

Reduktion systematischer und zufallsbedingter Fehler in Zielgebietsdaten: Theorie und empirische Ergebnisse einer Waldbesucherbefragung in Zürich

(Mit 6 Tabellen)

Von K. BERNATH¹⁾, P. ELSASSER²⁾ und A. ROSCHEWITZ³⁾

(Angenommen August 2006)

SCHLAGWÖRTER – KEY WORDS

Zielgebietsbefragung; Quellgebietsbefragung; Auswahlwahrscheinlichkeit; Selbstselektion; Gewichtung; Bias; Walderholung; ökonomische Bewertung; kontingente Bewertungsmethode.

On-site survey; off-site survey; selection probability; self selection; weighting; bias; forest recreation; economic valuation; contingent valuation method.

1. EINLEITUNG

Die Verwendung von Umfragedaten als Grundlage für Analysen und Entscheidungsprozesse ist weit verbreitet. So werden auch Informationen zur Nutzung der Wälder als Erholungsräume (z.B. LINDHAGEN, 1996; BÜRG et al., 1999; BUWAL, 1999) oder zum Wert der Erholungsfunktion des Waldes (z.B. LÖWENSTEIN, 1994; ELSASSER, 1996; TYRVÄINEN, 2001) durch Befragungen unter Waldbesuchenden gewonnen. Von zentraler Bedeutung für die Qualität der Daten ist die Auswahl der befragten Personen. Um auf Basis statistischer Modelle Populationsparameter schätzen zu können, muss die Stichprobenziehung den jeweiligen Modellvoraussetzungen entsprechen (vgl. z.B. COCHRAN, 1977). So müssen etwa bei einfachen Zufallsstichproben alle Elemente der Grundgesamtheit mit derselben Wahrscheinlichkeit in die Stichprobe aufgenommen werden. Dies ist bei Umfragen in der Praxis selten möglich. Ein wesentliches Problem besteht etwa darin, dass die Befragten oft selbst Einfluss auf den Auswahlprozess nehmen (Selbstselektion). Dies ist insbesondere der Fall, wenn ausgewählte Besucher und Besucherinnen eines bestimmten Ortes befragt werden, denn die Befragten bestimmen selbst darüber, ob und wie häufig sie diesen Ort aufsuchen. Derartige Befragungen sind neben der erwähnten Verwendung in Untersuchungen zur Walderholung auch in anderen Bereichen sehr verbreitet. Beispiele sind Pendlerbefragungen in Verkehrsmitteln, Besucherbefragungen auf Märkten und Ausstellungen, Kundenbefragungen in Marktzentren und Geschäften oder Gästebefragungen in Erholungs- und Tourismusgebieten. In letzterem Zusammenhang werden vor Ort durchgeführte Befragungen als *Zielgebietsbefragungen* bezeichnet. Wir verwenden diesen Begriff im Folgenden allgemein für Befragungen, bei denen die Teilnehmenden 'vor Ort' ausgewählt werden.

Ist eine Stichprobe durch Selbstselektion beeinflusst, dann sind die Ergebnisse verschiedener Schätzverfahren, wie zum Beispiel Mittelwertschätzungen, möglicherweise verzerrt. Zur Korrektur dieser systematischen Abweichung der geschätzten Werte von den 'wahren' Werten der Grundgesamtheit werden die Rohdaten der

Stichprobe häufig gewichtet (DIEKMANN, 2002, S. 365). Verbreitete Methoden der Gewichtung sind die *Designgewichtung* und das *Redressement*. Bei der Designgewichtung werden durch den Stichprobenplan gegebene Unterschiede in der Auswahlwahrscheinlichkeit berücksichtigt. Das Redressement entspricht einer Anpassung der Daten an externe Häufigkeitsverteilungen, die zum Beispiel amtlichen Statistiken entnommen werden. Bei dieser Vorgehensweise wird unterstellt, dass die zu untersuchenden Grössen mit den externen Anpassungsgrössen korreliert sind.

In der statistischen Literatur werden die theoretischen Grundlagen und die Wirkung der Gewichtung ausführlich dargestellt (z.B. GABLER et al., 1994; KISH und HEERINGA, 2003). Zudem werden Gewichtungen in verschiedenen Anwendungsbereichen verbreitet eingesetzt, wie zum Beispiel in regelmässig durchgeführten sozialwissenschaftlichen Bevölkerungsumfragen (ROTHER, 1994). Die besonderen Aspekte der Stichprobenauswahl und der Gewichtung bei Erhebungen im Zielgebiet werden jedoch sowohl in theoretischen als auch in empirischen Arbeiten sehr oft vernachlässigt, obwohl das Problem der stichprobenbedingten Verzerrung in Zielgebietsbefragungen hinlänglich bekannt ist (z.B. LUCAS, 1963; ROEDER, 1973).

Die vorliegende Untersuchung zur Designgewichtung von Zielgebietsdaten dient dazu, diese Lücke zwischen statistischer Theorie und praktischer Anwendung zu schliessen. Ausgehend von theoretischen Analysen der Effekte der Stichprobenwahl und der Gewichtung wird untersucht, unter welchen Bedingungen die Gewichtung von Daten aus Zielgebietsstichproben zu empfehlen ist und wann auf die Gewichtung verzichtet werden kann. Anschliessend wird die Praktikabilität einer Designgewichtung von Zielgebietsdaten unter Feldbedingungen empirisch überprüft. Dazu werden die Herleitung und die Anwendung der Gewichtungsfaktoren exemplarisch anhand zweier Befragungen zum Waldbesucherverhalten in der Stadt Zürich dargestellt. Für diese Befragungen wurden die Waldbesuchenden zum einen im Zielgebiet und zum anderen im Quellgebiet ausgewählt, so dass die Quellgebietsdaten zur Kontrolle der gewichteten Zielgebietsdaten herangezogen werden können. Die Ermittlung des Erholungswertes der Stadtwälder in monetären Grössen war ein zentraler Aspekt der Befragungen. Das Beispiel zeigt daher insbesondere die Bedeutung der Gewichtung von Zielgebietsdaten in ökonomischen Bewertungsstudien. Aus praktischen Gründen werden auch hier häufig Stichprobenziehungen im Untersuchungsgebiet bevorzugt (z.B. LUTTMANN und SCHRÖDER, 1995; HUTCHINSON et al., 2001; KLAPHAKE und MEYERHOFF, 2003).

Im folgenden Abschnitt 2 wird ein Modellrahmen zur Herleitung von Designgewichten in Zielgebietsbefragungen dargestellt. Zudem werden die Effekte der Gewichtung aus theoretisch-statistischer Sicht analysiert und Kriterien dafür formuliert, wann gewichtete Mittelwerte zu verwenden sind. In Abschnitt 3 wird die theoretische Analyse mit den empirischen Ergebnissen der Zielgebietsbefragung in den Zürcher Stadtwäldern illustriert; dies dient der Kontrolle, ob die Gewichtung auch unter Praxisbedingungen im Vergleich mit den Quellgebietsdaten zu akzeptablen Ergebnissen führt. Schliesslich werden die theoretischen und empirischen Erkenntnisse diskutiert und ein abschliessendes Fazit gezogen (Abschnitt 4).

¹⁾ Eidg. Forschungsanstalt für Wald, Schnee und Landschaft, Forschungsgruppe Umwelt- und Ressourcenökonomie, Zürcherstrasse 111, CH-8903 Birmensdorf (Tel. +41 (0) 44 739 23 84, Fax +41 (0) 44 739 22 15, E-Mail: katrin.bernath@wsl.ch)

²⁾ Institut für Ökonomie, Bundesforschungsanstalt für Wald- und Holzwirtschaft, Leuschnerstrasse 91, D-21031 Hamburg (Tel. +49 (0) 40 739 62 309, Fax ++49 (0) 40 739 62 399, E-Mail: p.elsasser@holz.uni-hamburg.de)

³⁾ Eidg. Forschungsanstalt für Wald, Schnee und Landschaft, Forschungsgruppe Umwelt- und Ressourcenökonomie, Zürcherstrasse 111, CH-8903 Birmensdorf (Tel. +41 (0) 44 739 23 84, Fax +41 (0) 44 739 22 15, E-Mail: anna.roschewitz@wsl.ch)

2. THEORIE ZUR GEWICHTUNG VON ZIELGEBIETS DATEN

2.1 Modellrahmen für die Herleitung von Gewichtungsfaktoren

Mit der Designgewichtung werden Rohdaten einer Stichprobe mittels der jeweiligen Auswahlwahrscheinlichkeiten gemäss Stichprobenplan gewichtet (DIEKMANN, 2002). Dabei werden die Gewichte umgekehrt proportional zu den individuellen Auswahlwahrscheinlichkeiten gebildet. In Zielgebietsbefragungen sind diese Wahrscheinlichkeiten jedoch nur teilweise im Voraus bekannt, da sie sowohl durch das Untersuchungsdesign als auch durch das Besucherverhalten beeinflusst werden.

Ein bereits früher entwickelter Modellrahmen (ELSASSER, 1996; 2001) bietet eine Grundlage für die systematische Berücksichtigung verschiedener Einflussgrössen auf die Auswahlwahrscheinlichkeit. Dabei wird eine Serie stochastischer Ereignisse betrachtet, die für das Zustandekommen einer auswertbaren Antwort zu einer Interviewfrage verantwortlich sind. Die Wahrscheinlichkeiten dieser einzelnen in *Tabelle 1* aufgeführten Ereignisse bilden die Grundlage für die Berechnung der Auswahlwahrscheinlichkeiten.

Die Wahrscheinlichkeiten der einzelnen Ereignisse sind nicht beobachtbar, können aber über relative Häufigkeiten näherungsweise hergeleitet werden. Zur Operationalisierung der Wahrscheinlichkeiten wird unterstellt, dass diese proportional zu den jeweiligen beobachteten Häufigkeiten sind (vgl. ELSASSER, 1996, S. 43ff.)

Die in *Tabelle 1* aufgeführten Ereignisse und ihre Eintretenswahrscheinlichkeit können in drei Typen eingeteilt werden. Der erste Typ umfasst die Ereignisse, die durch die Untersuchungsanlage steuerbar sind. Sowohl die *Antreff-Wahrscheinlichkeit des Interviewers* als auch die *Ansprech-Wahrscheinlichkeit* können so gesteuert werden, dass sie für alle Befragten konstant sind. Ist dies der Fall, so können diese beiden Grössen bei der Berechnung der Auswahlwahrscheinlichkeit vernachlässigt werden, wie weiter unten gezeigt wird.

Zum zweiten Typ gehören Ereignisse, deren Wahrscheinlichkeiten aufgrund der Daten einer Besucherbefragung nicht über relative

Häufigkeiten hergeleitet werden können. Dies betrifft die *globale und die spezifische Antwort-Wahrscheinlichkeit*. Die globale Ausschöpfung (der gesamten Stichprobe) sowie die spezifische Ausschöpfung (d.h. die Antwortraten auf die einzelnen Fragen) geben zwar Auskunft über die jeweiligen durchschnittlichen Antwort-Wahrscheinlichkeiten aller Stichprobeneinheiten. Auf der Ebene der Individuen sind jedoch nur dichotome Ereignisse beobachtbar (Antwort liegt vor oder nicht), nicht aber individuelle Wahrscheinlichkeiten, an der Befragung teilzunehmen bzw. die einzelnen Fragen zu beantworten. Die Antwort-Wahrscheinlichkeiten können deshalb bei der Berechnung der Gewichtungsfaktoren nicht berücksichtigt werden und allfällige durch die Teilnahmebereitschaft bedingte systematische Fehler können nicht durch Gewichtung ausgeglichen werden.

Schliesslich bleiben zwei Ereignisse des dritten Typs, die durch die Untersuchungsanlage nicht oder nur teilweise gesteuert werden können, deren relative Häufigkeiten aber mit den entsprechenden Daten berechnet werden können. Zum einen betrifft dies das Finden einer Person am Ort des Interviews und zum anderen die freie Kapazität des Interviewers. Die *Antreff-Wahrscheinlichkeit des Besuchers* wird proportional zur Anzahl der Besuche innerhalb eines bestimmten Zeitraumes ermittelt. Zwar kann die Antreffwahrscheinlichkeit je nach Stichprobendesign auch durch die Dauer der Besuche beeinflusst werden. Dies ist aber vermeidbar, wenn die Befragungsstandorte so gewählt werden, dass sie von allen Besuchenden nur einmal pro Besuch passiert werden – zum Beispiel an den Eingängen des Untersuchungsgebietes statt innerhalb des Gebietes. Die *Wahrscheinlichkeit freier Kapazität des Interviewers* kann anhand des Anteils der angesprochenen Personen an der Gesamtzahl der PassantInnen hergeleitet werden. Diese Grösse bezieht sich jeweils auf die einzelnen Befragungsstandorte und eine bestimmte Zeiteinheit. Für die praktische Anwendung darf diese nicht zu gross gewählt werden, damit Schwankungen der Besucherzahlen im Tagesverlauf berücksichtigt werden.

Aufgrund dieser Betrachtungen berechnen sich die Antreffwahrscheinlichkeit $P[Bes_i]$ der Besucher $i=1, \dots, n$ und die Wahrscheinlichkeit freier Kapazität des Interviewers am Ort (s) und zum Zeitpunkt (t) des Interviews $P[K_{s,t}]$ proportional zu den beobachteten Häufigkeiten. Sind die Ereignisse voneinander unabhängig, entspricht die Auswahlwahrscheinlichkeit $P[A_i]$ der einzelnen Besucher dem Produkt von $P[Bes_i]$ und $P[K_{s,t}]$:

$$P[A_i] = P[Bes_i] \cdot P[K_{s,t}], \text{ mit} \quad (1)$$

$$P[Bes_i] \propto \frac{\text{Anzahl Besuche von Person } i}{\text{Bezugszeitraum}}$$

$$P[K_{s,t}] \propto \frac{\text{Anzahl befragte Personen}}{\text{Anzahl passierende Personen}}$$

Der resultierende Gewichtungsfaktor w_i für Besucher i ist durch den Kehrwert der jeweiligen Auswahlwahrscheinlichkeit bis auf einen konstanten Faktor bestimmt. Eine verbreitete Vorgehensweise ist die Normierung der Gewichte, so dass deren Summe dem Stichprobenumfang n entspricht:

$$w_i = \frac{1}{P[A_i]} \cdot \frac{n}{\sum_{i=1}^n \frac{1}{P[A_i]}} \quad (2)$$

Da die Relation dieser individuellen Gewichte zueinander bei der Multiplikation der Auswahlwahrscheinlichkeiten $P[A_i]$ mit einem konstanten Faktor unverändert bleibt, können bei der Berechnung von $P[Bes_i]$ und $P[K_{s,t}]$ beliebige Proportionalitätsfaktoren verwendet werden.

Tab. 1

Beschreibung der Ereignisse, die einen Einfluss auf die Auswahlwahrscheinlichkeit haben.

Description of events influencing the selection probability.

Ereignis	korrespondierendes Element der Auswahlwahrscheinlichkeit
Der Besucher findet sich am Ort des Interviews ein.	Antreff-Wahrscheinlichkeit des Besuchers
Gleichzeitig ist am selben Standort ein Interviewer anwesend.	Antreff-Wahrscheinlichkeit des Interviewers
Der Interviewer hat Zeit, den Besucher anzusprechen.	Wahrscheinlichkeit freier Interviewer-Kapazität
Der Interviewer spricht den Besucher an.	Ansprech-Wahrscheinlichkeit
Der Besucher ist grundsätzlich bereit, befragt zu werden.	Globale Antwort-Wahrscheinlichkeit
Der Besucher gibt Auskunft zu der entsprechenden Frage.	Spezifische Antwort-Wahrscheinlichkeit

Quelle: ELSASSER (2001, S. 24), verändert.

2.2 Schätzer für den gewichteten Mittelwert und dessen Varianz

Mit der Gewichtung der Daten wird eine Reduktion der stichprobenbedingten Verzerrung angestrebt, die bei der Berechnung von statistischen Kenngrößen auftreten kann. Bei den folgenden Betrachtungen werden die praktischen Probleme bei der Herleitung der Gewichtungsfaktoren zunächst hintangestellt, und wir unterstellen, dass unterschiedliche Auswahlwahrscheinlichkeiten durch die Gewichtung vollständig ausgeglichen werden können.

Wichtige Kenngrößen für die Darstellung von Merkmalen einer Stichprobe sind Mittelwert und Varianz. Im Folgenden werden die Schätzer für ungewichtete und gewichtete Mittelwerte sowie deren jeweilige Varianz dargestellt.

Sei $y = (y_1, \dots, y_n)$ eine Realisation von unabhängigen, gleichverteilten Zufallsvariablen mit Mittelwert μ und Varianz σ_y^2 . Mit $w = (w_1, \dots, w_n)$ bezeichnen wir die individuellen Gewichte der einzelnen Beobachtungen $i = 1, \dots, n$, wobei w_i umgekehrt proportional zur Auswahlwahrscheinlichkeit von Beobachtung i gebildet wird. Unter Verwendung des üblichen Schätzers

$$\text{var}(y) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

für die Varianz resultieren die in *Tabelle 2* dargestellten Schätzer für die Mittelwerte und deren Varianzen.

Für Zufallsstichproben ist der ungewichtete Mittelwert-Schätzer erwartungstreu, d.h. die Schätzung entspricht stets dem ‚wahren‘ Wert der Grundgesamtheit (COCHRAN, 1977, S. 22). Liegen unterschiedliche Auswahlwahrscheinlichkeiten der einzelnen Stichprobeneinheiten vor, so ist die Mittelwertschätzung verzerrt. Hingegen ist der Schätzer für den gewichteten Mittelwert, mit den umgekehrt proportional zur Auswahlwahrscheinlichkeit berechneten individuellen Gewichten, konsistent (KISH, 1992, S. 189).¹ Das bedeutet, dass sich die Schätzung des gewichteten Mittelwertes mit wachsendem Stichprobenumfang dem zu schätzenden Wert der Grundgesamtheit annähert.

Ein Vorteil des hier dargestellten Schätzers für den gewichteten Mittelwert besteht darin, dass der geschätzte Wert unverändert bleibt bei einer Multiplikation der Gewichte w_i mit einem konstanten Faktor (ROTHE und WIEDENBECK, 1994, S. 49). Zudem wird dabei auch die Varianz des gewichteten Mittelwertes nicht verän-

¹ Ein Schätzer heisst konsistent, wenn er „in Wahrscheinlichkeit“ gegen den zu schätzenden Wert konvergiert („convergence in probability“). Mit zunehmender Stichprobengröße wird die Differenz zwischen einem konsistenten Schätzer und dem zu schätzenden Parameter reduziert (vgl. z.B. GREENE, 2003, S. 899).

dert. Daraus folgt, dass die Bestimmung der Auswahlwahrscheinlichkeiten bis auf eine Konstante ausreichend ist, und die in Abschnitt 2.1 beschriebene Normierung der Gewichte keinen Einfluss auf den gewichteten Mittelwert und dessen Varianz hat.

2.3 Effekte der Gewichtung

2.3.1 Reduktion des stichprobenbedingten Bias

Aufgrund unterschiedlicher Auswahlwahrscheinlichkeiten der Stichprobeneinheiten sind auf ungewichteten Zielgebietsdaten basierende Mittelwertschätzungen möglicherweise mit systematischen Fehlern behaftet. Die systematische Abweichung eines geschätzten Wertes vom ‚wahren‘ Wert der Grundgesamtheit wird als Verzerrung oder *Bias* bezeichnet. Der zu schätzende Wert der Grundgesamtheit ist in der Regel nicht bekannt; er wird jedoch mit steigendem Stichprobenumfang durch den gewichteten Mittelwert angenähert, da der oben dargestellte Schätzer für den gewichteten Mittelwert konsistent ist. Deshalb wird die Schätzung des gewichteten Mittelwertes als beste zur Verfügung stehende Referenzgröße für die Analyse der systematischen Fehler herangezogen. Danach entspricht der stichprobenbedingte Bias des ungewichteten Mittelwertes der Differenz zwischen dem ungewichteten und dem gewichteten Mittelwert: $\text{Bias}(\bar{y}_w) = 0$ und $\text{Bias}(\bar{y}) = \bar{y} - \bar{y}_w$.

Das Ausmass des Bias des Mittelwertschätzers wird durch die Standardabweichung der Gewichte (σ_w), deren Mittelwert (\bar{w}), die Standardabweichung des untersuchten Merkmals (σ_y) und die Korrelation (R_{wy}) zwischen den Gewichten w und dem Merkmal y bestimmt (KISH, 1987, S. 237):

$$\text{Bias}(\bar{y}) = \bar{y} - \bar{y}_w = -\frac{\sigma_w}{\bar{w}} \sigma_y R_{wy} \quad (3)$$

Für normierte Gewichtungsfaktoren ist der Mittelwert der individuellen Gewichte $\bar{w} = 1$, und somit gilt

$$\text{Bias}(\bar{y}) = -\sigma_w \sigma_y R_{wy}.$$

Die Abweichung des Schätzers vom wahren Wert kann positiv oder negativ sein, ausschlaggebend ist der absolute Betrag des Bias. Wie die obigen Formeln zeigen, wird der absolute Betrag des Bias größer, je höher

- die (relative) Varianz der Gewichte²
 - die Varianz des Untersuchungsmerkmals und
 - die Korrelation zwischen den Gewichten und dem Untersuchungsmerkmal
- sind.

Insbesondere ist der ungewichtete Schätzer dann unverzerrt (d.h. $\text{Bias}(\bar{y}) = 0$), wenn der Gewichtungsfaktor und das Untersuchungsmerkmal unkorreliert sind (d.h. $R_{wy} = 0$).

Tab. 2

Schätzer für Mittelwert und Varianz von ungewichteten und gewichteten Stichproben.
Mean and variance estimators for weighted and unweighted samples.

Ungewichteter Mittelwert	$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
Varianz des ungewichteten Mittelwertes	$\text{var}(\bar{y}) = \frac{\text{var}(y)}{n} = \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2$
Gewichteter Mittelwert	$\bar{y}_w = \frac{\sum_{i=1}^n y_i w_i}{\sum_{i=1}^n w_i}$
Varianz des gewichteten Mittelwertes	$\text{var}(\bar{y}_w) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \cdot \frac{\sum_{i=1}^n w_i^2}{\left(\sum_{i=1}^n w_i\right)^2}$

Quellen: COCHRAN (1977, S. 20, 26), KISH (1992, S. 188, 191).

2.3.2 Auswirkungen auf die Varianz des gewichteten Mittelwertes

Wie durch einfache Umformung aus *Tabelle 2* hervorgeht, unterscheidet sich die Varianz des gewichteten Mittelwertes von derjenigen des ungewichteten Mittelwertes um den Faktor

$$n \cdot \frac{\sum_{i=1}^n w_i^2}{\left(\sum_{i=1}^n w_i\right)^2},$$

den wir in Übereinstimmung mit KISH (1990; 1992) mit $1 + L$ bezeichnen. Es gilt also

$$\text{var}(\bar{y}_w) = \text{var}(\bar{y}) \cdot (1 + L),$$

wobei L ein Mass für die relative Veränderung der Varianz darstellt. Bei konstanten Gewichten ($w_i = c$ für alle i) wird die Varianz des Mittelwertes durch die Gewichtung nicht verändert. Sind die individuellen Gewichte w_i nicht konstant, so gilt $(1 + L) > 1$. Folglich wird die Varianz des geschätzten Mittelwertes durch die Gewichtung erhöht. Diese Zunahme der Varianz ist gleichbedeutend mit einer Verringerung der effektiven Stichprobengrösse (RÖSCH, 1994, S. 13) um den Faktor $(1 + L)$.²⁾ Der Effekt der Gewichtung auf die Varianz des Mittelwertes wird mit zunehmendem Stichprobenumfang durch die relative Varianz der Gewichte angenähert (KISH, 1992, S. 191):

$$\text{var}(\bar{y}_w) = \text{var}(\bar{y}) \cdot (1 + L) \approx \text{var}(\bar{y}) \cdot \left(1 + \frac{\sigma_w^2}{\bar{w}^2}\right), \quad (4)$$

wobei σ_w^2 die Varianz der Gewichte und \bar{w} wie oben deren Mittelwert bezeichnet. Für normierte Gewichtungsfaktoren mit $\bar{w} = 1$ gilt somit

$$\text{var}(\bar{y}_w) \approx \text{var}(\bar{y}) \cdot (1 + \sigma_w^2).$$

Diese Gleichungen zeigen, dass die Varianz des gewichteten Mittelwertes mit zunehmender (relativer) Varianz der Gewichte steigt. Dieser unerwünschte Effekt der Gewichtung (Verringerung der Aussagesicherheit) ist dem erwünschten Effekt (Verringerung des Bias) gegenüberzustellen.

2.4 Kriterien zur Verwendung gewichteter Mittelwerte

Mit dem hier betrachteten konsistenten Schätzer für den gewichteten Mittelwert kann der 'wahre' Mittelwert der Grundgesamtheit angenähert werden. Durch die Gewichtung wird daher der systematische stichprobenbedingte Fehler beseitigt, soweit die Gewichte die unterschiedlichen Auswahlwahrscheinlichkeiten korrekt abbilden. Allerdings weist der gewichtete Mittelwert eine grössere Varianz auf als der ungewichtete. Die Entscheidung, ob der gewichtete Mittelwert dem ungewichteten vorzuziehen ist, basiert auf einer Analyse dieser beiden Effekte. Dabei sind verschiedene Untersuchungsmerkmale einzeln zu betrachten, da sie unterschiedliche Korrelationen mit den Gewichten und damit unterschiedliche Biases aufweisen können (vgl. *Formel 3*). Der Einfluss der Gewichtung auf die Varianzen der Mittelwerte ist hingegen für alle Untersuchungsgrössen gleich (*Formel 4*).

Die Summe der Varianz und des quadrierten Bias eines Schätzers wird als mittlerer quadratischer Fehler (mean squared error, MSE) bezeichnet (STAHEL, 1995, S. 163). Die Verwendung dieses

²⁾ Als relative Varianz wird das Verhältnis der Varianz zum quadrierten Mittelwert bezeichnet: $\frac{\sigma_w^2}{\bar{w}^2}$.

³⁾ Die Varianz des Mittelwertes eines Stichprobenmerkmals y beträgt $\frac{\sigma_y^2}{n}$, wobei n den Stichprobenumfang bezeichnet. Die Varianz des gewichteten Mittelwertes ist gleich $\frac{\sigma_y^2}{n} = (1 + L) \cdot \frac{\sigma_y^2}{n}$ und entspricht deshalb der Varianz des Mittelwertschätzers in einer Zufallsstichprobe vom Umfang $n' = \frac{n}{1 + L}$. Das Verhältnis $\frac{n'}{n} = \frac{1}{1 + L}$ wird als *Effektivität* der Stichprobe bezeichnet (GABLER und HÄDER, 1997, S. 237).

Gütemasses als Kriterium für den Einsatz gewichteter Mittelwerte wird im folgenden Abschnitt vorgestellt. Anschliessend formulieren wir auf der Basis der *Formeln 3* und *4* weitere Bedingungen, wann gewichtete Mittelwerte gegenüber den ungewichteten vorzuziehen sind (Abschnitt 2.3.2).

2.4.1 Reduktion des mittleren quadratischen Fehlers (MSE)

Der mittlere quadratische Fehler (MSE), der die Qualität eines Schätzers in Abhängigkeit der Varianz und des Bias misst, kann als Kriterium für die Berechnung gewichteter Mittelwerte verwendet werden. KISH (1992) schlägt vor, auf die Gewichtung des Mittelwertes zu verzichten, wenn der MSE des gewichteten Mittelwertes grösser ist als derjenige des ungewichteten Mittelwertes.

Für einen Schätzer \hat{y} mit Varianz $\text{var}(\hat{y})$ und Bias $B(\hat{y})$ gilt $MSE(\hat{y}) = \text{var}(\hat{y}) + B(\hat{y})^2$ (KISH, 1965, S. 60). Aufgrund der Formeln für den Bias und die Varianz des ungewichteten Mittelwertes \bar{y} und des gewichteten Mittelwertes \bar{y}_w (Abschnitte 2.2 und 2.3) gilt:

$$MSE(\bar{y}) = \text{var}(\bar{y}) + B(\bar{y})^2 = \frac{\sigma_y^2}{n} + (\bar{y} - \bar{y}_w)^2 = \frac{\sigma_y^2}{n} + \frac{\sigma_w^2}{\bar{w}^2} \sigma_y^2 R_{wy}^2,$$

$$MSE(\bar{y}_w) = \text{var}(\bar{y}_w) = \sigma_y^2 \cdot \frac{\sum w_i^2}{\left(\sum w_i\right)^2} \approx \frac{\sigma_y^2}{n} \cdot \left(1 + \frac{\sigma_w^2}{\bar{w}^2}\right)$$

Nach dem Kriterium von KISH (1992) ist der gewichtete Mittelwert vorzuziehen, wenn der MSE durch die Gewichtung reduziert wird, d.h. wenn gilt: $\Delta MSE = MSE(\bar{y}) - MSE(\bar{y}_w) > 0$.

2.4.2 Korrelation zwischen Untersuchungsmerkmal und Gewichtungsfaktor

Da der Effekt der Gewichtung auf den Mittelwert eines Untersuchungsmerkmals von der Varianz der Gewichte (bzw. der Auswahlwahrscheinlichkeiten) und von der Korrelation der Gewichte mit dem Merkmal abhängig ist (*Tab. 2* und *Formel 3*) und sich zudem die Varianz der Gewichte auf die Varianz des gewichteten Mittelwertes auswirkt (*Formel 4*), formulieren wir ein Kriterium für die Verwendung gewichteter Mittelwerte, das auf der Varianz der Gewichte und der Korrelation zwischen Gewichten und Untersuchungsmerkmal basiert. Dabei unterscheiden wir die folgenden drei Fälle:

– *Fall A:* $\sigma_w^2 = 0$

Ist die Varianz der Gewichte gleich Null, so ist der ungewichtete Mittelwert unverzerrt (*Formel 3*).⁴⁾ Zudem ist die Varianz des gewichteten Schätzers gleich gross wie diejenige des ungewichteten (*Formel 4*). Folglich spielt es in diesem Fall keine Rolle, ob gewichtete oder ungewichtete Mittelwerte berechnet werden.

– *Fall B:* $\sigma_w^2 > 0$ und $R_{wy} = 0$

Sind die Gewichte und das Merkmal unkorreliert, so ist der Bias des ungewichteten Mittelwertes gleich Null. Die positive Varianz der Gewichte führt jedoch nach *Formel 4* dazu, dass die Varianz des Mittelwertes durch die Gewichtung vergrössert wird. Deshalb sind bei dieser Ausgangslage die ungewichteten Mittelwerte den gewichteten vorzuziehen.

– *Fall C:* $\sigma_w^2 > 0$ und $R_{wy} \neq 0$

Unter diesen Bedingungen führt die Gewichtung zu einer Reduktion des stichprobenbedingten Fehlers. Deshalb ist der gewichtete

⁴⁾ In diesem Fall sind die Gewichte konstant, d.h. die Auswahlwahrscheinlichkeit ist für alle Einheiten der Stichprobe gleich gross. Damit besteht kein Grund für eine Gewichtung, aber es stellt sich die Frage, welche Auswirkungen zu erwarten sind, wenn die Daten trotzdem gewichtet werden. Wie die hier dargestellten Analysen zeigen, hat die Anwendung konstanter Gewichte keinen Effekt auf Mittelwertschätzungen und deren Varianz.

Tab. 3

Gewichtung unter unterschiedlichen Bedingungen bezüglich der Varianz der Gewichte (σ_w^2) und deren Korrelation (R_{wy}) mit dem Untersuchungsmerkmal (y).

Weighting under different conditions regarding the variance of the weights (σ_w^2) and their correlation with the study parameter (y).

Bedingungen	Effekte der Gewichtung	Folgerungen
<i>Fall A</i> $\sigma_w^2 = 0$	$\bar{y}_w = \bar{y}$ und $\text{var}(\bar{y}_w) = \text{var}(\bar{y})$	<i>Gewichtung irrelevant</i>
<i>Fall B</i> $\sigma_w^2 > 0$ und $R_{wy} = 0$	$\bar{y}_w = \bar{y}$ und $\text{var}(\bar{y}_w) > \text{var}(\bar{y})$	<i>ungewichtete Werte verwenden</i>
<i>Fall C</i> $\sigma_w^2 > 0$ und $R_{wy} \neq 0$	$\bar{y}_w \neq \bar{y}$ und $\text{var}(\bar{y}_w) > \text{var}(\bar{y})$	<i>gewichtete Werte verwenden</i>

Mittelwert gegenüber dem ungewichteten vorzuziehen. Allerdings wird dabei ein Anstieg der Varianz in Kauf genommen.

Tabelle 3 fasst diese Ergebnisse zusammen und stellt die Kriterien für die Verwendung gewichteter Mittelwerte dar.

Tatsächlich ist die Entscheidung zwischen gewichteten und ungewichteten Mittelwerten nur möglich, wenn die Gewichte bereits berechnet sind. Daher ist die Untersuchungsanlage generell auf die Gewichtung der Daten auszurichten, und allfällige Zusatzdaten für die Berechnung der Gewichtungsfaktoren sind zu ermitteln. Für die Darstellung von Untersuchungsergebnissen ist es grundsätzlich sinnvoll, sowohl gewichtete als auch ungewichtete Mittelwerte mitzuteilen. Die hier genannten Kriterien ermöglichen zu entscheiden, welchem der beiden Werte bei der Interpretation der Schätzwerte der Vorzug zu geben ist.

3. EMPIRISCHE RESULTATE: BEFRAGUNG VON WALDBESUCHENDEN IN ZÜRICH

3.1 Datenerhebung

Im Rahmen einer Studie zur Erholungsnutzung der Wälder auf dem Gebiet der Stadt Zürich wurden im September 2004 eine Ziel- sowie eine vergleichbare Quellgebietsbefragung durchgeführt. Die *Quellgebietsbefragung* beruht auf einer Stichprobe von 1500 zufällig ausgewählten Einwohnern und Einwohnerinnen der Stadt Zürich. Die Befragung erfolgte postalisch nach der Total Design Methode von DILLMAN (1978). Mit einer Erinnerungskarte und einem zweiten Versand des Fragebogens innerhalb von vier Wochen nach dem ersten Versand wurde eine Rücklaufquote von 38% erreicht.⁵⁾

Die *Zielgebietsbefragung* wurde an zehn Standorten in den Wäldern rund um die Stadt Zürich durchgeführt. Bei der Auswahl der Standorte wurde eine möglichst breite Abdeckung der Wälder angestrebt, zusätzlich wurden verschiedene inhaltliche und praktische Kriterien berücksichtigt.⁶⁾

Die befragten Personen wurden nach einem vorgegebenen Schema ausgewählt, das eine weitgehend zufällige und von den Merkmalen der Besuchenden unabhängige Auswahl anstrebte. Die InterviewerInnen hatten den Auftrag, die jeweils erste Person anzu-

⁵⁾ Aufgrund ungültiger Adressen konnten 28 Fragebögen nicht zugestellt werden. Von den verbleibenden 1472 kontaktierten Personen schickten 558 den ausgefüllten Fragebogen zurück.

⁶⁾ So wurden zum Beispiel vier Standorte aus einer 1987 durchgeführten Befragung (SCHELBERT et al., 1988) übernommen, um Vergleiche mit dieser Studie ziehen zu können.

sprechen, die eine Markierungslinie beim Befragungsstandort überschritt. Stimmt diese Person der Befragung zu, so bekam sie einen Fragebogen zur selbständigen Beantwortung ausgehändigt; zusätzlich wurden Fragebogen-Nummer und Uhrzeit notiert. Bei einer Ablehnung wurden der Verweigerungsgrund sowie die Tätigkeit, das Geschlecht und das geschätzte Alter der ablehnenden Person protokolliert. Sobald diese Aufgaben abgeschlossen waren, wurde die nächste Person angesprochen, die die Markierung überschritt. Durch das selbständige Ausfüllen des Fragebogens ist die Befragungssituation im Zielgebiet vergleichbar mit der postalischen Quellgebietsbefragung. Insgesamt wurden an den zehn Standorten 3146 Personen angesprochen und eine Antwortquote von 49% erreicht.⁷⁾

Für die nachfolgenden Vergleiche werden nur diejenigen Fragebögen herangezogen, die zwischen Quell- und Zielgebietsbefragung direkt vergleichbar sind. Aus der Quellgebietsbefragung werden nur Personen berücksichtigt, welche innerhalb der vergangenen 12 Monate mindestens einmal einen Wald im Stadtgebiet besucht haben und somit zur Grundgesamtheit der Waldbesuchenden gehören (88% der Antwortenden). In der Zielgebietsbefragung wurden unterschiedliche Fragebogenversionen verwendet; für den vorliegenden Artikel wird nur die 'Grundvariante' (65% der Antwortenden) herangezogen, die identisch ist mit der im Quellgebiet eingesetzten Version. Zudem werden aus beiden Befragungen nur diejenigen Fragebögen berücksichtigt, welche hinsichtlich aller hier untersuchten Merkmale vollständig sind. Für die nachfolgenden Vergleiche stehen damit 435 Antworten aus dem Zielgebiet zur Verfügung und 235 unmittelbar vergleichbare Antworten aus dem Quellgebiet.

Der Fragebogen beider Befragungen ist bei BERNATH und ROSCHEWITZ (submitted) beschrieben und wird hier nur kurz zusammengefasst. Er beinhaltet neben den üblichen Fragen zu sozioökonomischen Merkmalen einen Abschnitt zum Waldbesuchverhalten. Darin wurden zum Beispiel die Besuchshäufigkeit, die Dauer des letzten Waldbesuchs und die dabei ausgeübte Tätigkeit ermittelt. Ein zentraler Aspekt war zudem die Schätzung eines monetären Wertes der Walderholung für die Besucher und Besucherinnen. Dazu wurde die 'Contingent Valuation Method', eine verbreitete Methode zur ökonomischen Bewertung von aussermarktlichen Gütern, eingesetzt. Die Grundidee dieser Methode ist die Schaffung eines hypothetischen Marktes für das zu bewertende Gut, um die Zahlungsbereitschaft der Befragten für dieses Gut zu ermitteln. In der hier beschriebenen Studie wurde dazu ein hypothetisches Szenario beschrieben, das den Kauf einer Jahreskarte als Berechtigung für Waldbesuche innerhalb des Stadtgebiets postulierte. Die Zahlungsbereitschaft der Waldbesuchenden für eine solche Jahreskarte diente als Mass für die Wertschätzung der Walderholung.⁸⁾

3.2 Anwendung des Modellrahmens zur Berechnung der Gewichte

Gemäß *Formel 1* wurde die *Antreff-Wahrscheinlichkeit des Besuchers* über die Häufigkeit der Waldbesuche im Unter-

⁷⁾ Gründe für die Ablehnung und weitere Angaben zu den beiden Befragungen sind in BERNATH et al. (2006) beschrieben.

⁸⁾ Ein typisches Problem von Contingent Valuation Studien ist ein hoher Anteil von Personen, die das hypothetische Szenario nicht akzeptieren. Mit entsprechenden Fragestellungen können diese Protestantworten identifiziert werden und eine vertiefte Analyse von verschiedenen Einflussgrößen zeigte, dass sowohl sozioökonomische Merkmale als auch das Waldbesuchverhalten keinen signifikanten Einfluss auf die Wahrscheinlichkeit einer Protestantwort hatten (BERNATH und ROSCHEWITZ, submitted). Deshalb kann angenommen werden, dass die Entfernung der Protestantworten aus der Stichprobe (30% im Zielgebiet, 35% im Quellgebiet) bei den nachfolgenden Auswertungen keine systematischen Fehler bewirkte.

suchungsgebiet hergeleitet. Sie ist proportional zur Anzahl der Besuche innerhalb des Bezugszeitraumes von einem Jahr. Die Besuchshäufigkeit wurde in 11 Klassen erhoben, deren Mittelwerte in die Berechnung der Gewichtungsfaktoren eingingen.⁹⁾ Auf die Berücksichtigung unterschiedlicher Aufenthaltsdauern im Wald konnte verzichtet werden, da Lage der Befragungsorte und Befragungszeiten einen Einfluss der Aufenthaltszeit auf die Auswahlwahrscheinlichkeit nahezu ausschlossen. Zur Bestimmung der *Wahrscheinlichkeit freier Kapazität des Interviewers* wurden alle Personen gezählt, die während der Befragungszeiten die verschiedenen Standorte passierten, und in Relation zur Anzahl der für das Interview angesprochenen Personen gesetzt (s. *Formel 1*).

Die *Antreff-Wahrscheinlichkeit des Interviewers* wird durch den Befragungsplan bestimmt. Durch die Abdeckung von unterschiedlichen Standorten, Wochentagen und Tageszeiten wurde dieses Element der Auswahlwahrscheinlichkeit in unserer Befragung weitgehend konstant gehalten;¹⁰⁾ es wird daher bei der Gewichtung nicht berücksichtigt. Auch die *Ansprech-Wahrscheinlichkeit* ist über die Untersuchungsanlage steuerbar. Durch das in unserer Befragung vorgegebene Auswahlverfahren war die Wahl der angesprochenen Personen nicht von den persönlichen Präferenzen der InterviewerInnen abhängig. Somit war die Ansprech-Wahrscheinlichkeit näherungsweise für alle Befragten konstant und konnte bei der Berechnung der Auswahlwahrscheinlichkeit vernachlässigt werden. Die *globale* und die *spezielle Antwort-Wahrscheinlichkeit* können nicht über relative Häufigkeiten hergeleitet werden (s. Abschnitt 2.1) und konnten folglich bei der Berechnung der Auswahlwahrscheinlichkeit nicht berücksichtigt werden.

Bei der Gewichtung der in den Zürcher Wäldern befragten Waldbesuchenden wurden also die unterschiedlichen Antreff-Wahrscheinlichkeiten der Besuchenden und die Wahrscheinlichkeit freier Kapazität des Interviewers berücksichtigt. Die individuellen Gewichte wurden gemäss *Formel 2* in Abschnitt 2.1 berechnet. Da es sich um normierte Gewichte handelt, weisen sie einen Mittelwert von 1 auf, weitere Kenngrößen sind in *Tabelle 4* ersichtlich. Der Vergleich zwischen Minimum und Maximum zeigt, dass die Spannbreite der aus der Auswahlwahrscheinlichkeit der BesucherInnen resultierenden Gewichte bei rund 1:1000 liegt. Die Spannbreite der einzelnen Faktoren liegt bei 1:400 für die Antreffwahrscheinlichkeit und bei 1:10 für die Interviewerkapazität.

Tab. 4

Designgewichte für die Stichprobe der Zürcher Waldbesuchenden (n = 435).

Design weights for the sample of forest visitors in Zurich (n = 435).

Minimum	Maximum	Median	Mittelwert	Standardabweichung
0,02	16,28	0,14	1	2,20

⁹⁾ Die folgenden Klassenmittelwerte wurden für die Herleitung der Besuchshäufigkeit verwendet: 1, 2, 3, 4,5, 8 (für 1, 2, 3, 4-5, 6-10 Besuche pro Jahr), 12, 30 (für 1 bzw. 2-3 Besuche pro Monat) und 52, 156, 312 (für 1, 2-4, 5-7 Besuche pro Woche). Für 7 Antwortende, die mehrmals täglich in den Wald gingen, wurde eine Besuchshäufigkeit von 400 Besuchen pro Jahr unterstellt. Damit wurde die tatsächliche Besuchshäufigkeit tendenziell unterschätzt; allerdings ist anzunehmen, dass es auch Tage gibt, an denen diese Personen den Wald nicht besuchen (z.B. wegen Ferien oder Krankheit).

¹⁰⁾ Auf die Berücksichtigung verschiedener Jahreszeiten haben wir verzichtet. Falls die Untersuchungsmerkmale mit der Verteilung der Waldbesuche über das Jahr korreliert sind, unterliegen die Schätzungen einem systematischen Fehler. Frühere Studien zum Erholungswert von Stadtwäldern (ELSASSER, 1996; KLEIBER, 2003) zeigten jedoch, dass die Jahreszeit keinen Einfluss hat auf die Zahlungsbereitschaft der Waldbesuchenden, d.h. auf das Hauptmerkmal der Untersuchung.

3.3 Gewichtung in der Befragung von Zürcher Waldbesuchenden

Das in Abschnitt 2.4.2 formulierte Kriterium zur Verwendung gewichteter Mittelwerte basiert auf der Varianz der Gewichte und auf der Korrelation der Gewichte mit dem Untersuchungsmerkmal.¹¹⁾ Die zur inversen Auswahlwahrscheinlichkeit proportionalen Gewichte für die Stichprobe der Zürcher Waldbesuchenden weisen eine Varianz auf, die deutlich über Null liegt ($\sigma_w^2 = 4,8$). Somit sind für diejenigen Merkmale gewichtete Mittelwerte vorzuziehen, die mit den Gewichten korreliert sind. Die Korrelation der Gewichte mit verschiedenen Merkmalen der befragten Waldbesuchenden wird in Abschnitt 3.3.1 dargestellt. Anschliessend werden ungewichtete und gewichtete Mittelwerte sowie deren Standardabweichungen berechnet (Abschnitt 3.3.2) und die Effekte der Gewichtung diskutiert (Abschnitt 3.3.3).

3.3.1 Korrelation zwischen Gewichtungsfaktoren und Untersuchungsmerkmalen

Die Daten der Besucherbefragung wurden mit Korrelationstests auf systematische Zusammenhänge zwischen den Gewichten und einzelnen Merkmalen der Stichprobe untersucht. Da sowohl die Gewichte als auch die meisten Merkmale keiner Normalverteilung folgten, wurden auf der Spermanschen Rangkorrelation basierende Tests durchgeführt (*Tabelle 5*). Die überprüfte Nullhypothese der Korrelationstests lautete jeweils „keine Korrelation zwischen dem Merkmal und den Gewichten“. Ein Merkmal wird im Folgenden als mit den Gewichten korreliert bezeichnet, wenn diese Nullhypothese im Test abgelehnt wurde.¹³⁾

Die Korrelationstests zeigen, dass die einzelnen sozioökonomischen Merkmale der BesucherInnen in der Regel schwächer mit den Gewichten korreliert sind (oberer Teil von *Tab. 5*) als die einzelnen besuchsbezogenen Merkmale (unterer Teil der *Tabelle*). Wird ein Signifikanzniveau von 0,05 unterstellt, sind die Korrelationen der meisten sozioökonomischen Merkmale mit den Gewichten nicht signifikant; einzig beim Alter und der Haushaltsgrösse wird die Nullhypothese „keine Korrelation“ abgelehnt. Die deutliche Korrelation zwischen den Gewichten und dem Alter ist darauf zurückzuführen, dass die Gewichte hauptsächlich durch die

¹¹⁾ Dieser Betrachtung liegt die Annahme zugrunde, dass der Faktor $(1 + L)$ durch die Varianz der Gewichte angenähert werden kann (vgl. *Formel 4*). Im hier dargestellten Beispiel mit den in Abschnitt 3.2 hergeleiteten Gewichten stimmen die beiden Grössen gut überein:

$$1 + L = n \cdot \sum_{i=1}^n w_i^2 / \left(\sum_{i=1}^n w_i \right)^2 = 5,82 \text{ und } 1 + \sigma_w^2 = 5,83.$$

¹²⁾ Unter Umständen könnte es adäquat sein, a priori vom Vorliegen relevanter Korrelationen auszugehen und folglich die gegenteilige Nullhypothese („es existiert eine Korrelation“) zu prüfen (vgl. ELSASSER, 1996, Anhang A2; s. auch KRISTOFFERSSON und NAVRUD, 2005). Hier wird aus zwei Gründen darauf verzichtet: Erstens sind entsprechende Tests komplizierter; da es sich nicht um Standardverfahren handelt, würde ihre Darstellung den Rahmen dieses Artikels sprengen. Zweitens unterscheiden sich die Ergebnisse der Prüfung beider möglicher Nullhypothesen nur dann, wenn beide nicht abgelehnt werden können. Im Folgenden wird sich jedoch zeigen, dass die hier geprüfte Nullhypothese tatsächlich in vielen Fällen und insbesondere bezüglich der Zahlungsbereitschaft für eine Jahreskarte abgelehnt wird. Auch die Prüfung der gegenteiligen Nullhypothese würde in diesen Fällen also zu keinem anderen Ergebnis führen können.

¹³⁾ Da hier mehrere Tests am gleichen Datensatz durchgeführt werden („multiples Testen“, vgl. STAHEL, 1995, S. 245), steigt die Wahrscheinlichkeit, dass von den Testergebnissen mindestens eines falsch-positiv oder falsch-negativ ist. Dieser Effekt wäre durch eine entsprechende Korrektur des nominellen Signifikanzniveaus zu berücksichtigen, wenn es hier um die Grundsatzentscheidung ginge, ob alle Merkmale gewichtet werden sollen oder nicht. Tatsächlich wird hier jedoch für jedes Merkmal separat entschieden, ob es gewichtet werden soll. Die Resultate der Korrelationstests dienen als Richtlinie für diese Entscheidungen, und auf dieser deskriptiven Ebene kann auf die Anpassung der Tests (z.B. nach der Regel von Bonferroni) verzichtet werden.

Tab. 5

Korrelation zwischen den Gewichten und einzelnen Merkmalen der Stichprobe (n = 435).
Correlation between weights and different sample characteristics (n = 435).

Untersuchungsmerkmal	Korrelation mit den Gewichten	
	Korrelationskoeffizient	p-Wert ^a
Alter	-0,16	0,001 **
Geschlecht (Frauenanteil)	0,05	0,311
Einkommen	-0,03	0,567
Haushaltgrösse	-0,10	0,034 *
Haushalte mit Kindern	-0,09	0,053
Beschäftigung		
Erwerbstätige	0,07	0,135
Personen in Ausbildung	0,07	0,132
RentnerIn	-0,09	0,065
Mitgliedschaft in Umweltorganisation	-0,04	0,370
Hundebesitz	-0,21	0,000 ***
Besuchshäufigkeit ^b	-0,93	0,000 ***
Dauer des Waldbesuchs	0,13	0,009 **
Haupttätigkeit		
SpaziergängerIn	0,33	0,000 ***
JoggerIn	-0,24	0,000 ***
BikerIn	-0,04	0,401
Reisezeit zum Wald	0,35	0,000 ***
Zahlungsbereitschaft für eine Jahreskarte	-0,16	0,001 **

^a Die p-Werte zeigen mit welcher Sicherheit die Nullhypothese „keine Korrelation“ abgelehnt wird. Signifikanzniveaus *** p-Wert < 0,001, ** 0,001 < p-Wert < 0,01, * 0,01 < p-Wert < 0,05.

^b Anzahl Besuche im Untersuchungsgebiet während eines Jahres; berechnet aus Klassenmittelwerten.

Besuchshäufigkeit bestimmt werden und diese mit zunehmendem Alter ansteigt.

Die Korrelation der Gewichte mit den im unteren Teil der Tabelle dargestellten besuchsbezogenen Charakteristika ist dagegen (mit einer Ausnahme) stets signifikant. Eine positive Korrelation mit dem Gewichtungsfaktor bedeutet, dass der Mittelwert (bzw. bei dichotomen Merkmalen das Auftreten des Merkmals) durch die Gewichtung erhöht wird. Umgekehrt werden bei einer negativen Korrelation tiefere Werte stärker gewichtet, d.h. die Gewichtung führt zu einer Reduktion des Mittelwertes.

Insbesondere zeigt sich eine negative Korrelation zwischen der Zahlungsbereitschaft und den Gewichten. Das bedeutet, dass Personen mit hoher Zahlungsbereitschaft eine höhere Besuchsfrequenz haben und/oder bei ihren Besuchen die Wahrscheinlichkeit freier Interviewerkapazität höher war als im Durchschnitt aller befragter Personen. Folglich wurden Personen mit hoher Zahlungsbereitschaft mit einer erhöhten Wahrscheinlichkeit befragt. Die Gewichtung proportional zur inversen Auswahlwahrscheinlichkeit führt dazu, dass der Anteil der besonders zahlungsbereiten BesucherInnen in der korrigierten Stichprobe sinkt, und damit auch der geschätzte Mittelwert der Zahlungsbereitschaft.

3.3.2 Resultate der beiden Befragungen

Unter Verwendung der in Abschnitt 2.2 aufgeführten Schätzer wurden die Mittelwerte von stetigen Stichprobenmerkmalen und

Anteilswerte von dichotomen Merkmalen sowie deren jeweiligen Standardabweichungen berechnet. Dabei wurden für alle Merkmale ungewichtete und gewichtete Werte berechnet, unabhängig von ihrer Korrelation mit den Gewichten. *Tabelle 6* zeigt die ungewichteten Werte der Quell- und der Zielgebietsbefragung sowie die gewichteten Werte der Zielgebietsbefragung. In der zweitletzten Spalte ist der Effekt der Gewichtung auf den mittleren quadratischen Fehler (MSE) dargestellt. Positive Werte bedeuten, dass der MSE des gewichteten Mittelwertes geringer ist als derjenige des ungewichteten. Schliesslich zeigt die letzte Spalte, ob sich die ungewichteten Mittelwerte zwischen den beiden Befragungen signifikant unterscheiden.¹⁴⁾

Für die im oberen Teil von *Tabelle 6* aufgeführten Merkmale konnte die Nullhypothese „keine Korrelation“ in den entsprechenden Tests nicht abgelehnt werden. Nach dem in Abschnitt 2.4.2 formulierten Kriterium wird für diese Merkmale die Verwendung ungewichteter Mittelwerte empfohlen.¹⁵⁾ Übereinstimmend mit dem Kriterium von Kish (Abschnitt 2.4.1) sind ungewichtete Mittelwerte zu bevorzugen, wenn die Differenz der MSE der Mittelwertschätzer positiv ist. Dies ist einzig für das Geschlecht und das Merkmal ‚BikerIn‘ nicht der Fall.

Die Gewichtung bewirkt in dieser Untersuchung eine Vergrösserung der Varianz des Mittelwertes jeweils um den Faktor 5,8. Die Standardabweichung ist dementsprechend jeweils 2,4mal höher. Die Unterschiede zwischen gewichteten und ungewichteten Mittelwerten sind im oberen Teil der Tabelle im Allgemeinen gering. Die gewichteten Mittelwerte der meisten Merkmale liegen im 95%-Vertrauensbereich der entsprechenden ungewichteten Mittelwerte. Grössere Abweichungen sind lediglich beim Geschlecht und beim Merkmal ‚BikerIn‘ feststellbar. Der Frauenanteil ist in der gewichteten Stichprobe deutlich höher als in der ungewichteten. Im Vergleich mit der Schätzung aus dem Quellgebiet führt die Gewichtung hier zu einer ‚Überkorrektur‘. Der Anteil der Personen, die mit dem Fahrrad im Wald unterwegs waren, wird durch die Gewichtung reduziert, was im Vergleich mit dem Quellgebiet eine Anpassung in die falsche Richtung bedeutet. Allerdings ist zu berücksichtigen, dass auch die Quellgebietsbefragung keine unstrittige Referenz bietet, weil es sich hier ebenfalls um eine Schätzung handelt, die Zufallsfehler und angesichts der hohen Verweigerungsquote auch systematische Fehler aufweisen kann.

Werden die beiden Befragungen hinsichtlich der diversen ungewichteten Mittelwerte miteinander verglichen (vgl. letzte Spalte in *Tab. 6*), so sind bei den im oberen Teil der Tabelle dargestellten Merkmalen keine signifikanten Unterschiede feststellbar. Damit bestätigt sich die Empfehlung, ungewichtete Mittelwerte zu verwenden, wenn die Merkmale nicht mit den Gewichten korreliert sind.

Der untere Teil von *Tabelle 6* zeigt die Merkmale, die mit den Gewichten korreliert sind. Nur bei der Haushaltgrösse ist nach dem Kriterium von Kish der ungewichtete Mittelwert vorzuziehen, da der MSE des Mittelwertes durch die Gewichtung erhöht wird. Bei den übrigen Merkmalen, die mehrheitlich einen direkten Bezug zu

¹⁴⁾ Zur Überprüfung der Gleichverteilung von diskreten Merkmalen wurde der Chi-Quadrat-Test (mit Yates-Korrektur) angewendet. Bei stetigen Merkmalen wurde in einem ersten Schritt die Übereinstimmung der Varianzen geprüft (F-Test). Bei einer Ablehnung gleicher Varianzen (d.h. bei einem p-Wert des F-Tests kleiner als 0,05) wurde ein Welch-Test (Welch Modified Two-Sample t-Test¹⁾ durchgeführt, andernfalls ein konventioneller t-Test.

¹⁵⁾ Allerdings ist zu beachten, dass von der Tatsache, dass die Nullhypothese „keine Korrelation“ nicht abgelehnt werden kann, nicht eindeutig auf Unkorreliertheit geschlossen werden kann. Die Beibehaltung der Nullhypothese kann grundsätzlich auch auf einen zu geringen Stichprobenumfang zurückzuführen sein. Insbesondere bei kleinen Stichproben ist deshalb zusätzlich die gegenteilige Hypothese zu testen (vgl. Fussnote 12).

Tab. 6

Mittelwerte und Standardabweichungen (SD) von Stichprobenmerkmalen, Effekt der Gewichtung auf den mittleren quadratischen Fehler (MSE) und Unterschiede zwischen Ziel- und Quellgebietsdaten.

Mean and standard deviation (SD) of sample characteristics, the impact of weighting on the mean squared error and differences between on-site and off-site data.

		Quellgebiet (n=235)	Zielgebiet (n=435)		Ziel- und Quellgebiet	
		Mittelwert (SD)	Mittelwert (SD)	Gewichteter Mittelwert (SD)	ΔMSE^a	Unterschiede ^b (ungewichtete Mittelwerte)
Korrelation mit den Gewichten nicht signifikant ^c	Geschlecht (Anteil Frauen)	0,54 (0,03)	0,51 (0,02)	0,60 (0,06)	0,0054	0,44
	Einkommen ^d	7894 (308)	7 984 (217)	7 734 (523)	-164127	0,81
	HH mit Kindern	0,27 (0,03)	0,25 (0,02)	0,22 (0,05)	-0,0015	0,59
	Erwerbstätige	0,69 (0,03)	0,67 (0,02)	0,69 (0,05)	-0,0021	0,80
	Personen in Ausbildung	0,07 (0,02)	0,09 (0,01)	0,09 (0,03)	-0,0009	0,46
	RentnerIn	0,17 (0,02)	0,16 (0,02)	0,12 (0,04)	-0,0001	0,95
	Mitgliedschaft in Umweltorganisation	0,31 (0,03)	0,34 (0,02)	0,35 (0,05)	-0,0025	0,39
	BikerIn	0,09 (0,02)	0,05 (0,01)	0,02 (0,03)	0,0007	0,06
Korrelation mit den Gewichten signifikant	Alter	44 (1,0)	45 (0,7)	43 (1,7)	3,4394	0,25
	Haushaltgrösse	2,2 (0,07)	2,3 (0,06)	2,2 (0,14)	-0,0094	0,45
	Hundebesitz	0,07 (0,02)	0,10 (0,01)	0,05 (0,04)	0,0022	0,21
	Besuchsfrequenz	65 (6,2)	102 (4,6)	11 (11,1)	8162	0,00
	Dauer des Waldbesuchs	1,4 (0,06)	1,3 (0,04)	1,5 (0,09)	0,0200	0,08
	SpaziergängerIn	0,51 (0,03)	0,45 (0,02)	0,72 (0,06)	0,0703	0,13
	JoggerIn	0,18 (0,03)	0,35 (0,02)	0,18 (0,03)	0,0560	0,00
	Reisezeit zum Wald	16 (0,8)	14 (0,6)	16 (0,8)	92,758	0,21
	Zahlungsbereitschaft für eine Jahreskarte ^e	123 (8,7)	178 (11,8)	118 (28,4)	2949	0,00

^a $\Delta MSE = MSE(\bar{y}) - MSE(\bar{y}_w)$: Ist das Vorzeichen der Differenz zwischen dem MSE des ungewichteten und dem MSE des gewichteten Mittelwertes positiv, so ist nach dem Kriterium von Kish der gewichtete Mittelwert vorzuziehen.

^b p-Werte der 'Tests for equal means', d.h. t-Tests für stetige und Chi-Quadrat-Tests für diskrete Merkmale.

^c Die Nullhypothese „Merkmal und Gewicht nicht korreliert“ wird im Spearmanschen Korrelationstest abgelehnt (Signifikanzniveau 0,05).

^d Monatliches Nettoeinkommen des Haushaltes.

^e Revidierte Zahlungsbereitschaft, ohne Protestantworten.

den Waldbesuchen aufzeigen, wird sowohl die Nullhypothese „keine Korrelation zwischen Merkmal und Gewichten“ deutlich abgelehnt als auch der MSE durch die Gewichtung reduziert. Beide Kriterien legen also übereinstimmend die Verwendung gewichteter Mittelwerte nahe.

Auch hier wird die Varianz des Mittelwertes durch die Gewichtung um den Faktor 5,8 vergrößert. Gleichzeitig weichen die gewichteten Mittelwerte deutlich von den ungewichteten ab und liegen nur bei der Haushaltgrösse innerhalb des 95%-Konfidenz-

intervalles des entsprechenden ungewichteten Mittelwertes. Der Vergleich der ungewichteten Mittelwerte aus dem Ziel- und dem Quellgebiet weist bei den besuchsbezogenen Merkmalen mehrheitlich auf signifikante Unterschiede hin.

Insbesondere zeigt sich bei der Zahlungsbereitschaft für eine Jahreskarte sowohl ein deutlicher Einfluss der Gewichtung als auch ein signifikanter Unterschied zwischen Ziel- und Quellgebiet. Der ungewichtete Wert aus dem Quellgebiet ist rund 1,5mal grösser als der gewichtete Wert und als die durchschnittliche Zahlungsbereit-

schaft in der Quellgebietsbefragung. Die Berechnung des Erholungswertes der Wälder auf der Basis der ungewichteten Zielgebietsdaten würde somit zu einer klaren Überschätzung führen.

3.3.3 Interpretation der Resultate

Fall 1: Schwache Korrelation zwischen Merkmal und Gewichten, MSE durch Gewichtung erhöht

Die meisten sozioökonomischen Untersuchungsmerkmale waren nur schwach und insignifikant mit den Gewichten korreliert und der MSE des Mittelwertes wurde durch die Gewichtung erhöht. Unter diesen Bedingungen sprechen beide in diesem Artikel vorgestellten Kriterien dafür, ungewichtete Mittelwerte zu verwenden. Werden trotzdem gewichtete Werte verwendet, erhöht sich die Varianz entsprechend der theoretischen Ausführungen zu den Effekten der Gewichtung; dagegen bleiben die Unterschiede zwischen den ungewichteten und den gewichteten Mittelwerten gering, wenn Merkmale und Gewichte nicht signifikant korreliert sind. Die empirischen Resultate stimmen weitgehend mit den theoretischen Erwartungen überein. Nur beim Merkmal ‚RentnerIn‘ liegt der gewichtete Mittelwert knapp ausserhalb des 95%-Vertrauensbereiches des ungewichteten Mittelwertes.

Fall 2: Schwache Korrelation zwischen Merkmal und Gewichten, MSE durch Gewichtung reduziert

Bei den beiden Merkmalen ‚Geschlecht‘ und ‚BikerIn‘ wurde ebenfalls kein signifikanter Zusammenhang mit den Gewichten festgestellt. Der MSE des Mittelwertes hingegen wurde durch die Gewichtung reduziert. Es ist hier also nicht eindeutig zu entscheiden, ob die Gewichtung die Schätzung dieser Merkmale verbessert. Werden die Mittelwerte der Quellgebietsbefragung als beste zur Verfügung stehende Referenzgrösse betrachtet, so ist die Gewichtung abzulehnen, da die ungewichteten Mittelwerte näher bei denjenigen der Quellgebietsbefragung liegen als die gewichteten. Wie oben erwähnt, ist bei diesen Vergleichen allerdings einschränkend zu berücksichtigen, dass auch auf Quellgebietsdaten basierende Schätzungen systematische wie auch zufällige Fehler aufweisen können.

Fall 3: Signifikante Korrelation zwischen Merkmal und Gewichten, MSE durch Gewichtung erhöht

Die Haushaltsgrosse ist mit den Gewichten korreliert, der MSE wird durch die Gewichtung jedoch erhöht. Allerdings ist der Effekt auf den Mittelwert gering und sowohl gewichteter als auch ungewichteter Mittelwert liegen innerhalb des 95%-Vertrauensbereiches der Mittelwertschätzung mit Quellgebietsdaten. Gegen den gewichteten Mittelwert spricht einzig die höhere Varianz.

Fall 4: Signifikante Korrelation zwischen Merkmal und Gewichten, MSE durch Gewichtung reduziert

Bei den besuchsbezogenen Charakteristika weisen beide Kriterien deutlich darauf hin, dass ungewichtete Mittelwerte stark verzerrt sind und daher gewichtete Mittelwerte zu verwenden sind. Die einzige Ausnahme bildet das Merkmal ‚BikerIn‘. Der Vergleich mit den Quellgebietsdaten als Referenzgrösse zeigt jedoch, dass auch die gewichteten Ergebnisse mit Vorsicht zu interpretieren sind. Bei allen Merkmalen führt die Gewichtung zu einer ‚Überkorrektur‘. Die Gewichtung bewirkt zwar eine Anpassung in die Richtung des Wertes aus dem Quellgebiet, aber der Effekt ist zu stark in dem Sinne, dass die ungewichteten und die gewichteten Mittelwerte des Zielgebiets diejenigen des Quellgebiets einschliessen. Bei einigen Untersuchungsmerkmalen wird die Abweichung vom Mittelwert der Quellgebietsdaten durch die Gewichtung sogar vergrössert. Allerdings gelten auch hier die oben erwähnten Einschränkungen bezüglich der Eignung der Quellgebietsbefragung als unstrittige Referenz. Als weitere mögliche Erklärung für diese Resultate kommt in Betracht, dass bei der Gewichtung Grössen vernach-

lässigt worden sein könnten, die gleichwohl einen Einfluss auf die Auswahlwahrscheinlichkeit haben, oder dass die Datenlage für deren Berechnung teilweise unzureichend gewesen sein könnte.

4. DISKUSSION UND FAZIT

Ein zentraler Aspekt der Gewichtung von Zielgebietsdaten ist die Bestimmung der Gewichtungsfaktoren, die einen Ausgleich unterschiedlicher Auswahlwahrscheinlichkeiten ermöglichen. Solche Designgewichte sind üblicherweise durch die Untersuchungsanlage bestimmt. Eine breite Verwendung finden sie zum Beispiel bei Quellgebietsbefragungen beim Ausgleich unterschiedlicher Auswahlwahrscheinlichkeiten von Einzelpersonen in Abhängigkeit der Haushaltsgrosse, wenn pro Haushalt jeweils eine Zielperson ausgewählt wird (DIEKMANN, 2002, S. 365). Die Schwierigkeit der Bestimmung von Designgewichten in Zielgebietsbefragungen besteht darin, dass die Auswahlwahrscheinlichkeit durch verschiedene Ereignisse bestimmt wird, die nur teilweise durch das Untersuchungsdesign kontrolliert werden können.

Während die Gewichtung von Zielgebietsdaten in Abhängigkeit von einzelnen Elementen der Auswahlwahrscheinlichkeit wie Besuchshäufigkeit und -dauer bereits seit längerem diskutiert wurde (z. B. LUCAS, 1963; ROEDER, 1973; NOWELL et al., 1988), wurde erst später versucht, die Einflüsse auf die Auswahlwahrscheinlichkeit zur Herleitung von Gewichtungsfaktoren systematisch zu erfassen (ELSASSER, 1996). Allerdings sind auch dort lediglich Vorschläge für Mittelwertkorrekturen zu finden, ohne dass der Einfluss von Gewichtungen auf Varianzen sowie deren Folgen für die Praktikabilität der Gewichtung berücksichtigt wurden. Diese Lücke konnte hier geschlossen werden. Mit dem Vergleich zwischen Ziel- und Quellgebietsdaten unserer Waldbefragungen konnten wir zudem die Eignung des Modellrahmens für die Herleitung der Gewichte unter Praxisbedingungen empirisch überprüfen. Als wichtigsten Bestimmungsfaktor der Auswahlwahrscheinlichkeit, die von verschiedenen Faktoren beeinflusst wird, identifizierten wir für den vorliegenden Fall die Besuchshäufigkeit. Die Gewichtung führte im Allgemeinen zu einer Anpassung an die Quellgebietsdaten. Die Überkorrektur bei manchen Merkmalen könnte jedoch darauf hinweisen, dass weitere Aspekte des Selektionsprozesses bei der Berechnung der Gewichte in der empirischen Umsetzung nicht hinreichend erfasst wurden.

Die verschiedenen Einflüsse auf die Auswahlwahrscheinlichkeit sind bereits bei der Planung einer Befragung zu berücksichtigen. Durch ein entsprechendes Untersuchungsdesign sind möglichst konstante Auswahlwahrscheinlichkeiten anzustreben. Ein nachträglicher Ausgleich ist verschiedenen Fehlerquellen ausgesetzt und erfordert die Erfassung zusätzlicher Grössen, wie zum Beispiel die Anzahl der PassantInnen für die Berechnung der Interviewerkapazität. Eine vollständige Beseitigung stichprobenbedingter systematischer Fehler ist mit der Verwendung von Designgewichten in Zielgebietsbefragungen praktisch kaum zu erreichen. Zum einen können die einzelnen Wahrscheinlichkeiten empirisch nur annähernd über beobachtete Häufigkeiten – oder wie im Falle der Antwortbereitschaft überhaupt nicht – bestimmt werden. Zum anderen können auch weitere, im Modellrahmen nicht berücksichtigte Grössen die Auswahlwahrscheinlichkeit beeinflussen. So ist zum Beispiel zu vermuten, dass in unserer Befragung von Waldbesuchenden die Auswahl von Personen, die mit dem Fahrrad unterwegs waren, nicht allein durch die hier berücksichtigten Faktoren ‚Besuchshäufigkeit‘ und ‚Interviewerkapazität‘ beeinflusst wurde. Ihr Anteil war in der Zielgebietsbefragung deutlich tiefer als in der Quellgebietsbefragung, und dieser Unterschied wurde durch die Gewichtung nicht ausgeglichen.

Sind die Gewichtungsfaktoren bestimmt, stellt sich die Frage nach den Auswirkungen der Gewichtung auf die Schätzung von Populationsmerkmalen. Basierend auf der statistischen Literatur

wurden die Effekte der Gewichtung in der vorliegenden Untersuchung analytisch dargestellt, um daraus Folgerungen für die praktische Anwendung abzuleiten. Die theoretische Analyse zeigte, dass die Reduktion des systematischen Fehlers bei der Schätzung von Mittelwerten generell mit einer Zunahme der Varianz der Schätzer einhergeht. Unterschiedliche individuelle Auswahlwahrscheinlichkeiten und daraus resultierende Gewichte führen immer zu einer Erhöhung der Varianz, während systematische Fehler nur bei Untersuchungsmerkmalen auftreten, die mit der Auswahlwahrscheinlichkeit korreliert sind. Deshalb ist von einer durchgehenden Verwendung gewichteter Mittelwerte abzusehen.

In praktischen Anwendungen kann die Entscheidung über die Verwendung gewichteter Mittelwerte auf den mittleren quadratischen Fehler gestützt werden (KISH, 1990; 1992). Dieses Gütekriterium berücksichtigt gleichzeitig zufallsbedingte und systematische Fehler eines Schätzers. Allerdings wird dabei unterstellt, dass der gewichtete Mittelwert dem ‚wahren‘ Wert der zugrunde liegenden Population entspricht. Basierend auf den Auswirkungen der Gewichtung auf den Mittelwert und dessen Varianz haben wir ein weiteres Kriterium für die Verwendung von gewichteten Mittelwerten hergeleitet, das auf der Varianz der Gewichte selbst und deren Korrelation mit den Untersuchungsmerkmalen basiert. Offensichtlich hat die Gewichtung keinen Einfluss, wenn die Varianz der Gewichte Null ist, denn dann werden alle Stichprobenelemente gleich stark gewichtet. Bei positiver Varianz der Gewichte wird die Veränderung des Mittelwertes durch die Korrelation der Gewichte mit dem jeweiligen Merkmal bestimmt. Existiert solch eine Korrelation, dann besteht ein systematischer Zusammenhang zwischen der Auswahlwahrscheinlichkeit und dem Untersuchungsmerkmal. Durch die Gewichtung werden die unterschiedlichen Auswahlwahrscheinlichkeiten und daraus entstehende systematische Fehler korrigiert. Gleichzeitig erhöht sich aber auch die Varianz des Mittelwertes, was gleichbedeutend ist mit einer Reduktion der effektiven Stichprobengröße.

Diese Reduktion kann sehr einschneidend sein. Bei den hier verwendeten empirischen Daten wird die Stichprobe durch die Gewichtung effektiv auf etwa ein Sechstel ihres ursprünglichen Umfangs reduziert;¹⁶⁾ um die gleiche Aussagesicherheit zu erreichen wie bei einer Quellgebietsbefragung, wäre also beispielsweise statt einer Stichprobe von $n = 1000$ im Quellgebiet eine von $n = 5820$ im Zielgebiet zu ziehen. Bei noch höherer Varianz der Gewichte, die unter anderen Untersuchungsbedingungen durchaus auftreten kann, wäre der Unterschied noch größer. Die möglichen Effizienzvorteile von Zielgebiets- gegenüber Quellgebietsbefragungen werden dadurch stark relativiert, falls die Untersuchungsmerkmale mit den Auswahlwahrscheinlichkeiten (bzw. den entsprechenden Gewichten) korreliert sind.

Die Prüfung der Korrelation zwischen Gewichten und Untersuchungsmerkmalen ist also für die Praxis von erheblicher Bedeutung. Daher ist auch von Belang, auf welche Weise diese Prüfung vorgenommen wird bzw. wie die Nullhypothese für diese Prüfung formuliert wird. Insbesondere wenn aus theoretischen Erwägungen eine Korrelation zu erwarten ist – wie dies etwa für Zahlungsbereitschaften gegeben ist, die kausal von der Besuchshäufigkeit beeinflusst werden – führt die Prüfung der üblichen Nullhypothese „keine Korrelation“ nur dann zu eindeutigen Ergebnissen, wenn sie im Korrelationstest abgelehnt wird. Wird sie dagegen nicht abgelehnt, so kann dies sowohl auf Unkorreliertheit der Merkmale als auch auf einen zu kleinen Stichprobenumfang zurückzuführen sein. In einem solchen Fall wäre also zusätzlich zu prüfen, ob die alternative Nullhypothese („eine Korrelation zwischen Gewichten und

¹⁶⁾ Da der effektive Stichprobenumfang $n' = \frac{n}{1+L}$ ist (vgl. Fußnote 3), und $1+L$ für die vorliegenden Daten 5,82 beträgt (vgl. Fußnote 11), ergibt sich $n' = n \frac{1}{5,82} = 0,178 n$.

Untersuchungsmerkmal existiert“) abgelehnt wird. Erst wenn diese Hypothese abgelehnt wird, kann eindeutig geschlossen werden, dass die ungewichteten Werte vorzuziehen sind.

Das Problem unterschiedlicher Auswahlwahrscheinlichkeiten kann in Zielgebietsbefragungen nicht umgangen werden. Daher ist bei der Interpretation und Verallgemeinerung von Ergebnissen aus Zielgebietsbefragungen generell Vorsicht geboten: Ungewichtete Schätzungen können aufgrund der durch unterschiedliche Auswahlwahrscheinlichkeiten entstehenden systematischen Fehler erheblich verzerrt sein. Die Gewichtung korrigiert solche systematischen Fehler, verringert die Aussagesicherheit aber u.U. erheblich; gewichtete Schätzungen können also umgekehrt durch erheblich vergrößerte Zufallsfehler gekennzeichnet sein.

Der Einfluss der Stichprobenauswahl ist insbesondere auch bei ökonomischen Bewertungsstudien zu berücksichtigen, wie am Beispiel des Erholungswertes des Zürcher Stadtwaldes gezeigt. Einerseits werden auf ungewichteten Mittelwertschätzungen basierende aggregierte Werte über- oder unterschätzt, wenn die individuellen Zahlungsbereitschaften mit der Auswahlwahrscheinlichkeit korreliert sind. Andererseits wird die Zuverlässigkeit der gewichteten Schätzungen durch die mit der Gewichtung verbundene Vergrößerung der Vertrauensbereiche eingeschränkt.

Die vorliegende Untersuchung zur Gewichtung von Zielgebietsdaten zeigt zum einen, dass die dargestellten Kriterien zur Verwendung gewichteter oder ungewichteter Kenngrößen die Kenntnis der Gewichte voraussetzt. In der praktischen Anwendung ist somit jede Untersuchung grundsätzlich so anzulegen, dass die aus den individuellen Auswahlwahrscheinlichkeiten basierenden Gewichte hergeleitet werden können. Zum anderen liefert die Darstellung der systematischen und zufallsbedingten Fehler eine Grundlage für die Beurteilung ungewichteter und gewichteter Mittelwerte.

5. ZUSAMMENFASSUNG

Befragungen von Waldbesuchenden werden in Untersuchungen zur Erholungsnutzung von Wäldern und zur Bewertung von Erholungsleistungen verbreitet eingesetzt. In Zielgebietsbefragungen, bei denen die Befragten vor Ort ausgewählt werden, zeichnen sich die einzelnen Stichprobeneinheiten durch unterschiedliche Auswahlwahrscheinlichkeiten aus. Anhand eines Modellrahmens wurden die systematische Berücksichtigung der einzelnen Elemente dieser Auswahlwahrscheinlichkeiten aufgezeigt (Tab. 1) und individuelle Gewichtungsfaktoren hergeleitet.

Ausgehend von der theoretischen Analyse der Auswirkungen der Gewichtung auf Mittelwerte und deren Varianzen (Tab. 2), wurden Kriterien für die Verwendung gewichteter Mittelwerte formuliert. Diese basieren entweder auf dem ‚mean squared error‘ oder auf der Korrelation zwischen den Gewichtungsfaktoren und dem zu untersuchenden Merkmal. Gewichtete Mittelwerte sind für diejenigen Untersuchungsmerkmale zu bevorzugen, die mit den Gewichten korreliert sind (Tab. 3).

In der empirischen Anwendung wurden die Kriterien zur Verwendung gewichteter Mittelwerte anhand einer Ziel- und einer Quellgebietsbefragung von Besucherinnen und Besuchern der Stadtwälder von Zürich unter Praxisbedingungen überprüft. Korrelationstests zeigten mehrheitlich signifikante Zusammenhänge zwischen Merkmalen des Waldbesuchsverhaltens und den Gewichtungsfaktoren (Tab. 5). Die Gewichtung von Zielgebietsdaten führte bei diesen Merkmalen in den meisten Fällen zu einer Annäherung an die auf einer Zufallsauswahl basierenden Daten der Quellgebietsbefragung (Tab. 6), also zu einer Reduktion des stichprobenbedingten Bias.

Die analytische Darstellung der Effekte der Stichprobenwahl und der Gewichtung liefert eine Grundlage für die Beurteilung unge-

wichteter und gewichteter Kenngrößen von Zielgebietsstichproben. Die empirischen Ergebnisse weisen insbesondere auf die Bedeutung der Stichprobenauswahl in ökonomischen Bewertungstudien hin. Bei der Interpretation und Verallgemeinerung von Schätzergebnissen ist zu berücksichtigen, dass einerseits ungewichtete Werte verzerrt sein können und andererseits die Zuverlässigkeit der gewichteten Schätzungen durch eine Vergrößerung der Vertrauensbereiche reduziert wird.

6. Summary

Title of the paper: *Reducing systematic and random errors in on-site data: Theory and empirical results of a survey among forest visitors in Zurich.*

Visitor surveys are widely used to collect information on recreational uses of forests and to estimate recreational values. If visitors are sampled on-site, their selection probabilities depend on external factors and differ, for example, according to visit frequencies. These inequalities are compensated with weights which are inversely proportional to selection probabilities. This article presents a framework for incorporating different elements which influence a visitor's selection probability in an on-site sample (table 1).

Based on a theoretical analysis of the weighting procedure's impact on estimates of mean and its variance (table 2), criteria for the use of weighted means were derived. The first criterion is based on the mean squared error (MSE) of the weighted and the unweighted mean estimates. Following the second criterion, the weighted mean has to be preferred for parameters which are correlated with the weighting variable (table 3).

These criteria were examined empirically by comparing an on-site and an off-site survey of Zurich's forest visitors. Correlation tests between several study parameters and the weighting variable revealed that most parameters related to forest visits are significantly correlated with the weighting variable (table 5). In these cases, weighting the on-site data according to individual selection probabilities reduced the difference between on-site and off-site estimates (table 6) and hence, the sampling bias.

The analytical description of the impact of the sampling process and of the data weighting procedure provides a basis for evaluating unweighted and weighted parameter estimates from on-site samples. Specifically, the empirical results underline the importance of the sampling strategy in economic valuation studies. When interpreting and generalizing estimates from such studies, it has to be considered that on one hand unweighted values may possibly be biased; on the other hand, the reliability of weighted estimates is reduced due to increased confidence intervals.

7. Résumé

Titre de l'article: *Réduire les erreurs systématiques et celles liées à l'aléatoire dans les données recueillies sur le terrain: théorie et résultats empiriques d'un sondage auprès des visiteurs en forêt de Zurich.*

Les enquêtes sur l'utilisation des forêts comme espaces de loisirs et l'estimation de la valeur des prestations récréatives recourent largement aux sondages auprès des visiteurs en forêt. Dans les sondages sur le terrain, c'est-à-dire lorsque les personnes interrogées sont sélectionnées sur place, les différentes unités d'échantillonnage se caractérisent par diverses probabilités de sélection. À l'aide d'un cadre modélisé, il a été possible de montrer la prise en considération systématique des éléments qui influencent ces probabilités de sélection (Tab. 1), et de déduire les facteurs individuels de pondération.

Sur la base de l'analyse théorique des effets de la pondération sur les valeurs moyennes et leurs variances (Tab. 2) ont été formulés

des critères pour l'application des valeurs moyennes pondérées. Ceux-ci reposent soit sur le principe du 'mean squared error', soit sur la corrélation entre les facteurs de pondération et le paramètre sujet à étude. On privilégiera les valeurs moyennes pondérées pour les paramètres d'étude corrélés aux variables de pondération (Tab. 3).

Dans la mise en application empirique, les critères pour l'utilisation de valeurs moyennes pondérées ont été examinés dans des conditions pratiques, en se fondant sur deux sondages de visiteuses et visiteurs des forêts urbaines de Zurich, l'un sur le terrain, l'autre en dehors du terrain (par le biais d'envois de questionnaires à des habitants zurichois). Dans leur majorité, les tests de corrélation ont démontré des liens significatifs entre les paramètres liés aux visites en forêt et les facteurs de pondération (Tab. 5). Dans la plupart des cas, la pondération des données recueillies sur le terrain a conduit, pour ces paramètres, à une convergence avec les données du sondage recueillies en dehors du terrain qui reposent sur une sélection aléatoire (Tab. 6). Le risque d'erreurs d'échantillonnage est alors réduit.

La représentation analytique des effets du processus d'échantillonnage et de la pondération des données fournit une base pour l'évaluation des valeurs paramétriques pondérées ou non pondérées des échantillons recueillis sur le terrain. Les résultats empiriques soulignent en particulier l'importance de la sélection des échantillons dans des études d'évaluation économique. Lorsque l'on interprète et généralise les résultats de telles études d'évaluation, il ne faut pas oublier que les valeurs non pondérées peuvent être faussées, et que la fiabilité des évaluations pondérées est réduite par un agrandissement des intervalles de confiance.

Traduction: JENNY SIGOT

8. Dank

Wir bedanken uns bei unseren Kollegen und Kolleginnen und zwei anonymen Gutachtern für kritische Anmerkungen zu einer früheren Version des Artikels. Unser Dank gilt auch MANUELA PETER und RITA GOSH für die hilfreiche Diskussion statistischer Fragen. Für die finanzielle Unterstützung des Forschungsprojektes 'Der Erholungswert des Zürcher Waldes' bedanken wir uns bei Grün Stadt Zürich und den WSL-Forschungsprogrammen 'Landschaft im Ballungsraum' und 'Walddynamik'.

9. Literatur

- BERNATH, K. und A. ROSCHEWITZ (submitted): Recreational Benefits of Urban Forests: Explaining Visitors' Willingness to Pay in the Context of the Theory of Planned Behavior. *Journal of Environmental Management*.
- BERNATH, K., A. ROSCHEWITZ und S. STUDHALTER (2006): Die Wälder der Stadt Zürich als Erholungsraum: Besuchsverhalten der Stadtbevölkerung und Bewertung der Walderholung. Eidg. Forschungsanstalt für Wald, Schnee und Landschaft (WSL), Zürich.
- BÜRG, J., A. OTTITSCH und M. PREGERNIG (1999): Die Wiener und ihre Wälder. Zusammenfassende Analyse sozioökonomischer Erhebungen über die Beziehung der Wiener Stadtbevölkerung zu Wald und Walderholung. Eigenverlag des Instituts für Sozioökonomik der Forst- und Holzwirtschaft, Wien.
- BUWAL (Hrsg.) (1999): Gesellschaftliche Ansprüche an den Schweizer Wald – Meinungsumfrage. Schriftenreihe Umwelt Nr. 309. Bundesamt für Umwelt Wald und Landschaft, Bern.
- COCHRAN, W. G. (1977): Sampling techniques. Wiley, New York.
- DIEKMANN, A. (2002): Empirische Sozialforschung Grundlagen, Methoden, Anwendungen. 8. Auflage. Rowohlt, Reinbek bei Hamburg.
- DILLMAN, D. A. (1978): Mail and Telephone Surveys. The Total Design Method. Wiley, New York.
- ELSASSER, P. (1996): Der Erholungswert des Waldes. Monetäre Bewertung der Erholungsleistung ausgewählter Wälder in Deutschland. Sauerländer, Frankfurt am Main.
- ELSASSER, P. (2001): Probleme der Stichprobenauswahl und der Repräsentativität bei KBM-Umfragen. In: ELSASSER, P. und J. MEYERHOFF (Hrsg.): Ökonomische Bewertung von Umweltgütern. Metropolis-Verlag, Marburg: 17–36.

- GABLER, S. und S. HÄDER (1997): Wirkung von Gewichtungen bei Face-to-Face- und Telefonstichproben. *In: GABLER, S. (Hrsg.): Stichproben in der Umfragepraxis*. Westdeutscher Verlag, Opladen: 221–245.
- GABLER, S., J. H. P. HOFFMEYER-ZLOTNIK und D. KREBS (1994): Gewichtung in der Umfragepraxis. Westdeutscher Verlag, Opladen.
- GREENE, W. H. (2003): *Econometric analysis*. Prentice Hall, Upper Saddle River, NJ.
- HUTCHINSON, W. G., R. SCARPA, S. M. CHILTON und T. MCCALLION (2001): Parametric and non-parametric estimates of willingness to pay for forest recreation in Northern Ireland: A discrete choice contingent valuation study with follow-ups. *Journal of Agricultural Economics* **52**(1): 104–122.
- KISH, L. (1965): *Survey sampling*. Wiley, New York.
- KISH, L. (1987): *Statistical design for research*. Wiley, New York.
- KISH, L. (1990): Weighting: Why, When and How? A Survey for Surveys. *Proceedings of the Survey Research Methods Section, American Statistical Association*: 121–130.
- KISH, L. (1992): Weighting for unequal Pi. *Journal of Official Statistic* **8**: 183–200.
- KISH, L. und S. HEERINGA (2003): *Selected papers*. Wiley-Interscience, Hoboken, New Jersey.
- KLAPHAKE, A. und J. MEYERHOFF (2003): Der ökonomische Wert städtischer Freiräume. Eine Anwendung der Kontingenten Bewertung auf eine städtische Parkanlage in Berlin. *Raumforschung und Raumordnung* **61**(1–2): 107–117.
- KLEIBER, O. (2003): Ökonomische Aspekte der Freizeitaktivitäten im Wald. *In: BAUR, B. (Hrsg.): Freizeitaktivitäten im Baselbieter Wald. Ökologische Auswirkungen und ökonomische Folgen*. Verlag des Kantons Basel-Landschaft, Liestal.
- LINDHAGEN, A. (1996): Forest recreation in Sweden. Four case studies using quantitative and qualitative methods. *Dissertation, Swedish University of Agricultural Sciences, Uppsala*.
- LÖWENSTEIN, W. (1994): Die Reisekostenmethode und die Bedingte Bewertungsmethode als Instrumente zur monetären Bewertung der Erholungsfunktion des Waldes: Ein ökonomischer und ökonometrischer Vergleich. Sauerländer, Frankfurt am Main.
- LUCAS, R. C. (1963): Bias in Estimating Recreationists' Length of Stay from Sample Interviews. *Journal of Forestry* **61**(12): 912–914.
- LUTTMANN, V. und H. SCHRÖDER (1995): Monetäre Bewertung der Fernerholung im Naturschutzgebiet Lüneburger Heide. Sauerländer, Frankfurt am Main.
- NOWELL, C., M. A. EVANS und L. McDONALD (1988): Length-Biased Sampling in Contingent Valuation Studies. *Land Economics* **64**(4): 367–371.
- ROEDER, A. (1973): Gewichtungsverfahren und ihre Anwendung bei der Ermittlung von Besucherzahlen. *Forstarchiv* **44**(2): 30–33.
- RÖSCH, G. (1994): Kriterien der Gewichtung einer nationalen Bevölkerungstichprobe. *In: GABLER, S., J. H. P. HOFFMEYER-ZLOTNIK und D. KREBS (Hrsg.): Gewichtung in der Umfragepraxis*. Westdeutscher Verlag, Opladen: 7–26.
- ROTHER, G. (1994): Wie (un)wichtig sind Gewichtungen? Eine Untersuchung am ALLBUS 1986. *In: GABLER, S., J. H. P. HOFFMEYER-ZLOTNIK und D. KREBS (Hrsg.): Gewichtung in der Umfragepraxis*. Westdeutscher Verlag, Opladen: 62–87.
- ROTHER, G. und M. WIEDENBECK (1994): Stichprobengewichtung: Ist Repräsentativität machbar? *In: GABLER, S., J. H. P. HOFFMEYER-ZLOTNIK und D. KREBS (Hrsg.): Gewichtung in der Umfragepraxis*. Westdeutscher Verlag, Opladen: 46–61.
- SCHELBERT, H., R. MAGGI, R. ITEN, C. NIELSEN, T. LANG, I. BUSE und J. HENZMANN (1988): *Wertvolle Umwelt. Ein wirtschaftswissenschaftlicher Beitrag zur Umwelteinschätzung in Stadt und Agglomeration* Zürich. Zürcher Kantonalbank, Zürich.
- STAHEL, W. A. (1995): *Statistische Datenanalyse. Eine Einführung für Naturwissenschaftler*. Vieweg, Braunschweig.
- TYRVÄINEN, L. (2001): Economic valuation of urban forest benefits in Finland. *Journal of Environmental Management* **62**(1): 75–92.