STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

Rapport S 358

Quotiëntschatters bij steekproeven uit eindige populaties

door

C. Visser

december 1965

Printed at the Mathematical Centre at Amsterdam,49,2nd Boerhaavestraat.
The Netherlands.

The Mathematical Centre, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Scientific Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

Inhoud

Inleiding

Gegeven zij een eindige populatie, bestaande uit de elementen

$$e_1, \ldots, e_N .$$

Aan het element e_i kan de grootheid y_1 $(i = 1, 2, \ldots, N)$ gemeten worden. In het volgende zullen we geen scherp onderscheid maken tussen het element en de daaraan gemeten grootheden.

Uit deze populatie wordt een aselecte steekproef van de omvang n getrokken. De n uitkomsten van de metingen kunnen voorgesteld worden door

$$\underline{y}_1, \ldots, \underline{y}_n .$$

Het steekproefgemiddelde $\bar{\underline{y}}$ is dan een zuivere schatter van het populatiegemiddelde $\bar{Y}$.

Indien aan de elementen nog andere grootheden dan $\underline{y}$, bijv. de grootheden

$$\underline{x}^{(1)}, \ldots, \underline{x}^{(k)},$$

gemeten kunnen worden, dan kunnen schatters van $\bar{Y}$ gevonden worden, die onder zekere voorwaarden een kleinere variantie bezitten dan de schatter $\bar{\underline{y}}$. Er wordt dan gebruik gemaakt van de correlatie tussen $\underline{y}$ en $\underline{x}^{(j)}$ $(j = 1, \ldots, k)$. Onder deze schatters bevinden zich de <u>quotiëntschatters</u> (ratio-estimators).

In dit rapport worden enkele typen quotiëntschatters behandeld. Daarbij is steeds $k = 1$ zodat aan elk element van de populatie slechts de grootheden $\underline{y}$ en $\underline{x}$ gemeten worden. Een verdere beperking is die tot aselecte steekproeven zonder teruglegging (simple random sampling), behalve in paragraaf 6, waar een eenvoudig geval van "probability sampling" wordt behandeld.

1. Definities

De populatie bestaat uit N elementen met de eigenschappen

$$(y_1, x_1), \ldots, (y_N, x_N).$$

Uit deze populatie wordt zonder teruglegging een aselecte steekproef van n ($n \leq N$) elementen getrokken. Iedere greep van n elementen heeft dan een kans $\binom{N}{n}^{-1}$ om getrokken te worden. De getrokken elementen worden voorgesteld door

$$(\underline{y}_1,\underline{x}_1),\ldots,(\underline{y}_n,\underline{x}_n).$$

Verder worden de volgende definities gegeven:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \; , \; \overline{\underline{x}} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i \; ;$$

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i \; , \; \overline{\underline{y}} = \frac{1}{n}\sum_{i=1}^{n} \underline{y}_i \; ;$$

$$r_i = \frac{y_i}{x_i} \quad (i = 1,\ldots, N), \; \underline{r}_i = \frac{\underline{y}_i}{\underline{x}_i} \; ,$$

$$\overline{R} = \frac{1}{N}\sum_{i=1}^{N} r_i \; , \; \overline{\underline{r}} = \frac{1}{n}\sum_{i=1}^{n} \underline{r}_i \; ;$$

$$\underline{s}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} (\underline{x}_i - \overline{\underline{x}})^2, \; \underline{s}_y^2 = \frac{1}{n-1}\sum_{i=1}^{n} (\underline{y}_i - \overline{\underline{y}})^2,$$

$$\underline{s}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n} (\underline{x}_i - \overline{\underline{x}})(\underline{y}_i - \overline{\underline{y}}) \; ,$$

$$\underline{s}_r^2 = \frac{1}{n-1}\sum_{i=1}^{n} (\underline{r}_i - \overline{\underline{r}})^2, \; \underline{s}_{xr} = \frac{1}{n-1}\sum_{i=1}^{n} (\underline{x}_i - \overline{\underline{x}})(\underline{r}_i - \overline{\underline{r}}) \; ;$$

$$\underline{s}_{\alpha\beta} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i^{\alpha}\underline{r}_i^{\beta};$$

$$\mu_{\alpha\beta} = \mathcal{E}(\underline{x} - \overline{X})^{\alpha}(\underline{y} - \overline{Y})^{\beta} \; ,$$

$$\mu_{\alpha\beta\gamma} = \mathcal{E}(\underline{x} - \overline{X})^{\alpha}(\underline{y} - \overline{Y})^{\beta}(\underline{r} - \overline{R})^{\gamma} \; , \; \underline{r} = \frac{\underline{y}}{\underline{x}}$$

Met $\underline{x}$ en $\underline{y}$ worden bedoeld stochastische grootheden met kansverdelingen $P[\underline{x} = x_i] = \frac{1}{N}$ en $P[\underline{y} = y_i] = \frac{1}{N}$, voor $i = 1, 2,\ldots, N$.

2. Quotiëntschatters

De bekendste en meest gebruikte quotiëntschatter is

$$\underline{\tilde{y}} = \bar{X} \cdot \frac{\bar{\underline{y}}}{\bar{\underline{x}}} \,. \tag{2.1}$$

Indien de waarde van $\bar{X}$ bekend is kunnen met $\underline{\tilde{y}}$ schattingen van $\bar{Y}$ verricht worden. Is $\bar{X}$ niet bekend, dan kan met de quotiëntschatter

$$\underline{\tilde{r}} = \frac{\bar{\underline{y}}}{\bar{\underline{x}}} \tag{2.2}$$

een schatting van $\bar{Y} \cdot \bar{X}^{-1}$ worden gemaakt.

Een andere voor de hand liggende quotiëntschatter is

$$\underline{\hat{y}} = \bar{X} \cdot \bar{\underline{r}} \tag{2.3}$$

Van deze schatter kan de onzuiverheid exact bepaald en zuiver geschat worden, waardoor de zuivere quotiëntschatter

$$\underline{\hat{y}}' = \bar{X}\bar{\underline{r}} + \frac{n(N-1)}{N(n-1)} (\bar{\underline{y}} - \overline{\underline{x}\underline{r}}) \tag{2.4}$$

ontstaat.

Met de boven gedefinieerde schatters van $\bar{Y}$ kan de som $N\bar{Y}$ van alle y_i geschat worden door deze schatters met N te vermenigvuldigen.

Er zijn meerdere quotiëntschatters bekend. Door MICKEY [11] en WILLIAMS [19] werden gehele klassen van (zuivere) quotiëntschatters geconstrueerd.

De schatters $\underline{\tilde{y}}$ en $\underline{\tilde{r}}$

3. De onzuiverheid van $\underline{\tilde{y}}$

De schatter $\underline{\tilde{y}}$ is in het algemeen onzuiver. Wel geldt, zoals uit (2.1) direct duidelijk zal zijn, dat $P[\underline{\tilde{y}} = \bar{Y}] = 1$ voor $n = N$, dus dat $\underline{\tilde{y}}$ "asymptotisch raak" (consistent) is.

Omdat $\underline{\tilde{y}}$ zowel in de noemer als in de teller een stochastische variabele bevat is het niet eenvoudig $\mathcal{E}\underline{\tilde{y}}$ te bepalen. Indien echter voor elke mogelijke steekproef geldt, dat $\bar{x} \neq 0$ en tevens, dat $\bar{X} \neq 0$, dan kan worden aangetoond, dat voor elk natuurlijk getal m

$$\mathcal{E}\tilde{y} = \bar{Y} + \bar{Y} \cdot \mathcal{E}(S\underline{y} - S\underline{x}) \sum_{\nu=0}^{m} (-S\underline{x})^{\nu} + \mathcal{E}(-S\underline{x})^{m+1}(\tilde{y} - \bar{Y}), \qquad (3.1)$$

met

$$S\underline{x} = \frac{\bar{\underline{x}} - \bar{X}}{\bar{X}}, \quad S\underline{y} = \frac{\bar{\underline{y}} - \bar{Y}}{\bar{Y}} \qquad (3.2)$$

(zie de Appendix). Verder wordt in de Appendix aangetoond, dat voor m = 1 de derde term van het tweede lid van (3.1) van de orde n^{-2} is. Voor een eerste benadering $B_1(\tilde{y})$ van de onzuiverheid $B(\tilde{y})$ van $\tilde{y}$ kan men dus nemen

$$B_1(\tilde{y}) = \bar{Y}\{\mathcal{E}(S\underline{x})^2 - \mathcal{E}(S\underline{x})(S\underline{y})\} =$$

$$= \frac{N-n}{N-1} \cdot \frac{\bar{Y}}{n}\left(\frac{\mu_{20}}{\bar{X}^2} - \frac{\mu_{11}}{\bar{X}\bar{Y}}\right) =$$

$$= \frac{N-n}{N-1} \cdot \frac{\bar{Y}}{n}\left\{\frac{\mu_{20}}{\bar{X}^2} - \zeta \cdot \frac{(\mu_{20} \cdot \mu_{02})^{\frac{1}{2}}}{\bar{X}\bar{Y}}\right\}, \qquad (3.3)$$

met

$$\zeta = \frac{\mathrm{Cov}\,(\underline{x},\underline{y})}{\{\mathrm{Var}(\underline{x}) \cdot \mathrm{Var}(\underline{y})\}^{\frac{1}{2}}} = \frac{\mu_{11}}{(\mu_{20} \cdot \mu_{02})^{\frac{1}{2}}} . \qquad (3.4)$$

In de Appendix worden de verwachtingen van $(S\underline{x})^{\alpha}(S\underline{y})^{\beta}$ voor $\alpha + \beta \leq 4$ vermeld.

Zowel B als B_1 nemen af met toenemende steekproefomvang n en verdwijnen geheel voor n = N, hetgeen in overeenstemming is met het feit, dat $\tilde{y}$ "asymptotisch raak" is. Verder is $B_1 = 0$ indien

$$\zeta = \frac{\bar{Y}}{\bar{X}} \cdot \left(\frac{\mu_{20}}{\mu_{02}}\right)^{\frac{1}{2}} . \qquad (3.5)$$

Indien de regressie van $\underline{y}$ op x gegeven wordt door een rechte door de oorsprong, dan is de schatter $\tilde{y}$ zuiver. Stel namelijk

$$\mathcal{E}(\underline{y}_i \mid x_i) = ax_i \tag{3.6}$$

en neem van beide leden van (3.6) de verwachting. Dan volgt

$$\bar{\bar{Y}} = a\bar{\bar{X}} \tag{3.7}$$

en dus

$$\mathcal{E}\underline{\tilde{y}} = \mathcal{E}\mathcal{E}\left(\bar{\bar{X}}\,\frac{\underline{\bar{\bar{y}}}}{\underline{\bar{\bar{x}}}} \mid \underline{x}_1, \ldots, \underline{x}_n\right) =$$

$$= \mathcal{E}\left(\bar{\bar{X}}\,\frac{a\underline{\bar{\bar{x}}}}{\underline{\bar{\bar{x}}}}\right) = a\bar{\bar{X}} = \bar{\bar{Y}}\,, \tag{3.8}$$

omdat

$$\mathcal{E}(\underline{y}_i \mid x_1, \ldots, x_n) = \mathcal{E}(\underline{y}_i \mid x_i). \tag{3.9}$$

Een zuivere schatter voor $B_1(\underline{\tilde{y}})$ is

$$\underline{b}_1(\underline{\tilde{y}}) = \frac{N - n}{nN\bar{\bar{X}}^2}\,(\underline{\bar{\bar{y}}}\underline{s}_x^2 - \underline{\bar{\bar{x}}}\underline{s}_{xy})\,, \tag{3.10}$$

waarbij dus de waarde van $\bar{\bar{X}}$ bekend wordt ondersteld.

In de Appendix wordt voor de onzuiverheid B nog de volgende ongelijkheid afgeleid:

$$|B(\underline{\tilde{y}})| \leq |\bar{\bar{X}}|^{-1}\left\{\frac{N - n}{N - 1} \cdot \frac{\mu_{20}}{n} \cdot \mathrm{Var}(\underline{\tilde{y}})\right\}^{\frac{1}{2}} \tag{3.11}$$

4. De variantie van $\underline{\tilde{y}}$

Men ziet direct in, dat

$$\mathrm{Var}(\underline{\tilde{y}}) = \mathcal{E}(\underline{\tilde{y}} - \bar{\bar{Y}})^2 - (\mathcal{E}\underline{\tilde{y}} - \bar{\bar{Y}})^2 = \mathcal{E}(\underline{\tilde{y}} - \bar{\bar{Y}})^2 - \{B(\underline{\tilde{y}})\}^2. \tag{4.1}$$

Omdat $B(\underline{\tilde{y}})$ van de orde n^{-1} is kan men volstaan met de eerste term van het derde lid van (4.1), de "mean square error". Gemakkelijk blijkt, dat

$$\underline{\tilde{y}} - \bar{\bar{Y}} = \bar{\bar{Y}}(S\underline{y} - S\underline{x}) - S\underline{x}(\underline{\tilde{y}} - \bar{\bar{Y}})\,, \tag{4.2}$$

zodat

$$\mathcal{E}(\tilde{\underline{y}} - \overline{Y})^2 = \overline{Y}^2 \cdot \mathcal{E}(S\underline{y} - S\underline{x})^2 +$$
$$+ 2\overline{Y} \cdot \mathcal{E}(S\underline{x})(S\underline{x} - S\underline{y})(\tilde{\underline{y}} - \overline{Y}) + \mathcal{E}(S\underline{x})^2(\tilde{\underline{y}} - \overline{Y})^2 . \quad (4.3)$$

De laatste twee termen van het tweede lid van (4.3) zijn weer van de orde n^{-2}. Voor een eerste benadering $\mathrm{Var}_1\ (\tilde{\underline{y}})$ van $\mathrm{Var}(\tilde{\underline{y}})$ vindt men dus

$$\mathrm{Var}_1(\tilde{\underline{y}}) = \overline{Y}^2\{\mathcal{E}(S\underline{x})^2 - 2\mathcal{E}(S\underline{x})(S\underline{y}) + \mathcal{E}(S\underline{y})^2\} =$$
$$= \frac{N - n}{N - 1} \cdot \frac{\overline{Y}^2}{n}\left(\frac{\mu_{20}}{\overline{X}^2} - 2\frac{\mu_{11}}{\overline{X}\overline{Y}} + \frac{\mu_{02}}{\overline{Y}^2}\right) =$$
$$= \frac{N - n}{N - 1} \cdot \frac{\overline{Y}^2}{n}\left\{\frac{\mu_{20}}{\overline{X}^2} - 2\zeta \cdot \frac{(\mu_{20} \cdot \mu_{02})^{\frac{1}{2}}}{\overline{X}\overline{Y}} + \frac{\mu_{02}}{\overline{Y}^2}\right\}, \quad (4.4)$$

met ζ gedefinieerd als in (3.4). Uit (4.4) blijkt, dat $\mathrm{Var}_1(\tilde{\underline{y}})$ afneemt met toenemende steekproefomvang n, terwijl $\mathrm{Var}_1(\tilde{\underline{y}}) = \mathrm{Var}(\tilde{\underline{y}}) = 0$ voor $n = N$.

Nu rijst direct de vraag wanneer deze variantie kleiner is dan die van de schatter $\overline{\underline{y}}$. Gemaakkelijk is in te zien, dat (zie ook A3.16)

$$\mathrm{Var}(\overline{\underline{y}}) = \frac{N - n}{N - 1} \cdot \frac{\mu_{02}}{n}, \quad (4.5)$$

zodat de relatieve doeltreffendheid (efficiency) van $\tilde{\underline{y}}$ t.o.v. $\overline{\underline{y}}$ in eerste benadering gegeven wordt door

$$\frac{\mathrm{Var}(\overline{\underline{y}})}{\mathrm{Var}_1(\tilde{\underline{y}})} = \frac{\mu_{02}}{\overline{Y}^2}\left\{\frac{\mu_{20}}{\overline{X}^2} - 2\zeta \cdot \frac{(\mu_{20} \cdot \mu_{02})^{\frac{1}{2}}}{\overline{X}\overline{Y}} + \frac{\mu_{02}}{\overline{Y}^2}\right\}^{-1}, \quad (4.6)$$

en dit is groter dan 1 indien

$$\zeta^2 > \tfrac{1}{4} \cdot \frac{\overline{Y}^2}{\overline{X}^2} \cdot \frac{\mu_{20}}{\mu_{02}} . \quad (4.7)$$

Een zuivere schatter $\mathrm{var}_1(\tilde{\underline{y}})$ van $\mathrm{Var}_1(\tilde{\underline{y}})$ is

$$\mathrm{var}_1(\tilde{\underline{y}}) = \frac{N - n}{nN\overline{X}^2} \cdot \{(\overline{\underline{y}}^2 \cdot \underline{s}_x^2 - 2\overline{\underline{x}}\overline{\underline{y}} \cdot \underline{s}_{xy} + \overline{\underline{x}}^2 \cdot \underline{s}_y^2) +$$
$$- 2\frac{(N - n)(n - 1)}{nN(n - 2)} \cdot (\underline{s}_x^2 \cdot \underline{s}_y^2 - \underline{s}_{xy}^2)\}. \quad (4.8)$$

5. De schatter $\tilde{\underline{r}}$

Tussen de schatters $\tilde{\underline{y}}$ en $\tilde{\underline{r}}$ bestaat de relatie

$$\tilde{\underline{y}} = \bar{X} \cdot \tilde{\underline{r}} , \tag{5.1}$$

zodat

$$\mathcal{E}\,\tilde{\underline{r}} = \bar{X}^{-1} \cdot \mathcal{E}\,\tilde{\underline{y}} , \tag{5.2}$$

$$\mathrm{Var}(\tilde{\underline{r}}) = \bar{X}^{-2} \cdot \mathrm{Var}(\tilde{\underline{y}}), \quad \mathrm{Var}_1(\tilde{\underline{r}}) = \bar{X}^{-2} \cdot \mathrm{Var}_1(\tilde{\underline{y}}), \tag{5.3}$$

en

$$B(\tilde{\underline{r}}) = \bar{X}^{-1} \cdot B(\tilde{\underline{y}}), \quad B_1(\tilde{\underline{r}}) = \bar{X}^{-1} \cdot B_1(\tilde{\underline{y}}), \tag{5.4}$$

waarin $B(\tilde{\underline{r}})$ de onzuiverheid van $\tilde{\underline{r}}$ en $B_1(\tilde{\underline{r}})$ een eerste benadering van deze onzuiverheid is. De waarde van $\bar{X}$ wordt bij deze schatter onbekend ondersteld.

Een, in het algemeen onzuivere, schatter van $B_1(\tilde{\underline{r}})$ is

$$b_1(\tilde{\underline{r}}) = \frac{N - n}{nN} \cdot \tilde{\underline{r}} \cdot \left(\frac{\underline{s}_x^2}{\bar{\underline{x}}^2} - \frac{\underline{s}_{xy}}{\bar{\underline{x}}\bar{\underline{y}}} \right). \tag{5.5}$$

Een, in het algemeen onzuivere, schatter van $\mathrm{Var}_1(\tilde{\underline{r}})$ wordt gegeven door

$$\mathrm{var}_1(\tilde{\underline{r}}) = \frac{N - n}{nN} \cdot \tilde{\underline{r}}^2 \left(\frac{\underline{s}_x^2}{\bar{\underline{x}}^2} - \frac{\underline{s}_{xy}}{\bar{\underline{x}}\bar{\underline{y}}} + \frac{\underline{s}_y^2}{\bar{\underline{y}}^2} \right) \tag{5.6}$$

De schatter $\tilde{\underline{r}}$ vindt toepassing bij het schatten van fracties. Stel dat de populatie bestaat uit een aantal klassen met N_i elementen in de i-de klasse, zodat

$$N_i = Np_i \text{ en } \sum_i p_i = 1. \tag{5.7}$$

De stochastische variabelen $\underline{x}$ en $\underline{y}$ worden als volgt gedefinieerd: $\underline{y}$ neemt in de i-de en j-de klasse de waarde 1 en in de overige klassen de waarde 0 aan; $\underline{x}$ neemt in de j-de en k-de klasse de waarde 1 en in de overige klassen de waarde 0 aan.

In een aselecte steekproef zonder teruglegging van n elementen bevinden zich nu $\underline{n}_i$, $\underline{n}_j$ en $\underline{n}_k$ elementen uit de i-de, j-de en k-de klasse.

Een schatting van

$$\frac{p_i + p_j}{p_j + p_k}$$

wordt dan verkregen met

$$\underline{\tilde{r}} = \frac{\underline{n}_i + \underline{n}_j}{\underline{n}_j + \underline{n}_k} = \frac{\sum_{\nu=1}^{n} \underline{y}_\nu}{\sum_{\nu=1}^{n} \underline{x}_\nu} = \frac{\overline{\underline{y}}}{\overline{\underline{x}}} . \qquad (5.8)$$

Gemakkelijk is in te zien, dat

$$\overline{X} = p_j + p_k \, , \; \overline{Y} = p_i + p_j \; ; \qquad (5.9)$$

en

$$\begin{aligned} \mu_{20} &= (p_j + p_k)(1 - p_j - p_k) \\ \mu_{11} &= p_j - (p_j + p_k)(p_i + p_j), \\ \mu_{02} &= (p_i + p_j)(1 - p_i - p_j). \end{aligned} \qquad (5.10)$$

Na enige herleiding volgt dan met (5.4) en (3.3), dat

$$B_1(\underline{\tilde{r}}) = \frac{N - n}{N - 1} \cdot \frac{p_i}{n(p_j + p_k)^2} , \qquad (5.11)$$

en uit (5.3) en (4.4), dat

$$Var_1(\underline{\tilde{r}}) = \frac{N - n}{N - 1} \cdot \frac{(p_i + p_j)(p_i + p_k)}{n(p_j + p_k)^3} . \qquad (5.12)$$

Met (5.5) en (5.6) kunnen, in het algemeen onzuivere, schatters van B_1 en Var_1 gevonden worden. Zij luiden:

$$b_1(\underline{\tilde{r}}) = \frac{N - n}{(n - 1)N} \cdot \frac{n\underline{n}_i}{(\underline{n}_j + \underline{n}_k)^2} ; \qquad (5.13)$$

$$\mathrm{var}_1(\tilde{\underline{r}}) = \frac{N - n}{(n - 1)N} \cdot \frac{n(\underline{n}_i + \underline{n}_j)(\underline{n}_i + \underline{n}_k)}{(\underline{n}_j + \underline{n}_k)^3} . \qquad (5.14)$$

Speciale gevallen verkrijgt men door één der p's gelijk aan 0 te stellen.

6. Eliminatie van de onzuiverheid door verandering van steekproefprocedure.

Het is duidelijk, dat de schatters $\tilde{\underline{y}}$ en $\tilde{\underline{r}}$ zuiver zouden zijn indien de kans op het trekken van de steekproef evenredig zou zijn met $\bar{\underline{x}}$. LAHIRI [9] heeft een methode aangegeven om dit op eenvoudige wijze te realiseren. Daarbij wordt steeds verondersteld, dat $x_i > 0$ voor $i = 1, 2, \ldots, N$.

De populatie bestaat weer uit de elementen (y_i, x_i) met $i = 1, 2, \ldots, N$. Stel M is de som van de n grootste waarden x_i in de populatie. Kies een aselect getal ξ met $0 \leq \xi < M'$ en $M' \geq M$. Trek verder een aselecte steekproef zonder teruglegging van n elementen. Indien $n\bar{x} \geq \xi$, dan is de procedure voltooid. Indien echter $n\bar{x} < \xi$, dan dient de gehele procedure herhaald te worden, d.w.z. er dient opnieuw een getal ξ gekozen en een aselecte steekproef getrokken te worden.

Een andere en eenvoudiger methode werd gegeven door MIDZUNO [12]. Hierbij wordt het eerste element getrokken met een kans, evenredig met de bijbehorende waarde x. Stel M is de waarde van de grootste van alle x_i in de populatie. Kies een aselect getal ξ met $0 \leq \xi < M'$ en $M' \geq M$. Trek aselect een element (y_i, x_i). Indien $x_i \geq \xi$, dan wordt dit element als eerste in de steekproef opgenomen. Indien $x_i < \xi$, dan wordt het getrokken element teruggelegd en wordt de procedure herhaald tot een element is getrokken. De overige n - 1 elementen worden op de normale wijze aselect uit de overgebleven N - 1 elementen van de populatie getrokken.

Nu geldt met de methode van MIDZUNO, dat (zie de Appendix):

$$P^*[\{\underline{x}_1,\ldots,\underline{x}_n\} = \{x_{i_1},\ldots,x_{i_n}\}] = \binom{N}{n}^{-1} \cdot \frac{\sum_{\nu=1}^{n} x_{i_\nu}}{n\overline{X}}, \qquad (6.1)$$

waarin $\{x_{i_1},\ldots,x_{i_n}\}$ een willekeurige greep van n elementen is. De schatter $\underline{y}$ is nu zuiver want

$$\mathcal{E}^*\underline{\tilde{y}} = \sum_{as} \tilde{y} \cdot P^*[\{\underline{x}_1,\ldots,\underline{x}_n\} = \{x_{i_1},\ldots,x_{i_n}\}] =$$

$$= \sum_{as} \left\{\overline{X}\frac{\overline{y}}{\overline{x}} \cdot \binom{N}{n}^{-1} \frac{\overline{x}}{\overline{X}}\right\} =$$

$$= \binom{N}{n}^{-1} \sum_{as} \overline{y} = \overline{Y}, \qquad (6.2)$$

waarin $\sum_{as}$ de sommatie betekent over alle $\binom{N}{n}$ mogelijke steekproeven van n uit N elementen.

Voor de variantie van $\underline{\tilde{y}}$ is geen exacte uitdrukking bekend. In de Appendix wordt aangetoond, dat voor een eerste benadering $Var_1^*(\underline{\tilde{y}})$ van $Var^*(\underline{\tilde{y}})$ geldt

$$Var_1^*(\underline{\tilde{y}}) = Var_1(\underline{\tilde{y}}). \qquad (6.3)$$

Het is echter mogelijk $Var^*(\underline{\tilde{y}})$ zuiver te schatten met

$$var^*(\underline{\tilde{y}}) = \underline{\tilde{y}}^2 - \frac{\overline{X}}{nN\underline{\overline{x}}}\left(\sum_{i=1}^{n} \underline{y}_i^2 + \frac{N-1}{n-1}\sum_{i\neq j=1}^{n} \underline{y}_i\underline{y}_j\right), \qquad (6.4)$$

hetgeen gemakkelijk te verifiëren is.

In [13] wordt uitvoeriger op een en ander ingegaan.

De schatters $\underline{\hat{y}}$ en $\underline{\hat{y}}'$

7. Verband tussen $\underline{\hat{y}}$ en $\underline{\hat{y}}'$

Bij de quotiëntschatter $\underline{\hat{y}}$, gegeven door (2.3), wordt de waarde van $\overline{X}$ bekend en $\neq 0$ en tevens $x_i \neq 0$ ($i = 1,\ldots,N$) ondersteld. Door HARTLEY en ROSS [7] werd voor de eerste maal de onzuiverheid $B(\underline{\hat{y}})$ van $\underline{\hat{y}}$ exact bepaald:

$$B(\hat{\underline{y}}) = \mathcal{E}\hat{\underline{y}} - \overline{Y} = \overline{X}\,.\,\mathcal{E}\overline{\underline{r}} - \mathcal{E}\overline{\underline{y}} =$$

$$= \mathcal{E}\overline{\underline{x}}\,.\,\mathcal{E}\overline{\underline{r}} - \mathcal{E}\overline{\underline{x}}\,.\,\overline{\underline{r}} = -\mathrm{Cov}\,(\overline{\underline{x}},\overline{\underline{r}}). \qquad (7.1)$$

Deze onzuiverheid kan verder zuiver geschat worden met

$$\underline{b}(\hat{\underline{y}}) = -\frac{N-1}{N}\,.\,\underline{s}_{xr} = -\frac{n}{n-1}\,.\,\frac{N-1}{N}\,(\overline{\underline{y}} - \overline{\overline{\underline{xr}}})\,, \qquad (7.2)$$

zodat een zuivere schatter

$$\hat{\underline{y}}' = \overline{X}\overline{\underline{r}} + \frac{n(N-1)}{N(n-1)}\,.\,(\overline{\underline{y}} - \overline{\overline{\underline{xr}}}) \qquad (7.3)$$

van $\overline{Y}$ gevonden kan worden.

De schatter $\hat{\underline{y}}$ is slechts dan "asymptotisch raak" indien $\overline{Y} = \overline{X}\overline{R}$, zoals direct uit (2.3) blijkt. Hij is zuiver indien $y_i = \overline{R}x_i$ $(i = 1, \ldots, N)$ en ook indien de regressie van $\underline{y}$ op x gegeven wordt door een rechte door de oorsprong.

8. De variantie van $\hat{\underline{y}}$

De variantie van $\hat{\underline{y}}$ kan exact bepaald worden:

$$\mathrm{Var}(\hat{\underline{y}}) = \overline{X}^2\,.\,\mathrm{Var}(\overline{\underline{r}}) = \overline{X}^2\,.\,\frac{N-n}{N-1}\,.\,\frac{\mu_{002}}{n}\,. \qquad (8.1)$$

Uit (8.1) blijkt direct, dat $\mathrm{Var}(\hat{\underline{y}})$ afneemt met toenemende n en een functie is van de correlatie tussen $\underline{x}$ en $\underline{y}$. Verder is $\mathrm{Var}(\hat{\underline{y}}) = 0$ voor $n = N$.

De doeltreffendheid van $\hat{\underline{y}}$ t.o.v. $\overline{\underline{y}}$ is

$$\frac{\mathrm{Var}(\overline{\underline{y}})}{\mathrm{Var}(\hat{\underline{y}})} = \frac{\mu_{020}}{\overline{X}^2\mu_{002}} \qquad (8.2)$$

en deze is groter dan 1 indien $\overline{X}^2\mu_{002} < \mu_{020}$.

Een zuivere schatter van $\mathrm{Var}(\hat{\underline{y}})$ is

$$\mathrm{var}(\hat{\underline{y}}) = \frac{N-n}{nN}\,.\,\overline{X}^2\underline{s}_r^2\,. \qquad (8.3)$$

De schatter $\tilde{\underline{y}}$ is in het algemeen te prefereren boven de schatter $\hat{\underline{y}}$ omdat de laatste meestal niet en de eerste altijd "asymptotisch raak" is. Verder wordt de onzuiverheid van $\hat{\underline{y}}$ niet kleiner voor toenemende n en is $\tilde{\underline{y}}$ eenvoudiger uit de waarnemingen te berekenen dan $\hat{\underline{y}}$. Voor de kwestie van de eventuele asymptotische normaliteit van deze schatters wordt verwezen naar paragraaf 11.

9. De variantie van $\hat{\underline{y}}'$

Een directe berekening van Var($\hat{\underline{y}}'$) is zeer tijdrovend en lastig. Hierbij vooral bewijzen de z.g.n. "symmetrische gemiddelden" (afk.: sg; sgn) goede diensten. ROBSON [15] berekende Var($\hat{\underline{y}}'$) met behulp van 3-dimensionale sgn, terwijl hij als nevenresultaat een zuivere schatter van deze variantie verkreeg.

De gebruikte sgn worden als volgt gedefinieerd:

$$\langle(\alpha_1\beta_1\gamma_1),\ldots,(\alpha_p\beta_p\gamma_p)\rangle =$$

$$= \frac{1}{n(n-1)\ldots(n-p+1)} \sum_{i_1\neq\ldots\neq i_p=1}^{n} (\underline{x}_{i_1}^{\alpha_1}\underline{y}_{i_1}^{\beta_1}\underline{r}_{i_1}^{\gamma_1}) \cdot \ldots \cdot (\underline{x}_{i_p}^{\alpha_p}\underline{y}_{i_p}^{\beta_p}\underline{r}_{i_p}^{\gamma_p}), \tag{9.1}$$

$$\langle(\alpha_1\beta_1\gamma_1),\ldots,(\alpha_p\beta_p\gamma_p)\rangle' =$$

$$= \frac{1}{N(N-1)\ldots(N-p+1)} \sum_{i_1\neq\ldots\neq i_p=1}^{N} (x_{i_1}^{\alpha_1}y_{i_1}^{\beta_1}r_{i_1}^{\gamma_1}) \cdot \ldots \cdot (x_{i_p}^{\alpha_p}y_{i_p}^{\beta_p}r_{i_p}^{\gamma_p}). \tag{9.2}$$

Verder is gemakkelijk na te gaan, dat

$$\mathcal{E}\langle(\alpha_1\beta_1\gamma_1),\ldots,(\alpha_p\beta_p\gamma_p)\rangle = \langle(\alpha_1\beta_1\gamma_1),\ldots,(\alpha_p\beta_p\gamma_p)\rangle' . \tag{9.3}$$

Door bijv. alle γ's gelijk 0 te stellen worden 2-dimensionale sgn verkregen terwijl men 1-dimensionale verkrijgt door voor bijv. alle β's en γ's 0 te substitueren. In de Appendix worden voor 1-dimensionale sgn enkele vermenigvuldigingsregels afgeleid. Voor verdere bijzonderheden wordt verwezen naar [17], [18] en [15].

ROBSON vond voor Var($\hat{\underline{y}}'$):

$$\begin{aligned}
\mathrm{Var}(\hat{\underline{y}}') = \frac{N-n}{nN} \Big[& \langle (020) \rangle' - \langle (010),(010) \rangle' + \\
& + \langle (200),(001),(001) \rangle' + 2 \langle (100),(010),(001) \rangle' + \\
& - 2 \langle (110),(001) \rangle' - \frac{N+1}{N} \langle (100),(100),(001),(001) \rangle' + \\
& + \frac{N-n}{(n-1)N} \{ \langle (010),(010) \rangle' - 2 \langle (100),(010),(001) \rangle' \} + \\
& + \frac{N-1}{(n-1)N} \{ \langle (200),(002) \rangle' - \langle (100),(100),(002) \rangle' + \\
& - \langle (200),(001),(001) \rangle' + 2 \langle (100),(100),(001),(001) \rangle' \} \Big].
\end{aligned} \tag{9.4}$$

Een zuivere schatter var($\hat{\underline{y}}'$) van Var($\hat{\underline{y}}'$) wordt eenvoudig gevonden door in (9.4) de accenten van de sgn te schrappen.

De in (9.4) voorkomende sgn dienen nog in de momenten $\mu_{\alpha\beta\gamma}$ (sgn zonder accent in de steekproefmomenten $\underline{s}_{\alpha\beta}$) "vertaald" te worden. In de Appendix wordt de daartoe te gebruiken methode aangeduid.

Voor oneindig grote populaties, dus voor $N \to \infty$ of bij steekproeven zonder teruglegging, neemt de schatter $\hat{\underline{y}}'$ de volgende gedaante aan:

$$\hat{\underline{y}}' = \bar{X}\bar{\underline{r}} + \frac{n}{n-1}(\bar{\underline{y}} - \overline{\underline{x}\underline{r}}). \tag{9.5}$$

Voor $N \to \infty$ geldt verder, dat (zie bijv. (A9.2))

$$\langle (\alpha_1\beta_1\gamma_1), \ldots, (\alpha_p\beta_p\gamma_p) \rangle' = \mathcal{E}\underline{x}^{\alpha_1}\underline{y}^{\beta_1}\underline{r}^{\gamma_1} \cdot \ldots \cdot \mathcal{E}\underline{x}^{\alpha_p}\underline{y}^{\beta_p}\underline{r}^{\gamma_p}. \tag{9.6}$$

(Voor sgn van andere dimensies geldt uiteraard een zelfde relatie als (9.6)). Met (9.6) ziet men direct, dat (9.4) voor $N \to \infty$ overgaat in

$$\begin{aligned}
\mathrm{Var}(\hat{\underline{y}}') = {} & \frac{1}{n}(\mathcal{E}\underline{y}^2 - \bar{Y}^2 + \bar{\bar{R}}^2\mathcal{E}\underline{x}^2 + 2\bar{X}\bar{Y}\bar{\bar{R}} - 2\bar{\bar{R}}\mathcal{E}\underline{x}\underline{y} - \bar{X}^2\bar{\bar{R}}^2) + \\
& + \frac{1}{n(n-1)}(\bar{Y}^2 - 2\bar{X}\bar{Y}\bar{\bar{R}} + \mathcal{E}\underline{x}^2 \cdot \mathcal{E}\underline{r}^2 - \bar{X}\mathcal{E}\underline{r}^2 - \bar{\bar{R}}\mathcal{E}\underline{x}^2 - 2\bar{X}^2\bar{\bar{R}}^2) = \\
= {} & \frac{1}{n}\{\mathrm{Var}(\underline{y}) + \bar{\bar{R}}^2\,\mathrm{Var}(\underline{x}) - 2\bar{\bar{R}}^2\mathrm{Cov}\,(\underline{x},\underline{y})\} +
\end{aligned}$$

$$+ \frac{1}{n(n-1)} \{Var(\underline{x}) \cdot Var(\underline{r}) + Cov^2(\underline{x},\underline{r})\} \qquad (9.7)$$

Door in (9.4) de accenten te schrappen en de factoren $\frac{N-n}{N}$, $\frac{N+1}{N}$ en $\frac{N-1}{N}$ door 1 te vervangen vindt men een zuivere schatter var($\underline{\hat{y}}'$) van Var($\underline{\hat{y}}'$). De sgn zonder accent moeten nu nog "vertaald" worden in de steekproefmomenten $\underline{s}_{\alpha\beta}$. In de Appendix wordt voor iedere in (9.4) voorkomende sg het resultaat van de vertaling vermeld.

10. Een eenvoudige schatter van de variantie van $\underline{\hat{y}}'$

Het schatten van Var($\underline{\hat{y}}'$) met behulp van (9.4) is bewerkelijk en tijdrovend. Een andere schattingsmethode, ontleend aan [4], wordt hieronder beschreven.

Uit de verzameling van alle $\binom{N}{m}$ steekproeven zonder teruglegging van m elementen worden aselect zonder teruglegging k steekproeven getrokken. Dit betekent, dat k maal zonder teruglegging m elementen (y_i, x_i) worden getrokken en dat deze elementen worden teruggelegd vóórdat de volgende steekproef van m elementen wordt getrokken. Indien een steekproef geheel identiek is aan een reeds eerder getrokkene, dan wordt deze steekproef als niet getrokken beschouwd.

Uit elk der k steekproeven wordt $\overline{\overline{Y}}$ geschat met

$$\underline{\hat{y}}'_j = \overline{\overline{X}} \cdot \overline{\overline{\underline{r}}}_j + \frac{m(N-1)}{(m-1)N} (\overline{\overline{\underline{y}}}_j - \overline{\overline{\underline{x}}}_j \overline{\overline{\underline{r}}}_j) \qquad (10.1)$$

voor j = 1,...., k. De schatter

$$\underline{\hat{y}}' = \frac{1}{k} \sum_{j=1}^{k} \underline{\hat{y}}'_j \qquad (10.2)$$

is een zuivere schatter van $\overline{\overline{Y}}$. De variantie van $\underline{\hat{y}}'$ is nu

$$Var(\underline{\hat{y}}') = \frac{p-k}{pk} \left\{ \frac{1}{p} \sum_{j=1}^{p} (\hat{y}'_j)^2 - \frac{1}{p(p-1)} \sum_{j_1 \neq j_2 = 1}^{p} \hat{y}'_{j_1} \cdot \hat{y}'_{j_2} \right\} \qquad (10.3)$$

met $p = \binom{N}{m}$. Een zuivere schatter van Var($\underline{\hat{y}}'$) is dus

$$\mathrm{var}(\hat{\underline{y}}') = \frac{p-k}{pk}\left\{\frac{1}{k}\sum_{j=1}^{k}(\hat{\underline{y}}'_j)^2 - \frac{1}{k(k-1)}\sum_{j_1\neq j_2=1}^{k}\hat{\underline{y}}'_{j_1}\cdot\hat{\underline{y}}'_{j_2}\right\} =$$

$$= \frac{p-k}{p(k-1)}\left\{\frac{1}{k}\sum_{j=1}^{k}(\hat{\underline{y}}'_j)^2 - (\hat{\underline{y}}')^2\right\}. \tag{10.4}$$

Voor k = 2 wordt dit

$$\mathrm{var}(\hat{\underline{y}}') = \frac{p-2}{4p}(\hat{\underline{y}}'_1 - \hat{\underline{y}}'_2)^2, \tag{10.5}$$

een vorm, die in [4] gegeven wordt. Voor $N \to \infty$ gaat de factor $\frac{p-2}{p}$ in 1 over.

11. Betrouwbaarheidsintervallen

Bij de practische toepassingen neemt men aan, dat de hier behandelde schatters voor grote steekproefomvang een bij benadering normale verdeling bezitten met verwachting $\overline{\overline{Y}}$ (voor de schatter $\tilde{\underline{r}}$ met verwachting $\overline{\overline{Y}} \cdot \overline{\overline{X}}^{-1}$ en voor de schatter $\hat{\underline{y}}$ met verwachting $\overline{\overline{Y}}$ - Cov $(\underline{x},\underline{r})$) en variantie Var (schatter).

MADOW [10], ERDÖS en RÉNYI [2] en HAJEK [5] bewezen, onder zekere voorwaarden, stellingen omtrent asymptotische normaliteit, welke direct op de schatters $\overline{\underline{y}}$ en $\tilde{\underline{y}}$ toegepast kunnen worden. MADOW [10] bewees verder dat, onder zekere voorwaarden, $\overline{\underline{y}}$ en $\overline{\underline{x}}$ asymptotisch een simultane normale verdeling bezitten. Met een stelling van CRAMÉR [1, p.366] kan dan aangetoond worden, dat de schatters $\tilde{\underline{y}}$, $\tilde{\underline{r}}$ en $\hat{\underline{y}}'$ asymptotisch normaal verdeeld zijn.

De voorwaarden, waarover hierboven wordt gesproken, worden hier niet vermeld, gedeeltelijk omdat moeilijk nagegaan kan worden of de te onderzoeken populatie in concrete gevallen aan deze voorwaarden voldoet.

Voor grote n worden benaderde tweezijdige betrouwbaarheidsgrenzen voor $\overline{\overline{Y}}$ gevonden met

$$\tilde{\underline{y}} \pm \xi_{\alpha/2}\sqrt{\mathrm{var}_1(\tilde{\underline{y}})} \tag{11.1}$$

(met analoge vormen voor de andere schatters), waarin ξ_α gevonden wordt uit

$$\frac{1}{\sqrt{2\pi}} \int_{\xi_\alpha}^{\infty} \exp(-\tfrac{1}{2}t^2)dt = \alpha. \qquad (11.2)$$

Met de schatter $\tilde{\underline{r}}$ vindt men betrouwbaarheidsgrenzen voor $\bar{Y} \cdot \bar{X}^{-1}$ en met de schatter $\hat{\underline{y}}$ betrouwbaarheidsgrenzen voor $\bar{Y} - \mathrm{Cov}(\underline{x},\underline{r})$.

Door FIELLER [3], PAULSON [14] en KLERK-GROBBEN [8] werd uitgegaan van de veronderstelling, dat $\underline{x}$ en $\underline{y}$ een simultane normale verdeling bezitten, zodat ook

$$\bar{\underline{y}} - \frac{\bar{Y}}{\bar{X}}\bar{\underline{x}}$$

normaal verdeeld is. Zij bewezen, dat

$$\underline{z} = \frac{(\bar{X}\bar{\underline{y}} - \bar{Y}\bar{\underline{x}})\sqrt{nN}}{(N-n)^{\frac{1}{2}}(\bar{Y}^2\underline{s}_x^2 - 2\bar{X}\bar{Y}\underline{s}_{xy} + \bar{X}^2\underline{s}_y^2)^{\frac{1}{2}}} \qquad (11.3)$$

voor grote n bij benadering $N(0;1)$ verdeeld is. Voor kleine n heeft $\underline{z}$ bij benadering een Student-verdeling met $n - 1$ vrijheidsgraden. Tweezijdige betrouwbaarheidsgrenzen voor $\bar{Y} \cdot \bar{X}^{-1}$, met onbetrouwbaarheid α, worden dan gevonden door het tweede lid van (11.3) gelijk aan $\pm\xi_\alpha$ te stellen en de zo ontstane vergelijkingen van de tweede graad in $\bar{Y} \cdot \bar{X}^{-1}$ op te lossen.

HAJEK [6] toonde aan, dat, indien de schatter $\tilde{\underline{y}}$ asymptotisch normaal verdeeld is en aan (3.6) (en dus ook aan (3.7)) is voldaan, de onbetrouwbaarheid van het betrouwbaarheidsinterval met de grenzen

$$\underline{y} \pm t_{\alpha/2} \cdot \tilde{\underline{y}} \sqrt{\frac{1}{n}\left(\frac{\underline{s}_x^2}{\bar{\underline{x}}^2} - 2\frac{\underline{s}_{xy}}{\bar{\underline{x}}\bar{\underline{y}}} + \frac{\underline{s}_y^2}{\bar{\underline{y}}^2}\right)}, \qquad (11.4)$$

waarin t_α gevonden wordt uit $P[\underline{t} \geq t_\alpha] = \alpha$, waar $\underline{t}$ een Student-verdeling met $n - 1$ vrijheidsgrenzen bezit, kleiner is dan α.

12. De doeltreffendheid van de behandelde schatters

De toepassing van de schatters $\tilde{\underline{y}}$ en $\hat{\underline{y}}$ is slechts dan aan te bevelen indien de regressiekromme van $\underline{y}$ op $\underline{x}$ een rechte door de oorsprong is.

Bij twijfel kan dit laatste eerst getoetst worden. Indien deze kromme niet door de oorsprong gaat bestaat de mogelijkheid van onzuiverheid en grote varianties. In deze paragraaf wordt dan ook aangenomen, dat de genoemde regressiekromme een rechte door de oorsprong is zodat de besproken schatters alle zuiver zijn. Verder wordt voor de eenvoud een zeer grote of oneindig grote populatie ($N \to \infty$) ondersteld. Alle trekkingen in de steekproef zijn dan onderling onafhankelijk.

Om de doeltreffendheid van de schatters te kunnen bepalen wordt eerst een zuivere quotiëntschatter, lineair in de $\underline{r}_i$, met minimale voorwaardelijke variantie afgeleid. Stel

$$\mathcal{E}(\underline{y} \mid x) = ax \quad (a > 0 \text{ en constant}) \tag{12.1}$$

en

$$\mathrm{Var}(\underline{r} \mid x) = g(x) \quad (g(x) > 0 \text{ voor alle } x). \tag{12.2}$$

Dan volgt

$$\bar{Y} = a\bar{X}\ , \ \mathcal{E}(\underline{r} \mid x) = a. \tag{12.3}$$

In de Appendix wordt verder aangetoond, dat de schatter

$$\underline{z} = \bar{X} \left\{ \sum_{i=1}^{n} \frac{1}{g(\underline{x}_i)} \right\}^{-1} \sum_{i=1}^{n} \frac{\underline{r}_i}{g(\underline{x}_i)} \tag{12.4}$$

met

$$\mathrm{Var}(\underline{z}) = \bar{X}^2 \cdot \mathcal{E}\left\{ \sum_{i=1}^{n} \frac{1}{g(\underline{x}_i)} \right\}^{-1} \tag{12.5}$$

aan de gestelde eisen voldoet.

Volgens een stelling van CRAMÉR [1, pp. 353-4] mag $\mathrm{Var}(\underline{z})$ voor zeer grote n geschreven worden als

$$\mathrm{Var}(\underline{z}) \approx \frac{1}{n} \bar{X}^2 \left\{ \mathcal{E} \frac{1}{g(\underline{x})} \right\}^{-1}. \tag{12.6}$$

De doeltreffendheid van de schatters $\bar{\underline{y}}$, $\tilde{\underline{y}}$, $\hat{\underline{y}}$ en $\hat{\underline{y}}'$ wordt gemeten t.o.v. de schatter $\underline{z}$. Hun varianties worden in de Appendix eerst in de x_i en de $g(x_i)$ uitgedrukt. De doeltreffendheid luidt nu bij benadering:

$$\mathrm{Eff}(\bar{\bar{\underline{y}}}) \approx \bar{\bar{X}}^2\left[\mathcal{E}\frac{1}{g(\underline{x})} \cdot \{\mathcal{E}\underline{x}^2 g(\underline{x}) + a^2\mathrm{Var}(\underline{x})\}\right]^{-1}; \qquad (12.7)$$

$$\mathrm{Eff}(\tilde{\underline{y}}) \approx \{\mathcal{E}\underline{x}^2 + (n - 1)\bar{\bar{X}}^2\} \cdot \{n\,\mathcal{E}\frac{1}{g(\underline{x})} \cdot \mathcal{E}\underline{x}^2 g(\underline{x})\}^{-1}; \qquad (12.8)$$

$$\mathrm{Eff}(\hat{\underline{y}}) \approx \{\mathcal{E}g(\underline{x}) \cdot \mathcal{E}\frac{1}{g(\underline{x})}\}^{-1}; \qquad (12.9)$$

$$\mathrm{Eff}(\hat{\underline{y}}') \approx \bar{\bar{X}}^2\{\mathcal{E}\frac{1}{g(\underline{x})} \cdot \mathcal{E}\underline{x}^2 g(\underline{x})\}^{-1}. \qquad (12.10)$$

Uit (12.7) en (12.8) leidt men direct af, dat $\mathrm{Eff}(\tilde{\underline{y}}) > \mathrm{Eff}(\bar{\bar{\underline{y}}})$, hoewel het verschil afneemt met toenemende n. Verder blijkt direct uit (12.7) en (12.10), dat $\mathrm{Eff}(\hat{\underline{y}}') > \mathrm{Eff}(\bar{\bar{\underline{y}}})$. De keuze tussen de schatters $\tilde{\underline{y}}$ en $\hat{\underline{y}}'$ is moeilijker, maar kan bepaald worden door de overweging, dat $\hat{\underline{y}}'$ altijd zuiver is, ook als niet aan (12.1) is voldaan.

Van de functie $g(x)$ worden drie vormen beschouwd:

$$g(x) = bx^{-2},\quad g(x) = bx^{-1} \text{ en } g(x) = b,$$

met $b > 0$. In onderstaand schema wordt de situatie voor deze drie vormen van $g(x)$ voor iedere schatter afzonderlijk weergegeven. Daaruit blijkt dat de doeltreffendheid van de schatter $\tilde{\underline{y}}$ voor $g(x) = bx^{-1}$ groter is dan 1. Dit is mogelijk omdat de schatter $\underline{z}$ wel een minimale voorwaardelijke variantie maar kennelijk geen minimale variantie onder alle mogelijke schatters van $\bar{\bar{Y}}$ bezit.

	$g(x) = \frac{b}{x^2}$	$g(x) = \frac{b}{x}$	$g(x) = b$
$\bar{\underline{y}}$	$\frac{b\bar{X}^2}{\mathcal{E}\underline{x}^2\{b + a^2 Var(\underline{x})\}}$	$1 - \frac{a^2 Var(\underline{x})}{b\bar{X} + a^2 Var(\underline{x})}$	$\frac{b\bar{X}^2}{b\mathcal{E}\underline{x}^2 + a^2 Var(\underline{x})}$
$\tilde{\underline{y}}$	$1 - \frac{n - 1}{n} \cdot \frac{Var(\underline{x})}{\mathcal{E}\underline{x}^2}$	$1 + \frac{Var(\underline{x})}{n\bar{X}^2}$	$1 - \frac{n - 1}{n} \cdot \frac{Var(\underline{x})}{\mathcal{E}\underline{x}^2}$
$\hat{\underline{y}}$	$\frac{1}{\mathcal{E}\underline{x}^2 \cdot \mathcal{E}\underline{x}^{-2}}$	$\frac{1}{\bar{X}\,\mathcal{E}\underline{x}^{-1}}$	1
$\hat{\underline{y}}'$	$1 - \frac{Var(\underline{x})}{\mathcal{E}\underline{x}^2}$	1	$1 - \frac{Var(\underline{x})}{\mathcal{E}\underline{x}^2}$

Appendix

De nummering van de Appendix loopt parallel met de nummering van de paragrafen waarop toelichting wordt gegeven.

A3. Met (3.2) werden reeds de grootheden $S\underline{x}$ en $S\underline{y}$ gedefiniëerd. Men verifiëert gemakkelijk de relatie

$$\tilde{\underline{y}} - \bar{Y} = \bar{Y}(S\underline{y} - S\underline{x}) - S\underline{x}(\tilde{\underline{y}} - \bar{Y}). \tag{A3.1}$$

Met volledige inductie in m kan dan verder de betrekking

$$\tilde{\underline{y}} - \bar{Y} = \bar{Y}(S\underline{y} - S\underline{x}) \sum_{\nu=0}^{m} (-S\underline{x})^{\nu} + (-S\underline{x})^{m+1}(\tilde{\underline{y}} - \bar{Y}) \tag{A3.2}$$

bewezen worden. Omdat hier slechts van een eindig aantal termen sprake is, kunnen links en rechts verwachtingen worden genomen waaruit dan direct (3.1) volgt.

Voor $m = 1$ geldt dus

$$\mathcal{E}\tilde{\underline{y}} = \bar{Y} + \bar{Y}\{\mathcal{E}(S\underline{x})^2 - \mathcal{E}(S\underline{x})(S\underline{y})\} + \mathcal{E}(S\underline{x})^2(\tilde{\underline{y}} - \bar{Y}). \tag{A3.3}$$

Blijft nu nog over aan te tonen, dat $|\mathcal{E}(S\underline{x})^2(\tilde{\underline{y}} - \bar{Y})|$ minstens van de orde n^{-2} is. Met (A3.2) en $m = 1$ vindt men

$$\begin{aligned}(S\underline{x})^2(\tilde{\underline{y}} - \bar{Y}) &= \bar{Y}\{(S\underline{x})^2(S\underline{y}) - (S\underline{x})^3\}\ (1 - S\underline{x}) + (S\underline{x})^4(\tilde{\underline{y}} - \bar{Y}) = \\ &= \bar{Y}\{(S\underline{x})^2(S\underline{y}) - (S\underline{x})^3 - (S\underline{x})^3(S\underline{y})\} + (S\underline{x})^4\tilde{\underline{y}}.\end{aligned} \tag{A3.4}$$

De verwachting van de eerste term van het laatste lid van (A3.4) is volgens (A3.11) en A3.12) van de orde n^{-2}. Omdat de populatie eindig is, heeft $|\tilde{\underline{y}}|$ zeker een maximum M. Dus

$$|\mathcal{E}(S\underline{x})^4\tilde{\underline{y}}| \leq \mathcal{E}(S\underline{x})^4|\tilde{\underline{y}}| \leq M\ \mathcal{E}(S\underline{x})^4 \tag{A3.5}$$

terwijl $\mathcal{E}(S\underline{x})^4$ volgens (A3.12) van de orde n^{-2} is.

Een bevredigender afschatting van $\mathcal{E}(S\underline{x})^4\tilde{\underline{y}}$ wordt verkregen indien men onderstelt, dat $x_i > 0$ en $y_i > 0$ voor $i = 1, \ldots, N$. Dan kan namelijk de volgende ongelijkheid afgeleid worden:

$$\tilde{\underline{y}} = \bar{X}\,\frac{\sum\limits_{i=1}^{n} \underline{y}_i}{\sum\limits_{i=1}^{n} \underline{x}_i} = \bar{X}\,\frac{\sum\limits_{i=1}^{n} \underline{x}_i\underline{r}_i}{\sum\limits_{i=1}^{n} \underline{x}_i} = \bar{X}\sum_{i=1}^{n}\frac{\underline{x}_i}{\sum\limits_{i=1}^{n} \underline{x}_i}\,\underline{r}_i \leq$$

$$\leq \bar{X}\sum_{i=1}^{n}\underline{r}_i = n\bar{X}\bar{\underline{r}}. \qquad \text{(A3.6)}$$

Dus, indien men op dezelfde wijze als $S\underline{x}$ en $S\underline{y}$ de grootheid $S\underline{r}$ definieert,

$$\mathcal{E}(S\underline{x})^4\tilde{\underline{y}} \leq n\bar{X}\,\mathcal{E}\bar{\underline{r}}(S\underline{x})^4 = n\bar{X}\bar{R}\,\mathcal{E}(1 + S\underline{r})(S\underline{x})^4, \qquad \text{(A3.7)}$$

waaruit volgt, dat $\mathcal{E}(S\underline{x})^4\tilde{\underline{y}}$ van minstens de orde n^{-2} is.

In het bovenstaande werden geen bijzondere eisen aan de populatie gesteld. Meestal wordt de volgende afleiding gegeven (zie bijvoorbeeld [16]):

$$\tilde{\underline{y}} = \bar{X}\,\frac{\bar{\underline{y}}}{\bar{\underline{x}}} = \bar{Y}\,\frac{1 + S\underline{y}}{1 + S\underline{x}} = \bar{Y}(1 + S\underline{y})\sum_{\nu=0}^{\infty}(-S\underline{x})^{\nu}. \qquad \text{(A3.8)}$$

Hierbij wordt dan de eis gesteld, dat $|S\underline{x}| < 1$ voor elke mogelijke steekproef van de omvang n. Verder neemt men in (A3.8) links en rechts verwachtingen zonder de zekerheid, dat dan het gelijkteken blijft gelden.

——— x ———

Hieronder volgen voor $\alpha + \beta \leq 4$ de verwachtingen van $(S\underline{x})^{\alpha}(S\underline{y})^{\beta}$. Alle niet vermelde momenten met $\alpha + \beta \leq 4$ kunnen door verwisseling of door identificatie van $\underline{x}$ en $\underline{y}$ uit de vermelde afgeleid worden.

$$\mathcal{E}(S\underline{x}) = \mathcal{E}(S\underline{y}) = 0; \qquad \text{(A3.9)}$$

$$\mathcal{E}(S\underline{x})(S\underline{y}) = \frac{N - n}{N - 1}\cdot\frac{\mu_{11}}{n\bar{X}\bar{Y}}; \qquad \text{(A3.10)}$$

$$\mathcal{E}(S\underline{x})^2(S\underline{y}) = \frac{(N - n)(N - 2n)}{(N - 1)(N - 2)}\cdot\frac{\mu_{21}}{n^2\bar{X}^2\bar{Y}}; \qquad \text{(A3.11)}$$

$$\mathcal{E}(S_x)^3(S_y) = \frac{N-n}{(N-1)(N-2)(N-3)}\left\{(N^2-6nN+6n^2+N)\frac{\mu_{31}}{n^3\bar{X}^3\bar{Y}} + 3(n-1)N(N-n-1)\frac{\mu_{20}\mu_{11}}{n^3\bar{X}^3\bar{Y}}\right\}; \tag{A3.12}$$

$$\mathcal{E}(S_x)^2(S_y)^2 = \frac{N-n}{(N-1)(N-2)(N-3)}\left\{(N^2-6nN+6n^2+N)\frac{\mu_{22}}{n^3\bar{X}^2\bar{Y}^2} + (n-1)N(N-n-1)\frac{\mu_{20}\mu_{02}+2\mu_{11}^2}{n^3\bar{X}^2\bar{Y}^2}\right\}. \tag{A3.13}$$

———— × ————

De afleiding van (3.11) gaat als volgt:

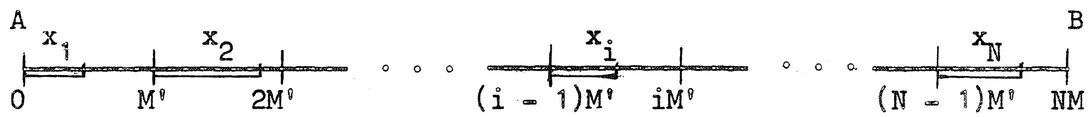
$$\mathrm{Cov}(\tilde{y},\bar{x}) = \bar{X}\bar{Y} - \bar{X}\,\mathcal{E}\tilde{y} = -\bar{X}B(\tilde{y}), \tag{A3.14}$$

$$\mathrm{Cov}(\tilde{y},\bar{x}) = \zeta(\tilde{y},\bar{x}) \cdot \{\mathrm{Var}(\tilde{y}) \cdot \mathrm{Var}(\bar{x})\}^{\frac{1}{2}}, \tag{A3.15}$$

$$\mathrm{Var}(\bar{x}) = \bar{X}^2 \cdot (S_x)^2 = \frac{N-n}{N-1} \cdot \frac{\mu_{20}}{n}. \tag{A3.16}$$

Met $|\zeta| \leq 1$ volgt hieruit direct (3.11).

A6. Het hieronder volgende bewijs van (6.1) is gebaseerd op een idee van LAHIRI [9]. Op een rechte wordt een lijnstuk AB ter lengte NM' in N gelijke delen ter lengte M' verdeeld. Elk van deze N delen wordt weer onderverdeeld in een stuk ter lengte x_i en een stuk ter lengte $M' - x_i$ $(i = 1,\ldots,N)$. Zie ook de bijgaande figuur.



Men kiest nu aselect een punt op AB op een afstand $(\nu_1 - 1)M' + \xi_1$ van A door eerst aselect een natuurlijk getal ν_1 $(1 \leq \nu_1 \leq N)$ en daarna aselect een reëel getal ξ_1 $(0 \leq \xi_1 < M',\ M' \geq M)$ te kiezen. Indien $\xi_1 > x_{\nu_1}$ kiest men een geheel nieuw tweetal getallen ν_2 $(1 \leq \nu_2 \leq N)$

en ξ_2 $(0 \leq \xi_2 < M')$, enz. Na $\underline{j} - 1$ herhalingen geldt voor de eerste maal $\xi_{\underline{j}} \leq x_{\nu_{\underline{j}}}$, met

$$P[\underline{j} = j] = \frac{\bar{X}}{M'}\left(1 - \frac{\bar{X}}{M'}\right)^{j-1} . \qquad (A6.1)$$

Men ziet gemakkelijk in, dat

$$P[\underline{\nu}_{\underline{j}} = i, \underline{j} = j] = \frac{\bar{X}}{M'}\left(1 - \frac{\bar{X}}{M'}\right)^{j-1} \cdot \frac{x_i}{N\bar{X}} = \left(1 - \frac{\bar{X}}{M'}\right)^{j-1} \cdot \frac{x_i}{NM'} \qquad (A6.2)$$

dus

$$P[\underline{\nu}_{\underline{j}} = i] = \sum_{j=1}^{\infty} P[\underline{\nu}_{\underline{j}} = i, \underline{j} = j] = \frac{x_i}{N\bar{X}} . \qquad (A6.3)$$

Er zijn, bij gegeven steekproefomvang n, $\binom{N}{n}$ verschillende steekproeven mogelijk. De werkelijk getrokken elementen worden hierna met $\{x_{i_1}, \ldots, x_{i_n}\}$ weergegeven. Dan geldt

$$P^*[\{\underline{x}_1, \ldots, \underline{x}_n\} = \{x_{i_1}, \ldots, x_{i_n}\}] =$$

$$= \sum_{\nu=1}^{n} P[\underline{\nu}_{\underline{j}} = i_\nu, \text{ daarna de andere } n - 1 \text{ elementen}] =$$

$$= \sum_{\nu=1}^{n} P[\underline{\nu}_{\underline{j}} = i_\nu] \cdot P[\text{de andere } n - 1 \text{ elementen} \mid \underline{\nu}_{\underline{j}} = \nu] =$$

$$= \sum_{\nu=1}^{n} \frac{x_{i_\nu}}{N\bar{X}} \cdot \binom{N-1}{n-1}^{-1} = \binom{N}{n}^{-1} \cdot \frac{\sum_{\nu=1}^{n} x_{i_\nu}}{n\bar{X}} . \qquad (A6.4)$$

Ter controle verifiëert men gemakkelijk, dat

$$\sum_{as} P^*[\{\underline{x}_1, \ldots, \underline{x}_n\} = \{x_{i_1}, \ldots, x_{i_n}\}] =$$

$$= \sum_{as} \binom{N}{n}^{-1} \cdot \frac{1}{\bar{X}} \cdot \frac{1}{n} \sum_{\nu=1}^{n} x_{i_\nu} = \frac{1}{\bar{X}} \cdot \bar{X} = 1 . \qquad (A6.5)$$

——— x ———

De variantie $\mathrm{Var}^*(\tilde{\underline{y}})$ van $\tilde{\underline{y}}$ wordt als volgt gevonden:

$$\mathcal{E}^*(\tilde{\underline{y}})^2 = \sum_{as} \left(\overline{X}\,\frac{\overline{\overline{y}}}{\overline{\overline{x}}}\right)^2 \cdot \binom{N}{n}^{-1} \frac{\overline{\overline{x}}}{\overline{X}} =$$

$$= \binom{N}{n}^{-1} \sum_{as} \overline{\overline{y}} \cdot \left(\overline{X}\,\frac{\overline{\overline{y}}}{\overline{\overline{x}}}\right) = \mathcal{E}\overline{\underline{y}}\tilde{\underline{y}}. \qquad (A6.6)$$

Dus

$$\mathrm{Var}^*(\tilde{\underline{y}}) = \mathcal{E}\overline{\underline{y}}\tilde{\underline{y}} - \overline{Y}^2 = \mathcal{E}\overline{\underline{y}}(\tilde{\underline{y}} - \overline{Y}) =$$

$$= \overline{Y} \cdot \mathcal{E}(1 + S\underline{y})\{\overline{Y}(S\underline{y} - S\underline{x}) \sum_{\nu=0}^{m} (-S\underline{x})^{\nu} + (-S\underline{x})^{m+1}(\tilde{\underline{y}} - \overline{Y})\}.$$

$$(A6.7)$$

Voor een eerste benadering $\mathrm{Var}_1^*(\tilde{\underline{y}})$ van $\mathrm{Var}^*(\tilde{\underline{y}})$, waarbij slechts verwachtingen van termen van de tweede graad in $S\underline{x}$ en $S\underline{y}$ worden meegeteld, vindt men derhalve:

$$\mathrm{Var}_1^*(\tilde{\underline{y}}) = \overline{Y}^2\{\mathcal{E}(S\underline{x})^2 - 2\mathcal{E}(S\underline{x})(S\underline{y}) + \mathcal{E}(S\underline{y})^2\} = \mathrm{Var}_1(\tilde{\underline{y}}). \qquad (A6.8)$$

A9. De vermenigvuldigingsregels voor sgn zijn niet moeilijk te vinden; ze worden uitvoerig behandeld in [18]. Hier volgen enkele voorbeelden:

$$(\sum_i x_i^{\,a}) \cdot (\sum_i x_i^{\,b}) = \sum_i x_i^{\,a+b} + \sum_{i\neq j} x_i^{\,a} x_j^{\,b}, \qquad (A9.1)$$

en dus

$$\langle a\rangle' \cdot \langle b\rangle' = \frac{1}{N}\langle a + b\rangle' + \frac{N-1}{N}\langle a, b\rangle' ; \qquad (A9.2)$$

$$(\sum_i x_i^{\,a}) \cdot (\sum_{i\neq j} x_i^{\,b} x_j^{\,c}) = \sum_{i\neq j} x_i^{\,a+b} x_j^{\,c} + \sum_{i\neq j} x_i^{\,b} x_j^{\,a+c} + \sum_{i\neq j\neq k} x_i^{\,a} x_j^{\,b} x_k^{\,c},$$

$$(A9.3)$$

en dus

$$\langle a\rangle' \cdot \langle b,c\rangle' = \frac{1}{N}\langle a + b, c\rangle' + \frac{1}{N}\langle a + c, b\rangle + \frac{N-2}{N}\langle a,b,c\rangle' ; \qquad (A9.4)$$

$$(\sum_{i\neq j} x_i^a x_j^b) \cdot (\sum_{i\neq j} x_i^c x_j^d) = \sum_{i\neq j} x_i^{a+c} x_j^{b+d} + \sum_{i\neq j} x_i^{a+d} x_j^{b+c} +$$

$$+ \sum_{i\neq j\neq k} x_i^{a+c} x_j^b x_k^d + \sum_{i\neq j\neq k} x_i^{a+d} x_j^b x_k^c + \sum_{i\neq j\neq k} x_i^a x_j^{b+c} x_k^d +$$

$$+ \sum_{i\neq j\neq k} x_i^a x_j^{b+d} x_k^c + \sum_{i\neq j\neq k\neq m} x_i^a x_j^b x_k^c x_m^d, \qquad (A9.5)$$

en dus

$$\langle a,b\rangle' \cdot \langle b,c\rangle' = \frac{1}{N(N-1)} \langle a+c, b+d\rangle' + \frac{1}{N(N-1)} \langle a+d, b+c\rangle' +$$

$$+ \frac{N-2}{N(N-1)} \langle a+c, b, d\rangle' + \frac{N-2}{N(N-1)} \langle a+d, b, c\rangle' +$$

$$+ \frac{N-2}{N(N-1)} \langle b+c, a, d\rangle' + \frac{N-2}{N(N-1)} \langle b+d, a, c\rangle' +$$

$$+ \frac{(N-2)(N-3)}{N(N-1)} \langle a,b,c,d\rangle'. \qquad (A9.6)$$

Voor sgn zonder accent gelden precies dezelfde regels, waarbij echter N vervangen dient te worden door n. Met behulp van deze regels kunnen vormen als bijv. $\langle(110),(001)\rangle'$ in de momenten μ omgerekend worden. Met (A9.2) vindt men:

$$\langle(110),(001)\rangle' = \frac{N}{N-1} \langle(110)\rangle' \cdot \langle(001)\rangle' - \frac{1}{N-1} \langle(111)\rangle' =$$

$$= \frac{1}{N-1} (N\Sigma \underline{\underline{xy}} \cdot \Sigma \underline{\underline{r}} - \Sigma \underline{\underline{xyr}}) =$$

$$= \frac{1}{N-1} (N\bar{\bar{R}}\mu_{11} + N\bar{\bar{X}}\bar{\bar{Y}}\bar{\bar{R}} - \mu_{02} - \bar{\bar{Y}}^2). \qquad (A9.7)$$

Voor alle in (9.4) voorkomende sgn (zonder accent) volgen hieronder de omzettingen in de steekproefmomenten $\underline{\underline{s}}_{\alpha\beta} = \langle(\alpha,0,\beta)\rangle$.

$$\langle(020)\rangle = \underline{\underline{s}}_{22} \; ; \qquad (A9.8)$$

$$\langle(010),(010)\rangle = \frac{1}{n-1} (n\underline{\underline{s}}_{11}^2 - \underline{\underline{s}}_{22}); \qquad (A9.9)$$

$$\langle(110),(001)\rangle = \frac{1}{n-1}\,(n\underline{s}_{21}\underline{s}_{01} - \underline{s}_{22}); \tag{A9.10}$$

$$\langle(200),(002)\rangle = \frac{1}{n-1}\,(n\underline{s}_{20}\underline{s}_{02} - \underline{s}_{22}); \tag{A9.11}$$

$$\langle(200),(001),(001)\rangle = \frac{1}{(n-1)(n-2)}\,(n^2\underline{s}_{20}\underline{s}_{01}^2 - 2n\underline{s}_{21}\underline{s}_{01} + -n\underline{s}_{20}\underline{s}_{02} + 2\underline{s}_{22}); \tag{A9.12}$$

$$\langle(100),(010),(001)\rangle = \frac{1}{(n-1)(n-2)}\,(n^2\underline{s}_{10}\underline{s}_{11}\underline{s}_{01} - n\underline{s}_{21}\underline{s}_{01} + -n\underline{s}_{11}^2 - n\underline{s}_{12}\underline{s}_{10} + 2\underline{s}_{22}); \tag{A9.13}$$

$$\langle(100),(100),(002)\rangle = \frac{1}{(n-1)(n-2)}\,(n^2\underline{s}_{10}^2\underline{s}_{02} - n\underline{s}_{20}\underline{s}_{02} + -2n\underline{s}_{12}\underline{s}_{10} + 2\underline{s}_{22}); \tag{A9.14}$$

$$\langle(100),(100),(001),(001)\rangle = \frac{1}{(n-1)(n-2)(n-3)}\,(n^3\underline{s}_{10}^2\underline{s}_{01}^2 + -n^2\underline{s}_{10}^2\underline{s}_{02} - n^2\underline{s}_{20}\underline{s}_{01}^2 - 4n^2\underline{s}_{11}\underline{s}_{10}\underline{s}_{01} + + n\underline{s}_{20}\underline{s}_{02} + 2n\underline{s}_{11}^2 + 4n\underline{s}_{12}\underline{s}_{10} + 4n\underline{s}_{21}\underline{s}_{01} - 6\underline{s}_{22}). \tag{A9.15}$$

——————— x ———————

A10. De variantie van $\hat{\underline{y}}'$ wordt als volgt gevonden. Stel

$$\langle a\rangle = \frac{1}{k}\sum_{j=1}^{k}(\hat{\underline{y}}_j')^a, \tag{A10.1}$$

dan is, wegens de zuiverheid van de schatter $\hat{\underline{y}}'$,

$$\langle 1\rangle' = \bar{\bar{Y}}. \tag{A10.2}$$

De variantie van $\hat{\underline{y}}' = \langle 1\rangle$ is nu

$$\mathrm{Var}(\hat{\underline{y}}') = \mathcal{E}(\langle 1\rangle - \langle 1\rangle')^2 = \mathcal{E}\langle 1\rangle\langle 1\rangle - \langle 1\rangle'\langle 1\rangle' =$$

$$= \frac{1}{k}\langle 2\rangle' + \frac{k-1}{k}\langle 1,1\rangle' - \frac{1}{p}\langle 2\rangle' - \frac{p-1}{p}\langle 1,1\rangle' =$$

$$= \frac{p - k}{pk} (\langle 2\rangle' - \langle 1,1\rangle'), \tag{A10.3}$$

hetgeen een andere schrijfwijze is voor (10.3).

A12. De zuivere quotientschatter, lineair in de r_i, met de kleinste (voorwaardelijke) variantie wordt, met (12.1) en (12.2), als volgt gevonden. Stel

$$\underline{z} = \sum_{i=1}^{n} \lambda_i \underline{r}_i . \tag{A12.1}$$

Dan geldt, indien $\underline{z}$ een zuivere schatter van $\overline{Y}$ is,

$$\mathcal{E}\underline{z} = \mathcal{E}\mathcal{E}(\underline{z} \mid \underline{x}) = \sum_{i=1}^{n} \lambda_i \mathcal{E}\mathcal{E}(\underline{r}_i \mid \underline{x}_i) = a \sum_{i=1}^{n} \lambda_i = \overline{Y}. \tag{A12.2}$$

De voorwaarde voor zuiverheid is dus

$$\sum_{i=1}^{n} \lambda_i - \overline{X} = 0. \tag{A12.3}$$

Verder geldt voor de variantie

$$\mathrm{Var}(\underline{z} \mid x) = \mathcal{E}\{(\underline{z} - \overline{Y})^2 \mid x\} =$$

$$= \mathcal{E}\left[\left\{\sum_{i=1}^{n} \lambda_i(\underline{r}_i - a)\right\}^2 \mid x\right] = \sum_{i=1}^{n} \lambda_i^2 g(x_i). \tag{A12.4}$$

Onder de voorwaarde (A12.3) dienen de λ_i zó bepaald te worden, dat $\mathrm{Var}(\underline{z} \mid x)$ minimaal is. Volgens Lagrange dient dan de functie

$$\emptyset = \sum_{i=1}^{n} \lambda_i^2 g(x_i) - 2\mu\left(\sum_{i=1}^{n} \lambda_i - \overline{X}\right) \tag{A12.5}$$

geminimaliseerd te worden. Dit geeft, met partiële differentiatie naar de λ_i,

$$\lambda_i = \overline{X}\left\{g(x_i) \sum_{i=1}^{n} \frac{1}{g(x_i)}\right\}^{-1}. \tag{A12.6}$$

Dus

$$\underline{z} = \overline{X}\left\{\sum_{i=1}^{n} \frac{1}{g(\underline{x}_i)}\right\}^{-1} \sum_{i=1}^{n} \frac{\underline{r}_i}{g(\underline{x}_i)} \tag{A12.7}$$

en

$$\mathrm{Var}(\underline{z} \mid x) = \overline{X}^2\left\{\sum_{i=1}^{n} \frac{1}{g(x_i)}\right\}^{-1}. \tag{A12.8}$$

Bij het bepalen van Var($\underline{z}$) wordt gebruik gemaakt van de volgende bekende relatie tussen voorwaardelijke en onvoorwaardelijke variantie:

$$\mathrm{Var}(\underline{z}) = \mathcal{E}\,\mathrm{Var}(\underline{z} \mid \underline{x}) + \mathrm{Var}\{\mathcal{E}(\underline{z} \mid \underline{x})\} \tag{A12.9}$$

Met (A12.9) vindt men dan Var($\underline{z}$), zoals gegeven in (12.5).

De varianties van de schatters $\overline{\underline{y}}$, $\tilde{\underline{y}}$, $\hat{\underline{y}}$ en $\hat{\underline{y}}'$ dienen nu nog in x_i en $g(x_i)$ uitgedrukt te worden.

De schatter $\overline{\underline{y}}$:

$$\overline{\underline{y}} = \frac{1}{n}\sum_{i=1}^{n} \underline{y}_i = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i\underline{r}_i\ , \tag{A12.10}$$

$$\mathcal{E}\overline{\underline{y}} = \mathcal{E}\mathcal{E}(\overline{\underline{y}} \mid \underline{x}) = \mathcal{E}\frac{1}{n}\sum_{i=1}^{n} \underline{x}_i\, \mathcal{E}(\underline{r}_i \mid \underline{x}_i) = \mathcal{E}\, a\overline{\underline{x}} = \overline{Y}, \tag{A12.11}$$

$$\begin{aligned}\mathrm{Var}(\overline{\underline{y}} \mid x) &= \mathcal{E}\{(\overline{\underline{y}} - a\overline{\underline{x}})^2 \mid x\} = \\ &= \frac{1}{n^2}\,\mathcal{E}\left[\left\{\sum_{i=1}^{n} \underline{x}_i(\underline{r}_i - a)\right\}^2 \mid x\right] = \frac{1}{n^2}\sum_{i=1}^{n} x_i^2 g(x_i),\end{aligned} \tag{A12.12}$$

dus met (A12.9):

$$\mathrm{Var}(\overline{\underline{y}}) = \frac{1}{n}\,\mathcal{E}\underline{x}^2 g(\underline{x}) + \frac{1}{n}\, a^2\mathrm{Var}(\underline{x}). \tag{A12.13}$$

De schatter $\tilde{\underline{y}}$:

$$\tilde{\underline{y}} = \bar{X}\,\frac{\bar{\underline{y}}}{\bar{\underline{x}}} = \frac{\bar{X}}{n\bar{\underline{x}}}\sum_{i=1}^{n} \underline{x}_i\underline{r}_i\;, \tag{A12.14}$$

$$\mathcal{E}\tilde{\underline{y}} = \mathcal{E}\mathcal{E}(\tilde{\underline{y}} \mid \underline{x}) = \mathcal{E}\,\frac{\bar{X}}{n\bar{\underline{x}}}\sum_{i=1}^{n} \underline{x}_i\,\mathcal{E}(\underline{r}_i \mid \underline{x}) = \mathcal{E}\,a\bar{X} = \bar{Y}, \tag{A12.15}$$

$$\mathrm{Var}(\tilde{\underline{y}} \mid x) = \frac{\bar{X}^2}{n^2\bar{x}^2}\,\mathcal{E}\left[\left\{\sum_{i=1}^{n} \underline{x}_i(\underline{r}_i - a)\right\}^2 \mid x\right] =$$

$$= \frac{\bar{X}^2}{n^2\bar{x}^2}\sum_{i=1}^{n} x_i^2 g(x_i), \tag{A12.16}$$

$$\mathrm{Var}(\tilde{\underline{y}}) = \frac{\bar{X}^2}{n^2}\,\mathcal{E}\left\{\bar{\underline{x}}^{-2}\sum_{i=1}^{n} \underline{x}_i^2 g(\underline{x}_i)\right\}. \tag{A12.17}$$

Volgens een stelling van CRAMÉR [1, pp. 353-4] mag voor $n \to \infty$ $\mathrm{Var}(\tilde{\underline{y}})$ geschreven worden als

$$\mathrm{Var}(\tilde{\underline{y}}) \approx \frac{\bar{X}^2}{n^2}\,\frac{\mathcal{E}\sum_{i=1}^{n} \underline{x}_i^2 g(\underline{x}_i)}{\mathcal{E}\bar{\underline{x}}^2} = \bar{X}^2\,\frac{\mathcal{E}\,\underline{x}^2 g(\underline{x})}{\mathcal{E}\underline{x}^2 + (n-1)\bar{X}^2}\,. \tag{A12.18}$$

De schatter $\hat{\underline{y}}$:

$$\hat{\underline{y}} = \frac{1}{n}\bar{X}\sum_{i=1}^{n} \underline{r}_i\;, \tag{A12.19}$$

$$\mathcal{E}\hat{\underline{y}} = \mathcal{E}\mathcal{E}(\hat{\underline{y}} \mid \underline{x}) = \mathcal{E}\,\frac{\bar{X}}{n}\sum_{i=1}^{n} \mathcal{E}(\underline{r}_i \mid \underline{x}) = \mathcal{E}\,a\bar{X} = \bar{Y}, \tag{A12.20}$$

$$\mathrm{Var}(\hat{\underline{y}} \mid x) = \frac{\bar{X}^2}{n^2}\,\mathcal{E}\left[\left\{\sum_{i=1}^{n} (\underline{r}_i - a)\right\}^2 \mid x\right] = \frac{\bar{X}^2}{n^2}\sum_{i=1}^{n} g(x_i), \tag{A12.21}$$

$$\mathrm{Var}(\hat{\underline{y}}) = \frac{\bar{X}^2}{n}\,\mathcal{E}g(\underline{x}). \tag{A12.22}$$

De schatter $\underline{\hat{y}}'$:

$$\underline{\hat{y}}' = \bar{X}\bar{\underline{r}} + \frac{n}{n-1}(\bar{\underline{y}} - \bar{\underline{x}}\bar{\underline{r}}) = \sum_{i=1}^{n}\left(\frac{\bar{X}}{n} + \frac{\underline{x}_i - \bar{\underline{x}}}{n-1}\right)\underline{r}_i \,, \qquad \text{(A12.23)}$$

$$\mathcal{E}\underline{\hat{y}}' = \mathcal{E}\mathcal{E}(\underline{\hat{y}}' \mid \underline{x}) = \mathcal{E}\sum_{i=1}^{n}\left(\frac{\bar{X}}{n} + \frac{\underline{x}_i - \bar{\underline{x}}}{n-1}\right)\mathcal{E}(\underline{r}_i \mid \underline{x}) =$$

$$= \mathcal{E}\, a\bar{X} = \bar{Y}, \qquad \text{(A12.24)}$$

$$\mathrm{Var}(\underline{\hat{y}}' \mid x) = \mathcal{E}\left[\left\{\sum_{i=1}^{n}\left(\frac{\bar{X}}{n} + \frac{x_i - \bar{x}}{n-1}\right)(\underline{r}_i - a)\right\}^2 \;\middle|\; x\right] =$$

$$= \sum_{i=1}^{n}\left(\frac{\bar{X}}{n} + \frac{x_i - \bar{x}}{n-1}\right)^2 g(x_i), \qquad \text{(A12.25)}$$

$$\mathrm{Var}(\underline{\hat{y}}') = \mathcal{E}\sum_{i=1}^{n}\left(\frac{\bar{X}}{n} + \frac{\underline{x}_i - \bar{\underline{x}}}{n-1}\right)^2 g(\underline{x}_i) =$$

$$= \frac{1}{n}\mathcal{E}\underline{x}^2 g(\underline{x}) - \frac{1}{n(n-1)}\mathcal{E}g(\underline{x}) \cdot \mathrm{Var}(\underline{x}), \qquad \text{(A12.26)}$$

en voor $n \to \infty$:

$$\mathrm{Var}(\underline{\hat{y}}') = \frac{1}{n}\mathcal{E}\underline{x}^2 g(\underline{x}). \qquad \text{(A12.27)}$$

Literatuur

[1] Cramér, H., "Mathematical Methods of Statistics", Princeton (1946).

[2] Erdös, P., Rényi, A., "On the central limit theorem for samples from a finite population", Magyar Tud. Akad. Mat. Kutató Int. Közl. 4 (1959), pp. 49-61.

[3] Fieller, E.C., "A fundamental formula in the statistics of biological assay and some applications", Quart. J. Pharm. 17 (1944), pp. 117-123.

[4] Goodman, L.A., Hartley, H.O., "The precision of unbiased ratio estimators", J.A.S.A. 53 (1958), pp. 491-508.

[5] Hájek, J., "Limiting distributions in simple random sampling from a finite population", Magyar Tud. Akad. Mat. Kutató Int. Közl. 5 (1960), pp. 361-374.

[6] Hájek, J., "On the theory of ratio estimates", Aplikace Matematiky 3 (1958), pp. 384-398.

[7] Hartley, H.O., Ross, A., "Unbiased ratio estimators", Nature 174 (1954), p. 270.

[8] Klerk-Grobben, G., "Een betrouwbaarheidsgebied voor het quotiënt van de verwachtingen van twee stochastische grootheden, die een simultane normale verdeling bezitten", Rapport S90 (M36), Math. Centrum (1952).

[9] Lahiri, D.B., "A method of sample selection providing un-

biased ratio estimates", Bull. Intern. Stat. Inst. 33 (2) (1951), pp. 133-140.

[10] Madow, W.G., "On the limiting distributions of estimates based on samples from finite universes", Ann. Math. Stat. 19 (1948), pp. 535-545.

[11] Mickey, M.R., "Some finite population unbiased ratio and regression estimators", J.A.S.A. 54 (1959), pp. 594-612.

[12] Midzuno, H., "On the sampling system with probability proportional to sum sizes", Annals Inst. Stat. Math. 3 (1952), pp. 99-107.

[13] Nanjamma, N.S., Murthy, M.N., Sethi, V.K., "Some sampling systems providing unbiased ratio estimators", Sankhyā 21 (1959), pp. 299-314.

[14] Paulson, E., "A note on the estimation of some mean values for a bivariate distribution", Ann. Math. Stat. 13 (1942), pp. 440-445.

[15] Robson, D.S., "Applications of multivariate polykays to the theory of unbiased ratio-type estimation", J.A.S.A. 52 (1957), pp. 511-522.

[16] Sukhatme, P.V., "Sampling Theory of Surveys with Applications", New Delhi 1954.

[17] Tukey, J.W., "Some sampling simplified", J.A.S.A. 45 (1950), pp. 501-519.

[18] Tukey, J.W., "Keeping moment-like sampling computations

simple", Ann. Math. Stat. 27 (1956), pp. 37-54.

[19] Williams, W.H., "Generating unbiased ratio and regression estimators", Biometrics 17 (1961), pp. 267-274.

Niet geciteerde literatuur

Cochran, W.G., "Sampling Techniques", New York 1953.

Deming, W.E., "Some Theory of Sampling", New York 1950.

Hansen, M.H., Hurwitz, W.N., Madow, W.G., "Sample Survey Methods and Theory", I, II, New York 1953.

Murthy, M.N., Nanjamma, N.S., "Almost unbiased ratio estimates based on interpenetrating sub-sample estimates", Sankhyā 21 (1959), pp. 381-392.

Olkin, D.S., "Multivariate ratio estimation for finite populations", Biometrika 45 (1958), pp. 154-165.

Robson, D.S., Vithayasai, C., "Unbiased componentwise ratio estimation", J.A.S.A. 56 (1961), pp. 350-358.