2	3	4	4,01	5	6	7	Totaal
Frequenties in %	10,81	21,62	35,14	8,785	8,785	9,46	4,05	1,35	100

Het gemiddelde is 3,1% fouten; 76,4% van alle populaties heeft ten hoogste 4% fouten, 23,6% 4,01% fouten of meer. Deze laatste moeten bij $P_1 = 0{,}04$ afgekeurd worden. Bij steekproefvoorschrift 12, echter niet verder voortgezet dan tot en met de derde steekproef, vindt Vermaas dat het aantal ten onrechte geaccepteerde populaties stijgt van 0,7% bij één steekproef tot 1% bij twee en 1,3% bij drie steekproeven. Triomfantelijk concludeert de auteur: „Het is duidelijk dat het uitbreiden van de steekproeven de betrouwbaarheidsgrens van 95% in het geheel niet aantast". Mijn commentaar luidt:

a. 95% is geen betrouwbaarheidsgrens maar een „betrouwbaarheid", i.c. $1 - \beta_o$.

b. Als de berekeningen correct zouden zijn (wat ik niet geloof) hebben zij niet de minste bewijskracht. Ik zie niet in dat deze a priori-verde-

ling met 3,1% fouten gemiddeld en 23,6% onaanvaardbare populaties een „reële verdeling" zou zijn.

In de kwaliteitszorg heten deze verdelingen „procescurven" (beter zou zijn *procesfuncties*, omdat er behalve continue ook discrete verdelingen - zoals hier ten tonele gevoerd - voorkomen). Hoewel er goede voorbeelden van zulk soort verdelingen uit de fabricage zijn - zie Wetherill (1969) - zijn er voor de accountantscontrole voor zover mij bekend geen empirische gegevens, waarop zij gebaseerd zouden kunnen zijn. Opgemerkt zij dat ik het niet eens ben met Blokdijk (1979, blz. 540) dat zo'n verdeling „volmaakt oninteressant" zou zijn omdat er maar weinig populaties bij de steekproefcontrole van één jaarrekening worden gecontroleerd. Het is immers een, hoewel kleine, steekproef uit de superpopulatie der populaties. Op den duur zou men gemiddeld wel een uit de procesfunctie te berekenen fractie der populaties ten onrechte niet afkeuren. Dit voorbehoud neemt echter niet weg dat ik geen kans zie zo'n fractie „realistisch" te berekenen. Ik acht dan ook Vermaas' betoog in dezen verre van overtuigend.

Een laatste opmerking: Het foutieve gebruik dat Vermaas van de Poisson-tabellen maakt berust vermoedelijk op het niet onderkennen van het feit dat deze in wezen *binomiale* (benaderde) tabellen zijn, waarbij de steekproef enkelvoudig is, zodat er met één vaste steekproefgrootte gewerkt dient te worden. Zij gelden niet bij uitbreiding, zoals reeds vaker werd gesteld.

3.2. Goed- of afkeuren: welke „criteria"?

De door Vermaas gekritiseerde publikaties Kriens en Dekkers (1979) en Blokdijk (1979) gaan uit van enkelvoudige steekproefschema's met $c = 0$ of van meervoudige met $c_1 = 0$. Waarom?

3.2.1. De kritiek van Vermaas bij onderzoek naar ernstige fouten

Kriens en Dekkers (*op. cit.* blz. 188) toetsen bij controle op ernstige fouten de nulhypothese $P = P_2 = 0$ tegen de alternatieve hypothese $P > 0$. Zij hanteren een enkelvoudig voorschrift met criterium $c = 0$ (blz. 28, 53) en motiveren dit o.m. door te stellen dat één ernstige fout tot afkeuring dient te leiden. Wij komen hierop in het vervolg terug; redeneert men alléén statistisch, dan is daar niets op aan te merken.

Vermaas (1979b, o.a. blz. 527) vindt zo'n schema onlogisch: er is bij goedkeuren tolerantie, maar niet bij afkeuren. Hij illustreert dit op blz. 532-534 met het volgende steekproefschema.

Zij $P_1 = 0{,}005$ en $P_2 = 0$; stel de eis dat $S(P_1) = \beta_o = 0{,}025$. Dan volgt uit de betrouwbaarheidsgrenzen voor de Poisson-verdeling dat voor $c = 0$ de steekproefomvang 738 moet bedragen. Bij nul fouten wordt de populatie goedgekeurd. Bij $r = 1, 2, \ldots$ beschouwt de auteur de onderste betrouwbaarheidsgrens $p_- = m_-/738$. Deze loopt op van 0,000034 bij $r = 1$ tot 0,0047 bij $r = 8$ en 0,0056 bij $r = 9$. Aangezien de laatste waarde groter dan P_1 is, wordt de populatie bij 9 fouten afgekeurd, maar bij

1, 2, . . . , 8 fouten volgt tenminste eenmaal uitbreiding.

Als na een uitbreiding de bovenste betrouwbaarheidsgrens p_+ gelijk aan P_1 is, vindt goedkeuring plaats. Ook hier is het risico van de procedure veel groter dan het aanvaarde bedrag β_o (vgl. paragraaf 3.1.). Van belang is echter nog welke rol de tolerantiebegrippen spelen in het betoog.

Kriens en Dekkers (1979, o.a. blz. 28) noemen (a) bij *toetsing* van $P = P_2$ tegen $P = P_1$ het verschil $P_1 - P_2$ de tolerantie. Die slaat dus op goed- èn afkeuren. Het kiezen van $P_2 = 0$ verandert daar niets aan. Bij *schatting* definiëren zij de halve breedte van het betrouwbaarheidsinterval, $(p_+ - p_-)/2$, als de onnauwkeurigheid van de uitkomst (*op. cit.* blz. 32) en merken op dat deze overeenkomt met (a); in de paragrafen 2.4. en 2.5. gaan zij uitvoerig op de verschillen in. Derhalve hebben wij hier (b), de tolerantie bij schatten = de halve breedte van het interval (men kan uiteraard ook de breedte zelf nemen).

Wat doet Vermaas nu? Hij stelt voor goedkeuring de eis $p_+ = P_1$ en kiest n_1 zodanig dat voor $r = 0$ hieraan voldaan is. $(n_1 + n_2)$ wordt gevonden door te eisen dat bij een enkelvoudig (!) schema voor $r = 1$ deze relatie geldt(enz.). Dus (c), tolerantie bij *goedkeuring*: $p_+ = P_1$. Verder voert hij in (d), de tolerantie bij *afkeuring*: $p_- \geqslant P_1$. Tenslotte oppert hij de mogelijkheid dat $(p_+ - p_-)$ „dicht genoeg" tot P_1 nadert.

Samenvattend: de heer Vermaas opereert (vooral) met de twee nieuwe begrippen (c) en (d) en bestempelt die als toleranties. Hij gebruikt de statistische taal hier onzorgvuldig.

Een tweede opmerking van Vermaas (blz. 530) snijdt m.i. meer hout: waarom wordt er met gelijke kansen getrokken? Stel dat men een *scherpe* scheiding kan maken tussen ernstige en niet-ernstige fouten; als men bovendien weet dat er N_1 grote posten in de populatie zijn waarbij deze ernstige fouten aanwezig kunnen zijn, terwijl er N_2 kleine posten zijn die hoogstens niet-ernstige fouten kunnen bevatten, dan komt de volgende stratificatie in aanmerking, zie De Wolff (1956): controleer de grote posten volledig (trekkingskans dus = 1) en de kleine steekproefsgewijs (trekkingskans bijv. 5%). In het bedoelde artikel wordt aangegeven waar men de grens tussen klein en groot moet leggen, en welke steekproeffractie bij de kleine posten moet worden gehanteerd.

Nog verder gaat men op deze weg als men bijv. zeer ernstige, vrij ernstige en niet-ernstige fouten kan onderscheiden. De Wolff (1959) neemt voor controle op fraude aan dat men een meer gedetailleerde onderscheiding in de populatie kan aanbrengen, zodat men naast volledige controle van grote posten nog twee of meer steekproeffracties (< 1) kan toepassen.

Men vergelijke echter ook de kritiek van Kriens (1963, 1968). De uiterste consequentie van het gebruik van uiteenlopende steekproeffracties is het werken met trekkingskansen evenredig aan de omvang der afzonderlijke posten. Verwant hiermee is de „guldensrangnummermethode", afkomstig van Van Heerden, en de „monetary unit sampling". Dan gaat het echter over een populatie van *guldens* (c.q. dollars); daarbij en

ook bij trekking van *posten* met kansen evenredig aan het aantal guldens mag men de theorie voor trekking van posten met gelijke kansen niet toepassen. Als men posten met ongelijke kansen trekt levert weging met het omgekeerde van de trekkingskansen een zuivere schatter op, zie bijv. Moors en Muilwijk (1975, hoofdstuk 5).

3.2.2. Over hypothesen en criteria

Vermaas (1979b) wil wel (onmiddellijk) goedkeuren als het aantal fouten in de eerste steekproef van n_1 posten nul is, maar hij wil niet afkeuren als $r_1 = 1$ of 2. Kriens en Dekkers (1979, blz. 189) concentreren zich op fouten van de tweede soort, het niet verwerpen van $P = P_2 = 0$ als de alternatieve hypothese $P > 0$ juist is. Omdat zij (bij een enkelvoudig schema) uitgaan van een kritiek gebied $r/n \geqslant 1/n$, dus $r \geqslant 1$ of $c = 0$, kan zich zo'n fout alleen maar voordoen als $r = 0$.

Ik heb in paragraaf 2.1.5. een andere aanpak gevolgd. Bij onderzoek naar ernstige fouten stel ik P_2 (fractie fouten in de slechtste goede populatie) wel klein maar niet nul, zodat α_o eveneens niet nul (maar wel klein) is. Ik toets dan de nulhypothese $P = P_2$ tegen de alternatieve hypothese $P = P_1$ (in de beste slechte populatie).[13]) Verder vind ik dat de keuringskarakteristiek bij $c = 0$, $S(P) = (1 - P)^n$, een minder gunstig verloop heeft, de steilheid h_i is dan te klein. Meer hierover vindt men in hoofdstuk 4; het gaat erom de fouten van de eerste en van de tweede soort op de beste manier tegen elkaar af te wegen, en dan is het de vraag of afkeuren bij één of meer fouten wel optimaal is. Als ik Vermaas goed begrijp is voor hem een belangrijk bezwaar gericht tegen het absoluut stellen van het begrip ernstige fout; een genuanceerder standpunt in dezen verdient zeker overweging, zoals ik ook in de vorige paragraaf bepleitte.

3.3. Betrouwbaarheid

In Vermaas (1980a) wordt het voorbeeld besproken van Tuitjer en Zuijdervliet (1975, blz. 492) waarbij $P_1 = 0{,}01$ en $\beta_o = 0{,}01$; $n = 1005$ en $c = 3$, zodat $S(P_1) = 0{,}010$. Stel nu, aldus Vermaas, dat $r = 0$. Wanneer dan $P > 0{,}01$ zou zijn is $S(P) < 0{,}00041$ en *dus* is de betrouwbaarheid $1 - S(P) > 0{,}99959$ en niet $1 - P_1 = 0{,}99$.

Deze redenering is door Blokdijk (1980) weerlegd en ik kan daarom kort zijn. De betrouwbaarheid is een eigenschap van de procedure, onafhankelijk van de steekproefuitkomst. Anders geformuleerd: P en het steekproefvoorschrift bepalen samen de kansverdeling in de steekproevenruimte. Er is bij iedere P een kans op $r = 0$, 1, 2 of 3 en dan volgt goedkeuren; bij $P = P_1$ is *die* kans naar behoren 1,0%, bij $P < P_1$ groter dan 1,0% en bij $P > P_1$ kleiner. Het gaat niet aan te werken met alleen de kans onder de voorwaarde dat $r = 0$. Men kan de betrouwbaarheid (anders dan P) niet uit de steekproef schatten.

---

[13]) In paragraaf 4.2.3. wordt een m.i. nog betere mogelijkheid genoemd.

4. Hoe dan wel?

Wij hebben enkele methoden voor de accountantscontrole en de kritiek van Vermaas daarop de revue laten passeren. Het entameren van een aantal problemen is te waarderen, maar zijn aanbevelingen kunnen meestal niet als verbeteringen worden beschouwd. Welke conclusies kunnen nu voor de praktijk worden getrokken? Kan het beter en zo ja, hoe?

4.1. Probleemstelling

Men kan de volgende keuzemogelijkheden onderscheiden:

a. Moet men keuren, toetsen of schatten? Wanneer is het laatste aan te bevelen?

b. Als men voor keuring of toetsing kiest rijst de vraag welke technieken men moet gebruiken. Moet men daarbij uitgaan van te toetsen hypothesen of kan men andere grootheden gebruiken voor de bepaling van een techniek?

c. In hoeverre is bij *b.* het onderscheid tussen ernstige en niet-ernstige fouten relevant? Moet men uitgaan van een geleidelijke overgang tussen beide categorieën?

Geheel uitgewerkte oplossingen kunnen in het bestek van dit artikel niet worden aangedragen. Daartoe moet er nog meer discussie en studie plaatsvinden. Sommige vraagstukken en methoden komen niet aan bod; ik volsta met het signaleren daarvan.

4.2. Keuren, toetsen of schatten?

4.2.1. Schatten van P

Als men ervan afziet tevoren normen op te stellen waaraan de populatie moet voldoen is het schatten van foutenpercentages aangewezen. Dat zal bij niet-ernstige fouten eerder het geval zijn dan bij ernstige. Als men de fractie *ernstige* fouten wil schatten dan moet men zich realiseren dat deze fractie redelijkerwijs klein geacht kan worden. De linker betrouwbaarheidsgrens p_- ligt dan dicht bij nul (als $r = 0$ is die grens er niet) en is zelden van belang. Grote betekenis heeft de rechter grens p_+; hoe nauwkeurig is die?

Wij beschouwen ter illustratie enkelvoudige steekproeven van 1000 en vergelijken in tabel 4.1. de schatting r/n met p_+; in kolom 4 wordt het quotiënt van bovengrens en schatting gegeven. Wij gebruiken de Poissonverdeling, zodat bijv. voor $n = 100$ dezelfde (benaderde) verhoudingen gelden.

Tabel 4.1. Schatting en rechter betrouwbaarheidsgrens bij enkelvoudige steekproeven; $n = 1000$; eenzijdige betrouwbaarheid 97,5%

Aantal fouten r	Schatting r/n	Bovengrens p_+	Quotiënt np_+/r
0	0	0,0037	∞
1	0,001	0,0056	5,6
2	0,002	0,0072	3,6
3	0,003	0,0088	2,9
4	0,004	0,0102	2,6
5	0,005	0,0117	2,3
10	0,01	0,0184	1,8
15	0,015	0,0247	1,6
20	0,02	0,0309	1,5
25	0,025	0,0369	1,5
30	0,03	0,0428	1,4
40	0,04	0,0545	1,4
50	0,05	0,0659	1,3

Zoals te verwachten was daalt de verhouding tussen bovengrens en schatting bij toenemend aantal fouten. Voor $r \leqslant 5$ (ook voor $r = 6$ of 7) is de betrouwbaarheidsgrens meer dan tweemaal zo groot als de schatting; pas bij $r \geqslant 24$ is het quotiënt minder dan 1,5.

Het is duidelijk dat zowel de schatting als de rechter betrouwbaarheidsgrens sterk van het toeval afhankelijk zijn, hoe kleiner r, des te sterker. Indien men genoegen neemt met zulke betrekkelijk ruime grenzen, dan komt schatten bij ernstige fouten in aanmerking. Men kan echter de schatting ook als een *bijprodukt* van de keuring meenemen; zoals gezegd mag men dan een enkelvoudige steekproef c.q. de eerste van een meervoudige niet afknotten.

Bij *niet-ernstige* fouten kan het geconstateerde percentage hoger zijn, zodat de bovengrens daar relatief dichter bij ligt. In dat geval komt keuring alleen in aanmerking als men bereid is de populatie af te keuren bij een te hoge fractie fouten in de steekproef.

4.2.2. Toetsen van hypothesen over P

Kriens en Dekkers (1979, hoofdstuk 15) onderscheiden enkele gevallen waarvan wij de volgende beschouwen:

1. Nulhypothese $P = P_2 > 0$, alternatieve hypothese $P > P_2$;
2. $P \leqslant P_2$ resp. $P \geqslant P_1$;
3. $P = P_2 = 0$ resp. $P > 0$.

Het kiezen van een bepaald stel hypothesen is in de eerste plaats de verantwoordelijkheid van de accountant. Wij komen later op 1. en 3. terug en concentreren ons nu op 2.

In de kwaliteitszorg gaat men vaak van dit tweetal hypothesen uit, waarbij P_2 als de grenskwaliteit voor de producent en P_1 als die voor de

consument wordt geïnterpreteerd. De keuze van de waarden van P_2, P_1, α_o en β_o - zodanig dat $S(P_2) = 1 - \alpha_o$ en $S(P_1) = \beta_o$ geldt - is eveneens primair een aangelegenheid van de accountant c.q. de bedrijfsleiding (al zal de statisticus er het zijne toe kunnen bijdragen), de keuze van de techniek is vooral een zaak van de statisticus (waarbij de accountant natuurlijk ook meespreekt).

4.2.3. Keuren van populaties

Dit beschouw ik als de belangrijkste aanpak. Nu is er op de in paragraaf 4.2.2. vermelde gedachtengang een fundamentele kritiek geleverd door Wetherill (1969), die ik thans weergeef, vgl. ook Muilwijk (1977, blz. 135 en 141): Als men een keuringskarakteristiek wil bepalen door P_2, P_1, α_o en β_o vast te leggen, geeft men blijk van een zekere bewustzijnsvernauwing: het gecontroleerde bedrijf is alleen maar geïnteresseerd in het aanvaarden van administraties waarvoor $P \leqslant P_2$ met een kans die groter is dan bijv. 95%, de controleur let alleen maar op het grensgeval en wil bij $P \geqslant P_1$ afkeuren met een eveneens grote kans. Dergelijke opmerkingen zijn ook in de artikelen van Vermaas te vinden.

Stel nu dat men deze vier grootheden vastgelegd heeft en dat dit leidt tot een zeer grote steekproefomvang. De neiging zal dan opkomen om één of meer van de gekozen getallen - fracties of kansen - te wijzigen tot men bijv. bij een enkelvoudig voorschrift een n en een c vindt die als „redelijk" worden beschouwd. In feite is men dan bezig de *keuringskosten* af te wegen tegen de *kosten van verkeerde beslissingen*, echter op een niet-geformaliseerde manier. Tot zover de kritiek.

Wetherill past nu een besliskundige methode toe en voert allereerst het begrip *evenwichtskwaliteit* in, dat is een zodanige waarde P_3 voor de fractie fouten dat het verlies bij goed- of afkeuren even groot is. Stelt men k_1 de opbrengst van het aanvaarden van de administratie per goede post en k_2 $(> k_1)$ het verlies per foute post bij goedkeuring, dan voldoet zoals hierna blijkt de evenwichtskwaliteit aan

$$P_3 = \frac{k_1}{k_1 + k_2}.$$

De juistheid van een beslissing wordt dus als een bedrijfseconomisch probleem beschouwd. Het aantal goede posten is $N(1-P)$ en het aantal foute NP. Als men goedkeurt is de netto-opbrengst dus
$N(1-P)k_1 - NPk_2 = Nk_1 - NP(k_1 + k_2) = N(k_1 + k_2)(P_3 - P)$ (dus nul voor $P = P_3$). Voor $P > P_3$ is de opbrengst negatief en het verlies $N(k_1 + k_2)(P - P_3) > 0$; bij $P < P_3$ is de beslissing juist en het verlies nul. Bij afkeuring is de netto-opbrengst:
$NPk_2 - N(1-P)k_1 = N(k_1 + k_2)(P - P_3)$; bij $P > P_3$ is de beslissing juist en het verlies nul, bij $P < P_3$ bedraagt het verlies
$N(k_1 + k_2)(P_3 - P)$. De „verliesmatrix" vat dit samen:

Fractie foute posten P / Beslissing	$P < P_3$	$P > P_3$
Goedkeuren	0	$N(k_1 + k_2)(P - P_3)$
Afkeuren	$N(k_1 + k_2)(P_3 - P)$	0

Vervolgens beschouwt Wetherill de procesfunctie, die de verdeling van P over alle administraties beschrijft (vgl. paragraaf 3.1.2.). Als men daarvoor de bèta-verdeling kiest, zie bijv. Rijken van Olst (1974, paragraaf 3.10.) en men kent de kosten k_3 per gecontroleerde post, dan is het mogelijk de optimale n en c te vinden.

Hoe relevant is deze fraaie theoretische analyse voor de accountantspraktijk? De optima zijn, aldus Wetherill, tamelijk robuust voor de vorm van de procesfunctie, maar zeer sterk afhankelijk van k_1/k_3 en k_2/k_3, in het bijzonder van P_3. Dit laatste geldt ook voor de minimax-methode, zie bijv. Kriens, Göbel en Molenaar (1968, hoofdstuk I), waarbij de procesfunctie geen rol speelt. Als men P_3 redelijk kan ramen zijn er dus wel perspectieven.

Het begrip evenwichtskwaliteit doet denken aan het *controlepunt* P_i (men gaat gemakkelijk na dat P_3 tussen P_2 en P_1 gelegen moet zijn bij redelijke keuze van P_1 en P_2; dezelfde eigenschap heeft P_i). De definities zijn verschillend maar er is wel analogie: $S(P_i) = 0{,}5$ - de kansen op goed- of afkeuren zijn gelijk bij $P = P_i$ - terwijl $P_3 = k_1/(k_1 + k_2)$ een zodanige kwaliteit is dat het bedrijfseconomisch „indifferent" is of men goedkeurt of niet. Men kan zelfs stellen dat P_i een soort operationalisatie van P_3 is, omdat zij vaak dicht bij elkaar liggen. Dit blijkt als volgt.

Stel dat men een optimaal enkelvoudig schema bepaald heeft, dan geldt volgens Wetherill:

$$c = P_3(n + A + B + 2) - A - 1,$$

waarin $(A + 1)$ en $(B + 1)$ de twee parameters zijn die de vorm van de bèta-verdeling bepalen.[14]) Voor negatieve A en B is deze verdeling U-vormig en dat zou betekenen dat zeer goede en zeer slechte populaties vaker voorkomen dan minder extreme. Wij nemen A en B dus positief aan en verder A zeer klein ten opzichte van B, maar B niet al te groot, bijv. kleiner dan 10.[15])

Anderzijds geldt, vgl. Appendix 4, voor het controlepunt:

$$c = nP_i - 0{,}67$$

zodat

$$P_i - P_3 = P_3 \frac{A + B + 2}{n} - \frac{A + 0{,}33}{n}.$$

[14]) In de literatuur vindt men α in plaats van $(A + 1)$ en β voor $(B + 1)$; deze symbolen zouden echter verwarring met α_o en β_o kunnen teweegbrengen.

[15]) Het maximum van de procesfunctie ligt bij $P = A/B$ en dit zal in de praktijk een kleine fractie zijn. Als B echter groot is zou de top van de verdeling hoog zijn, wat evenmin realistisch is.

Het rechterlid ligt onder de genoemde veronderstellingen bij grote n dicht bij nul (het kan positief of negatief zijn) zodat P_i en P_3 elkaar dan niet veel ontlopen. Zo gezien is de keus van P_i en de *steilheid* h_i (paragraaf 2.1.4.) het overwegen waard; in ieder geval is daarop de bovenvermelde kritiek van Wetherill niet van toepassing.

Wij keren nu terug op de drie gevallen uit paragraaf 4.2.2. Als men P_2 wil specificeren kan men daar nul voor kiezen; daar is iets voor te zeggen, want welke andere geschikte waarde moet men nemen: 0,1%, 0,01% of 0,001%? En wat kan men zich voorstellen bij een kans op verwerpen voor $P \leqslant P_2$ die bijv. hoogstens 5% bedraagt? Kriens en Dekkers kiezen $P_2 = 0$ en toetsen 3. de nulhypothese $P = 0$ tegen de alternatieve hypothese $P > 0$. Onder deze strikte nulhypothese kan er echter geen fout optreden en het is dan ook volkomen logisch af te keuren bij $r \geqslant 1$; c is dan nul.

Wat gebeurt er als men bij $c = 0$ de steekproefomvang laat toenemen? De keuringskarakteristiek komt in haar geheel lager te liggen en het controlepunt P_i schuift naar links, vgl. figuur 2.1. Tegelijkertijd neemt de helling in het controlepunt toe, maar het produkt $P_i S'(P_i) = -h_i/2$ blijft vrijwel constant.[16])

Bij $c > 0$ is de steilheid groter[16]); dit betekent dat de helling groter wordt bij gelijk controlepunt.

Tabel 4.2. Steilheid en helling in het controlepunt bij enkelvoudige schema's met $P_i = 0{,}0167$

n	c	h_i	$S'(P_i)$
41	0	0,70	− 21
100	1	1,06	− 32
160	2	1,33	− 40
220	3	1,54	− 46
280	4	1,74	− 52

De grafiek van de keuringskarakteristiek vertoont bij $c > 0$ de omgekeerde S-vorm (vgl. figuur 2.2.). Voor goede populaties is de goedkeurkans dan groot, voor slechte klein. Voor $c = 0$ wordt de helling bij toenemende P (afgezien van het minteken) voortdurend kleiner en dit brengt met zich mee dat de kromme bij grotere P altijd vlakker verloopt. Door $c > 0$ te kiezen heeft men meer keuzemogelijkheden.

Daarom vind ik het stel hypothesen 3., waarbij $c = 0$, minder geschikt voor de praktijk.

---

[16]) Zie Appendix 4: $h_i = n\,(2^{1/n} - 1)$. Voor $n = 41$ levert deze formule 0,699, voor $n = 100$: 0,696 en voor $n = 200$: 0,694. Bij grote n heeft men de benadering $h_i = \sqrt{(2c + 1{,}46)/\pi}$; in tabelvorm (benaderd):

c	0	1	2	3	4
h_i	0,682	1,05	1,32	1,54	1,74

Dus is h_i bij grote n (vrijwel) onafhankelijk van de steekproefomvang en alleen door het criterium c bepaald.

In de gevallen 1. en 2. staat men zoals gezegd voor het probleem hoe klein men de foutenfractie P_2 moet kiezen. Dat pleit ervoor *de keuringskarakteristieken bij accountantscontrole niet via twee hypothesen vast te leggen* maar op een andere manier.

De door Schaafsma en Willemze (1978) aanbevolen specificatie van P_i en h_i heeft anderzijds het bezwaar dat h_i voor accountantscontrole vrij moeilijk te interpreteren is, wat een doelmatige keuze in de weg staat. Daarom beveel ik een (voorzover ik weet) nog niet in de literatuur vermelde combinatie aan, waarop voetnoot 7 in paragraaf 2.1.4. doelde: men basere zich op het *controlepunt* P_i, de *beste slechte populatie* (fractie fouten P_1) en het *risico* β_o dat men zo'n populatie accepteert. Dat heeft de volgende voordelen:

a. P_i is een gemakkelijk te begrijpen waarde voor P;

b. P_1 en β_o zijn reeds ingeburgerd in de accountantspraktijk.

Het enige nadeel van het drietal P_i, P_1 en β_o is dat de berekeningen iets lastiger zijn dan wanneer men uitgaat van P_i en h_i. Men moet immers behalve van

$$nP_i = c + 0{,}67$$

gebruik maken van

$$S(P_1) = (1 - P_1)^n + nP_1(1 - P_1)^{n-1} + \ldots$$

$$+ \frac{n(n-1)\ldots(n-c+1)}{1.2.\ldots.c} P_1^c (1 - P_1)^{n-c} = \beta_o.$$

Maar dit is altijd nog eenvoudiger dan wanneer men in plaats van de eerste formule gebruik maakt van P_2 en α_o, waarbij men een formule hanteert die ook $(c + 1)$ termen omvat. Bovendien heeft $S(P_1)$ het voordeel dat de termen een stijgende rij vormen[17]) zodat men kan beginnen met de laatste term en daarna doorgaan met de voorlaatste enz. Bij toepassing van een eenvoudige electronische calculator verloopt het rekenproces zonder moeite. Voor β_o kunnen de waarden 0,5%, 1%, 2,5%, 5% en 10% gekozen worden, *al naar gelang van de ernst van de te controleren fouten.*

4.2.4. Meervoudige steekproefschema's

Wij gingen in de vorige paragraaf uit van een enkelvoudig steekproefvoorschrift. Zoals in paragraaf 2.1.3. werd uiteengezet besparen dubbele schema's voor goede populaties gemiddeld 30% of meer van de steekproefomvang en als $P = P_i$, dus bij een middelmatige fractie fouten, nog altijd 15%. Voor grote foutenpercentages zagen wij in tabel 2.4. dat de verwachte

---

[17]) Bewijs: Deel de kans op r fouten door die op $(r-1)$ fouten. Er komt dan $P_1(n-r+1)/r(1-P_1)$, hetgeen een dalende functie van r is. Voor $r > (n+1)P_1$ wordt deze verhouding kleiner dan 1 en dan dalen de termen. Maar $(n+1)P_1$ is groter dan c, want $(n+1)P_1 > nP_i - 0{,}67 = c$ omdat $P_1 > P_i$ is. Dus is de laatste term van $S(P_1)$ nog niet het maximum en alle voorafgaande termen zijn kleiner.

steekproefgrootte weer daalt, omdat de eerste steekproef steeds vaker uitsluitsel geeft zodat de tweede steekproef achterwege blijft. Maar dan moet wel voldaan zijn aan de eisen dat n_2 twee à drie maal zo groot is als n_1 en dat $c_2 > c_1 (n_1 + n_2)/n_1$ is. Voor $n_2 = 2n_1$ betekent dit $c_2 > 3c_1$ en voor $n_2 = 3n_1$: $c_2 > 4c_1$.

Drievoudige, viervoudige enz. steekproefvoorschriften geven nog grotere besparingen: 25% bij $P = P_i$, 45% als $P = P_2$. Ik vermoed dat men zich wel kan beperken tot drievoudige schema's. De ontwikkeling daarvan voor accountantscontroles vereist echter nog een behoorlijk rekenwerk, waaraan wel een computer te pas zal moeten komen. Opgemerkt zij nog dat dubbele en drievoudige schema's vooral van belang zijn bij grote populaties.

Afknotten van de tweede en derde steekproef (liever niet van de eerste, omdat men dan P niet kan schatten) komt zeker in aanmerking; met name bij slechte populaties geeft dit nog meer besparing.

4.3. Wat niet aan de orde kwam

1. Het trekken met ongelijke kansen, in het bijzonder van posten met kansen evenredig aan de grootte, verdient bepaald overweging. Vgl. paragraaf 3.2.1.
2. Sequente steekproeven kunnen onder omstandigheden van belang zijn; zie de in paragraaf 2.1.3. genoemde literatuur.
3. Wij hebben vrijwel steeds de binomiale of de Poissonverdeling gebruikt. Als $n > 0{,}1N$ is moet men de hypergeometrische verdeling hanteren. Zie paragraaf 2.2.3. en Schaafsma en Willemze (1978, blz. 231).
4. Een graduele overgang van ernstige naar niet-ernstige fouten kan soms worden aangenomen. Dit kan zoals gezegd tot gelding gebracht worden in de keuze van een steekproefschema - hoe ernstiger een fout is, des te kleiner zal de kans daarop in het algemeen zijn - maar ook door stratificatie, vgl. nogmaals paragraaf 3.2.1.
5. Wij beperkten ons tot de tweedeling: goed of fout. Bij andere controleproblemen kan men normale of scheve verdelingen van een kwantitatieve eigenschap aannemen. Zie Kriens en Dekkers (1979) of Tuitjer en Zuijdervliet (1975).

Appendix 1 *Terminologie en notatie*

Term	Notatie		
	Dit artikel	Kriens en Dekkers	Andere publikaties
Producentenrisico, onbetrouwbaarheidsdrempel (eenzijdig)	α_o	α_o *)	α
Consumentenrisico, maximaal toegestane kans op het accepteren van een slechte populatie	β_o	β_o *)	β
Criterium (enkelvoudig)	c	k_o	
Idem eerste steekproef	c_1		
Steilheid	h_i		h_o
Gemiddelde van een Poissonverdeling	M	λ	μ, m
Omvang van een afgeknotte steekproef	m		
Idem van een populatie	N	N	
Idem van een steekproef (enkelvoudig)	n	n	
Idem van steekproef nr. 1 (2, . . .)	$n_1(n_2, \ldots)$		
Fractie fouten in de populatie	$P = R/N$	p	
Idem in de steekproef (enkelvoudig)	$p = r/n$	f	
Linker (rechter) betrouwbaarheidsgrens	p_- (p_+)	p_* (p^*)	
Idem voor gemiddelde (van Poissonverdeling)	m_- (m_+)	μ_* (μ^*)	
Beste slechte P, grenskwaliteit voor de consument	P_1	p_1 **)	p_t, θ_2
Slechtste goede P, grenskwaliteit voor de producent	P_2	p_o ***)	p_a, θ_1
Evenwichtskwaliteit	P_3		
Controlepunt	P_i		p_o, θ_o
Aantal fouten in de populatie	R	K	
Cumulatief idem (meervoudige steekproef)	$R_1, R_2, \ldots$		
Idem in de steekproef (enkelvoudig)	r	k	
Idem in steekproef nr. 1 (2, . . .)	r_1 $(r_2, \ldots)$		
Aantal goede exemplaren in een enkelvoudige afgeknotte steekproef	s		
Keuringskarakteristiek	$S(P)$	$Q(p)$	

*) Enigszins andere definities
**) Alternatieve hypothese
***) Nulhypothese

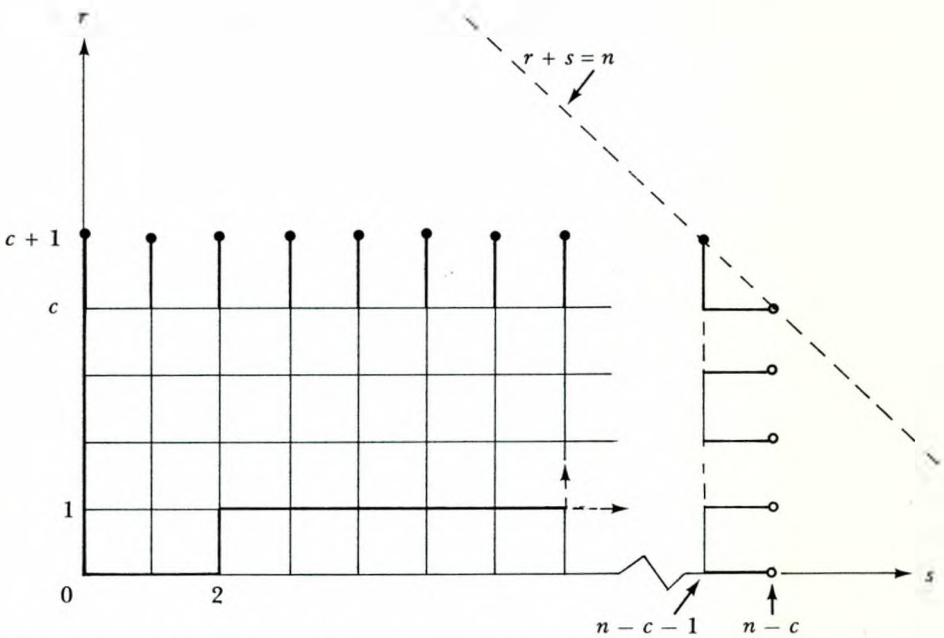
Appendix 2 *Afgeknot (verkort) steekproefvoorschrift*

In plaats van het enkelvoudige voorschrift, waarbij n posten gecontroleerd worden met als criterium ten hoogste c fouten, beschouwen wij nu een afgeknot schema dat wij *pseudo-enkelvoudig* noemen: er worden *maximaal* n waarnemingen verricht, maar men houdt op zodra er $(c + 1)$ fouten gevonden zijn - waarna afkeuring volgt - dan wel $(n - c)$ correcte posten aangetroffen zijn - er vindt dan goedkeuring plaats, omdat er niet meer dan c fouten kunnen zijn in de steekproef. De mogelijke uitkomsten zijn dus:

Tabel A2.1. Afgeknot pseudo-enkelvoudig schema

Aantal goede s	Aantal foute r	Resultaat	Aantal waarnemingen m
0	$c + 1$	afkeuren	$c + 1$
1	$c + 1$	afkeuren	$c + 2$
...	$c + 1$	afkeuren	...
$n - c - 1$	$c + 1$	afkeuren	n
$n - c$	0	goedkeuren	$n - c$
$n - c$	1	goedkeuren	n − c + 1
$n - c$	...	goedkeuren	...
$n - c$	$c - 1$	goedkeuren	$n - 1$
$n - c$	c	goedkeuren	n

Men kan dit eveneens overzichtelijk voorstellen in een z.g. *roosterdiagram* dat de mogelijke *toevalspaden* (random walks) laat zien.



Figuur A2.1. Toevalspaden bij afgeknot pseudo-enkelvoudig schema

Verticaal leest men het aantal fouten r af, horizontaal het aantal goede exemplaren s. Als de waarnemingen bijv. achtereenvolgens g(oed)gf(out)ggggg . . . zijn, verloopt het pad vanaf de oorsprong tweemaal naar rechts, eenmaal omhoog, vijfmaal naar rechts . . . (dikke lijn in figuur A2.1.). De paden eindigen hetzij in de stippen, waarbij de laatste waarneming een fout is, hetzij in de cirkeltjes, waarbij de laatste waarneming een goed exemplaar betreft; deze mogelijke einden zijn ook aangegeven.

Alle toevalspaden die bij *vaste* steekproefomvang ($r + s = n$) mogelijk zouden zijn verlopen via (of eindigen in) de stippen of de cirkeltjes; vanuit de stippen kan men uitsluitend de punten met $r \geqslant c + 1$ bereiken, vanuit de cirkeltjes alleen de punten met $r \leqslant c$ op de gestippelde lijn. Hieruit volgt dat de *kansen op goed- of afkeuren* bij dit afgeknotte systeem dezelfde zijn als bij vaste steekproefomvang n.

De kansen op het bereiken van de stippen met een *fout als laatste waarneming* zijn:

$$P^{c+1} \qquad (s = 0)$$

$$\frac{(c+s)\,(c+s-1)\ldots(c+1)}{1.\,2.\ldots.s}\,P^{c+1}\,(1-P)^s \qquad (s = 1, \ldots, n-c-1)$$

zoals gemakkelijk blijkt, zie bijv. Stam (1963, paragraaf XI.4). Evenzo zijn de kansen op het bereiken van de cirkeltjes met een *goed exemplaar als laatste waarneming*:

$$(1-P)^{n-c} \qquad (r = 0)$$

$$\frac{(n-c+r-1)\,(n-c+r-2)\ldots(n-c)}{1.\,2.\ldots.r}\,P^r\,(1-P)^{n-c} \qquad (r = 1, \ldots, c)$$

Door deze kansen te vermenigvuldigen met de bijbehorende aantallen waarnemingen $m = r + s$ verkrijgt men de verwachting van het aantal waarnemingen (het gemiddelde van m in de steekproevenruimte) indien men deze produkten optelt (wij laten een gedetailleerde afleiding, die vrij moeilijk is, achterwege). Het resultaat is:

$$\frac{1-(1-P)^n}{P} \qquad (c = 0)$$

$$(c+1)\,\frac{1-(1-P)^{n-c}}{P} - (n-c)\,(1-P)^{n-c}A_c \qquad (c \geqslant 1)$$

waarin

$$A_c = c + \frac{(c-1)\,(n-c+1)}{1.\,2.}\,P + \frac{(c-2)\,(n-c+2)\,(n-c+1)}{1.\,2.\,3.}\,P^2 +$$

$$\ldots + \frac{1.\,(n-1)\,(n-2)\ldots(n-c+1)}{1.\,2.\ldots.c}\,P^{c-1}$$

m a b blz. 660

Enkele voorbeelden:

$c = 1 \qquad 2\,\frac{1-(1-P)^{n-1}}{P} - (n-1)\,(1-P)^{n-1};$

$c = 4 \qquad 5\,\frac{1-(1-P)^{n-4}}{P} - (n-4)\,(1-P)^{n-4}$

$$\{4 + \frac{3\,(n-3)}{2}P + \frac{(n-2)\,(n-3)}{3}P^2 +$$

$$\frac{(n-1)\,(n-2)\,(n-3)}{24}P^3\}.$$

Een aantal uitkomsten is in de volgende tabel weergegeven.

Tabel A2.2. Verwachte steekproefomvang bij afgeknotte pseudo-enkelvoudige steekproefvoorschriften

Maximale steekproefomvang n	P / c	0,0001	0,001	0,01
100	0	99,5	95	*63*
	1	99,0	98,9	*89*
	2	98	98	97
	3	97	97	98
	4	96	96	97
200	0	198	181	*87*
	1	199	198	*146*
	2	198	198	*178*
	3	197	197	191
	4	196	196	196
500	0	488	*394*	*99*
	1	499	483	*195*
	2	498	497	*283*
	3	497	497	*356*
	4	496	497	414
1000	0	952	*632*	*100*
	1	998	*896*	*200*
	2	998	976	*300*
	3	997	994	*399*
	4	996	996	*496*

De gecursiveerde aantallen corresponderen met meer dan 10% besparing ten opzichte van de maximale steekproefomvang.

Appendix 3 *Dubbele steekproefschema's*

Men begint met een steekproef van de omvang n_1, waarin niet meer dan c_2 fouten mogen worden aangetroffen. Bedraagt het aantal fouten daarin c_1 of minder, dan wordt de populatie direkt goedgekeurd; geldt voor dit aantal $c_1 + 1 \leqslant r_1 \leqslant c_2$, dan wordt een tweede steekproef van de omvang n_2 genomen. In beide steekproeven samen mogen niet meer dan c_2 fouten worden aangetroffen. De formule voor de goedkeurkans $S(P)$ (P is de fractie fouten in de populatie) wordt als volgt gevonden.

a. De kans op onmiddellijk goedkeuren is, vgl. paragraaf 2.1.1.:

$$S_a(P) = (1-P)^{n_1} + n_1 P (1-P)^{n_1-1} + \ldots$$

$$+ \frac{n_1 (n_1-1) \ldots (n_1 - c_1 + 1)}{1.2.\ldots.c_1} P^{c_1} (1-P)^{n_1-c_1}$$

mits $c_1 \geqslant 1$; voor $c_1 = 0$ alleen de eerste term.

b. De kans op uitbreiding tot $n_1 + n_2$ bedraagt:

$$S_{b1}(P) = \frac{n_1 (n_1-1) \ldots (n_1-c_1)}{1.2.\ldots.(c_1+1)} P^{c_1+1} (1-P)^{n_1-c_1-1} + \ldots$$

$$+ \frac{n_1 (n_1-1) \ldots (n_1-c_2+1)}{1.2.\ldots.c_2} P^{c_2} (1-P)^{n_1-c_2}$$

$$= T_{c_1+1} + T_{c_1+2} + \ldots + T_{c_2}.$$

De indices van T in deze som geven de aantallen fouten in de eerste steekproef aan. Er mogen dus in de tweede steekproef resp. ten hoogste $(c_2 - c_1 - 1)$, ten hoogste $(c_2 - c_1 - 2)$ enz. fouten worden gevonden, anders overschrijdt het totaal aantal fouten de grens c_2. De termen van de som $S_{b1}(P)$ moeten dus met de kansen op de laatstbedoelde gebeurtenissen worden vermenigvuldigd, waarbij T_{c_2} met de kans op nul fouten, nl. $(1-P)^{n_2}$ wordt vermenigvuldigd.

De totale kans op goedkeuring na uitbreiding wordt dus:

$$S_b(P) = T_{c_1+1} \{ (1-P)^{n_2} + n_2 P (1-P)^{n_2-1} + \ldots$$

$$+ \frac{n_2 (n_2-1) \ldots (n_2 - c_2 + c_1 + 2)}{1.2.\ldots.(c_2-c_1-1)} P^{c_2-c_1-1} (1-P)^{n_2-c_2+c_1+1} \}$$

$$+ T_{c_1+2} \{ (1-P)^{n_2} + \ldots$$

$$+ \frac{n_2 (n_2-1) \ldots (n_2 - c_2 + c_1 + 3)}{1.2.\ldots.(c_2-c_1-2)} P^{c_2-c_1-2} (1-P)^{n_2-c_2+c_1+2} \} +$$

$$\ldots + T_{c_2} (1-P)^{n_2}.$$

Voorbeeld. Zij $c_1 = 0$ en $c_2 = 2$. Dan is $S_a(P) = (1-P)^{n_1} = T_o$;

$$S_{b1}(P) = T_1 + T_2 = n_1 P (1-P)^{n_1-1} + \frac{n_1 (n_1-1)}{2} P^2 (1-P)^{n_1-2}.$$

T_1 is de kans op één fout in de eerste steekproef; deze moet met de (voorwaardelijke) kans om in de tweede steekproef 0 of 1 fouten te vinden worden vermenigvuldigd, nl.

$$\{(1-P)^{n_2} + n_2 P (1-P)^{n_2-1}\}.$$

Verder moet T_2 met de kans op nul fouten in de tweede steekproef worden vermenigvuldigd. Hieruit volgt:

$$S_b(P) = T_1 \{(1-P)^{n_2} + n_2 P (1-P)^{n_2-1}\} + T_2 (1-P)^{n_2}.$$

De totale goedkeurkans voor beide steekproeven samen is dus

$$S(P) = S_a(P) + S_b(P) = (1-P)^{n_1} + n_1 P (1-P)^{n_1+n_2-1}$$

$$+ \frac{n_1 (n_1 + 2n_2 - 1)}{2} P^2 (1-P)^{n_1+n_2-2}.$$

In *voorbeeld* 7 is $n_1 = 20$ en $n_2 = 40$; de keuringskarakteristiek is dus:

$$S_7(P) = (1-P)^{20} + 20P(1-P)^{59} + 990P^2(1-P)^{58};$$

tabel 2.4. geeft een aantal getalwaarden.

Wij berekenen nu de gemiddelde omvang. De kans op $(n_1 + n_2)$ bedraagt $S_{b1}(P)$ en de kans op n_1 dus $1 - S_{b1}(P)$. Voor willekeurige c_1 en c_2 (waarbij uiteraard $c_1 < c_2$ is) komt er:

Verwachte steekproefomvang

$$= (n_1 + n_2) S_{b1}(P) + n_1 \{1 - S_{b1}(P)\} = n_1 + n_2 S_{b1}(P).$$

Voor $c_1 = 0$ en $c_2 = 2$ heeft men de volgende uitdrukking:

Verwachte steekproefomvang

$$= n_1 + n_1 n_2 P \left(1 + \frac{n_1 - 3}{2} P\right) (1-P)^{n_1-2}.$$

Hiermee is de laatste kolom van tabel 2.4. berekend. Door de afgeleide naar P van deze laatste uitdrukking nul te stellen blijkt dat (voor elke n_2) de verwachting van de omvang maximaal is voor

$$P = \frac{-2 + \sqrt{2(n_1 - 1)(n_1 - 2)}}{n_1 (n_1 - 3)},$$

hetgeen voor $n_1 = 20$ leidt tot $P = 0{,}071$. In voorbeeld 7 is het maximum 44,2.

Voorbeeld 10: $c_1 = 0$, $c_2 = 2$, $n_1 = 75$, $n_2 = 150$. De goedkeurkans is nu: $S(P) = (1-P)^{75} + 75P(1-P)^{224} + 14025P^2(1-P)^{223}$;
de gemiddelde steekproefomvang is $75 + 11250P(1 + 36P)(1-P)^{73}$, waarvan het maximum 163,7 bereikt wordt voor $P = 0{,}0189$. Zie tabel A3.1. voor de numerieke details.

Tabel A3.1. Keuringskarakteristiek en verwachte steekproefgrootte in voorbeeld 10

P	$S(P)$	Verwachte steekproefomvang
$P_2 = 0{,}0041$	0,952	114,2
0,005	0,923	121,0
0,010	0,699	148,5
$P_i = 0{,}0142$	0,502	160,0
0,015	0,468	161,2
0,020	0,298	163,6
0,030	0,118	151,0
$P_1 = 0{,}040$	0,050	130,8
0,050	0,022	112,2
0,060	0,010	98,3
0,100	0,0004	77,4

Behoudens afronding is dus $\alpha_o = \beta_o = 0{,}05$.

Appendix 4 *Controlepunt, steilheid en buigpunt*

Bij een *enkelvoudig* steekproefvoorschrift is de afgeleide van de goedkeurkans gelijk aan, zie Muilwijk (1977, paragraaf 6.2.1):

$$S'(P) = - \frac{n\,(n-1)\ldots(n-c)}{1.\,2.\ldots.c}\,P^c\,(1-P)^{n-c-1}$$

waaruit volgt

$$h_i = -\,2P_iS'(P_i) =$$

$$+\,\frac{2n\,(n-1)\ldots(n-c)}{1.\,2.\ldots.c}\,P_i{}^{c+1}\,(1-P_i)^{n-c-1},$$

bijv. voor $c = 0$: $2nP_i\,(1-P_i)^{n-1} = nP_i/(1-P_i) = n(2^{1/n}-1)$;
$c = 1$: $2n\,(n-1)P_i^2\,(1-P_i)^{n-2}$;
$c = 2$: $n\,(n-1)\,(n-2)P_i^3\,(1-P_i)^{n-3}$.

Verder blijkt door nog eens te differentiëren:

$$S''(P) = +\,\frac{n\,(n-1)\ldots(n-c)}{1.\,2.\ldots.c}\,P^{c-1}\,(1-P)^{n-c-2}$$

$$\{\,(n-1)\,P-c\,\},$$

welke uitdrukking nul is voor $P = 0$ $(c \geqslant 2)$, voor $P = 1$ $(c \leqslant n-3)$ en voor $P = c/(n-1)$. Voor $1 \leqslant c \leqslant n-2$ is er dus een buigpunt met $0 < P < 1$. Links van het buigpunt verloopt de kromme bij toenemende P steeds steiler, rechts ervan steeds minder steil. Men ziet dit in figuur 2.2., waaruit ook blijkt dat het controlepunt dicht bij het buigpunt ligt; meestal ligt het controlepunt meer naar rechts. De raaklijn in het controlepunt is (uiteraard) iets minder steil dan die in het buigpunt, vgl. de volgende tabel.

Tabel A4.1. Buigpunt, controlepunt en steilheid bij enkelvoudige schema's

Voorbeeld	n	c	$c/(n-1)$	P_i	$S'(c/(n-1))$	$S'(P_i)$	h_i
3	100	1	0,0101	0,0167	− 37,0	− 31,7	1,06
4	160	2	0,0126	0,0167	− 43,6	− 39,8	1,33
8	45	2	0,0455	0,059	− 12,5	− 11,5	1,36
9	50	2	0,0408	0,054	− 13,8	− 12,6	1,36

Als men de Poisson-benadering toepast gelden de volgende relaties, vgl. Schaafsma en Willemze (1978, blz. 230 en 291); voor de binomiale en de hypergeometrische verdeling vindt men de analoga op blz. 231:

$$nP_i = c + 0{,}67$$

$$\frac{\pi}{2} h_i^2 = nP_i + 0{,}06 \qquad \text{(waarin } \pi = 3{,}1416\text{)}.$$

Als men bepaalde waarden voor P_i en h_i voorschrijft, kan men uit de tweede vergelijking n oplossen en daarna uit de eerste vergelijking c. Omgekeerd kan men bij gegeven n en c ook snel P_i en h_i vinden, zie *op. cit.* blz. 232.

Ter illustratie van de nauwkeurigheid van de zojuist gegeven benaderingsformules voor de binomiale verdeling beschouwen wij de voorbeelden 1-4 uit paragraaf 2.1.1.

Tabel A4.2. Controlepunt en steilheid bij enkelvoudige schema's

Voorbeeld	1	2	3	4
P_i	0,0135	0,0069	0,0167	0,0167
h_i	0,698	0,696	1,060	1,331
$nP_i - 0{,}67$ *)	0,018	0,021	1,000	2,002
$\frac{\pi}{2} h_i^2$	0,765	0,760	1,765	2,781
$nP_i + 0{,}06$ **)	0,748	0,751	1,730	2,732

*) Vergelijk deze waarden met die van c (resp. 0; 0; 1; 2)
**) Vergelijk met de regel erboven.

Voor *dubbele* steekproefvoorschriften geldt (*op. cit.* blz. 291):

$$n_1 P_i = \psi\,(c_2 + 0{,}67)$$

$$\frac{\pi}{2} h_i^2 = c_2 + 0{,}73\psi + 0{,}1 c_2\,(1 - \psi - 0{,}4\varphi^4/\psi^2)$$

waarin $\psi = n_1/(n_1 + n_2)$ en $\varphi = c_1/c_2$.

Voor $c_1 = 0$ en $c_2 = 2$ komt er:

$$(n_1 + n_2)\,P_i = 2{,}67$$

$$\frac{\pi}{2} h_i^2 = 2{,}2 + \frac{0{,}53 n_1}{n_1 + n_2}$$

Deze relaties zijn voor $c_1 = 0$ wat minder nauwkeurig dan de overeenkomstige bij enkelvoudige steekproeven. Voorbeeld 7 heeft een waarde $(n_1 + n_2)\,P_i = 3{,}17$ en $\frac{\pi}{2} h_i^2 = 2{,}28$; de formule geeft 2,67 resp. 2,38. Voorbeeld 10: resp. 3,20 en 2,32 in plaats van 2,67 en 2,38.

Appendix 5 *Enkele meervoudige steekproefschema's*

In *voorbeeld* 11 (paragraaf 3.1.1.) gaf ik een viervoudig steekproefvoorschrift. De kansen op goedkeuren zijn in de negen mogelijke gevallen:

Tabel A5.1. Goedkeurkansen bij voorbeeld 11

	r_1	r_2	r_3	r_4	Kansen
1	0	–	–	–	$(1-P)^{n_1}$
2	1	0	–	–	$n_1 P\,(1-P)^{n_1-1}\,(1-P)^{n_2}$
3	1	1	0	–	$n_1 P\,(1-P)^{n_1-1}\,n_2 P\,(1-P)^{n_2-1}\,(1-P)^{n_3}$
4	1	1	1	0	$n_1 P\,(1-P)^{n_1-1}\,n_2 P\,(1-P)^{n_2-1}\,n_3 P\,(1-P)^{n_3-1}\,(1-P)^{n_4}$
5	1	2	0	0	$n_1 P\,(1-P)^{n_1-1}\,\frac{n_2\,(n_2-1)}{2}\,P^2\,(1-P)^{n_2-2}\,(1-P)^{n_3}\,(1-P)^{n_4}$
6	2	0	0	–	$\frac{n_1\,(n_1-1)}{2}\,P^2\,(1-P)^{n_1-2}\,(1-P)^{n_2}\,(1-P)^{n_3}$
7	2	1	0	0	$\frac{n_1\,(n_1-1)}{2}\,P^2\,(1-P)^{n_1-2}\,n_2 P\,(1-P)^{n_2-1}\,(1-P)^{n_3}\,(1-P)^{n_4}$
8	2	0	1	0	$\frac{n_1\,(n_1-1)}{2}\,P^2\,(1-P)^{n_1-2}\,(1-P)^{n_2}\,n_3 P\,(1-P)^{n_3-1}\,(1-P)^{n_4}$
9	3	0	0	0	$\frac{n_1\,(n_1-1)\,(n_1-2)}{6}\,P^3\,(1-P)^{n_1-3}\,(1-P)^{n_2}\,(1-P)^{n_3}\,(1-P)^{n_4}$

Hierin komen de factoren P en $(1-P)$ slechts in vier combinaties voor. Sommeert men de bijbehorende binomiaalcoëfficiënten dan komt er:

$$S(P) = (1-P)^{n_1} + n_1 P\,(1-P)^{n_1+n_2-1} +$$

$$\frac{n_1}{2}\,(n_1 + 2n_2 - 1)\,P^2\,(1-P)^{n_1+n_2+n_3-2} +$$

$$\frac{n_1}{6}\,\{(n_1-1)\,(n_1-2) + 3n_2\,(n_1+n_2+2n_3-2) +$$

$$3n_3\,(n_1-1)\}\,P^3\,(1-P)^{n_1+n_2+n_3+n_4-3}.$$

Substitutie van $n_1 = 96$, $n_2 = 51$, $n_3 = 35$, $n_4 = 32$ levert

$$S(P) = (1-P)^{96} + 96P\,(1-P)^{146} + 9456P^2\,(1-P)^{180} +$$

$$828\,335P^3\,(1-P)^{211}$$

en als men hierin $P = 0{,}04$ substitueert vindt men 0,04946.

m a b blz. 667

Gezien de enorme hoeveelheid rekenwerk dat *voorbeeld* 12 vereist geef ik daarvoor niet alle termen, maar slechts zoveel dat blijkt dat S (0,04) in dat geval duidelijk groter is dan 0,05. Ik volsta met vier gevallen waarin de populatie wordt goedgekeurd bij ten hoogste $n_1 + n_2 + n_3 = 158$ waarnemingen:

Tabel A5.2. Enkele goedkeurkansen in voorbeeld 12

r_1	r_2	r_3	Steekproefomvang	Kansen
0	–	–	75	$(1-P)^{75}$
1	0	–	119	$75P\,(1-P)^{118}$
1	1	0	158	$75 \times 44P^2(1-P)^{156}$
2	0	0	158	$75 \times 37P^2(1-P)^{156}$

Substitutie van $P = 0{,}04$ levert:

$$0{,}0468 + 0{,}0243 + 0{,}0091 + 0{,}0076 = 0{,}0878.$$

Toevoeging van de bijdragen van alle andere gevallen waarin de populatie goedgekeurd wordt levert een grotere uitkomst op, zodat blijkt dat

$$S\,(0{,}04) > 0{,}0878 > 0{,}05,$$

hetgeen te bewijzen was.

Literatuurlijst

Arkin, H., 1963, Handbook of sampling for auditing and accountancy, 2 dln, McGraw-Hill, New York

Blokdijk, J. H., 1979, Uitbreiding van de steekproef, MAB november 1979, blz. 537-541

Blokdijk, J. H., 1980, Terug naar de steekproef, MAB maart 1980, blz. 127-131

Cochran, W. G., 1977, Sampling Techniques, 3e druk, Wiley, New York

Fisher, R. A., and F. Yates, 1957, Statistical tables for biological, agricultural and medical research, 5e druk, Oliver and Boyd, Edinburgh

De Jonge, H., en G. Wielinga, 1973, Statistische methoden voor psychologen en sociologen, 3e druk, Tjeenk Willink, Haarlem

Kriens, J., 1963, De methoden van De Wolff en Van Heerden (. . .), Statistica neerlandica 1963, blz. 215-231

Kriens, J., 1968, Het verdelen van steekproeven over subpopulaties bij accountantscontroles, Statistica neerlandica 1968, blz. 151-157

Kriens, J., F. Göbel en W. Molenaar, 1968, Leergang Besliskunde deel 8, Mathematisch Centrum, Amsterdam

Kriens, J., en A. C. Dekkers, 1979, Steekproeven in de accountantscontrole, Stenfert Kroese, Leiden

Moors, J. J. A., en J. Muilwijk, 1975, Steekproeven, Agon Elsevier, Amsterdam

Muilwijk, J., en J. H. Schouten, 1960, Inleiding tot de wiskundige statistiek deel III, Tabellen, Staatsbedrijf der PTT, Den Haag

Muilwijk, J., 1977, Dictaat Toegepaste statistiek deel D, Econometrisch instituut van de RUG, Groningen

Rijken van Olst, H., 1974, Algemene statistiek, Van Gorcum, Assen

Schaafsma, A. H., en F. G. Willemze, 1978, Modern kwaliteitsbeleid, 3e druk, Kluwer, Deventer

Stam, A. J., 1963, Inleiding tot de waarschijnlijkheidsrekening, Stam, Haarlem

Tuitjer, E. B., en J. M. Zuijdervliet, 1975, De steekproef, De Accountant april 1975, blz. 485-508

Vance, L. L., and J. Neter, 1956, Statistical sampling for auditors and accountants, Wiley, New York

Vermaas, M., 1979a, Uitbreiden van de steekproef, MAB mei 1979, blz. 203-215

Vermaas, M., 1979b, Steekproeven bij controles op ernstige fouten, MAB november 1979, blz. 525-536

Vermaas, M., 1980a, Van steekproef tot steekspel, MAB januari 1980, blz. 36-40

Vermaas, M., 1980b. De steekproef nogmaals uitgebreid, MAB juni 1980, blz. 255-256.

Wetherill, G. B., 1969, Sampling inspection and quality control, Methuen, London

Wolff, P. de, 1956, Steekproeven bij administratieve controles, Statistica neerlandica 1956, blz. 35-44

Wolff, P. de, 1959, Produktiviteitsverhoging bij accountantscontroles, Statistica neerlandica 1959, blz. 215-232

Wolff, P. de, 1973, Bedrijfsstatistiek, 10e druk, Samsom, Alphen aan den Rijn