

**Arbeitsberichte der
Hochschule für Wirtschaft FHNW – Nr. 33****Repräsentativität von Unternehmens-Gruppen als Vertreter
von Branchen**

Bericht im Auftrag des Eidgenössischen Institutes für Geistiges Eigentum

Beat Hulliger, Marc Bill

ISSN	Nr. 1662-3266 (Print) Nr. 1662-3274 (Online)
ISBN	Nr. 978-3-03724-149-3
Institut	Institute for Competitiveness and Communication ICC
Datum	März 2014

© 2014 Hochschule für Wirtschaft FHNW und der Autor bzw. die Autorin. Jede Reproduktion, auch von Teilen und unabhängig vom Medium, ist nur mit Genehmigung der Hochschule für Wirtschaft FHNW und des Autors bzw. der Autorin gestattet.

Kontaktangaben

Prof. Dr. Beat Hulliger
Fachhochschule Nordwestschweiz FHNW
Hochschule für Wirtschaft
Institute for Competitiveness and Communication
Riggenbachstrasse 16
4600 Olten
T +41 62 957 23 06
beat.hulliger@fhnw.ch

Marc Bill
Fachhochschule Nordwestschweiz FHNW
Hochschule für Wirtschaft
Institute for Competitiveness and Communication
Riggenbachstrasse 16
4600 Olten
T +41 62 957 23 37
marc.bill@fhnw.ch

Inhaltsverzeichnis

Kontaktangaben.....	i
Inhaltsverzeichnis	ii
Abbildungs- und Tabellenverzeichnis.....	iii
Zusammenfassung.....	1
1 Einleitung.....	3
2 Statistische Definitionen von Repräsentativität.....	3
2.1 Zufallsstichproben.....	4
2.2 Modelle	4
2.3 Definition für die Repräsentativität eines Branchenverbandes.....	5
3 Grundlagen zur Repräsentativität eines Branchenverbands.....	5
3.1 Abstandsmass und Schwellenwert.....	7
3.2 Grösse der Gruppe	8
4 Vorschlag für das Vorgehen zur Feststellung der Repräsentativität	9
4.1 Bildung der Gruppe und Erstellen von Kriterienkatalog (Schritt 1.1).....	11
4.2 Liste der Mitglieder (Schritt 1.2)	11
4.3 Überprüfung der Mitglieder des BV (Schritt 1.3).....	12
4.4 Definition der Grundgesamtheit aus der Sicht BV (Schritt 2.1)	12
4.5 Erste Prüfung anhand des BUR (Schritt 2.2).....	12
4.6 Öffentliche Ausschreibung (Schritt 2.3)	12
4.7 Bereinigung der Grundgesamtheit (Schritt 2.4)	12
4.8 Festlegung der Grundgesamtheit (Branche) (Schritt 2.5)	12
4.9 Überprüfung der Repräsentativität anhand BUR-Variablen (Schritt 3.1).....	13
4.10 Information für zusätzliche Kovariablen (Schritt 3.2)	13
4.11 Überprüfung der Qualität (Schritt 3.3)	13
4.12 Überprüfung der Repräsentativität (Schritt 3.4)	13
5 Repräsentativitätsüberprüfung anhand eines fiktiven Beispiels mit Hilfe des Chiquadrat-Anpassungstests	14
5.1 Definition verschiedener Gruppen (Branchenverbände).....	14
5.2 Univariate Chiquadrat-Testergebnisse der verschiedenen Gruppen	14
5.3 Bivariate Chiquadrat-Testergebnisse der verschiedenen Gruppen	15
5.4 Diskussion	16
6 Schlussbemerkungen.....	17
Literaturverzeichnis	18
Anhang	19

Abbildungs- und Tabellenverzeichnis

Tabelle 1: Notwendige Schritte zur Definition von Repräsentativität eines Branchenverbands (BV)	10
Tabelle 2: Struktur der Grundgesamtheit in den untersuchten Variablen Grossregion und Unternehmensgrösse	14
Tabelle 3: Chiquadrat-Anpassungstest - Ergebnisse nach Gruppen und Kovariablen.....	16
Tabelle 4: Struktur Grundgesamtheit	19
Tabelle 5: Struktur Branchenverband Version 1.....	19
Tabelle 6: Struktur Branchenverband Version 2.....	19
Tabelle 7: Struktur Branchenverband Version 3.....	20

Zusammenfassung

Statistische Definition von Repräsentativität

Repräsentativität im weitesten Sinn bedeutet in der Statistik, dass aus einer Teilmenge (Stichprobe) ein gültiger Schluss auf Eigenschaften einer Gesamtheit (Grundgesamtheit) möglich ist. Der Begriff der Repräsentativität ist unscharf und muss für die Erstellung einer Branchenverordnung genauer definiert und operationalisiert werden. Es muss eine Definition gefunden werden, mit welcher entschieden werden kann, ob die Unternehmen in einem Branchenverband repräsentativ für die Unternehmen in einer Branche sind. In der Statistik werden im Wesentlichen zwei Definitionen verwendet.

Die erste Definition basiert auf der zufälligen Auswahl einer Stichprobe nach einem festgelegten Stichprobenplan. Dank der Zufallsauswahl besteht eine theoretische Grundlage, auf der eine Aussage über die Grundgesamtheit aufgebaut werden kann. Diese Definition kann aber für die Untersuchung der Repräsentativität eines Branchenverbandes nicht herangezogen werden, weil die Unternehmen des Branchenverbandes nicht zufällig aus der Branche ausgewählt wurden.

Die zweite Definition für Repräsentativität basiert auf dem Zusammenhang zwischen einer interessierenden Eigenschaft (Variablen) und erklärenden Eigenschaften (Kovariablen) der Unternehmen. Der Zusammenhang wird als statistisches Modell formuliert, bei dem die interessierende Variable durch die Kovariablen erklärt wird. Die interessierende Variable liegt nur für die Stichprobe (Branchenverband) vor, während die Kovariablen für die Grundgesamtheit (Branche) vorliegen müssen. Eine interessierende Variable kann zum Beispiel die Zustimmung eines Unternehmens zu einer Branchenverordnung sein. Eine Kovariable kann zum Beispiel die Grösse eines Unternehmens sein. Diese Modell-basierte Repräsentativität ist gegeben, wenn folgende zwei Bedingungen erfüllt sind:

1. Die interessierende Eigenschaft wird genügend gut durch die Kovariablen erklärt. Dabei ist das Modell (Zusammenhang) in der Stichprobe dasselbe wie in der Grundgesamtheit.
2. Die Zusammensetzung der Stichprobe (Verteilung) bezüglich der Kovariablen ist gleich wie in der Grundgesamtheit.

Unter diesen Bedingungen kann die interessierende Eigenschaft in der Grundgesamtheit durch die Stichprobe geschätzt werden und damit ist die Repräsentativität erfüllt. Nachteilig ist bei dieser Definition, dass das Modell für die Grundgesamtheit nicht überprüft werden kann, eben weil die interessierende Eigenschaft für die Grundgesamtheit nicht bekannt ist. Die Überprüfung des Modells muss sich normalerweise auf die Stichprobe beschränken.

Nachweis der Repräsentativität für einen Branchenverband

Für die Operationalisierung der Modell-basierten Repräsentativität muss die Grundgesamtheit abgegrenzt, die Kovariablen und das Modell bestimmt und die Übereinstimmung der Verteilung der Kovariablen zwischen Stichprobe und Grundgesamtheit überprüft werden.

Die Abgrenzung der Grundgesamtheit (Branche) ist ein operationelles und rechtliches Problem, welches im vorgeschlagenen Vorgehen berücksichtigt wird.

Die Überprüfung des Modells ist im Fall eines Branchenverbands oft auch für die Stichprobe nicht möglich, weil die interessierende Variable (z.B. Zustimmung zu einer noch zu erstellenden Branchenverordnung) oft auch nicht für den Branchenverband vorliegt. Die Überprüfung muss also auf inhaltliche Plausibilität des Modells abstellen. Bei der Festlegung der Kovariablen im Modell muss berücksichtigt werden, welche Informationen für die Grundgesamtheit vorliegen. Informationen guter Qualität über eine Branche stehen im Betriebs- und Unternehmensregister (BUR) des Bundesamts für Statistik zur Verfügung. Allerdings sind die Informationen im BUR auf die Anzahl Beschäftigte, die Wirtschaftsaktivität, den Standort, die Rechtsform und unter Umständen den Gesamtumsatz beschränkt. Weitere Informationen wie der Umsatz für eine bestimmte Produkte-Gruppe müssen aus anderen Quellen oder mit eigens durchgeführten Erhebungen eruiert werden.

Die Übereinstimmung der Verteilung der Kovariablen zwischen Stichprobe (Branchenverband) und Grundgesamtheit (Branche) kann mit Hilfe eines statistischen Abstandsmasses, der Chi-Quadrat-Statistik, untersucht werden: Übersteigt die Chi-Quadrat-Statistik einen kritischen Wert, wird die Repräsentativität abgelehnt. Damit das Abstandsmass genügend gut ist, und das unterliegende Modell nicht überdehnt wird, sollte der Branchenverband eine absolute und relative Mindestgrösse haben.

Mehrstufiges Vorgehen zum Nachweis der Repräsentativität eines Branchenverbandes

Das vorgeschlagene Vorgehen sieht drei Phasen vor:

1. Der Branchenverband legt einen Kriterienkatalog vor, in dem die zu regelnden Produkte, die Branche und die zu überprüfenden Kovariablen enthalten sind. Das IGE überprüft den Kriterienkatalog.
2. Der Branchenverband legt eine Liste der Unternehmen der Branche (Grundgesamtheit) vor, welche mit Hilfe einer öffentlichen Ausschreibung bereinigt wird. Die Repräsentativität des Branchenverbandes wird anhand der im BUR gegebenen Kovariablen durch das IGE mit Unterstützung des BFS überprüft.
3. Allenfalls notwendige weitere Kovariablen werden durch den Branchenverband eruiert. Die Qualität der gewonnenen Informationen und die Repräsentativität bezüglich der weiteren Kovariablen werden durch das IGE überprüft.

Das Vorgehen kann nach jeder der obigen Phasen beendet werden, sei es mit einem positiven oder negativen Entscheid bezüglich der Repräsentativität. Als weitere Eskalationsmöglichkeit oder als Alternative für die Phase 3 kann mit einer Urabstimmung in der Grundgesamtheit die interessierende Variable direkt erhoben werden. Die Modalitäten einer solchen Urabstimmung sind nicht Gegenstand dieses Berichts

1 Einleitung

Das Eidgenössische Institut für Geistiges Eigentum (IGE) ist neben dem Erteilen von Marken, Patenten und Designs auch für die Vorbereitung der Gesetzgebung im Immaterialgüterrecht verantwortlich. In dieser Funktion hatte es in den letzten Jahren die Federführung beim Gesetzgebungsprojekt "Swissness" inne.

Nach jahrelangem Seilziehen wurde Mitte Juni 2013 die Swissness-Vorlage vom Parlament verabschiedet. Das neue Gesetz dient zum Schutze der Marke Schweiz und regelt im Groben, wie viel Schweiz in einem Produkt stecken muss, damit es als "Swiss Made" vermarktet werden darf. Im Anschluss an den parlamentarischen Entscheid, müssen die notwendigen Verordnungen des Bundesrats verfasst werden. Eine Kategorie dieser Verordnungen sind sogenannte Branchenverordnungen. In einer solchen können Branchenvertreter (i.d.R. Verbände) die vom Parlament definierten Swissness-Kriterien für ihre Branche je nach Bedarf branchenspezifisch konkretisieren. Eine Anforderung an die Branchenvertreter ist der Nachweis ihrer Repräsentativität für die gesamte Branche.

Dieser Bericht beschreibt ein Verfahren, das es dem IGE ermöglicht die Repräsentativität von Branchenvertretern in einem statistisch korrekten Sinne zu überprüfen.

Der Bericht ist folgendermassen aufgebaut. In Abschnitt 2 wird Repräsentativität aus statistischer Sicht definiert und zwei Methoden kurz vorgestellt. Abschnitt 3 wendet den in Abschnitt 2 vorgestellten Modellansatz auf die Situation der Branchenverordnung an. Abschnitt 4 beschreibt einen schrittweisen Prozess um die Repräsentativität eines Branchenverbandes schlussendlich nachzuweisen. In Abschnitt 5 zeigen wir beispielhaft die statistische Überprüfung der vorher erarbeiteten Kriterien. Abschnitt 6 dient der Schlussreflexion.

2 Statistische Definitionen von Repräsentativität

Repräsentativität im statistischen Sinn ist seit der Einführung des Begriffs Ende 19. Jahrhundert mit vielen verschiedenen Interpretationen versehen worden (vgl. die Besprechung in Kruskal und Mosteller 1979). Wir stellen hier kurz zwei Definitionslinien für eine statistische Definition von Repräsentativität vor: i) Repräsentativität aufgrund von Zufallsstichproben und ii) Repräsentativität aufgrund von Modellen. Die zweite Methode wird im nächsten Kapitel für die Definition der Repräsentativität eines Branchenverbands verwendet.

Bei der Diskussion der Repräsentativität in diesem Kapitel wird davon ausgegangen, dass die Grundgesamtheit bekannt ist. Deren Definition ist aber ein eigenes Problem bei den Branchenverordnungen und wird im Abschnitt 4 besprochen. Wir gehen hier auch davon aus, dass die Einheit der Grundgesamtheit das Unternehmen (mit einer UID¹) ist. Da es sich um eine Rechtsverordnung handelt scheint dies auch die geeignete Einheit zu sein.

Mit Repräsentativität wird angestrebt, dass keine systematische Abweichung zwischen den Schätzungen und den tatsächlichen Werten besteht. Eine systematische Abweichung wird

¹ Jedem in der Schweiz aktiven Unternehmen wird seit 2011 eine eindeutige und übergreifende Unternehmens-Identifikationsnummer (UID) zugeordnet. Diese ermöglicht es den Unternehmen, sich bei allen Behördenkontakten mit ein und derselben Nummer zu identifizieren. Ein zentrales Unternehmensregister - UID-Register - wird der Öffentlichkeit zur Abfrage der UID zur Verfügung gestellt. Bei einem Unternehmen handelt es sich um die "kleinste Kombination rechtlicher Einheiten zur Erzeugung von Waren und Dienstleistungen, die in Bezug auf die Verwendung der ihr zufließenden Mittel über eine gewisse Entscheidungsfreiheit verfügt".

auch Bias oder Verzerrung genannt. Die absolute und relative Grösse der Stichprobe spielt für die Repräsentativität aus statistischer Sicht eine sekundäre Rolle, weil sich die Repräsentativität auf mögliche Verzerrungen fokussiert. Die Grösse der Stichprobe hat zwar keinen Einfluss auf Verzerrungen, wohl aber auf die Genauigkeit von Aussagen. Wir kommen in Kapitel 3 auf die Grösse der Stichprobe zurück.

2.1 Zufallsstichproben

Die erste, auf Zufallsstichproben basierende Methode definiert streng genommen die Repräsentativität von Erhebungen, nicht von Stichproben. Die Voraussetzungen sind im strengen Sinn:

1. Die Stichprobe wurde gemäss einem Stichprobenplan² gezogen, der allen Elementen der Population eine strikt positive und berechenbare Einschlusswahrscheinlichkeit gibt.
2. Das Verfahren für die Schätzung von Populations-Charakteristiken berücksichtigt den Stichprobenplan so, dass keine Verzerrung entsteht.

In Lehrbüchern wird dabei oft angemerkt, dass alle Populationselemente die gleiche Einschlusswahrscheinlichkeit haben müssen. Das ist falsch, d.h. es ist weder eine notwendige noch eine hinreichende Bedingung für eine repräsentative Erhebung.

In der Praxis muss mit Antwortausfällen gerechnet werden und die obigen zwei Bedingungen müssen ergänzt werden durch eine etwas schwache Aussage:

3. Antwortausfälle dürfen keinen verzerrenden Effekt haben.

2.2 Modelle

Die zweite Definition von Repräsentativität, die hier betrachtet werden soll, basiert auf dem Zusammenhang zwischen einer interessierenden Variable und sogenannten Kovariablen. Die interessierende Variable sei hier mit Y bezeichnet und die Kovariablen (meistens mehrere) mit X . Der Zusammenhang wird in der statistischen Fachsprache als Modell bezeichnet. Dieses Modell erlaubt es, die interessierende Variable mit Hilfe der Kovariablen vorherzusagen. Die Auswahl der Stichprobe ist nur insofern von Interesse, dass das gleiche Modell für die Grundgesamtheit gelten muss wie für die Stichprobe. Die zweite Bedingung besteht darin, dass die Kovariablen in der Stichprobe gleich verteilt sind wie in der Grundgesamtheit. Dadurch kann die Verteilung der interessierenden Variablen in der Population durch die Verteilung in der Stichprobe geschätzt werden. Allenfalls kann zusätzlich eine Gewichtung der Stichprobe angesetzt werden. Dies ist aber im Zusammenhang mit der Repräsentativität von Branchenverbänden nicht relevant.

Die Bedingungen für Repräsentativität einer Stichprobe sind damit:

1. Es gibt ein genügend gutes Modell, mit dessen Hilfe die interessierende Variable aus den Kovariablen vorhergesagt werden kann. Das Modell gilt sowohl für die Stichprobe als auch für die Grundgesamtheit.

² Ein Stichprobenplan definiert das Verfahren für die Ziehung einer Zufallsstichprobe. Der Stichprobenplan definiert u.a. auch die Einschlusswahrscheinlichkeiten, welche nach der Erhebung die Grundlage für Gewichtungen bilden.

2. Die Verteilung der Kovariablen in der Grundgesamtheit ist bekannt und die (ebenfalls bekannte) Verteilung der Kovariablen in der Stichprobe ist gleich deren Verteilung in der Grundgesamtheit.

In der Praxis kann das Modell nur anhand der Stichprobe überprüft werden und es muss genauer definiert werden, wie gut das Modell sein muss. Ausserdem muss mit Abweichungen zwischen der Verteilung der Kovariablen in der Stichprobe und in der Grundgesamtheit gerechnet werden.

Anzumerken ist auch, dass die erste Bedingung für jede interessierende Variable erfüllt werden muss. M.a.W. bezieht sich die Repräsentativität aufgrund eines Modells immer auf spezifische Variablen.

Diese zwei Bedingungen werden zwar selten explizit gemacht, liegen aber den Definitionen von Repräsentativität zugrunde, welche verlangen, dass die Stichprobe ein Abbild der Grundgesamtheit (bezüglich überprüfbarer Kovariablen) sei. Beispielsweise bauen Quotenstichproben auf dieser Definition auf. Die erste Bedingung, dass es ein genügend gutes Modell gibt, welches sowohl für die Stichprobe als auch für die Grundgesamtheit gilt, kann allerdings selten überprüft werden. Typischerweise kann das Modell für die Stichprobe abgesichert werden, aber die Überprüfung an der Grundgesamtheit ist nicht möglich.

2.3 Definition für die Repräsentativität eines Branchenverbandes

Wenn die Mitglieder eines Branchenverbandes als Stichprobe betrachtet werden, kann nicht von Repräsentativität im Sinne einer Zufallsstichprobe gesprochen werden, da die Auswahl der Branchenmitglieder nicht einem geplanten zufälligen Vorgehen entspricht. M.a.W. es fehlt der Stichprobenplan und der Branchenverband ist keine Zufallsstichprobe. Es muss also auf die Repräsentativität im Sinne von Modellen zurückgegriffen werden. Um Verwechslungen zu vermeiden werden wir im Folgenden bei den Mitgliedern eines Branchenverbandes nicht von einer Stichprobe sprechen, sondern von einer Gruppe. Wir werden also im Folgenden von der Repräsentativität einer Gruppe für eine Grundgesamtheit sprechen. Die Gruppe entspricht dabei dem Branchenverband und die Grundgesamtheit der Branche.

3 Grundlagen zur Repräsentativität eines Branchenverbands

In diesem Abschnitt wird die Methode zur Definition von Repräsentativität aufgrund von Modellen auf einen Branchenverband angewendet. Die Überprüfbarkeit der in Abschnitt 2.2 definierten Bedingungen wird diskutiert und ein konkretes Verfahren zur Repräsentativitätsüberprüfung vorgestellt.

Die Repräsentativität einer Gruppe wird sich auf Informationen über die Verteilung von Eigenschaften in der Gruppe und in der Grundgesamtheit abstützen. Dabei baut die Argumentation auf einem Modell zwischen der interessierenden Variable und Kovariablen (Bedingung 1). Anhand eines Branchenverbandes kann die Situation beispielhaft wie folgt dargestellt werden: Die Zustimmung eines Unternehmens zu einer Branchenverordnung sei eine Variable Y , die für jedes Unternehmen der Grundgesamtheit einen unbekanntem Wert (hier Ja oder Nein) annimmt. Da die Zustimmung unbekannt ist, wird versucht, sie mit einer (oder mehreren) Kovariablen X vorherzusagen. Mögliche Kandidaten für Kovariablen sind zum Beispiel die Grösse einer Unternehmung (z.B.: Die grossen Unternehmen stimmen eher zu), die regionale Zugehörigkeit (z.B.: Die Unternehmen einer bestimmten Region stimmen eher zu), das Produktionsvolumen eines betroffenen Produkts (z.B.: Die Unternehmen mit kleinem Produktionsvolumen stimmen eher zu). Es wird jetzt postuliert, dass es ein genügend gutes

Modell zwischen der interessierenden Variable Y als abhängige und den Kovariablen X als unabhängige Variablen (Prädiktoren) gibt. Das Modell kann in der Form $Y = f(X) + E$ geschrieben werden, wobei f eine (unbekannte) Funktion ist und E ein (kleiner) Fehler. In dem Fall kann die Variable Y (z.B. die Zustimmung) aus den Kovariablen X (z.B. Unternehmensgrösse, Region etc.) vorhergesagt werden. Die erste Bedingung der Repräsentativität aufgrund eines Modells verlangt ausserdem, dass das Modell für die Grundgesamtheit (Branche) und für die Gruppe (Branchenverband) gilt. Während im Branchenverband die Zustimmung (Y) mit eher kleinem Aufwand erhoben werden könnte, ist das für die Grundgesamtheit nicht vorgesehen, weil dies einer Urabstimmung a-priori gleichkäme. Damit kann das Modell nur anhand der Gruppe überprüft werden. M.a.W. bleibt es ein a-priori nicht nachprüfbares Postulat, dass dasselbe Modell auch in der Grundgesamtheit gilt. Damit ist die erste Bedingung für die Repräsentativität empirisch nicht nachprüfbar und muss aufgrund von Plausibilitätsbetrachtungen postuliert werden. Trotz diesem Nachteil wird die modell-basierte Repräsentativität oft herangezogen, insbesondere in der Marktforschung.

Zur Erfüllung der zweiten Bedingung der Repräsentativität wird die Verteilung der Kovariablen in der Gruppe und in der Grundgesamtheit untersucht. Es wird vorausgesetzt, dass beide Verteilungen bekannt sind, aber dies muss natürlich im Anwendungsfall geprüft werden. Um die Verteilung in der Stichprobe mit der Verteilung in der Grundgesamtheit zu vergleichen, wird ein Abstand zwischen den beiden Verteilungen berechnet. Als Abstandsmass schlagen wir vor, die Chiquadrat-Statistik oder den Kontingenzkoeffizienten nach Pearson zu verwenden. Es muss ein Schwellenwert für den Abstand festgelegt werden, oberhalb derer die Gruppenverteilung nicht mehr als gleich der Grundgesamtheits-Verteilung angesehen wird. Wenn der Abstand kleiner als der Grenzwert ist, dann kann angenommen werden, dass die Vorhersage der Werte von Y in der Grundgesamtheit ausserhalb der Gruppe die gleiche Verteilung wie die (beobachteten) Werte in der Gruppe haben. Damit ist gesichert, dass z.B. die Zustimmung zu einer Branchenverordnung in der Gruppe als gute Vorhersage für die Zustimmung zu einer Branchenverordnung in der Grundgesamtheit herangezogen werden kann.

Liegt also der Abstand der Verteilungen in der Gruppe und in der Grundgesamtheit unter dem festgelegten Schwellenwert, dann ist die Gruppe repräsentativ. Der Schwellenwert und die Gewichtung der verschiedenen Kovariablen werden unter Berücksichtigung branchenspezifischer Eigenschaften vom IGE festgelegt. Ein Beispiel für diese Berechnungen wird in Abschnitt 5 aufgezeigt.

Wie oben erwähnt, werden meistens mehrere Kovariablen im Modell verwendet. Entsprechend müsste jeweils die gemeinsame Verteilung der Kovariablen X überprüft werden. Dazu ist die Überprüfung der Verteilungen der einzelnen Variablen je für sich genommen (univariat) ein erster Schritt. Da in der Praxis oft die gemeinsame Verteilung der Kovariablen in der Grundgesamtheit nicht bekannt ist, beschränken wir uns hier auf Ausführungen zur univariaten Überprüfung. Diese können aber bei Vorliegen von genügend Informationen auf gemeinsame Verteilungen ausgedehnt werden.

Bezüglich der konkreten Kovariablen ist die Grösse der Unternehmen in der Branche einzubeziehen. Besteht die Branche beispielsweise aus vielen kleinen und wenig grossen Unternehmen, der Branchenverband jedoch nur aus einigen grossen Unternehmen, wird das obige Distanzmass aufgrund des Überhangs von grossen Produzenten einen zu hohen Wert anzeigen, obwohl die Mitglieder des Branchenverbands u.U. die Mehrheit der Beschäftigten stellen. Die Wirtschaftsbranche nach der NOGA Nomenklatur könnte eine weitere wichtige Variable sein. Es kann durchaus vorkommen, dass sich eine Branchenverordnung über mehrere Wirtschaftszweige nach NOGA erstreckt und dabei kann die Verteilung der Branche auf

die NOGA von Bedeutung sein. Neben der Grösse und der Wirtschaftsklasse könnte auch die Standortregion wichtig sein, obwohl der Standort bei Mehrbetriebsunternehmen mit Vorsicht zu behandeln ist³. Der Gesamtumsatz der Unternehmen ist eine weitere mögliche Kovariable. Anzahl Beschäftigte, Wirtschaftsklasse, Standort und Umsatz sind im Betriebs- und Unternehmensregister des BFS enthalten (Umsatz allerdings nicht überall in konsolidierter Form). Die Verteilungen dieser Kovariablen könnten also aus dem BUR abgeleitet werden, wenn die Grundgesamtheit bekannt ist.

Je nach Branchen müssen allenfalls weitere Kovariablen für die Repräsentativitätsdefinition hinzugezogen werden. Die Produktionsvolumen des zu regelnden Produkts könnten zum Beispiel wichtig werden. Oder der Anteil eines Teilprodukts (z.B. Funkuhren) an der Produktion. Diese Variablen können nicht aus dem BUR eruiert werden und es wird u.U. die Aufgabe des Branchenverbands sein, hier genügend Information über die Verteilung in der Grundgesamtheit zu liefern. Das IGE wird die Aufgabe haben, die Qualität dieser Informationen zu beurteilen.

3.1 Abstandsmass und Schwellenwert

Der Abstand zwischen zwei Verteilungen kann mit dem Chiquadrat-Anpassungstest beurteilt werden.

Der **Chiquadrat-Anpassungstest** postuliert in der Null-Hypothese, dass die Verteilung einer Kovariablen x in der Gruppe $F_G(x)$ der Grösse n identisch ist mit der Verteilung in der Grundgesamtheit $F_U(x)$:

$$H_0: F_G(x) = F_U(x)$$

Der Wertebereich der Kovariablen wird für den Test in Klassen aufgeteilt (z.B. Unternehmensgrössenklassen nach VZÄ: 0-49, 50-249 und ≥ 250). Zur Berechnung der Teststatistik wird aus der Verteilung in der Grundgesamtheit eine erwartete Häufigkeit pro Klasse in der Gruppe berechnet. Die Distanz zwischen diesen neu kalkulierten, zu erwartenden Häufigkeiten der Gruppe (falls H_0 zutrifft) und den effektiv beobachteten Häufigkeiten der Gruppe bildet die Grundlage der Teststatistik, die knapp als $\hat{\chi}^2$ bezeichnet und folgendermassen berechnet wird:

$$\hat{\chi}^2 = \sum_{i=1}^k \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$$

, wobei

- k = Anzahl der Klassen in der untersuchten Kovariablen
- n_i = Anzahl Unternehmen (Häufigkeit) in der Gruppe in Klasse i

³ Bei Mehrbetriebsunternehmen entspricht u.U. der Standort des Hauptbetriebs nicht dem wesentlichen Produktionsstandort des Produkts. Ähnliches kann bei Unternehmensgruppen vorkommen.

- p_i = Anteil der Unternehmen in Klasse i in der Grundgesamtheit. Wenn N die Grösse der Grundgesamtheit und N_i die Grösse der Klasse i in der Grundgesamtheit bezeichnet, ist $p_i = \frac{N_i}{N}$.
- $n \cdot p_i$ = (unter H_0) erwartete Anzahl Unternehmen in der Gruppe in Klasse i

Der Test verwirft die Nullhypothese, dass beide Verteilungen identisch sind, sobald die berechnete Teststatistik $\hat{\chi}^2$ grösser ist als ein Quantil der χ^2_v Verteilung mit $v = k - 1$ Freiheitsgraden⁴. Das Quantil $\chi^2_{v;1-\alpha}$ wird durch das Signifikanzniveau α des Tests festgelegt (Hedrich und Sachs 2012). Als Signifikanzniveau schlagen wir $\alpha = 0.05$ vor.

Bei der Bildung der Klassen ist die Voraussetzung des Chiquadrat-Anpassungstests, dass nicht mehr als 20% der Klassen eine erwartete Häufigkeit unter 5 haben, zu beachten. Hier kommt also auch ein Aspekt der Genauigkeit bzw. der Grösse der Stichprobe ins Spiel.

Ausserdem hat der Chiquadrat-Anpassungstest die Eigenschaft, dass die Teststatistik mit wachsender Grösse der Gruppe den Schwellenwert eher übersteigt. Dies ist einsichtig, weil zufällige Abweichungen der Gruppe von der Grundgesamtheit mit wachsender Grösse der Gruppe kleiner sein sollten⁵.

Als Abstand zwischen der Verteilung in der Gruppe und der Verteilung in der Grundgesamtheit dient also die Chiquadrat-Statistik $\hat{\chi}^2$ und als Schwellenwert dient der sogenannte kritische Wert $\chi^2_{v;1-\alpha} = \chi^2_{v;0.95}$. Im Abschnitt 5 werden einige Beispiele für die Berechnung dieses Repräsentativitätsindikators vorgestellt.

3.2 Grösse der Gruppe

Ein Branchenverband sollte eine absolute Mindestgrösse haben, um eine genügende Genauigkeit der Vorhersagen für die nicht organisierten Unternehmen der Grundgesamtheit zu gewährleisten. Ausserdem werden auch Chiquadrat-Anpassungstests mit etwas grösserer Anzahl von Klassen bei kleinen Gruppen unzuverlässig. Bei einer Grösse von $n = 20$ können 4 ungefähr gleich grosse oder 3 leicht unterschiedlich grosse Klassen gebildet werden, damit der Test noch zuverlässig ist. Wir schlagen vor, dass diese Grenze ($n = 20$) nur unterschritten werden darf, wenn die Grundgesamtheit weniger als 20 Unternehmen umfasst.

Ein weiterer Aspekt ist die implizite Modellierung, welche für die nicht in der Gruppe vertretenen Unternehmen der Grundgesamtheit gelten sollte. Damit das Modell wenigstens für die Mehrheit der Unternehmen überprüft werden kann, sollte die Gruppe mindestens 50% der Unternehmen der Grundgesamtheit umfassen, also auch eine relative Mindestgrösse haben. Mit anderen Worten, wir schlagen vor, dass als Bedingung für die Grösse der Gruppe gilt $20 \leq n \leq 0.5 \cdot N$.

⁴ Da die Klassen-Häufigkeiten auf n addieren müssen, reduziert sich die Anzahl Freitsgrade um 1.

⁵ Um den Effekt zu kompensieren könnte der Wert der Teststatistik mit n_0/n multipliziert werden, wenn die Gruppengrösse n eine Standardgrösse n_0 übersteigt. Dabei könnte z.B. $n_0 = 100$ gesetzt werden.

4 Vorschlag für das Vorgehen zur Feststellung der Repräsentativität

Um bereits für den Verhandlungsprozess die Repräsentativität der Gruppe nachzuweisen, werden Informationen über die Grundgesamtheit und die Gruppe bereits am Anfang des Prozesses notwendig. Dadurch kann eine nachträgliche Verifizierung der Zustimmung der Grundgesamtheit zur Branchenverordnung vermieden werden oder zumindest vorausgenommen werden, um so die Akzeptanz zu erhöhen.

Der erste Schritt im Prozess ist die Abgrenzung der Grundgesamtheit selbst. Die Einheit der Grundgesamtheit ist die Unternehmung. Welche Unternehmen tatsächlich zur Grundgesamtheit gehören wird in einem mehrstufigen Prozess eruiert. Nach Abschluss dieses Prozesses, sollten alle notwendigen Informationen vorliegen, damit die Repräsentativitätsüberprüfung anhand der in Abschnitt 3 konkretisierten Methode vollzogen werden kann.

Unser Vorschlag beginnt mit einem Kriterienkatalog, welcher diese Abgrenzung ermöglichen soll. Der Branchenverband muss in einem ersten Schritt einen Kriterienkatalog festlegen, der die Branche und die Produkte, für welche eine Branchenverordnung erstellt werden soll, exakt definiert. Der Kriterienkatalog muss auch einen Vorschlag für die Kovariablen enthalten, welche für die Überprüfung der Repräsentativität herangezogen werden sollen. Ein Modell für die Zustimmung zur Branchenverordnung muss nicht explizit angegeben werden. Hingegen müssen die Kovariablen so ausgewählt sein, dass sie die wichtigen Faktoren, welche einen Einfluss auf die Zustimmung zu einer angestrebten Branchenverordnung haben, umfassen. Die Variablen Anzahl Beschäftigte, Grossregion und Wirtschaftsbranche nach NOGA sollten obligatorisch aufgenommen werden, weitere Kovariablen, welche branchen- und produktspezifisch sein können, müssen fallweise abgeklärt werden.

Im zweiten Schritt wird die Grundgesamtheit individuell definiert. Die Grundgesamtheit besteht aus allen Unternehmen, die den Kriterienkatalog erfüllen. Der Branchenverband muss eine Liste der Unternehmen (Namen und UID) vorlegen, welche aus seiner Sicht die Branche (Grundgesamtheit) bilden. Diese Liste muss anschliessend bereinigt werden. Das heisst, es muss ein Prozess definiert werden, der es erlaubt Kandidaten für die Grundgesamtheit entsprechend dem Kriterienkatalog einzuschliessen oder auch auszuschliessen. Dabei müssen Unternehmen die Möglichkeit haben, sowohl den Ausschluss als auch den Einschluss zu beantragen.

In einem dritten Schritt muss die Repräsentativität der Gruppe überprüft werden. Dazu müssen die Verteilungen der im Kriterienkatalog enthaltenen Kovariablen sowohl für die Grundgesamtheit (Branche) als auch für die Gruppe (Branchenverband) eruiert werden. Das BUR wird eine wichtige Unterstützung sein. Um spezifische Kovariablen zu erheben, muss entweder eine eigene Erhebung konzipiert und durchgeführt werden oder es kann auf zuverlässige bestehende Erhebungen/Informationsquellen zurückgegriffen werden. Hier stellt sich allerdings die Frage, ob es nicht zielführender ist, die Branchenverordnung direkt zu erarbeiten und einer Urabstimmung in der Branche (Grundgesamtheit) zu unterwerfen. Dies hängt u.a. davon ab, wie gross die Branche ist und wie gross eine allfällige Stichprobe wäre.

Das vorgeschlagene Verfahren zum Nachweis der Repräsentativität eines Branchenverbandes basiert demzufolge auf drei Stufen. Die Beweislast liegt beim Branchenverband. Eine Zusammenfassung der notwendigen Schritte ist in Tabelle 1 dargestellt und in den folgenden Abschnitten ausgeführt.

Tabelle 1: Notwendige Schritte zur Definition von Repräsentativität eines Branchenverbands (BV)

Schritt	Beschreibung	Wer
1.1	Kriterienkatalog mit einer Beschreibung der Grundgesamtheit (Branche). Der Kriterienkatalog enthält einen Vorschlag für die Kovariablen, welche für die Überprüfung der Repräsentativität verwendet werden.	BV
1.2	Mitgliederliste des BV (UID*) mit der Angabe ob das betreffende Mitglied aus der Sicht des BV unter die Branchenverordnung fällt, d.h. zur Grundgesamtheit gehört.	BV
1.3	Überprüfung: Wird der Kriterienkatalog und die Kovariablen akzeptiert? Erfüllen die bezeichneten Mitglieder des Branchenverbandes den Kriterienkatalog? Falls Nein: Zurück zu Schritt 1.1 oder Sprung zu 3.4	IGE
Kriterienkatalog und Branchenverband (Gruppe) sind definiert		
2.1	BV legen dem IGE eine Liste mit allen Schweizerischen Unternehmen, die unter den formulierten Kriterienkatalog fallen (Grundgesamtheit), vor (UID).	BV
2.2	Erste Überprüfung der Liste anhand des BUR auf Plausibilität.	BFS
2.3	Öffentlicher Aushang des Kriterienkatalogs und der Liste. Sicherstellen, dass für sämtliche Unternehmen auf der Liste und allenfalls betroffene Unternehmen die Ausschreibung einsehbar ist.	IGE
2.4	Unternehmen informieren sich über ihre Zugehörigkeit zur Grundgesamtheit gemäss Liste und Kriterienkatalog und stellen allenfalls einen Antrag auf Löschung von oder Aufnahme in die Liste	Unternehmen
2.5	Bereinigung: Erfüllen die Aufnahmeantragssteller den Kriterienkatalog? (Hier ist u.U. eine Anpassung des Kriterienkatalogs nötig) Anpassung der Grundgesamtheit (löschen/hinzufügen)	IGE
Grundgesamtheit definiert		
3.1	BFS überprüft die Repräsentativität anhand der Informationen im BUR. Falls Repräsentativität nicht gewährleistet: Sprung zu Schritt 3.4. Falls im Kriterienkatalog keine weiteren Kovariablen vorgesehen sind, Sprung zu Schritt 3.4.	BFS
3.2	Die Verteilung der Kovariablen, welche nicht im BUR enthalten sind, werden anhand von bestehenden Informationsquellen oder einer spezifischen Erhebung berechnet und dem IGE eingereicht.	BV
3.3	Überprüfung der Informationsquellen bzw. der Erhebung Falls ungenügende Qualität: Zurück zu Schritt 3.2 oder Anordnung einer Urabstimmung über die Branchenverordnung.	IGE
3.4	Entscheid über Repräsentativität des BV aufgrund der erhobenen Daten	IGE
Repräsentativität nachgewiesen bzw. abgelehnt		

*UID ist die Unternehmens-Identifikationsnummer des Betriebs- und Unternehmensregisters (BUR) des Bundesamts für Statistik

4.1 Bildung der Gruppe und Erstellen von Kriterienkatalog (Schritt 1.1)

In der ersten Stufe muss die Gruppe das von der Branchenverordnung betroffene Produkt bzw. die betroffene Branche durch einen Kriterienkatalog abgrenzen. Dieser Kriterienkatalog kann je nach Antrag zum Beispiel folgende Punkte enthalten:

- Definition des Produktionsprozesses mit allen gängigen/notwendigen Produktionsschritten
- Abgrenzung des Produktionsprozess' von Beginn bis Fertigstellung
- Tätigkeit laut NOGA⁶ (4 oder 6-stellig), CPA⁷ oder den Warennomenklaturen der EZV⁸
- Regionale Abgrenzung (hier wäre der Standort der Arbeitsstätten relevant)
- Spezifische Anforderungen an das Produktionsverfahren
- Spezifische Produktattribute
- Substituierbarkeit der Produkte
- Definition der Kovariablen welche für den Repräsentativitätsnachweis verwendet werden sollen.

Da der Kriterienkatalog ab diesem Zeitpunkt grundsätzlich nicht mehr angefochten werden kann, sollte der exakten Ausarbeitung des Katalogs besonderes Interesse gewidmet werden. Eine Expertenüberprüfung wäre in diesem Zeitpunkt angebracht, z.B. um den Produktionsprozess und die Produktionsverfahren und die Definition des Endproduktes zu überprüfen. Auch juristische Aspekte, wie z.B. die Koexistenz von verwandten Produkten (siehe Bündner Bergkäse) müssen überprüft und definiert werden. Wird z.B. keine Koexistenz erlaubt, ergibt sich daraus eine massiv grössere Grundgesamtheit, wodurch die Chance für Repräsentativität der Gruppe sinkt. So könnte der Gruppe ein Anreiz für eine möglichst detaillierte Definition gesetzt werden.

4.2 Liste der Mitglieder (Schritt 1.2)

Der Branchenverband erstellt eine Liste seiner Mitglieder, in der die Mitglieder welche aus seiner Sicht unter die Branchenverordnung fallen, speziell bezeichnet sind. Es ist möglich, dass nicht alle Mitglieder unter die Branchenverordnung fallen. Damit der Verband eine Stellungnahme über die Meinung der Grundgesamtheit abgeben kann, muss er die Zugehörigkeit der Mitglieder individuell angeben können.

⁶ Die NOGA steht kurz für "Nomenclature Générale des Activités économiques" (Allgemeine Systematik der Wirtschaftszweige). In der amtlichen Statistik werden Unternehmen und Arbeitsstätten aufgrund ihrer Haupttätigkeit systematisch in einen Wirtschaftszweig eingeteilt. Die Klassifizierung dient dazu statistische Informationen zu strukturieren, zu analysieren und darzustellen. Die NOGA Klassifizierung wird in verschiedenen Granularitäten, von 2-stelligen bis 6-stelligen Codes angegeben. Die 6-stelligen NOGA-Codes sind die feinste Unterteilung der Wirtschaftszweige.

⁷ CPA steht kurz für "Classification statistique des produits associée aux activités". Sie ist eine Güterklassifikation (für Waren und Dienstleistungen) auf Ebene der Europäischen Union und bezieht sich auf die in der Systematik der Wirtschaftszweige der Europäischen Union (NACE) definierten Wirtschaftszweige. Die NOGA berücksichtigt die vorgegebenen Rahmenbedingungen der NACE.

⁸ Warennomenklaturen der Eidg. Zollverwaltung 2013

4.3 Überprüfung der Mitglieder des BV (Schritt 1.3)

Das IGE überprüft die Übereinstimmung der Unternehmenstätigkeit aller Mitglieder des BV anhand des Kriterienkatalogs. Das IGE kann die Gruppe auffordern, gewisse Unternehmen aus der Grundgesamtheit auszuschliessen oder den Kriterienkatalog entsprechend anzupassen.

Ab diesem Zeitpunkt ist der Kriterienkatalog nicht weiter anfechtbar und gilt als Grundlage für die beantragte Branchenverordnung.

4.4 Definition der Grundgesamtheit aus der Sicht BV (Schritt 2.1)

Der BV reicht dem IGE eine Liste mit sämtlichen Unternehmen der Grundgesamtheit ein. Es sollten alle Unternehmen, die am Produktionsprozess beteiligt sind, aufgeführt werden, unabhängig von der Produktionsstufe. Auf der Liste werden die Mitglieder des BV entsprechend markiert.

4.5 Erste Prüfung anhand des BUR (Schritt 2.2)

Die Liste wird durch das IGE mithilfe des BUR auf ihre Plausibilität überprüft. Welche Wirtschaftsklassen sind an der Branche beteiligt, sind die Unternehmen aktiv bzw. welche Rechtsform besitzen sie. Allenfalls wird bereits hier klar, dass die Repräsentativität nicht erreicht werden kann.

4.6 Öffentliche Ausschreibung (Schritt 2.3)

Eine vollständige Identifikation der Grundgesamtheit ist aus dem BUR oder anderen Registern leider nicht möglich. Deshalb muss die Öffentlichkeit über das Begehren und die Liste informiert werden. Die Liste der Unternehmen und der Kriterienkatalog müssen der Öffentlichkeit zugänglich gemacht werden (z.B. im SHAB). Es muss darauf verwiesen werden, dass die Unternehmen selbst in der Pflicht sind, ihre Zugehörigkeit zu überprüfen. Ausserdem muss eine Frist festgelegt werden, nach deren Ablauf die Grundgesamtheit unanfechtbar ist.

4.7 Bereinigung der Grundgesamtheit (Schritt 2.4)

Falls ein Unternehmen mit seinem Eintrag (oder Fehlen) nicht einverstanden ist, kann es sich entweder von der Liste entfernen lassen oder einen Antrag für Aufnahme in die Liste stellen. Die Beweislast der Übereinstimmung der Unternehmenstätigkeit mit dem Kriterienkatalog liegt bei den Unternehmen selbst.

Die Tätigkeit eines Unternehmens muss aufgrund von Selbstdeklaration erfasst werden. Bestehen erhebliche Zweifel oder erhebt die Gruppe Einspruch, ist ein Audit der Unternehmung nicht ausschliessbar. Der Kriterienkatalog müsste in diesem Falle auch auditfähige Kriterien enthalten.

Da Unternehmen im Extremfall gegen ihren Willen durch die Gruppe auf die Populationsliste gesetzt werden können, ist hier die Notwendigkeit eines Briefversandes an alle von vornherein durch die Gruppe definierten Populationsmitglieder zu überprüfen. Ein solcher Versand würde je nach Listengrösse erhebliche Kosten mit sich bringen, ist jedoch unter Umständen, u.A. aus juristischen Gründen unvermeidbar.

4.8 Festlegung der Grundgesamtheit (Branche) (Schritt 2.5)

Ist der Prozess der Populationsdefinition abgeschlossen, d.h. bestehen keine weiteren Anträge auf Aufnahme oder Löschung von der Liste, ist die Grundgesamtheit abschliessend de-

finiert und kann nicht mehr angefochten werden. Gemessen an den im BUR erfassten Variablen kann nun diese Grundgesamtheit mit Verteilungsfunktionen beschrieben werden. Dies könnte z.B. mithilfe des Umsatzes, Anzahl Beschäftigte oder Vollzeitäquivalenten geschehen.

4.9 Überprüfung der Repräsentativität anhand BUR-Variablen (Schritt 3.1)

Das BFS berechnet die Verteilungen der Kovariablen des Kriterienkatalog, welche im BUR erfasst sind. Dies sind insbesondere die wirtschaftliche Aktivität, die Grösse und die regionale Verteilung der Unternehmen, eventuell auch Rechtsform und/oder Umsatz. Die Verteilungen für die Grundgesamtheit und die Gruppe werden verglichen und damit wird eine erste Überprüfung der Repräsentativität erreicht. Falls diese negativ verläuft kann hier der Prozess abgebrochen werden. Falls keine weiteren Kovariablen notwendig sind, um die Repräsentativität zu belegen, kann direkt zum Entscheid (Schritt 3.4) gesprungen werden.

4.10 Information für zusätzliche Kovariablen (Schritt 3.2)

Für die Kovariablen, die nicht im BUR erfasst sind, müssen externe Informationsquellen herangezogen werden oder es muss eine spezifische Erhebung durchgeführt werden.

Da oft keine Daten zu diesen Kriterien vorliegen und die Grundgesamtheit neu definiert wurde, muss allenfalls eine spezifische Erhebung durchgeführt werden. Der Erhebungsgegenstand muss demzufolge bereits im Kriterienkatalog definiert werden. Da die Beweislast auf Seiten der Gruppe liegt, müssen auch die Kosten von ihnen getragen werden. Um diese Kosten in einem zumutbaren Rahmen zu halten, muss allerdings nur eine Stichprobe der Grundgesamtheit befragt werden. Die Stichprobengrösse wird durch das IGE definiert. Die Glaubwürdigkeit der Resultate könnte hier ein grosser Diskussionspunkt sein. Um die Glaubwürdigkeit der Erhebung zu erhöhen, wird der Stichprobenplan durch das IGE/BFS bewilligt und die Stichprobe durch das IGE/BFS aus dem BUR direkt gezogen. Eine Schichtung wird aufgrund der Zugehörigkeit erfolgen, so dass für Mitglieder und Nicht-Mitglieder der Gruppe separate Hochrechnungen gemacht werden können. Falls die Gruppe bereits über zuverlässige Zahlen verfügt, müsste eventuell die Erhebung nur für die Nicht-Mitglieder durchgeführt werden. Möglicherweise bestehen bereits Statistiken welche herangezogen werden können. Hier entscheidet auch das IGE über die Qualität.

4.11 Überprüfung der Qualität (Schritt 3.3)

Die ausgewertete Erhebung wird mit einer ausführlichen Dokumentation dem IGE übergeben. Bei Beizug von anderen Informationsquellen ist ebenfalls eine ausführliche Dokumentation beizubringen.

Das IGE überprüft die Qualität und entscheidet ob sie genügend ist für den Nachweis der Repräsentativität.

Falls der Aufwand für eine Erhebung zu gross ist oder die Qualität der Erhebung bzw. der Informationsquellen ungenügend ist, kann als Alternative auch eine Urabstimmung in der Grundgesamtheit über die Branchenverordnung angesetzt werden.

4.12 Überprüfung der Repräsentativität (Schritt 3.4)

Das IGE entscheidet aufgrund der Verteilungen der relevanten Variablen und dem in Abschnitt 3 erläuterten Distanzmass über die Repräsentativität der Gruppe für die zuvor definierte Grundgesamtheit.

5 Repräsentativitätsüberprüfung anhand eines fiktiven Beispiels mit Hilfe des Chiquadrat-Anpassungstests

In diesem Abschnitt zeigen wir anhand eines fiktiven Beispiels, wie die Überprüfung der Repräsentativität im Schritt 3.1 und 3.4 konkret ablaufen könnte. Wir verwenden dazu Daten aus dem BUR. Wir überprüfen mit einem Chiquadrat-Anpassungstest, ob der (imaginäre) Branchenverband in den zwei Charakteristiken Grossregion und Unternehmensgrössenklasse eine ähnliche Verteilung aufweist wie die Grundgesamtheit. Es wird hier also die zweite Bedingung für die Modell-Repräsentativität überprüft. Als Teststatistik dient die in Abschnitt 3 eingeführte Teststatistik des Chiquadrat Anpassungstests.

5.1 Definition verschiedener Gruppen (Branchenverbände)

Im folgenden Beispiel setzt sich die Grundgesamtheit aus allen im BUR eingetragenen Unternehmen eines vierstelligen NOGA-Codes, also einer relativ fein definierten Branche, zusammen. Es handelt sich um eine Branche mit 159 Unternehmen. Die Gruppe (Branchenverband) wird hier fiktiv erstellt und zu Illustrationszwecken zweimal verändert. In Tabelle 2 ist die Verteilung der Grundgesamtheit in den Kovariablen Grossregion und Unternehmensgrösse (klassifiziert nach Anzahl Vollzeitäquivalenten (VZÄ)) dargestellt. Die erste Gruppe unterscheidet sich nur im Merkmal "Grossregion" von der Grundgesamtheit, die zweite Gruppe in beiden Merkmalen (Grossregion und Unternehmensgrössenklassen). Gruppe drei unterscheidet sich zu Gruppe zwei nur in der Anzahl ihrer Mitglieder. Die Verteilung der Unternehmen der Gruppe nach Grossregion und VZÄ ist im Anhang aufgeführt.

Tabelle 2: Struktur der Grundgesamtheit in den untersuchten Variablen Grossregion und Unternehmensgrösse

	SW	EM	NW	ZH	OS	ZS	TI	Summe
0-1	11	8	8	7	15	13	3	65
2-9	16	9	3	10	8	3	5	54
10-	12	7	4	5	7	4	1	40
Summe	39	24	15	22	30	20	9	159

(SW=Genferseeregion, EM=Espace Mittelland,
 NW=Nordwestschweiz, ZH=Zürich, OS=Ostschweiz,
 ZS=Zentralschweiz, TI=Tessin)

5.2 Univariate Chiquadrat-Testergebnisse der verschiedenen Gruppen

Die Resultate der Chiquadrat-Analyse sind in Tabelle 3 zusammengefasst.

Wir betrachten zuerst Gruppe 1, die im Tessin gar nicht und in der West- und Zentralschweiz untervertreten, dafür im Mittelland übervertreten ist. Die verschiedenen Unternehmensgrössenklassen sind identisch verteilt wie in der Grundgesamtheit (prozentualer Anteil der kleinen und grossen Unternehmen stimmt in allen Regionen überein). Das Tessin ist hier zu klein für eine zuverlässige Chiquadrat Statistik und wird zur Region Genfersee geschlagen. Die strukturelle Abweichung der Gruppe spiegelt sich auch im Resultat des Chiquadrat-

Anpassungstest wider. Das Chi-Quadrat für die Kovariable Grossregion beträgt 12.353 und ist daher über der Schwelle von 11.070. Wir schliessen also, dass die Gruppe bezüglich Grossregionen, wegen den oben genannten Abweichungen, nicht repräsentativ ist.

Für die Vollzeitäquivalente wurden 3 Klassen⁹ gebildet, nämlich 0 bis unter 2 VZÄ, 2 bis unter 10 VZÄ und 10 oder mehr VZÄ. Das Chi-Quadrat für die Kovariable Vollzeitäquivalente beträgt 0.045 und zeigt, dass Repräsentativität hier gegeben ist. So wurde in diesem Beispiel die Gruppe natürlich auch konstruiert.

Gruppe 2 unterscheidet sich sowohl in der regionalen Struktur signifikant (Chi-Quadrat 12.353 wie bei Gruppe 1) als auch in der Grössenstruktur (Chi-Quadrat 6.811 bei einem Schwellenwert von 5.991) signifikant von der Grundgesamtheit. In Gruppe 2 sind Unternehmen mit mehr als 10 Vollzeitäquivalenten (VZÄ) massiv untervertreten, solche mit weniger als 10 VZÄ übervertreten. Die Verteilung in den Grossregionen stimmt mit derer von Gruppe 1 überein.

Gruppe 3 hat nur noch halb so viele Mitglieder wie Gruppe 2, ist ansonsten jedoch identisch mit Gruppe 2. Da die Gruppe kleiner ist, muss etwas mehr Spielraum für Abweichungen von der Verteilung der Grundgesamtheit gewährt werden. Darum sind die Chi-Quadrat-Werte proportional zur Gruppengrösse (12.353 Gruppe 2 vs. 6.177 Gruppe 3). Es zeigt sich, dass dann die Gruppe nicht nur bezüglich der Grossregionen sondern auch (knapp) bezüglich den Vollzeitäquivalenten als repräsentativ angesehen werden muss (Chi-Quadrat ist 3.405 bei einem Schwellenwert von 5.991).

Wir haben hier zwei Kovariablen separat getestet und auf Repräsentativität begutachtet. Je mehr Variablen getestet werden, desto öfter kann sich ein grosser Chi-Quadrat-Wert rein zufällig ergeben. Dieses Problem des mehrfachen Testens kann durch verschiedene Korrekturen (z.B. Bonferroni Korrektur) kompensiert werden. Da für die Repräsentativität von Branchen meistens nur wenige Variablen betrachtet werden, schlagen wir vor, auf diese Korrekturen zu verzichten.

5.3 Bivariate Chi-Quadrat-Testergebnisse der verschiedenen Gruppen

Im Normalfall sollten die univariaten Chi-Quadrat-Statistiken für die Beurteilung der Repräsentativität genügen. Es kann aber vorkommen, dass auch zwei (oder theoretisch auch mehr Dimensionen) gleichzeitig berücksichtigt werden müssen, zum Beispiel weil kleine Betriebe auf eine bestimmte Region konzentriert sind. Dies ist prinzipiell machbar. Hier werden die zwei Variablen Grossregion und Grössenklasse gekreuzt betrachtet, aber immer noch mit der Verteilung in der Grundgesamtheit (Tabelle 2) verglichen¹⁰. Damit aber die Anzahl zu erwartende Unternehmen pro Klasse noch gross genug ist, werden nur noch drei Regionen (SW+TI, MW+NW+ZH, OS+ZS) betrachtet.

Die Gruppe 1 ist bei gekreuzten Kovariablen repräsentativ mit einem Chi-Quadrat von 9.761 bei einer Schwelle von 15.507. Die Gruppe 2 ist auch bei Kreuzung der Kovariablen nicht repräsentativ. Bei Gruppe 3 zeigt es sich, dass trotz der Reduktion auf nur drei Grossregionen zu viele Zellen schwach besetzt sind (weniger als 5 erwartete Unternehmen). Der Vergleich des Chi-Quadrat-Wert von 8.648 mit dem Schwellenwert 15.507 muss also mit Vorsicht betrachtet werden, würde aber anzeigen, dass kein Problem mit der Repräsentativität besteht.

⁹ Im BUR ist eine sehr feine Klassifizierung der Unternehmensgrösse vorhanden. Zu Illustrationszwecken beschränken wir uns hier auf eine sehr grobe Klassifizierung (siehe Tabelle 2).

¹⁰ Es handelt sich also hier immer noch um einen Anpassungs-Test, d.h. nicht um einen Unabhängigkeits-Test wie er oft in Kreuztabellen durchgeführt wird.

5.4 Diskussion

Die Beispiele zeigen, dass der Chiquadrat-Anpassungstest genügend grosse Verteilungsunterschiede von Kovariablen erkennt. Es wurde auch deutlich, dass es essentiell ist, welche Kovariablen in Betracht gezogen werden. Unterscheidet sich die Gruppe in einer unbeobachteten Kovariable signifikant von der Grundgesamtheit, ist dies mit dem Test nicht nachweisbar. Zusätzlich hat die Mitgliederanzahl der Gruppe einen entscheidenden Einfluss. Je kleiner die Gruppe, desto mehr Spielraum muss bezüglich der Repräsentativität gewährt werden.

Um die Übertragbarkeit des impliziten Modells (Vorhersagbarkeit der interessierenden Variable anhand der Kovariablen) von der Gruppe auf die Grundgesamtheit zu gewährleisten, schlagen wir im Abschnitt 3.2 eine relative Mindestgrösse der Gruppe von 50% vor. In unserem Beispiel erfüllt Gruppe 3 dieses Kriterium klar nicht und gelte somit von vornherein als nicht repräsentativ, bzw. würde als Antragssteller von Beginn abgelehnt.

Tabelle 3: Chiquadrat-Anpassungstest - Ergebnisse nach Gruppen und Kovariablen

Untersuchte Kovariable		Gruppe 1 (n=80)	Gruppe 2 (n=80)	Gruppe 3 (n=40)
Grossregion	Testgrösse	12.353	12.353	6.177
$(\nu = 5)$: Schwelle=11.070				
Unternehmensgrösse	Testgrösse	0.045	6.811	3.405
$(\nu = 2)$: Schwelle=5.991				
Grossregion x Unternehmensgrösse	Testgrösse	9.761	17.295	8.648
$(\nu = 8)$: Schwelle=15.507				

Es wird nicht zu vermeiden sein, dass sich die Repräsentativitätsindikatoren (Chiquadrat-Werte für verschiedene Variablen) manchmal widersprechen, d.h. dass die Gruppe bezüglich einer (oder mehrerer) Variablen repräsentativ ist aber nicht für andere. Dies ist der Fall im obigen Beispiel bei Gruppe 2. Damit der Nachweis der Repräsentativität nicht verwässert wird schlagen wir vor, maximal 4 Kovariablen zu überprüfen und für jede Kovariable die Signifikanzschwelle gleich anzusetzen, nämlich bei $\alpha = 0.05$. Damit die Repräsentativität insgesamt gegeben ist, muss sie bei allen Kovariablen vorhanden sein.

6 Schlussbemerkungen

Die empirische Überprüfung der Plausibilität des Modells benötigt u.U. eine Urabstimmung bei den Mitgliedern des Verbands. Ob diese eingefordert werden kann, bleibt abzuklären.

Der Prozess zur Abklärung der Repräsentativität kann an mehreren Stellen vorzeitig abgeschlossen werden, bzw. die Tiefe der Repräsentativitäts-Überprüfung kann dem Problem angepasst werden. Bis zu welcher Stufe die Abklärungen gehen müssen, kann eventuell erst im Lauf des Prozesses entschieden werden, z.B. erst nachdem eine allfällige Opposition Gelegenheit hatte, sich zu äussern.

Die vorgeschlagenen Schwellenwerte ($\alpha = 0.05, n \geq \max(20, 0.5 \cdot N)$) konnten empirisch für den Fall der Repräsentativität bis jetzt nicht überprüft werden. Diese Schwellenwerte basieren auf Erfahrungswerten aus Erhebungen.

Literaturverzeichnis

Hedderich, J. und Sachs, L. (2012): *Angewandte Statistik*, Berlin, Springer Verlag Berlin Heidelberg.

Kruskal, W. und Mosteller, F.(1979): Representative Sampling, III: The Current Statistical Literature, *International Statistical Review / Revue Internationale de Statistique, International Statistical Institute (ISI)*, 47, pp. 245-265.

Bundesamt für Statistik (2013): *Anwenderhandbuch zum UID-Register für die Öffentlichkeit V1.1*.

Anhang

Tabelle 4: Struktur Grundgesamtheit

	SW	EM	NW	ZH	OS	ZS	TI	Sum
0-1	11	8	8	7	15	13	3	65
2-9	16	9	3	10	8	3	5	54
10-	12	7	4	5	7	4	1	40
Sum	39	24	15	22	30	20	9	159

Tabelle 5: Struktur Branchenverband Version 1

	SW	EM	NW	ZH	OS	ZS	TI	Sum
0-1	5	6	5	6	7	4	0	33
2-9	7	6	4	6	3	2	0	28
10-	4	4	3	4	2	2	0	19
Sum	16	16	12	16	12	8	0	80

Tabelle 6: Struktur Branchenverband Version 2

	SW	EM	NW	ZH	OS	ZS	TI	Sum
0-1	6	7	5	7	8	5	0	38
2-9	8	7	5	7	3	2	0	32
10-	2	2	2	2	1	1	0	10
Sum	16	16	12	16	12	8	0	80

Tabelle 7: Struktur Branchenverband Version 3

	SW	EM	NW	ZH	OS	ZS	TI	Sum
0-1	3	3	3	4	4	2	0	19
2-9	4	4	2	3	2	1	0	16
10-	1	1	1	1	0	1	0	5
Sum	8	8	6	8	6	4	0	40