Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2006-12

Optimale Schichtung in Theorie und Praxis
-
Möglichkeiten zur häufigeren Anwendung dieses kostensparenden Stichprobenverfahrens

Andreas Quatember

Mai 2006

Optimale Schichtung in Theorie und Praxis – Möglichkeiten zur häufigeren Anwendung dieses kostensparenden Stichprobenverfahrens

Andreas Quatember,
Johannes Kepler Universität Linz
IFAS – Institut für Angewandte Statistik

Zusammenfassung: Für die Verwendung kostensparender bzw. genauigkeitserhöhender geschichteter Stichproben sind Vorkenntnisse über verschiedene Hilfsvariable notwendig. Fehlen diese, so können die Konzeptionen der Schichtenverfahren dennoch verwendet werden, wenn Schätzungen der Parameter dieser Hilfsvariablen vorliegen. Solche Schätzungen beeinträchtigen natürlich die Genauigkeit der Stichprobenergebnisse. Diese kann jedoch immer noch jene bei uneingeschränkt zufälliger Ziehung der Erhebungseinheiten übertreffen. Dabei können die Informationen über die notwendigen Vorkenntnisse sowohl aus einer unabhängig von der eigentlichen Stichprobe gezogenen Stichprobe als auch – als praktischere Lösung – in der laufenden Erhebung eingeholt werden. Dem Praktiker werden mehrere Möglichkeiten aufgezeigt, mit deren Hilfe er geschichtete Zufallsstichproben ziehen kann. Es wird dabei eine harmonisierte Darstellungsform gewählt, die deutlich macht, worauf es bei den einzelnen Verfahren ankommt, wenn man deren Vorteile ausnützen möchte.

Schlagwörter: Stichproben, Stichprobentheorie, Stichprobenpraxis, Schichtenverfahren, geschichtete Zufallsauswahl, proportionale Schichtung, optimale Schichtung

I Einleitung

Die empirische Sozialforschung steht gegenwärtig wie viele andere sich auf die Empirie stützende Disziplinen vor der Herausforderung, einem immer größer werdenden Aufgabenbereich bei gleichzeitig vor allem an den Universitäten immer knapper werdender finanzieller Ausstattung nachkommen zu müssen. Dies hat zur Folge, dass etwa im Bereich der Datengewinnung und -verarbeitung neue, kostengünstigere Methoden wie Internetbefragungen, Online-Datenverarbeitung etc. auf ihre Tauglichkeit geprüft werden und häufig im Mittelpunkt von wissenschaftlichen Tagungen stehen.
Zu gering wird m.E. bei dieser Suche nach Einsparungsmöglichkeiten die Rolle der Stichprobentheorie wahrgenommen. Gemeint sind dabei jene Einsparungspotenziale, die durch die Wahl geeigneter Stichprobenverfahren eröffnet werden könnten. Effizient eingesetzte Vorab-Informationen über die betreffenden Untersuchungspopulationen bieten die Möglichkeit, bei gleichen Stichprobenumfängen und Kosten genauere Stichprobenergebnisse zu erhalten bzw. zur gleichen Genauigkeit niedrigere Stichprobenumfänge zu benötigen als bei Verwendung herkömmlicher uneingeschränkter Zufallsauswahlen.
In der praktischen Anwendung ist dabei jedoch zu beachten, dass die meisten statistischen Programmpakete die erhobenen Daten standardmäßig auf Basis uneinge-

schränkter Zufallsauswahlen verarbeiten und die Verwendung eines anderen Stichprobendesigns somit bei Bereichschätzungen oder etwa Ergebnissen statistischer Tests ignoriert wird und somit etwa Genauigkeitsgewinne (genauso natürlich auch -verluste) in Folge der Verwendung dieser anderen Verfahren „verloren gehen".
Oftmals droht die Anwendung „höherer" Stichprobenverfahren jedoch einfach an der Unkenntnis dieser oder daran zu scheitern, dass die für ihren Einsatz benötigten Vorab-Informationen über die Untersuchungspopulation nicht vorhanden sind. Letzteres ist jedoch – wie sich zeigen lässt – kein Hinderungsgrund, wenn die benötigten Informationen geschätzt werden können.

II Geschichtete Stichproben

Als Stichprobenmethode wird das bewusste Schlussfolgern von einer Teilmenge auf die Gesamtheit bezeichnet. Wir bedienen uns dieser Methode täglich, wenn wir Speisen abschmecken, Weine verkosten, Parfüms testen oder Prüfungen ablegen. Für den wahrscheinlichkeitstheoretischen Schluss vom Ergebnis einer solchen Stichprobe auf die ihr zu Grunde liegende Grundgesamtheit werden Verfahren verwendet, welche die konkrete Auswahl der dafür benötigten Erhebungseinheiten aus der Grundgesamtheit sorgsam regeln. Die Entwicklung solcher „Zufallsverfahren" und die Untersuchung der grundlegenden statistischen Kennzahlen – des Erwartungswertes und der Varianz – der damit erhobenen Schätzer für Mittelwerte, Anteile und Merkmalssummen sind Gegenstand der statistischen Stichprobentheorie, die fast zur Gänze im 20. Jahrhundert entwickelt worden ist.
Zu Beginn der Entwicklung dieser Theorie wurden Zufallsauswahlen ausschließlich unter dem Aspekt gleicher Auswahlwahrscheinlichkeiten betrachtet. Eine wirkliche Revolution in Hinsicht auf das Verständnis des Begriffes Zufallsauswahl brachte ein Aufsatz des in Russland als Sohn polnischer Eltern geborenen Statistikers Jerzy Neyman, der im Jahr 1934 im *Journal of the Royal Statistical Society* erschien. Dieser Meilenstein der Stichprobentheorie trug den Titel: „On the Two Different Aspects of the Representiative Method: The Method of Stratified Sampling and the Method of Purposive Selection". Entgegen den bisherigen Vorstellungen entwickelte Neyman in seinem Aufsatz eine Konzeption mit ungleichen Auswahlwahrscheinlichkeiten. Neyman wies darin nach, dass bei einer geschichteten Zufallsauswahl eine höhere Genauigkeit der Stichprobenergebnisse dann erreicht werden kann, wenn ein überproportional großer Teil des Gesamtstichprobenumfanges auf jene Schichten entfällt, in denen das Untersuchungsmerkmal stärker als in anderen streut.
Dieser Ansatz war indes in Osteuropa nicht neu. Im großen zaristischen Russland war ab Beginn des 20. Jahrhunderts ebenfalls intensiv theoretisch auf dem Gebiet der Stichprobenverfahren gearbeitet worden. Die dabei gewonnenen Erkenntnisse fasste A.J. Kowalsky 1924 zusammen. In einem Kapitel über geschichtete Zufallsauswahlen beschrieb Kowalsky genau jene optimale Aufteilung des Stichprobenumfanges auf die einzelnen Schichten, die – wahrscheinlich unabhängig davon – schon im Jahr 1923 auch Alexander A. Tschuprow und 1934 eben Neyman veröffentlicht haben (vgl. Zarkovich, 1962, S.487).
Soll aus einer Grundgesamtheit eine geschichtete Zufallsauswahl (≡S) gezogen werden, so wird die Grundgesamtheit aller Erhebungseinheiten nach einem ein- oder mehrdi-

mensionalen Merkmal Y, dem Schichtmerkmal, in K disjunkte Teilgesamtheiten, die Schichten, zerlegt. Jede Erhebungseinheit gehört damit einer und nur einer dieser Schichten an. Innerhalb jeder dieser Schichten erfolgt eine uneingeschränkte Zufallsauswahl der Erhebungseinheiten vom Umfang n_h (h=1,2,...K). Für den Gesamtstichprobenumfang n gilt: $n = \Sigma n_h$.
Dieses Modell zeichnet sich nun dadurch aus, dass bei geeigneter Aufteilung des Gesamtstichprobenumfanges n auf die einzelnen Schichten eine Erhöhung der Genauigkeit des damit erhobenen Mittelwertschätzers $\overline{x}_S$ für den Parameter μ, den Mittelwert eines interessierenden Merkmals X in der Grundgesamtheit, im Vergleich zu seiner Schätzung durch den Mittelwert einer uneingeschränkten Zufallsauswahl erzielt werden kann. Äquivalent dazu gilt, dass auf diese Weise der Stichprobenumfang und die Erhebungskosten bei gleicher Genauigkeit des Stichprobenergebnisses reduziert werden können.
Erhoben wird die für μ, den interessierenden Mittelwert des Merkmals X in der Grundgesamtheit, erwartungstreue Statistik (Notationen nach: Cochran, 1972, S. 111ff.)

$$\overline{x}_S = \sum_{h=1}^{K} W_h \cdot \overline{x}_h \qquad (1)$$

(W_h ... der Anteil der Erhebungseinheiten der h-ten Schicht in der Grundgesamtheit).

$$\overline{x}_h = \frac{1}{n_h} \cdot \sum_{i=1}^{n_h} x_{hi} \qquad (2)$$

(n_h ... der Stichprobenumfang in der h-ten Schicht; x_{hi} ... die Ausprägung des Merkmals X bei der i-ten Erhebungseinheit der Stichprobe in der h-ten Schicht; h=1,2,...K) ist das für μ_h, den bedingten Mittelwert von X in der h-ten Schicht, erwartungstreue Stichprobenmittel (Die Statistiken für interessierende Merkmalssummen oder Anteile folgen direkt aus (1) und (2)). Hinsichtlich der Varianz von $\overline{x}_S$ gilt bei Ziehen mit Zurücklegen:

$$\mathrm{Var}\,\overline{x}_S = \sum_{h=1}^{K} W_h^{\,2} \cdot \mathrm{Var}\,\overline{x}_h = \sum_{h=1}^{K} W_h^{\,2} \cdot \frac{\sigma_h^{\,2}}{n_h} \qquad (3)$$

(s_h^2 ... die Varianz des Merkmals X in der h-ten Schicht) und bei Ziehen ohne Zurücklegen:

$$\mathrm{Var}\,\overline{x}_S = \sum_{h=1}^{K} W_h^{\,2} \cdot \frac{\sigma_h^{\,2}}{n_h} \cdot \frac{N_h - n_h}{N_h - 1} \qquad (4)$$

(N_h ... die Anzahl von Erhebungseinheiten in der h-ten Schicht). In (3) und (4) wird deutlich, dass die Varianz von $\overline{x}_S$ von der Wahl des Schichtmerkmals Y insofern abhängt, als die damit gebildeten Schichten gewährleisten sollen, dass die Schichtvarianzen σ_h^2 des Untersuchungsmerkmals X möglichst klein sind.
Für im Vergleich zu den Schichtstichprobenumfängen n_h großen Schichtgrößen N_h, also für kleine „Auswahlsätze“ n_h/N_h, ist Formel (3) auch Näherungslösung von (4). Da diese Voraussetzung in den allermeisten Anwendungsfällen zutrifft, wollen wir uns der Einheitlichkeit der Darstellung wegen im Folgenden auf Modelle mit Zurücklegen beschränken.
Ein erwartungstreuer Schätzer für diese Varianz ist:

$$\widehat{\mathrm{Var}}\ \overline{x}_S = \sum_{h=1}^{K} W_h^{\,2} \cdot \frac{s_h^{\,2}}{n_h} \tag{5}$$

mit

$$s_h^{\,2} = \frac{1}{n_h - 1} \cdot \sum_{i=1}^{n_h} (x_i - \overline{x})^2 \,,$$

der erwartungstreuen Stichprobenvarianz des Merkmals X in der h-ten Schicht.

III Proportionale Schichtung

Nach der Wahl des Schichtmerkmals ist zunächst einmal die Frage nach der Aufteilung des Stichprobenumfanges n auf die einzelnen Schichten zu klären. Bei gegebenem Schichtmerkmal Y hängt nämlich die Varianz (3) des Mittelwertschätzers $\overline{x}_S$ nur mehr von der Wahl der Stichprobenumfänge n_h in den Schichten ab (h=1,2,...K).
Bestimmt man diese Umfänge nun mit

$$n_h = W_h \cdot n \tag{6}$$

so spricht man von einer proportionalen Aufteilung (≡P) des Stichprobenumfanges n auf die Schichten.
Für den mit dieser Aufteilung mit (1) und (2) bestimmten Mittelwertschätzer $\overline{x}_P$, ergibt sich eine Varianz von:

$$\mathrm{Var}\ \overline{x}_P = \frac{1}{n} \cdot \sum_{h=1}^{K} W_h \cdot \sigma_h^{\,2} \tag{7}$$

Die Varianz dieses Mittelwertschätzers entspricht somit dem (1/n)-fachen des (arithmetischen) Mittels der Schichtvarianzen (≡M σ_h^2) des Merkmals X über alle Schichten und lässt sich somit darstellen als:

$$\mathrm{Var}\ \overline{x}_P = \frac{1}{n} \cdot \mathrm{M}\ \sigma_h^{\,2} \tag{7'}$$

Die Differenz Δ_P der Varianzen uneingeschränkter und proportional geschichteter Zufallsauswahlen wird als Schichtungseffekt oder Genauigkeitsgewinn der proportionalen Aufteilung bezeichnet: Mit der Varianz für uneingeschränkte Zufallsauswahl (≡U), dargestellt als das (1/n)-fache der Summe aus dem Mittel der Schichtvarianzen und der Varianz der Schichtmittelwerte Var μ_h

$$\begin{aligned} \mathrm{Var}\ \overline{x}_U &= \frac{\sigma^2}{n} = \frac{1}{n} \cdot \sum_{h=1}^{K} W_h \cdot \sigma_h^{\,2} + \frac{1}{n} \cdot \sum_{h=1}^{K} W_h \cdot (\mu_h - \mu)^2 \\ &= \frac{1}{n} \cdot \mathrm{M}\ \sigma_h^{\,2} + \frac{1}{n} \cdot \mathrm{Var}\ \mu_h \end{aligned} \tag{8}$$

(σ^2 ... die Varianz von X) ergibt sich als Schichtungseffekt:

$$\begin{aligned} \Delta_P = \mathrm{Var}\ \overline{x}_U - \mathrm{Var}\ \overline{x}_P &= \frac{1}{n} \cdot \sum_{h=1}^{K} W_h \cdot (\mu_h - \mu)^2 \\ &= \frac{1}{n} \cdot \mathrm{Var}\ \mu_h \end{aligned} \tag{9}$$

Der Schichtungseffekt entspricht somit dem (1/n)-fachen der Varianz der Schichtmittelwerte μ_h. Sobald die Schichtmittelwerte also streuen, werden mittels einer proportional

geschichteten Zufallsauswahl wirksamere Mittelwertschätzer erzielt als mit einer uneingeschränkten Zufallsauswahl gleichen Umfangs. Ungenauere Ergebnisse sind bei Verwendung dieses Stichprobenverfahrens unmöglich.
Gibt man umgekehrt die erwünschte Genauigkeit in Form der maximalen Abweichung ε vor, mit der das Stichprobenergebnis $\overline{x}$ vom Parameter μ mit einer Wahrscheinlichkeit von mindestens $1-\alpha$ abweichen darf, dann lässt sich jener Mindeststichprobenumfang berechnen, der für diese Anforderung nötig ist.
Für genügend große Stichprobenumfänge (schon ab $n = 30$) gilt für uneingeschränkte Zufallsauswahlen aus der Grundgesamtheit

$$\varepsilon = z \cdot \frac{\sigma}{\sqrt{n}}$$

(z ... das $(1-\alpha/2)$-Fraktil der Standardnormalverteilung) und damit als Mindeststichprobenumfang

$$n_U = \frac{z^2}{\varepsilon^2} \cdot \sigma^2 \tag{10}$$

Bei proportionaler Aufteilung des Stichprobenumfanges gilt analog

$$\varepsilon = z \cdot \frac{\sqrt{\sum_{h=1}^{K} W_h \cdot \sigma_h^{\ 2}}}{\sqrt{n}}$$

und damit

$$n_P = \frac{z^2}{\varepsilon^2} \cdot \sum_{h=1}^{K} W_h \cdot \sigma_h^{\ 2} = \frac{z^2}{\varepsilon^2} \cdot M\,\sigma^2 \tag{11}$$

Somit beträgt die relative Ersparnis im Stichprobenumfang und damit bei den variablen Erhebungskosten bei proportionaler Aufteilung im Vergleich zu einer uneingeschränkt zufälligen Ziehung der Erhebungseinheiten

$$1 - \frac{n_P}{n_U} = 1 - \frac{M\,\sigma_h^{\ 2}}{\sigma^2} \tag{12}$$

Beispiel:
Gegeben sei ein stetiges Merkmal X, das sich innerhalb eines dichotomen Schichtmerkmals Y (etwa dem Merkmal Geschlecht) mit folgendem Mittelwert und folgender Varianz verteilt:

Schicht	W_h	μ_h	σ_h
1	0,4	2	3
2	0,6	4	1

Es sei einerseits $n = 1.000$ und andererseits die erwünschte Genauigkeit $\varepsilon = 0{,}1$ und $1-\alpha = 0{,}95$.
Bei gegebenem Stichprobenumfang n = 1.000 gelten folgende Schätzervarianzen:
Bei uneingeschränkter Zufallsauswahl nach (8):

$$\text{Var}\,\overline{x}_U = \frac{1}{1000} \cdot \left(4{,}2 + 0{,}96\right) = 5{,}16 \cdot 10^{-3};$$

bei proportional geschichteter Zufallsauswahl (Schichtstichprobenumfänge nach (6): $n_1 = 400$, $n_2 = 600$) nach (7):

$$\operatorname{Var} \overline{x}_P = \frac{1}{1000} \cdot 4{,}2 = 4{,}2 \cdot 10^{-3};$$

Der Genauigkeitsgewinn Δ_P beträgt somit $0{,}96 \cdot 10^{-3}$.

Bei einer erwünschten Genauigkeit von $\varepsilon = 0{,}1$ gilt für die notwendigen Stichprobenumfänge:

Bei uneingeschränkter Zufallsauswahl nach (10):

$$n_U = \frac{1{,}96^2}{0{,}1^2} \cdot 5{,}16 = 1982{,}3$$

bei proportional geschichteter Zufallsauswahl nach (11):

$$n_P = \frac{1{,}96^2}{0{,}1^2} \cdot 4{,}2 = 1613{,}5$$

Die relative Ersparnis im Stichprobenumfang beträgt nach (12) demnach 0,186, also 18,6 %. ♣

Bei Unkenntnis der Varianzen σ_h^2 (h=1,2,...K) können diese Varianzen aus vergangenen Erhebungen geschätzt werden, so dass damit die erforderlichen Stichprobenumfänge bestimmbar werden.

III.1 Geschätzte Schichtgrößen

Für die Anwendung dieser Konzeption werden allerdings − in (6) − die relativen Schichtgrößen W_h benötigt (h=1,2,...K). Fehlen diese Vorkenntnisse, so kann dennoch nach dem Schichtmerkmal Y geschichtet werden, wenn die diesbezüglichen Informationen aus einer Zufallsstichprobe vom Umfang m verwendet werden. Die in der – auf diese Schätzungen basierenden – proportional geschichteten Zufallsauswahl (≡P*) vom Umfang n zu verwendenden Schichtstichprobenumfänge n_h werden dann bestimmt durch:

$$n_h = w_h \cdot n \tag{13}$$

(w_h ... der Anteil der Erhebungseinheiten der h-ten Schicht in der „Basisstichprobe“).
Die damit erhobene erwartungstreue Statistik

$$\overline{x}_{P*} = \sum_{h=1}^{K} w_h \cdot \overline{x}_h \tag{14}$$

besitzt, wenn die Basisstichprobe eine uneingeschränkte Zufallsauswahl war, die Varianz (vgl. etwa: Cochran, 1972, S. 385):

$$\begin{aligned} \operatorname{Var} \overline{x}_{P*} &= \frac{1}{n} \cdot \sum_{h=1}^{K} W_h \cdot \sigma_h^2 + \frac{1}{m} \cdot \sum_{h=1}^{K} W_h \cdot (\mu_h - \mu)^2 \\ &= \frac{1}{n} \cdot M\, \sigma_h^2 + \frac{1}{m} \cdot \operatorname{Var} \mu_h \end{aligned} \tag{15}$$

und den Schichtungseffekt (vgl. Quatember, 1996, S. 49):

$$\Delta_{P*} = \text{Var}\,\overline{x}_U - \text{Var}\,\overline{x}_{P*} = \left(\frac{1}{n} - \frac{1}{m}\right) \cdot \sum_{h=1}^{K} W_h \cdot \left(\mu_h - \mu\right)^2$$
$$= \left(\frac{m-n}{m \cdot n}\right) \cdot \text{Var}\,\mu_h \qquad (16)$$

Δ_{P*} ist also positiv, wenn die Varianz der Schichtmittelwerte nicht null ist und der Stichprobenumfang der Basisstichprobe m größer als jener der darauf basierenden geschichteten Stichprobe n ist. Bei Verwendung von Basisstichproben vom Umfang m, die genauere Schätzungen als uneingeschränkte liefern – etwa proportional geschichtete –, geben (15) und (16) die maximale Schätzervarianz bzw. den Mindestschichtungseffekt an, die bei Schätzung der Schichtgrößen für eine proportional geschichtete Stichprobe erzielt werden können (vgl. etwa: Quatember, 1996, S. 51ff).

Somit bieten sich die in großen Zufallsstichproben erhobenen Variablen als Schichtmerkmale für eine proportional geschichtete Zufallsauswahl mit geschätzten Schichtgrößen an. Wenn die Eigenschaften der Schätzer w_h jenen aus uneingeschränkten Zufallsauswahlen entsprechen (oder die Schätzer sogar genauer sind), eignen sich deshalb nationale Erhebungen der amtlichen Stichproben wie der Mikrozensus ganz besonders für diesen Zweck. Und dies auch deshalb, weil die damit gewonnenen Informationen jedenfalls aktueller sind als die etwa aus der Volkszählung vorliegenden, möglicherweise schon veralteten Populationsergebnisse!

IV Optimale Schichtung

Neyman und andere (siehe Abschnitt I) haben jene Aufteilung eines gegebenen Stichprobenumfanges n bestimmt, welche die Varianz von $\overline{x}_S$ minimiert (Die insgesamt genaueren Ausführungen in diesem Abschnitt veranschaulichen die diesbezüglichen Ausführungen in Abschnitt IV.6 von Quatember, 2001, durch Beispiele, decken sich aber über weite Strecken damit). Als Ergebnis einer Extremwertaufgabe mit Nebenbedingung gibt

$$n_h = \frac{W_h \cdot \sigma_h}{\sum_{h=1}^{K} W_h \cdot \sigma_h} \cdot n \qquad (17)$$

die in diesem Sinne (varianz)optimale Aufteilung (≡O) an (h=1,2,...K).

Für den mit diesen Schichtstichprobenumfängen mit (1) und (2) berechneten Mittelwertschätzer $\overline{x}_O$ wird die Varianz in (3) zu:

$$\text{Var}\,\overline{x}_O = \frac{1}{n} \cdot \left(\sum_{h=1}^{K} W_h \cdot \sigma_h\right)^2$$
$$= \frac{1}{n} \cdot \left(M\,\sigma_h\right)^2 \qquad (18)$$

Die Varianz dieses Mittelwertschätzers entspricht somit dem (1/n)-fachen des Quadrats des Mittels der Schichtstandardabweichungen des Merkmals X.

Der Schichtungseffekt Δ_O besitzt folgende Darstellung:

$$\begin{aligned}\Delta_O = \mathrm{Var}\,\overline{x}_U - \mathrm{Var}\,\overline{x}_O &= \frac{1}{n}\cdot\sum_{h=1}^{K} W_h\cdot\left(\sigma_h - \mathrm{M}\,\sigma_h\right)^2 + \frac{1}{n}\cdot\sum_{h=1}^{K} W_h\cdot\left(\mu_h-\mu\right)^2\\ &= \frac{1}{n}\cdot\mathrm{Var}\,\sigma_h + \frac{1}{n}\cdot\mathrm{Var}\,\mu_h \qquad (19)\\ &= \frac{1}{n}\cdot\mathrm{Var}\,\sigma_h + \Delta_P\end{aligned}$$

Die optimale Allokation des Stichprobenumfanges auf die Schichten gewährleistet also einen noch größeren Schichtungseffekt als die proportionale, sobald die Schichtstandardabweichungen des Merkmals X streuen. Geringer als bei proportionaler Schichtung kann der Schichtungseffekt nicht werden.

Damit ein nennenswerter weiterer Genauigkeitsgewinn gegenüber proportionaler Schichtung eintritt, sollte das Maximum der Schichtstandardabweichungen jedoch mindestens doppelt so groß sein wie deren Minimum. In diesem Falle gilt nämlich immer noch: $\mathrm{Var}\,\overline{x}_O \geq 0{,}889\cdot\mathrm{Var}\,\overline{x}_P$ (vgl.: Sadooghi-Alvandi, 1988, S. 197), d.h. die optimale Schichtung kann höchstens eine Reduzierung um 11,1 % bei der Schätzervarianz bewirken. Ist das Maximum mindestens dreimal so groß wie das Minimum, dann ist die Varianz des Mittelwertschätzers bei optimaler Aufteilung immer noch mindestens drei Viertel jener bei proportionaler Aufteilung.

Hinsichtlich des erforderlichen Stichprobenumfanges gilt bei optimaler Aufteilung bei Festlegung der erwünschten maximalen Abweichung ε des Stichprobenergebnisses vom Parameter (vergleiche Abschnitt III):

$$\varepsilon = z\cdot\frac{\sum_{h=1}^{K} W_h\cdot\sigma_h}{\sqrt{n}}$$

Damit errechnet sich der notwendige Stichprobenumfang, mit dem diese Forderung mit einer Wahrscheinlichkeit von $1-\alpha$ eingehalten wird, folgendermaßen:

$$n_O = \frac{z^2}{\varepsilon^2}\cdot\left(\sum_{h=1}^{K} W_h\cdot\sigma_h\right)^2 = \frac{z^2}{\varepsilon^2}\cdot(\mathrm{M}\,\sigma_h)^2 \qquad (20)$$

Somit beträgt die relative Ersparnis im Stichprobenumfang bei optimaler Aufteilung im Vergleich zu einer uneingeschränkt zufälligen Ziehung der Erhebungseinheiten

$$1-\frac{n_O}{n_U} = 1-\frac{(\mathrm{M}\,\sigma_h)^2}{\sigma^2} \qquad (21)$$

Beispiel (Fortsetzung):

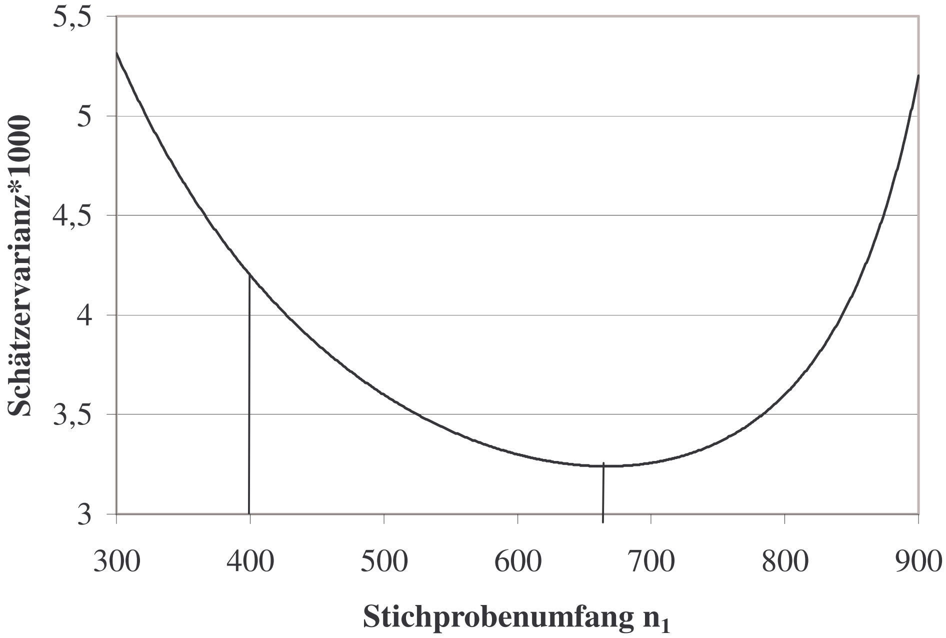
Bei optimal geschichteter Zufallsauswahl (Stichprobenumfänge nach (17): $n_1 = 667$, $n_2 = 333$) gilt nach (18):

$$\mathrm{Var}\,\overline{x}_O = \frac{1}{1000}\cdot 1{,}8^2 = 3{,}24\cdot 10^{-3}$$

Die relative Ersparnis beträgt nach (21) 37,2 %. Das ist etwa doppelt so viel wie bei proportionaler Schichtung.

Betrachten wir nun die Varianz des Mittelwertschätzers $\overline{x}_S$ mit den Daten aus unserem Beispiel in der nachfolgenden Abbildung 1 für beliebige Aufteilungen von n = 1.000 auf die Schichten.

Abbildung 1:
Vergleich der Varianzen der Mittelwertschätzer des Beispiels bei verschiedenen Aufteilungen des Stichprobenumfanges 1.000 auf die beiden Schichten



Aus Abbildung 1 geht hervor, dass auch bei von der optimalen Aufteilung ($n_1 = 667$) abweichenden Aufteilungen in unserem Beispiel gegenüber proportionaler Aufteilung ($n_1 = 400$) ein weiterer Schichtungseffekt erzielt werden kann. Wird durch die Schätzung von σ_h in (17) die optimale Aufteilung verfehlt, so reagiert in unserem Beispiel die Varianz des Mittelwertschätzers $\overline{x}_S$ nicht dramatisch. Es gilt z.B. für $525 < n_1 < 784$: $\text{Var}\,\overline{x}_S < 3{,}5 \cdot 10^{-3} \ll \text{Var}\,\overline{x}_P$. Dies lässt erkennen, dass hier für die unbekannten Parameter σ_h auch durchaus ungenaue Schätzungen s_h verwendet werden dürfen, weil die diese Parameter stark streuen. ♣

Für die Verwendung der optimalen Konzeption wird in (17) neben der Kenntnis der Verteilung des Schichtmerkmals Y (in W_h), die schon für die proportionale Aufteilung benötigt wurde, zusätzlich jene über die Schichtstandardabweichungen σ_h (h=1,2,...K) des Untersuchungsmerkmals X vorausgesetzt. Sind diese Vorkenntnisse nicht vorhanden, so müssen wiederum Schätzungen dieser Parameter weiterhelfen.

IV.1 Geschätzte Schichtstandardabweichungen

Der häufigste Fall der Unkenntnis der für die Anwendung der optimalen Schichtung benötigten Parameter wird wohl jener der Unkenntnis der Schichtstandardabweichungen σ_h von X bei gleichzeitiger Kenntnis der Anteile W_h von Y sein, setzen die Kenntnis der ersteren doch Kenntnisse über das Untersuchungsmerkmal X voraus, die eine Stichprobenerhebung von X eigentlich unnötig machen würden. Will man sich dennoch nicht

mit einer proportionalen Aufteilung begnügen, so kann man die Schätzer s_h für die Parameter σ_h aus uneingeschränkten Basisstichproben vom Umfang m mit Schichtumfängen m_h und $m = \Sigma m_h$ verwenden. Dabei sei unterstellt, dass die Schichtvarianzen von X tatsächlich gleich geblieben sind. Gleichzeitig sollen sich aber die Schichtmittelwerte verändert haben, so dass eine neuerliche Erhebung von X Sinn macht oder es wird eine genauere Mittelwertschätzung als in der Basisstichprobe gewünscht.
Die auf diese Schätzer basierende geschichtete Zufallsstichprobe mit optimaler Aufteilung (≡O**) des Stichprobenumfanges n setzt sich folgendermaßen zusammen:

$$n_h = \frac{W_h \cdot s_h}{\sum_{h=1}^{K} W_h \cdot s_h} \cdot n \tag{22}$$

Für den darauf basierenden erwartungstreuen Mittelwertschätzer

$$\overline{x}_{O**} = \sum_{h=1}^{K} W_h \cdot \overline{x}_h \tag{23}$$

lässt sich die Varianz folgendermaßen darstellen:

$$\text{Var}\, \overline{x}_{O**} = E\left[\text{Var}(\overline{x}_{O**} | s_h)\right]$$

Aus der Varianz des bedingten Mittels

$$\text{Var}(\overline{x}_{O**} | s_h) = \sum_{h=1}^{K} W_h^{\ 2} \cdot \frac{\sigma_h^{\ 2}}{n_h}$$

folgt durch Einsetzen von (22):

$$\begin{aligned}\text{Var}(\overline{x}_{O**} | s_h) &= \frac{1}{n} \cdot \left(\sum_{h=1}^{K} W_h \cdot \frac{\sigma_h^{\ 2}}{s_h} \right) \cdot \left(\sum_{h=1}^{K} W_h \cdot s_h \right) \\ &= \frac{1}{n} \cdot \left(\sum_{h=1}^{K} W_h^{\ 2} \cdot \sigma_h^{\ 2} + \sum_{h \neq i} W_h \cdot W_i \cdot \sigma_h^{\ 2} \cdot \frac{s_i}{s_h} \right)\end{aligned}$$

woraus folgt:

$$\begin{aligned}\text{Var}\, \overline{x}_{O**} &= \frac{1}{n} \cdot E\left[\left(\sum_{h=1}^{K} W_h^{\ 2} \cdot \sigma_h^{\ 2} + \sum_{h \neq i} W_h \cdot W_i \cdot \sigma_h^{\ 2} \cdot \frac{s_i}{s_h} \right) \right] \\ &= \frac{1}{n} \cdot \sum_{h=1}^{K} W_h^{\ 2} \cdot \sigma_h^{\ 2} + \frac{1}{n} \cdot \sum_{h \neq i} W_h \cdot W_i \cdot \sigma_h^{\ 2} \cdot E\left(\frac{s_i}{s_h} \right)\end{aligned}$$

Dayal (1979) zeigt, dass gilt (ebd., S. 172f):

$$E\left(\frac{s_i}{s_h} \right) \approx \frac{\sigma_i}{\sigma_h} \cdot \left(1 + V_h^{\ 2}\right)$$

mit

$$V_h = \frac{\sqrt{\text{Var}(s_h)}}{\sigma_h} \ ,$$

dem Variationskoeffizienten der Stichprobenstandardabweichungen in der h-ten Schicht. Daraus folgt schließlich:

$$\begin{aligned}\operatorname{Var}\bar{x}_{O**} &\approx \frac{1}{n}\cdot\sum_{h=1}^{K} W_h^{\ 2}\cdot\sigma_h^{\ 2} + \frac{1}{n}\cdot\sum_{h\neq i} W_h\cdot W_i\cdot\sigma_h\cdot\sigma_i\cdot\left(1+V_h^{\ 2}\right)\\ &= \frac{1}{n}\cdot\left(\sum_{h=1}^{K} W_h\cdot\sigma_h\right)^2 + \frac{1}{n}\cdot\sum_{h\neq i} W_h\cdot W_i\cdot\sigma_h\cdot\sigma_i\cdot V_h^{\ 2}\end{aligned}$$

$$\operatorname{Var}\bar{x}_{O**} \approx \operatorname{Var}\bar{x}_O + \frac{1}{n}\cdot\sum_{h\neq i} W_h\cdot W_i\cdot\sigma_h\cdot\sigma_i\cdot V_h^{\ 2} \qquad (24)$$

V_h^2 geht mit zunehmendem Stichprobenumfang m_h sehr schnell gegen null, weshalb für große Schichtstichprobenumfänge m_h demnach gilt:

$$\operatorname{Var}\bar{x}_{O**} \approx \operatorname{Var}\bar{x}_O$$

Die Schätzung der Schichtstandardabweichungen s_h durch diesbezügliche Informationen aus großen, früheren Erhebungen wirkt sich auf die Genauigkeit der Stichprobenergebnisse vernachlässigbar gering aus. Die optimale Konzeption behält ihren Genauigkeitsvorteil.

Beispiel (Fortsetzung):
Es sei nun der Basisstichprobenumfang m = 1.000 und m_1 = 400, m_2 = 600 (eine zufällig proportionale Aufteilung von m auf die beiden Schichten). Wenn das Merkmal X normalverteilt ist, dann ist der Quotient

$$\frac{(m-1)\cdot s^2}{\sigma^2} = \chi^2$$

χ^2-verteilt mit (n–1) Freiheitsgraden.
Betrachten wir nun folgende Fälle: Wir erhalten für den Schätzer s_1 sein 0,05-Fraktil und für den Schätzer s_2 sein 0,95-Fraktil (sozusagen das andere Extrem), so gilt:

$$\chi^2_{399;0,05} \approx 353{,}7 \text{ und } \chi^2_{599;0,95} \approx 657{,}0$$

und daraus

$$s_1^{\ 2} = \frac{353{,}7\cdot 9}{399} = 7{,}98 \text{ und } s_2^{\ 2} = \frac{657{,}0\cdot 1}{599} = 1{,}10$$

Schätzt man damit die optimale Aufteilung nach (22), dann ergibt dies folgende Aufteilung des Umfanges n = 1.000 auf die beiden Schichten dieses Beispiels:

$$n_1 = 0{,}642 \cdot 1.000 = 642$$
$$n_2 = 0{,}358 \cdot 1.000 = 358$$

Sogar bei m = 100 (also einer sehr kleinen Basisstichprobe zur Schätzung der Schichtstandardabweichungen von X) werden durch deutliche Unterschätzung von σ_1 einerseits und ebensolche Überschätzung von σ_2 andererseits nur folgende Abweichungen von der optimalen Aufteilung erzielt ($\chi^2_{39;0,05} \approx 25{,}7$ und $\chi^2_{59;0,95} \approx 77{,}9$):

$$n_1 = 0{,}586 \cdot 1.000 = 586$$
$$n_2 = 0{,}414 \cdot 1.000 = 414$$

Bei diesen Aufteilungen ergeben sich folgende Schätzervarianzen (siehe Abbildung 1):

$$\operatorname{Var}(\bar{x}_S \mid n_1 = 642, n_2 = 358) = 3{,}25\cdot 10^{-3}$$

$$\operatorname{Var}(\bar{x}_S \mid n_1 = 586, n_2 = 414) = 3{,}33\cdot 10^{-3}$$

Vergleicht man diese Ergebnisse mit der Varianz bei optimaler Aufteilung (siehe oben), so wird deutlich, wie geringfügig ungenauer die Schätzervarianz selbst bei Schätzung in einer sehr kleinen Basisstichprobe wird. ♣

Zu beachten ist, dass dazu $\Delta_{OPT} >> \Delta_P$ sein muss, dass also die Schichtstandardabweichungen σ_h wie in unserem Beispiel stark streuen müssen. Aber dies ist ja sowieso die Voraussetzung für den gewinnbringenden Einsatz der optimalen Schichtung. Ist jedoch $\Delta_{OPT} \approx \Delta_P$, so sollte man sich von Haus aus für eine proportionale Aufteilung des Stichprobenumfanges entscheiden. Falls auch noch $\Delta_P \approx 0$ gilt, so ist für eine uneingeschränkte Zufallsauswahl zu empfehlen, da in diesem Fall durch die Schätzung einer optimalen Aufteilung der geringe zusätzliche Schichtungseffekt vernachlässigbar gering ausfällt oder überhaupt verschwindet.
An Stelle der Stichprobenstandardabweichungen s_h eignet sich natürlich auch jede andere Variable t_h (h=1,2,...K) als Ersatz für σ_h, wenn diese annähernd proportional dazu ist: $t_h \approx k \cdot \sigma_h$ mit $k \in \mathbb{R}$ (vgl. Hansen et al., 1953, S. 213). Z.B. ist vorstellbar, dass bei einer Erhebung der Mietkosten die Streuung der Wohnungsgrößen bei regionaler Schichtung (annähernd) proportional zur Streuung der Mietkosten ist. Kennt man diese Kennzahlen, dann darf man sie an Stelle der Streuung des eigentlichen Untersuchungsmerkmals in (17) bzw. (22) – je nachdem, ob die Parameter oder Schätzer derselben bekannt sind – verwenden.

IV.2 Geschätzte Schichtgrößen

Ein weiterer möglicher – wenngleich wahrscheinlich seltenerer – Fall tritt auf, wenn die Varianzen σ_h bekannt sind, die Anteile eines geeigneten Schichtmerkmals Y jedoch nicht. So kann etwa in einem Produktionsprozess die Varianz von Messungen des Merkmals X als in allen Schichten gleichbleibend gegenüber früheren Erhebungen betrachtet werden. Bei Unkenntnis der relativen Schichtgrößen W_h kann man – wie in Abschnitt III.1 – deren Schätzer w_h aus einer uneingeschränkten Basisstichprobe vom Umfang m verwenden. Für die darauf basierende geschichtete Zufallsstichprobe mit optimaler Aufteilung des Stichprobenumfanges n bei geschätzten Schichtgrößen (≡O*) gilt:

$$n_h = \frac{w_h \cdot \sigma_h}{\sum_{h=1}^{K} w_h \cdot \sigma_h} \cdot n \tag{25}$$

Für die daraus resultierende Varianz des erwartungstreuen Mittelwertschätzers

$$\bar{x}_{O*} = \sum_{h=1}^{K} w_h \cdot \bar{x}_h \tag{26}$$

gilt:

$$\text{Var } \bar{x}_{O*} = E\left[\text{Var}(\bar{x}_{O*} | w_h)\right]$$

Zunächst ist also der Erwartungswert von $(\bar{x}_{O*} - \mu)^2$ über alle möglichen Stichproben mit konstanten Schichtanteilen w_h (h=1,2,...K) zu bilden und danach jener der Erwartungswerte über alle möglichen w_h. Es gilt:

$$E\left((\overline{x}_{O*}-\mu)^2 \,\middle|\, w_h\right)=\sum_{h=1}^{K} w_h{}^2\cdot\frac{\sigma_h{}^2}{n_h}+\left(\sum_{h=1}^{K}\left(w_h-W_h\right)\cdot\mu_h\right)^2$$

Der zweite Summand auf der rechten Seite der letzten Gleichung ist jener Beitrag an der Varianz, den die Verzerrung der Schichtanteile w_h liefert. Setzen wir nun n_h nach (25) in den ersten Summanden ein:

$$\sum_{h=1}^{K} w_h{}^2\cdot\frac{\sigma_h{}^2}{n_h}=\frac{1}{n}\cdot\left(\sum_{h=1}^{K} w_h\cdot\sigma_h\right)^2$$

Bilden wir nun noch den Erwartungswert der Schätzervarianzen über alle möglichen w_h. Es gilt:

$$E\left(\frac{1}{n}\cdot\left(\sum_{h=1}^{K} w_h\cdot\sigma_h\right)^2\right)=\frac{1}{n}\cdot E\left(\sum_{h=1}^{K} w_h\cdot\sigma_h\right)^2$$

und

$$E\left(\sum_{h=1}^{K} w_h\cdot\sigma_h\right)^2=\operatorname{Var}\left(\sum_{h=1}^{K} w_h\cdot\sigma_h\right)+\left(E\left(\sum_{h=1}^{K} w_h\cdot\sigma_h\right)\right)^2$$

Hierin gilt:

$$\begin{aligned}\operatorname{Var}\left(\sum_{h=1}^{K} w_h\cdot\sigma_h\right)&=\sum_{h=1}^{K}\sigma_h{}^2\cdot\operatorname{Var} w_h+\sum_{h\neq i}\sigma_h\cdot\sigma_i\cdot\operatorname{Kov}\left(w_h,w_i\right)\\&=\frac{1}{m}\cdot\sum_{h=1}^{K}\sigma_h{}^2\cdot W_h\cdot(1-W_h)-\frac{1}{m}\cdot\left(\sum_{h=1}^{K}\sigma_h\cdot W_h\right)^2+\frac{1}{m}\cdot\sum_{h=1}^{K}\sigma_h{}^2\cdot W_h{}^2\end{aligned}$$

Es ist weiters:

$$E\left(\sum_{h=1}^{K} w_h\cdot\sigma_h\right)=\sum_{h=1}^{K} W_h\cdot\sigma_h$$

Schließlich gilt noch:

$$\begin{aligned}E\left(\sum_{h=1}^{K}\left(w_h-W_h\right)\cdot\mu_h\right)^2&=\underbrace{\left(E\left(\sum_{h=1}^{K}\left(w_h-W_h\right)\cdot\mu_h\right)\right)^2}_{0}+\operatorname{Var}\left(\sum_{h=1}^{K}\left(w_h-W_h\right)\cdot\mu_h\right)\\&=\sum_{h=1}^{K}\mu_h{}^2\cdot\operatorname{Var} w_h+\sum_{h\neq i}\operatorname{Kov}\left(w_h\cdot\mu_h,w_i\cdot\mu_i\right)\\&=\frac{1}{m}\cdot\sum_{h=1}^{K}\mu_h{}^2\cdot W_h\cdot(1-W_h)-\frac{1}{m}\cdot\sum_{h\neq i}\mu_h\cdot\mu_i\cdot W_h\cdot W_i\\&=\frac{1}{m}\cdot\sum_{h=1}^{K}\mu_h{}^2\cdot W_h\cdot(1-W_h)-\frac{1}{m}\cdot\left(\sum_{h=1}^{K}W_h\cdot\mu_h\right)^2+\frac{1}{m}\cdot\sum_{h=1}^{K}W_h{}^2\cdot\mu_h{}^2\\&=\frac{1}{m}\cdot\sum_{h=1}^{K}W_h\cdot\mu_h{}^2-\frac{1}{m}\cdot\left(\sum_{h=1}^{K}W_h\cdot\mu_h\right)^2\\&=\frac{1}{m}\cdot\sum_{h=1}^{K}W_h\cdot(\mu_h-\mu)^2\end{aligned}$$

Wir fassen die Ergebnisse zusammen (vgl. etwa Dayal, 1979, S.162):

$$\begin{aligned}\operatorname{Var}\overline{x}_{O*} &= \frac{m-1}{m\cdot n}\cdot\left(\sum_{h=1}^{K} W_h\cdot\sigma_h\right)^2+\frac{1}{m\cdot n}\cdot\sum_{h=1}^{K} W_h\cdot\sigma_h{}^2+\frac{1}{m}\cdot\sum_{h=1}^{K} W_h\cdot(\mu_h-\mu)^2 \\ &= \frac{1}{n}\cdot\left(\sum_{h=1}^{K} W_h\cdot\sigma_h\right)^2+\frac{1}{m\cdot n}\cdot\sum_{h=1}^{K} W_h\cdot(\sigma_h-M\,\sigma_h)^2+\frac{1}{m}\cdot\sum_{h=1}^{K} W_h\cdot(\mu_h-\mu)^2\end{aligned} \tag{27}$$

Die Varianz des so erhobenen Schätzers $\overline{x}_{O*}$ nach (21) besteht demnach aus drei Summanden, die sich folgendermaßen darstellen lassen:

$$\operatorname{Var}\overline{x}_{O*}=\frac{1}{n}\cdot(M\,\sigma_h)^2+\frac{1}{m\cdot n}\cdot\operatorname{Var}\sigma_h+\frac{1}{m}\cdot\operatorname{Var}\mu_h \tag{27'}$$

Der Mittelwertschätzer $\overline{x}_{O*}$ konvergiert − wie sich mit einem Blick bestätigen lässt − mit wachsendem Umfang m der zur Schätzung der Schichtgrößen verwendeten uneingeschränkten Zufallsstichprobe gegen die Varianz des Mittelwertschätzers $\overline{x}_O$ nach (18):

$$\lim_{m\to\infty}\operatorname{Var}\overline{x}_{O*}=\operatorname{Var}\overline{x}_O$$

Die Varianz (27) lässt sich im Vergleich zur Verwendung proportionaler Aufteilung bei geschätzten Schichtgrößen folgendermaßen darstellen:

$$\begin{aligned}\operatorname{Var}\overline{x}_{O*} &= \frac{1}{n}\cdot\left(\sum_{h=1}^{K} W_h\cdot\sigma_h\right)^2+\frac{1}{n}\cdot\sum_{h=1}^{K} W_h\cdot\sigma_h{}^2-\frac{1}{n}\cdot\sum_{h=1}^{K} W_h\cdot\sigma_h{}^2+ \\ &\quad +\frac{1}{m\cdot n}\cdot\sum_{h=1}^{K} W_h\cdot(\sigma_h-M\,\sigma_h)^2+\frac{1}{n}\cdot\sum_{h=1}^{K} W_h\cdot(\mu_h-\mu)^2 \\ &= \operatorname{Var}\overline{x}_{P*}-\frac{m-1}{m\cdot n}\cdot\sum_{h=1}^{K} W_h\cdot(\sigma_h-M\,\sigma_h)^2\end{aligned}$$

$$\operatorname{Var}\overline{x}_{O*}=\operatorname{Var}\overline{x}_{P*}-\frac{m-1}{m\cdot n}\cdot\operatorname{Var}\sigma_h \tag{28}$$

Der Subtrahend auf der rechten Seite von (28) gibt die Verminderung der Schätzervarianz an, wenn bei Verwendung geschätzter Schichtgrößen optimal statt proportional geschichtet wird. Der Schichtungseffekt Δ_{O*} lässt sich mit (27) folgendermaßen darstellen:

$$\begin{aligned}\Delta_{O*} &= \operatorname{Var}\overline{x}_U-\operatorname{Var}\overline{x}_{O*} \\ &= \frac{m-1}{m\cdot n}\cdot\sum_{h=1}^{K} W_h\cdot(\sigma_h-M\,\sigma_h)^2+\frac{m-n}{m\cdot n}\cdot\sum_{h=1}^{K} W_h\cdot(\mu_h-\mu)^2 \\ &= \frac{m-1}{m}\cdot\Delta_O-\frac{n-1}{m}\cdot\Delta_P\end{aligned}$$

$$\Delta_{O*}=\frac{1}{m}\cdot\left((m-1)\cdot\Delta_O-(n-1)\cdot\Delta_P\right) \tag{29}$$

Der Schichtungseffekt Δ_{O*} ist demnach genau dann nicht positiv, wenn entweder $\Delta_O=\Delta_P=0$ ist, wenn also auch bei Kenntnis der Parameter σ_h kein Schichtungseffekt erzielt werden würde oder für $\Delta_O>0$, wenn $(m-1)\cdot\Delta_O<(n-1)\cdot\Delta_P$ ist, wenn also die Basisstichprobe zu klein ausgefallen ist.

Sind die Informationen hinsichtlich der Schichtstandardabweichungen σ_h des Merkmals X vorhanden (h=1,2,...K), dann sollte man also auch bei Verwendung geschätzter

Schichtgrößen optimal schichten, um den Schichtungseffekt zu maximieren und dabei darauf achten, dass m > n ist. Das garantiert jedenfalls einen positiven Schichtungseffekt!

Beispiel (Fortsetzung):
Bei Schätzung der Schichtgrößen in einer uneingeschränkten Zufallsauswahl vom Umfang m = 1.000 gilt für eine proportionale Schichtung mit geschätzten Schichtgrößen nach (15):

$$\operatorname{Var}\overline{x}_{P*} = \operatorname{Var}\overline{x}_{U} = 5{,}16 \cdot 10^{-3}$$

bzw. für eine optimale Schichtung mit geschätzten Schichtgrößen nach (27):

$$\operatorname{Var}\overline{x}_{O*} = 4{,}20096 \cdot 10^{-3}$$

IV.3 Schätzung der Schichtgrößen und der Schichtstandardabweichungen

Liegen weder die Parameter W_h noch σ_h vor, dann ist eine optimale Aufteilung mit den in der Basisstichprobe vom Umfang m gewonnenen Schätzern w_h und s_h durch

$$n_h = \frac{w_h \cdot s_h}{\sum_{h=1}^{K} w_h \cdot s_h} \cdot n \tag{30}$$

nur dann sinnvoll, wenn m im Vergleich zu n sehr groß ist und sich die geschätzten Schichtstandardabweichungen s_h deutlich voneinander unterscheiden. Dies ist notwendig, um nicht im Vergleich zu einer uneingeschränkten Zufallsauswahl einen Genauigkeitsverlust hinnehmen zu müssen. Andernfalls sollte man uneingeschränkt zufällig ziehen.

V Schichtung in der laufenden Erhebung

Die eleganteste Lösung für die Problematik der Unkenntnis der für eine proportionale bzw. optimale Schichtung nötigen Vorkenntnisse wäre natürlich deren Schätzung in der laufenden Erhebung: Erhoben wird ein erster Teil vom Umfang m der Gesamtstichprobe vom Umfang n mit m < n. Nach diesen m Erhebungseinheiten werden

- die fehlenden Parameter geschätzt
- die Schichtstichprobenumfänge n_h mit diesen Informationen festgelegt und
- die restlichen m – n Erhebungseinheiten so gezogen, dass sich die Gesamtstichprobe am Ende mit den im zweiten Schritt bestimmten Umfängen n_h auf die K Schichten aufteilt.

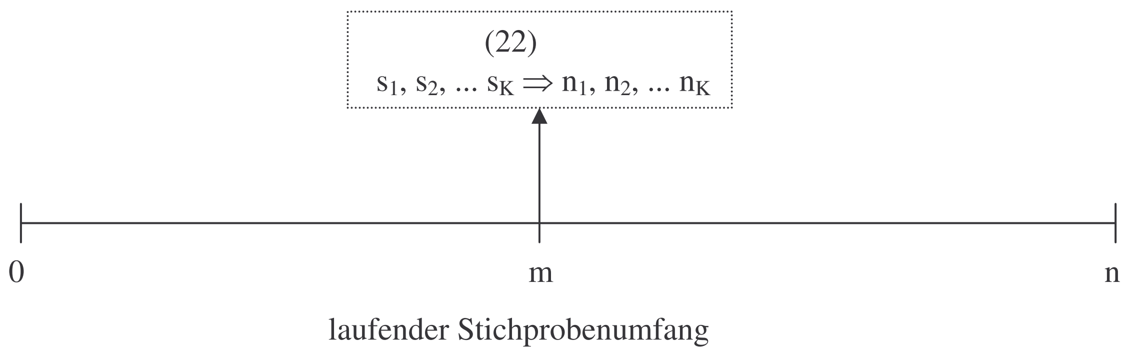
V.1 Proportionale Schichtung

Die Varianz von $\overline{x}_{P*}$ nach (15) bzw. der Schichtungseffekt Δ_{P*} nach (16) machen deutlich, dass für eine proportionale Schichtung die benötigten Schichtgrößen nicht in der laufenden Erhebung geschätzt werden können, da die uneingeschränkt zufällig gezogene Basisstichprobe einen Umfang m > n aufweisen müsste, um genauere Resultate zu erzielen als mit einer uneingeschränkt zufälligen Gesamtstichprobe vom Umfang n. Dies ist bei dem Modell, das hier betrachtet werden soll, per definitionem unmöglich.

V.2 Optimale Schichtung

Einen positiven Schichtungseffekt kann es bei Schätzung der fehlenden Parameter in der laufenden Erhebung demnach nur bei „optimalen“ Konzeptionen geben. Der in Abschnitt IV.1 geschilderte, häufigste Fall fehlender Schichtstandardabweichungen des Merkmals X bietet sich geradezu für diese Vorgangsweise an. Die Stichprobenziehung folge dabei nachstehendem Modell: Zunächst werden, da die Kenntnisse über die Schichtanteile W_h des Schichtmerkmals Y vorliegen, m Erhebungseinheiten aus der Grundgesamtheit proportional geschichtet nach einem Schichtmerkmal Y gezogen. An diesen m Erhebungseinheiten wird das Untersuchungsmerkmal X beobachtet und die Schichtstandardabweichungen s_h der Stichprobe berechnet. Von einer weiteren proportionalen Aufteilung nach (6) kann dann, wenn die Schätzer s_h stark streuen, auf eine optimale übergegangen werden. Mit diesen Schätzern lassen sich nach (22) die Gesamtstichprobenumfänge n_h der Schichten für eine optimale Aufteilung bestimmen (siehe dazu Abbildung 2).

Abbildung 2:
Die optimale Schichtung in der laufenden Erhebung:



Für die geeignete Wahl des Stichprobenumfanges m ist dabei einerseits entscheidend, dass in der Folge noch eine Abweichung von der proportionalen in Richtung einer optimalen Aufteilung des Gesamtstichprobenumfanges n erreicht werden kann. Andererseits aber ist mit zunehmendem m eine genauere Schätzung der σ_h möglich. Auch organisatorische Gründe können für die Festlegung des Umfanges m maßgeblich sein: Bei Befragungen lassen im Allgemeinen nur telefonische Interviews eine sich nicht auf die statistischen Eigenschaften des Schätzers auswirkende Unterbrechung der Erhebung nach beliebig vielen Einheiten zu. Bei Befragungen von Angesicht zu Angesicht oder brieflichen Befragungen aber muss die Erhebung von vornherein in zwei Teile zerlegt werden können.

Im Grunde aber ist m frei wählbar, wobei sich − organisatorisch betrachtet − möglicherweise die Hälfte des geplanten Gesamtstichprobenumfanges anbietet.

Wenn für alle errechneten Schichtstichprobenumfänge n_h gilt: $n_h \geq m_h$, dann sind aus jeder Schicht noch weitere $n_h - m_h$ Einheiten uneingeschränkt zufällig zu ziehen. Wenn aber für mindestens eine Schicht gilt: $m_h > n_h$, dann ist in der Schicht bzw. den Schichten, für die dies zutrifft, der optimale Stichprobenumfang bereits überschritten.

Unter dieser Nebenbedingung sind dann die restlichen Schichtstichprobenumfänge festzulegen, die die Schätzervarianz minimieren und in diesem Sinne wiederum optimal sind.
Als Lösung dieser Extremwertaufgabe ergibt sich für die noch unvollständigen Schichten:

$$n_h = \frac{W_h \cdot s_h}{\sum\limits_{h \in A} W_h \cdot s_h} \cdot \left(n - \sum_{h \in A^C} m_h \right) \tag{31}$$

(A ... die Menge jener Schichten, für die gilt: $n_h \geq m_h$; A^C ... Komplementärmenge von A; die beiden Mengen A und A^C bilden also eine Zerlegung aller Schichten).
Mit den Informationen s_h und n_h lässt sich mit (5) nach dem ersten Teil von m Erhebungseinheiten die Schätzervarianz erwartungstreu schätzen. Bei $\widehat{\mathrm{Var}}\, \overline{x}_S \approx \widehat{\mathrm{Var}}\, \overline{x}_P$ sollte man dann proportional weiterziehen. Ist $\widehat{\mathrm{Var}}\, \overline{x}_S << \widehat{\mathrm{Var}}\, \overline{x}_P$, dann wird man mit optimaler Schichtung in der laufenden Erhebung einen weiteren Schichtungseffekt erzielen können. Der endgültige Varianzschätzer errechnet sich erst am Ende der Stichprobe vom Umfang n wiederum mit (5).
Bei dieser Vorgangsweise der Schätzung der Schichtstandardabweichungen in der laufenden Erhebung gilt es zu beachten, dass der Mittelwertschätzer $\overline{x}_{O**}$ in der Gesamtstichprobe vom Umfang n verzerrt ist, soferne die Statistiken $\overline{x}_P$ und s_h^2 beim Stichprobenumfang m nicht statistisch unabhängig voneinander sind. Diese sind aber dann und nur dann unabhängig voneinander, wenn X normalverteilt ist (vgl. etwa Fisz, 1980, S.402ff). Bei der Erhebung von Anteilen etwa sind die Anteilschätzer und die Varianzschätzer für diese Anteilschätzer offenbar nicht unabhängig voneinander – man betrachte nur deren Formeln –, so dass die Varianzschätzer, weil sie die Anzahl der noch zu ziehenden Elemente bestimmen, dafür sorgen, dass am Ende der Anteil einer interessierenden Eigenschaft in der Stichprobe vom Umfang n nicht erwartungstreu für den unbekannten Anteil in der Grundgesamtheit ist. Allerdings ist die Verzerrung – wie ebenso leicht nachvollziehbar ist – vernachlässigbar gering.
Bei der in Abschnitt IV.2 beschriebenen Parameterausgangssituation sind die Standardabweichungen gegeben. Streuen diese nur gering, so gilt: $\mathrm{Var}\, \overline{x}_{O*} = \mathrm{Var}\, \overline{x}_{P*}$ und die in diesem Abschnitt beschriebene Vorgangsweise ist nicht anwendbar, da – wie oben bereits beschrieben wurde – eine proportionale Schichtung während der laufenden Erhebung keine genaueren Ergebnisse erzielen kann als eine durchgehend uneingeschränkte Ziehung aller n Erhebungseinheiten. Var σ_h muss also sehr groß sein, damit mit einer Schätzung der Schichtanteile nach $m << n$ Einheiten noch ein Genauigkeitsgewinn verbunden ist. Insgesamt wird diese Voraussetzung nur selten erfüllt sein. Trifft sie aber zu, so ist mit unserer Vorgangsweise in diesem Fall jedenfalls eine unverzerrte Schätzung des Mittelwertes μ möglich.
Bei der in Abschnitt IV.3 beschriebenen Parametersituation schließlich muss m – wie oben beschrieben – sehr groß gewählt werden und dies ist per definitionem bei der hier beschriebenen Vorgangsweise während der laufenden Erhebung nicht möglich.

VI Zusammenfassung

Die Wahl einer einheitlichen Darstellungsform für die mit den verschiedenen Schichtenverfahren erzielbaren Schätzervarianzen mit Hilfe der Varianz der Schichtmittelwerte und den Mittelwerten der Schichtstandardabweichungen und -varianzen macht deutlich, worauf es bei den einzelnen Verfahren ankommt, wenn man die Genauigkeit von Stichprobenergebnissen erhöhen will.
Die Varianz eines Mittelwertschätzers aus einer Stichprobe vom Umfang n ist

$$\text{Var}\,\overline{x}_U = \frac{1}{n}\cdot \text{M}\,\sigma_h^{\;2} + \frac{1}{n}\cdot \text{Var}\,\mu_h$$

bei uneingeschränkter Zufallsauswahl,

$$\text{Var}\,\overline{x}_P = \frac{1}{n}\cdot \text{M}\,\sigma_h^{\;2}$$

bei proportional geschichteter Zufallsauswahl, und

$$\text{Var}\,\overline{x}_{P*} = \frac{1}{n}\cdot \text{M}\,\sigma_h^{\;2} + \frac{1}{m}\cdot \text{Var}\,\mu_h$$

bei proportional geschichteter Zufallsauswahl mit geschätzten Schichtanteilen. Die Schätzung der Schichtanteile beruht dabei auf einer von der eigentlichen Stichprobe unabhängigen, uneingeschränkten Basiszufallsauswahl vom Umfang m.
Weiters ist

$$\text{Var}\,\overline{x}_O = \frac{1}{n}\cdot \left(\text{M}\,\sigma_h\right)^2$$

bei optimal geschichteter Zufallsauswahl,

$$\text{Var}\,\overline{x}_{O*} = \frac{1}{n}\cdot \left(\text{M}\,\sigma_h\right)^2 + \frac{1}{m\cdot n}\cdot \text{Var}\,\sigma_h + \frac{1}{m}\cdot \text{Var}\,\mu_h$$

bei optimal geschichteter Zufallsauswahl mit geschätzten Schichtanteilen aus einer davon unabhängigen, uneingeschränkten Basiszufallsauswahl vom Umfang m.
Schließlich ist

$$\text{Var}\,\overline{x}_{O**} \approx \frac{1}{n}\cdot \text{E}\left(\text{M}\,\sigma_h\right)^2 + \frac{1}{n}\cdot \sum_{h\neq i} W_h \cdot W_i \cdot \sigma_h \cdot \sigma_i \cdot V_h^{\;2}$$

bei optimal geschichteter Zufallsauswahl mit geschätzten Schichtstandardabweichungen und einer für diese Schätzung verwendeten, davon unabhängigen, uneingeschränkten Zufallsauswahl vom Umfang m.
Die Anwendung der Schichtenverfahren muss somit keineswegs an der Unkenntnis der dafür benötigten Parameter scheitern. Schätzungen z.B. aus den laufenden, großen nationalen Erhebungen wie dem Mikrozensus können die fehlenden Vorkenntnisse ersetzen und sind zudem sehr genau und aktueller als z.B. Volkszählungsergebnisse.
Werden – für die optimale Schichtung – auch die Schichtstandardabweichungen des Untersuchungsmerkmals selbst benötigt, dann zeigt sich, dass oftmals auch sehr kleine Basisstichproben für den Zweck der Schätzung dieser Parameter genügen. Dies gewährleistet immer noch einen Genauigkeitsgewinn gegenüber uneingeschränkter Zufallsauswahl, wenn nur die Schichtstandardabweichungen stark streuen.
Die Schätzung der benötigten Parameter in der laufenden Erhebung ist bei fehlenden Schichtstandardabweichungen für eine optimale Allokation des Stichprobenumfanges zu empfehlen – auch deshalb, weil man bei geringen Unterschieden in den Standardab-

weichungen immer noch eine proportionale Schichtung ohne zusätzlichen organisatorischen Aufwand vornehmen kann. Für die anderen Fälle fehlender Parameter ist diese Lösung jedoch ungeeignet.

Literatur

Cochran, W.G. (1972). *Stichprobenverfahren*. Walter de Gruyter. Berlin.

Dayal, S. (1979). Use of estimates of proportions of stratum sizes and standard deviations in allocation of sample to different strata under stratified random sampling. *Sankhya: The Indian Journal of Statistics*. Volume 41. Series C. Part 2. S. 159-175.

Fisz, M. (1980). *Wahrscheinlichkeitsrechnung und mathematische Statistik*. 10. Auflage. VEB Deutscher Verlag der Wissenschaften. Berlin.

Hansen, M.H., Hurwitz, W.N., Madow, W.G. (1953). *Sample Survey Methods and Theory. Volume I: Methods and Applications*. John Wiley & Sons. New York.

Quatember, A. (1996). Schichten mit geschätzten Schichtgrößen. *Österreichische Zeitschrift für Statistik*. 25. Jahrgang. Heft 1. S. 43-58.

Quatember, A. (2001). *Die Quotenverfahren – Stichprobentheorie und -praxis*. Shaker Verlag, Aachen.

Sadooghi-Alvandi, S.M. (1988). A Note on the Efficiency of Proportional Stratification. *Australian Journal of Statistics*. 30(2). S. 196-199.

Sukhatme, P.V. (1935). Contributions to the Theory of the Representative Method. *Journal of the Royal Statistical Society, Supplement*. S. 253-268.

Zarkovich, S.S. (1962). A Supplement to „Note on the History of Sampling Methods in Russia“. In: Kendall, M., Plackett, R.L. (eds.) (1977). *Studies in the History of Statistics and Probability*. Volume II. Charles Griffin & Company Limited. London. S. 486-488.