# Heterogentitätsindizes zur Messung der Pluralität von Lebensformen und ihre Berechnung in SPSS

Franzmann, Gabriele; Wagner, Michael

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

# Heterogentitätsindizes zur Messung der Pluralität von Lebensformen und ihre Berechnung in SPSS[1]

**von Gabriele Franzmann und Michael Wagner[2]**


## *Zusammenfassung:*

*In der Diskussion um die Pluralisierung von Lebensformen ist bislang nicht versucht worden, die Heterogenität von Lebensformen quantitativ zu bestimmen. Ziel dieses Beitrages ist es, dieses Defizit der familiensoziologischen Forschung auszugleichen. Es werden drei Maße der Heterogenität vorgestellt, nämlich das Entropie-Maß, der Diversity-Index und der Dissimilaritätsindex. Ihre Eigenschaften werden anhand idealtypischer Verteilungen diskutiert, dem sich dann die Erläuterung ihrer Umsetzung in SPSS anschließt.*

## *Abstract:*

*Despite of a long debate about an increasing heterogeneity of household types no attempt was made to measure this heterogeneity. This article aims to overcome this deficit. Three measures of heterogeneity are explained: the measure of entropy, the diversity index, and the index of dissimilarity. These measures are explained by an illustrative example. Furthermore, their computation in SPSS is shown.*


## 1 Fragestellung

In den vergangenen Jahren wurde von vielen Familiensoziologen die These vertreten, es habe eine Pluralisierung der Lebensformen bzw. -stile stattgefunden (***Wehrspaun*** 1988:157ff.; ***Zapf*** et al. 1987; ***Tyrell*** 1979). Einige Autoren führen sie auf gesellschaftliche Veränderungen zurück, die seit den 60er Jahren Einfluß auf die Familie nehmen, wie zum Beispiel die Anhebung des materiellen Lebensstandards, die Bildungsexpansion, die zunehmende Erwerbstätigkeit der Frauen sowie ein Rückgang von Klassenbindungen und

---

1 Dieser Beitrag ist im Rahmen unseres Projektes ‘Messung der Heterogenität von Lebensformen’ angefertigt worden.

2 Prof. Dr. ***Michael Wagner*** und ***Gabriele Franzmann***, M.A., Forschungsinstitut für Soziologie, Universität zu Köln, Greinstraße 2, 50939 Köln.

daran geknüpften Denk- und Verhaltensweisen (***Beck*** 1989: 130ff.). Im Rahmen dieser Entwicklung kommt es - so die These - zu einem Legitimitäts- oder Plausibilitätsverlust der traditionellen Kern- oder Gattenfamilie als institutionalisierte Lebensform (***Tyrell*** 1979: 59ff.). Die biographische Selbstverständlichkeit von Ehe und Elternschaft bestehe nicht mehr, sondern die Gestaltung der eigenen Lebensform werde zunehmend zur Frage der persönlichen Wahl. Diese neuen Entscheidungsmöglichkeiten, aber auch der veränderte Anpassungsdruck anderer gesellschaftlicher Teilsysteme, wie zum Beispiel des Arbeitsmarktes (***Kaufmann*** 1988: 407f.), führten dazu, daß sich neue Lebensformen herausbilden, die funktional besser auf die Bedürfnisse ihrer Mitglieder spezialisiert sind (***Nave-Herz*** 1997: 48). Infolge dieser Entwicklungen setze eine Diversifizierung und Individualisierung von Lebenslagen und -wegen ein. Dieser Prozeß wird auf der gesellschaftlichen Ebene als eine „Zunahme von gruppen-, milieu- und situationsspezifischen Ordnungsmustern zur Organisation von Lebenslagen, Ressourcen und Lebensplanung“, als „Pluralisierung der Lebensstile“ sichtbar (***Zapf*** et al. 1987: 18).

Wir beschränken uns im folgenden auf die Untersuchung von Lebensformen, genauer auf Haushaltsformen. Behauptet wird, daß die Pluralität der Haushaltsformen gegenwärtig hoch sei und etwa seit dem Ende der 60er Jahre zugenommen habe (***Kaufmann*** 1995). Es sind jedoch nach unserer Kenntnis im Rahmen dieser Diskussion mit Ausnahme des Beitrags von ***Huinink*** und ***Wagner*** (1998) keine Heterogenitätsmaße angewendet worden.

Unter dem Begriff ‘Heterogenitätsmaße’ verbirgt sich eine Vielfalt von verschiedenen Konzepten zur Messung von Verteilungen, die in unterschiedlichen wissenschaftlichen Disziplinen entwickelt wurden. Zunächst wird in Abschnitt 2 die bisherige Verwendung von Heterogenitätsmaßen in soziologischen Studien behandelt. Anschließend werden in Abschnitt 3 die Maße in ihren formalen Eigenschaften vorgestellt und anhand beispielhafter, idealtypischer Verteilungen kurz diskutiert. In Abschnitt 4 wird die Berechnung dieser Maße mit SPSS sowie die Anwendung eines SPSS-Macros für Heterogenitätsmaße erläutert. Abschließend wird in Abschnitt 5 anhand der Diskussion um die Heterogenität von Haushaltsformen mit Hilfe des kumulierten ALLBUS-Datensatzes die Aussagekraft dieser Maße beschrieben.

## 2 Die Verwendung von Heterogenitätsmaßen

Zur Messung der Heterogenität von Lebensformen bieten sich Maße an, deren Anwendung in der Soziologie nicht neu ist. So haben ***Gibbs*** und ***Martin*** (1962) ein Maß für den Grad der Arbeitsteilung vorgeschlagen, um so Rückschlüsse auf die Urbanisierung eines Gebietes ziehen zu können. ***Herfindahl*** (1950) entwickelte ein Homogenitäts- oder Konzentrationsmaß, um das Ausmaß der Monopolisierung in der amerikanischen Stahl-Industrie zu messen. ***Lieberson*** (1969) stellte ein Maß für die Diversität innerhalb einer Population

bezüglich eines Merkmals vor, welches die Wahrscheinlichkeit wiedergibt, daß zwei zufällig ausgewählte Mitglieder einer Population bezüglich eines Merkmals eine unterschiedliche Ausprägung besitzen. Um das Ausmaß der ungleichen Verteilung von Bevölkerungsgruppen über städtische Teilgebiete (Segregation) zu messen, sind mehrere Indizes entwickelt worden, wobei der von ***Duncan*** und ***Duncan*** (1955a und 1955b) vorgeschlagene Dissimilaritätsindex am erfolgreichsten war (***Friedrichs*** 1983: 219f.). Das auf dem Gebiet der Kommunikationstheorie von ***Shannon*** entwickelte Entropie-Maß wurde übernommen, um zum Beispiel Fragen einer angemessenen parlamentarischen Repräsentation von Parteien beantworten zu können (***Theil*** 1969).

Das Konzept der Heterogenität wird in jüngster Zeit wieder verstärkt in der sozialwissenschaftlichen Forschung aufgenommen. Beispielhaft seien hier die Beiträge von ***Jagodzinski*** und ***Quandt*** (1997) sowie von ***Huinink*** und ***Wagner*** (1998) erwähnt, in denen das Entropie-Maß zur Anwendung kommt. In diesen Beiträgen wird der Versuch unternommen, den Individualisierungsbegriff zu präzisieren - siehe dazu auch ***Friedrichs*** (1998) - sowie den Zusammenhang zwischen Individualisierung und Pluralisierung zu klären. Ausgangspunkt ist die These, daß aufgrund der sinkenden Bedeutung von sozio-kulturellen Handlungsvorgaben ein Zuwachs an Handlungsoptionen eintreten kann. Gruppennormen beeinflussen immer weniger individuelle Verhaltensweisen. Diese Vergrößerung der Optionen kann „mit einer Zunahme der Vielfalt von realisierten Handlungsweisen einhergehen“ (***Huinink/ Wagner*** 1998: 87), so daß letztlich „das Handeln der Individuen im Zuge der Modernisierung immer schlechter vorhersagbar“ wird (***Jagodzinski/Quandt*** 1997: 762). Handlungs- und Einstellungswahrscheinlichkeiten verschiedener Individuen aus verschiedenen Gruppen gleichen sich im Verlauf dieses Prozesses aneinander an (***Jagodzinski/Quandt*** 1997: 762). Derartige Sachverhalte lassen sich in Anlehnung an das Entropiegesetz der Thermodynamik, welches sich auf den Austausch von Elementen eines Systems mit Elementen aus seiner Umgebung bezieht, als einen wechselseitigen Austausch von Mitgliedern verschiedener Gruppen verstehen. Pluralisierung - als mögliche gesellschaftliche Folge von Individualisierung[3] - bedeutet dann die Vergrößerung der Anzahl der realisierten Handlungsalternativen auf einem Gebiet, sei es die Wahl von Parteien oder die Wahl einer Lebensform. „(Dabei) muß sich nicht ... die Zahl realisierter Zustände, sondern es kann sich auch die Verteilung realisierter Zustände verändern“ (***Huinink/Wagner*** 1998: 88), so daß sich eine Tendenz zur Gleichverteilung von realisierten Handlungsalternativen in der Bevölkerung durchsetzt.

3 ***Huinink*** und ***Wagner*** (1998: 91f.) weisen darauf hin, daß Individualisierungsprozesse nicht zu einer Pluralisierung von Lebensformen führen müssen. Ebenso wird die Wirkung von Traditionen falsch eingeschätzt, die nicht zwangsläufig zu einer Homogenisierung der Lebensformen führen müssen.

***Jagodzinski*** und ***Quandt***, die die Individualisierungsthese am Beispiel des Wahlverhaltens in Abhängigkeit von Konfessionszugehörigkeit und Kirchgangshäufigkeit zwischen 1953 und 1996 prüfen, wenden zusätzlich noch den Dissimilaritätsindex von ***Duncan*** und ***Duncan*** und den Herfindahl-Index an. Mit dem Dissimilaritätsindex sollen die Unterschiede zwischen zwei Konfessionsgruppen im Zeitverlauf herausgearbeitet werden. Der ***Herfindahl***-Index dient zur Bestimmung der Veränderungen der Konzentration auf eine bestimmte Merkmalsausprägung innerhalb einer Wählergruppe (***Jagodzinski/Quandt*** 1997: 769f.).

## 3 Das Konzept der Heterogenität

### 3.1 Definitionen ausgewählter Heterogenitätsmaße

***Shannon*** stellt sich bei einer gegebenen Verteilung von Ereignissen die Frage: „Can we find a measure of how much 'choice' is involved in the selection of the event or of how uncertain we are of the outcome?" (***Shannon/Weaver*** 1959: 18). Das Entropie-Maß, welches den Informationsgehalt einer Nachricht oder die Unsicherheit des Eintreffens eines Ereignisses messen soll, wird definiert als die Summe der mit ihrem Logarithmus gewichteten Anteile einer Verteilung. Merkmalsausprägungen mit geringen Anteilswerten in einer Population weisen einen hohen Logarithmuswert auf, womit zum Ausdruck gebracht wird, daß die Unsicherheit, einer Person zu begegnen, die diesen schwach besetzten Merkmalskategorien angehört, groß ist. Damit ist auch der Informationsbedarf hoch, um ein solches Ereignis vorhersagen zu können. Die Formel lautet wie folgt (***Shannon/Weaver*** 1959: 20):

$$H = -\sum_{i=1}^{k} p_i \log p_i \ .$$

Das Entropie-Maß nimmt den Wert Null an, wenn alle Einheiten sich in einer Kategorie wiederfinden und erreicht seinen Maximalwert bei einer gleichmäßigen Verteilung aller Einheiten über alle Kategorien hinweg, was den größten Grad an Unsicherheit und Unvorhersagbarkeit anzeigt (***Shannon/Weaver*** 1959: 21). Da es sich auf die Verteilung der Mitglieder einer Gruppe über die Kategorien eines Merkmals bezieht, ist es ein Maß der Intragruppenheterogenität. Um dieses Maß zu standardisieren, schlägt ***Coulter*** (1989: 107) die Division des Entropie-Wertes durch den Logarithmus der Anzahl der Kategorien[4] vor:

$$H_{st} = \left( -\sum_{i=1}^{k} p_i \log_2 p_i \right) \Big/ \left( \log_2 K \right).$$

4 Der Logarithmus aus der Anzahl der Kategorien (log(k)) ist zugleich der Maximalwert, den das unstandardisierte Entropie-Maß erreichen kann.

***Coulter*** zieht den Logarithmus auf der Basis 2 dem natürlichen Logarithmus vor, da so der Prozeß der Selektion zwischen richtigen und falschen Informationen zur Vorhersage des Eintreffens eines Ereignisses in der Formel besser berücksichtigt wird. Er verdeutlicht diese Überlegung am Beispiel eines Feldes mit 64 möglichen Positionen, die auf diesem Feld eingenommen werden können. Um herauszufinden, welche Position besetzt ist, benötigt man insgesamt 6 Informationen. Jede Information auf eine gestellte Frage schließt sukzessiv durch die Angabe 'richtig' oder 'falsch' die Hälfte aller Möglichkeiten aus, bis schließlich nur noch eine bestimmte Position als richtige Lösung in Frage kommt. Der Logarithmus von 64 auf der Basis 2 beträgt 6 und gibt somit den Informationsbedarf wieder. Durch die Gewichtung der Häufigkeiten mit ihrem Logarithmus hat dieses Maß die Eigenschaften, daß es schwach besetzte Merkmalskategorien stärker berücksichtigt. Zieht man das Entropie-Maß für einen Vergleich der Verteilungen einer Stichprobe zu verschiedenen Zeitpunkten heran, so reagiert das Maß sensibler auf Veränderungen in schwach besetzten Kategorien (***Peet*** 1974: 296).

Der *Diversity-Index*[5] ist ein Heterogenitätsmaß, welches das Ausmaß der Verteilung einer Stichprobe über die verschiedenen Kategorien eines Merkmals - die Intragruppenheterogenität - mißt. Er gibt die Wahrscheinlichkeit an, mit der zwei zufällig ausgewählte Individuen aus einer Stichprobe verschiedenen Kategorien eines Merkmals angehören (***Agresti/ Agresti*** 1978: 206; ***Lieberson*** 1969: 851). Zu diesem Wert kommt man durch das Aufsummieren der quadrierten Anteile einer Verteilung, die dann von 1 subtrahiert werden [6]. Der Index erreicht - wie auch das Entropie-Maß - den Wert Null, wenn alle Mitglieder einer Population in der gleichen Kategorie eines Merkmals zu finden sind, so daß keine Heterogenität in der Population vorliegt. Die Formel für den Diversity-Index lautet:

$$D = 1 - \sum_{i=1}^{k} p_i^2 \ .$$

Den Höchstwert, den der Diversity-Index im Falle einer vollkommenen Gleichverteilung erreichen kann, beträgt (1-1/k), wobei k die Anzahl der Kategorien eines Merkmals ist. Die standardisierte Form des Index erhält man durch die Division des Indexwertes durch den erreichbaren Höchstwert:

---

5 In diesem Beitrag wird der Diversity-Index dem ***Herfindahl***-Index vorgezogen, da er - wie auch das Entropie-Maß - ein Maß für Heterogenität ist.

6 Der ***Herfindahl***-Index ist das Gegenstück zum Diversity-Index; die aufsummierten quadrierten Anteile geben das Ausmaß der Konzentration in einer Stichprobe wieder. Sie werden nicht von 1 abgezogen, so daß die Formel lautet:

$$C = \sum_{i=1}^{k} p_i^2 \ .$$

$$D_{st} = \left(1 - \sum_{i=1}^{k} p_i^2\right) \Big/ (1 - 1/k),$$

so daß der Index Werte zwischen 0 und 1 annimmt (***Agresti/Agresti*** 1978: 208)[7]. Durch die Aufsummierung der quadrierten Anteile werden die stark besetzten Merkmalsausprägungen im Maß stärker berücksichtigt, was dazu führt, daß der Diversity-Index bei einem Vergleich der gleichen Stichprobe zu zwei verschiedenen Zeitpunkten sensibler auf Veränderungen in stark besetzten Gruppen reagiert.

Der *Dissimilaritäts-Index* ermittelt die Differenz zwischen zwei Gruppen bezüglich ihrer Verteilung über die Kategorien eines Merkmals, so daß er einer etwas anderen Logik folgt wie die beiden zuvor besprochenen Indizes. Er ist ein Maß für die Intergruppenheterogenität. ***Duncan*** und ***Duncan*** (1955a: 211; 1955b: 494) entwickelten diesen Index, um zwei Bevölkerungsgruppen im Hinblick auf ihre Verteilung über die Teilgebiete eines Gebietes vergleichen zu können. Die berechneten Werte zwischen zwei Gruppen geben an, wieviel Prozent der Personen aus einer der beiden Gruppen in das jeweils andere Gebiet wechseln müßten, damit eine identische Verteilung zwischen diesen beiden Gruppen erreicht werden kann. Zur Berechnung des Wertes werden die absoluten prozentualen Differenzen summiert und durch zwei dividiert:

$$Diss = \tfrac{1}{2} \sum \left| p_i - p_j \right|.$$

Dieser Index erreicht seinen Maximalwert von 100%, wenn sich die Gruppen bezüglich ihrer Verteilung völlig unterscheiden, und einen Minimalwert von 0%, wenn das jeweilige Merkmal in beiden Gruppen identisch verteilt ist.

### 3.2 Darstellung der Heterogenitätsmaße anhand fiktiver Datenkonstellationen

Zunächst beschreiben wir am Beispiel fiktiver idealtypischer Verteilungen die Eigenschaften der oben besprochenen Maße. Die folgenden Tabellen zeigen vier konstruierte Verteilungsmuster und die Reaktion der besprochenen Maße auf diese Verteilungen, wobei X die Gruppierungsvariable ist und Y die interessierende abhängige Verteilung.

7 Die Berechnung inferenzstatistischer Werte des Diversity-Index haben ***Agresti*** und ***Agresti*** (1978) ausführlich beschrieben.

**Tab. 1a) bis 4c)**: Berechnungen von Heterogenitätsmaßen bei verschiedenen Annahmen über die Verteilung zweier Variablen X und Y.

1a) Gleichverteilung

| | X=1 | X=2 | |
|---|---|---|---|
| Y=1 | 50 (.33) | 50 (.33) | 100 |
| Y=2 | 50 (.33) | 50 (.33) | 100 |
| Y=3 | 50 (.33) | 50 (.33) | 100 |
| Summe | 150 (1.0) | 150 (1.0) | 300 |

1b) Berechnung des Dissimilaritätsindex

| | \|(X=1) - (X=2)\| |
|---|---|
| | 0.0<br>0.0<br>0.0 |
| Summe | 0.0 |
| Summe / 2 | 0.0 |

1c) Heterogenitätsindizes

| Maße der Heterogenität | X=1 | X=2 | Maximum | Minimum |
|---|---|---|---|---|
| unstandardisierte Entropie | 1,58 | 1,58 | 1,58 | 0 |
| standardisierte Entropie | 1,00 | 1,00 | 1,00 | 0 |
| unstandardisierte Diversity | 0,67 | 0,67 | 0,67 | 0 |
| standardisierte Diversity | 1,00 | 1,00 | 1,00 | 0 |
| Dissimilarität | 0,0% | | 100% | 0% |

2a) Maximale Konzentration

| | X=1 | X=2 | |
|---|---|---|---|
| Y=1 | 150 (1.0) | 0 (.00) | 150 |
| Y=2 | 0 (.00) | 0 (.00) | 0 |
| Y=3 | 0 (.00) | 150 (1.0) | 150 |
| Summe | 150 (1.0) | 150 (1.0) | 300 |

2b) Berechnung des Dissimilaritätsindex

| | \|(X=1) - (X=2)\| |
|---|---|
| | 1.0<br>0.0<br>1.0 |
| Summe | 2.0 |
| Summe / 2 | 1.0 |

2c) Heterogenitätsindizes

| Maße der Heterogenität | X=1 | X=2 | Maximum | Minimum |
|---|---|---|---|---|
| unstandardisierte Entropie | 0,00 | 0,00 | 1,58 | 0 |
| standardisierte Entropie | 0,00 | 0,00 | 1,00 | 0 |
| unstandardisierte Diversity | 0,00 | 0,00 | 0,67 | 0 |
| standardisierte Diversity | 0,00 | 0,00 | 1,00 | 0 |
| Dissimilarität | 100% | | 100% | 0% |

3a) ungleiche Verteilung mit gleichsinnigem Verlauf zwischen X=1 und X=2

| | X=1 | X=2 | |
|---|---|---|---|
| Y=1 | 10 (.07) | 10 (.07) | 20 |
| Y=2 | 70 (.47) | 70 (.47) | 140 |
| Y=3 | 70 (.47) | 70 (.47) | 140 |
| Summe | 150 (1.0) | 150 (1.0) | 300 |

3b) Berechnung des Dissimilaritätsindex

| | \|(X=1) - (X=2)\| |
|---|---|
| | 0.0<br>0.0<br>0.0 |
| Summe | 0.0 |
| Summe / 2 | 0.0 |

3c) Heterogenitätsindizes

| Maße der Heterogenität | X=1 | X=2 | Maximum | Minimum |
|---|---|---|---|---|
| unstandardisierte Entropie | 1,29 | 1,29 | 1,58 | 0 |
| standardisierte Entropie | 0,81 | 0,81 | 1,00 | 0 |
| unstandardisierte Diversity | 0,56 | 0,56 | 0,67 | 0 |
| standardisierte Diversity | 0,84 | 0,84 | 1,00 | 0 |
| Dissimilarität | 0% | | 100% | 0% |

4a) ungleiche Verteilung mit gegensinnigem Verlauf zwischen X=1 und X=2

| | X=1 | X=2 | |
|---|---|---|---|
| Y=1 | 10 (.07) | 70 (.47) | 80 |
| Y=2 | 70 (.47) | 70 (.47) | 140 |
| Y=3 | 70 (.47) | 10 (.07) | 80 |
| Summe | 150 (1.0) | 150 (1.0) | 300 |

4b) Berechnung des Dissimilaritätsindex

| | \|(X=1) - (X=2)\| |
|---|---|
| | 0.4<br>0.0<br>0.4 |
| Summe | 0.8 |
| Summe / 2 | 0.4 |

4c) Heterogenitätsindizes

| Maße der Heterogenität | X=1 | X=2 | Maximum | Minimum |
|---|---|---|---|---|
| unstandardisierte Entropie | 1,29 | 1,29 | 1,58 | 0 |
| standardisierte Entropie | 0,81 | 0,81 | 1,00 | 0 |
| unstandardisierte Diversity | 0,56 | 0,56 | 0,67 | 0 |
| standardisierte Diversity | 0,84 | 0,84 | 1,00 | 0 |
| Dissimilarität | 40% | | 100% | 0% |

In den Ergebnistabellen zu den einzelnen Verteilungen sind in den Tabellen 1a) bis 4c) zu jedem Maß zum einen das unstandardisierte Ergebnis, das erreichbare Maximum für die Variable Y mit drei Ausprägungen sowie die standardisierten Maße angegeben. Im Falle der Gleichverteilung (Tab. 1a) nehmen der Entropie- wie auch der Diversity-Index ihren maximalen Wert an (Tab. 1c). Der Dissimilaritätsindex zeigt mit dem Wert 0% an, daß zwischen X=1 und X=2 des Gruppierungsmerkmals über die Ausprägungen von Y eine vollkommen identische Verteilung besteht (Tab. 1b). Das Gegenbeispiel der maximalen Konzentration (Tab 2a) führt zu den erwarteten Ergebnissen für das Entropie- und Diversity-Maß von 0 sowohl für X=1 als auch für X=2, welches besagt, daß weder in X=1 noch in X=2 eine Verteilung über die Ausprägungen von Y besteht (Tab 2c). Da sich allerdings die Fälle von X=1 und X=2 auf unterschiedliche Kategorien von Y konzentrieren, nimmt der Dissimilaritätsindex den Wert von 100% an. 100% der Fälle aus X=1 oder X=2 müßten in eine andere Kategorie eingeordnet werden, damit die Verteilungen zwischen X=1 und X=2 vollkommen identisch sind.

Betrachtet man nun eine ungleiche Verteilung (Tab. 3a), sieht man, daß sowohl das Entropie-Maß, als auch der Diversity-Index einen deutlich hohen Wert annehmen (Tab. 3c). Diese Werte verändern sich nicht für die Daten in Tab. 4a, welche die gleichen Häufigkeiten beinhaltet, aber in der ein negativer Zusammenhang zwischen X und Y besteht, d.h. unter X=2 befindet sich die schwächste Zellenbesetzung in der dritten Kategorie von Y. Sowohl in der Gruppe X=1 als auch in der Gruppe X=2 besteht eine Entropie von 0,81 bzw. eine Heterogenität von 0,84 (Tab. 4c). Hier wird deutlich, daß diese Indizes die inhaltliche Qualität einer Verteilung (welche Ausprägungen von Y sind stark oder schwach besetzt) nicht wiedergeben, sondern nur die numerische Größe der gruppeninternen Heterogenität einer Verteilung. Daher ist es wichtig, neben den Werten für die Indizes auch immer die Häufigkeitsverteilungen zu betrachten, um weitere inhaltliche Aussagen treffen zu können. Demgegenüber reagiert der Dissimilaritätsindex deutlich auf diese Veränderung. Während in dem Zahlenbeispiel der Tab. 3a 0% in eine andere Y-Kategorie eingeordnet werden müssen, sind es für Tab. 4a 40% einer Gruppe (entweder aus X=1 oder aus X=2), die neu geordnet werden müßten, um eine identische Verteilung für X=1 und X=2 zu erreichen. Der auf Gruppenvergleiche angelegte Dissimilaritätsindex verliert seine Anschaulichkeit, sobald die X-Variable mehr als drei Ausprägungen hat, da es dann zu zahlreichen Vergleichen kommt (bei 4 Kategorien kommt es zu 6 Vergleichen).

# 4 Berechnung der Maße mit SPSS

## 4.1 Anwendung des Macros qvarianz

Für die Berechnung der Heterogenitätsmaße haben wir ein Macro[8] mit dem Namen 'qvarianz' definiert. Bevor die in dem Macro enthaltenen Rechenoperationen in den Abschnitten 4.2 und 4.3 im einzelnen erläutert werden, soll in einem ersten Schritt die Anwendung des Macros sowie der durch das Macro erstellte SPSS-Output erklärt werden.

Nachdem der Arbeitsdatensatz in SPSS aufgerufen wurde, kann mit der Eingabe von

```
INCLUDE FILE = 'C:\QVARIANZ.INC'.
```

das Macro aktiviert werden. Der nächste Schritt besteht in dem Aufrufen des Macros QVARIANZ, dessen Syntax sich aus folgenden Komponenten zusammensetzt:

```
QVARIANZ  Y = VARY
        /X = VARX
        /PREDCAT = 2
        /DEPCAT = 3.
```

Mit den Subkommandos Y und X werden die Variablen VARY und VARX für die Kreuztabelle angegeben, wobei X als unabhängige Variable die Spaltenvariable bildet. PREDCAT gibt die Anzahl der Kategorien für die unabhängige Prediktor-Variable X an. So hat zum Beispiel die X-Variable in den besprochenen fiktiven Verteilungen in Abschnitt 3.2 zwei Kategorien. Durch DEPCAT (*Cat*egories of the *Dep*endent Variable) wird die Anzahl der Kategorien für die abhängige Y-Variable eingetragen.

Das Macro erstellt eine Kontingenztabelle mit den angegebenen Variablen und schreibt diese als Rohdatensatz in die Datei 'c:\crosstab.dat'. Anschließend bewirkt das Macro, daß dieser Datensatz wieder eingelesen wird, um dann das Entropiemaß und den Diversity-Index zu berechnen. Die Ergebnisse werden in dem Output-Fenster[9] mit Hilfe des im Macro enthaltenen LIST VARIABLES-Kommandos wie folgt angezeigt (die Heterogenitätsmaße beziehen sich auf Tab. 3a bis 3c):

---

8 Das Macro kann im Anhang dieses Artikels eingesehen werden. Es ist bei den Autoren des Artikels erhältlich. Wir weisen darauf hin, daß das Programm in seiner gegenwärtigen Fassung nur Variablen mit maximal 9 Kategorien verarbeiten kann und Signifikanzwerte zu den berechneten Indizes nicht ausgerechnet werden. Wir arbeiten an einer Verbesserung des Macros.

9 Für unbesetzte Zellen erhält man im Output eine Warnung (Warning#602), die darauf hinweist, daß diesen Zellen bei der Berechnung des Logarithmus ein System-Missing zugeordnet wurde. Diese Meldung braucht nicht beachtet zu werden, da die System-Missings durch das Macro zu Null recodiert werden, um einen fehlerlosen Ablauf des Programms zu ermöglichen.

**LIST**

| PREDICT | ENTRO | ENTROMAX | ENTROSTD |
|---|---|---|---|
| X1 | 1,29 | 1,58 | 0,81 |
| X2 | 1,29 | 1,58 | 0,81 |

**LIST**

| PREDICT | D | DMAX | DSTAND |
|---|---|---|---|
| X1 | 0,56 | 0,67 | 0,84 |
| X2 | 0,56 | 0,67 | 0,84 |

In der ersten Spalte PREDICT stehen die Ausprägungen der unabhängigen Gruppierungsvariablen X. ENTRO und D geben jeweils das unstandardisierte Maß der Entropie bzw. des Diversity-Indexes wieder. Die folgende Spalte ENTROMAX bzw. DMAX geben den maximalen Wert an, den die unstandardisierten Maße annehmen können. In der letzten Spalte ENTROSTD und DSTAND werden die standardisierten Werte des Entropiemaßes bzw. des Diversity-Index angegeben.

In den folgenden zwei Abschnitten werden die in dem Macro enthaltenen Arbeitsschritte erläutert.

### 4.2 Einrichtung des Arbeitsdatensatzes durch das Macro qvarianz[10]

Um die gewünschten Indizes berechnen zu können, müssen die erzeugten Kreuztabellen zunächst als Arbeitsdatensatz herausgeschrieben werden. Dieses ermöglicht die SPSS-Anweisung PROCEDURE OUTPUT OUTFILE= 'FILENAME', so daß die gesamte SPSS-Anweisung im Syntax-Fenster wie folgt zu schreiben ist:

```
PROCEDURE OUTPUT OUTFILE='C:\CROSSTAB.DAT'.
CROSSTABS VARIABLES=Y(1,3) X (1,2)
        /TABLES=Y BY X
        /FORMAT=AVALUE TABLES
        /CELLS=COUNT
        /WRITE=ALL.
```

10 Wir bedanken uns hier bei Dipl.-Soz. ***Jürgen Sensch*** und ***Ralf Ponemereo*** vom Zentralarchiv für die intensive und sehr hilfreiche Beratung zu Fragen der programmtechnischen Umsetzung.

Die auf CROSSTABS VARIABLES folgenden Angaben geben an, wie groß die Tabelle ist, nämlich 3 mal 2 Zellen. Das FORMAT-Subkommando bewirkt mit AVALUE, daß die Werte der Zeilen- und Spaltenvariablen im herausgeschriebenen Rohdatensatz in aufsteigender Reihenfolge wiedergegeben werden. Die Angabe TABLES im Subkommando FORMAT gibt die Kreuztabelle im SPSS-Outputfenster aus. Da die Prozentwerte nicht in den Rohdatensatz übernommen werden, steht in dem Subkommando CELLS nur die Anweisung COUNT für die Zellenhäufigkeiten, nicht aber COLUMN für die spaltenweise Prozentuierung. Die Anweisung WRITE=ALL dient der Übernahme der Zellenhäufigkeiten in die Rohdatenmatrix. In unserem Beispiel besteht der Rohdatensatz aus drei Spalten: Die erste Spalte enthält die Zellenhäufigkeiten, in der zweiten Spalte werden die Ausprägungen von Y in aufsteigender Reihenfolge wiedergegeben und die dritte Spalte beinhaltet die Ausprägungen von X ebenfalls in aufsteigender Reihenfolge. Jetzt muß die herausgeschriebene Tabelle wieder als Datensatz in SPSS eingelesen werden. Dies geschieht durch das DATA LIST Kommando, welches die Variablennamen der oben definierten Output-Datei 'CROSSTAB.DAT' und die in ihr enthaltene Datenstruktur beschreibt.

```
DATA LIST
        FILE='C:\CROSSTAB.DAT'  FIXED  RECORDS=1
        /1 anzahl 9-16 Y 17-24 X 25-32.
EXECUTE.
```

Da die Werte für die gleiche Variable immer in der gleichen Spalte eingetragen sind, kommt das Subkommando FIXED zur Anwendung. Das Subkommando RECORDS=1 zeigt an, daß für jeden Fall im Datensatz nur eine Zeile angelegt ist. In der folgenden Zeile werden nun die Variablennamen und die Positionen der Variablen angegeben. Durch sie erkennt das SPSS-Programm, daß in der ersten Zeile pro Fall die Werte der Variablen anzahl in den Spalten 9-16 stehen, die Werte der abhängigen Variablen Y in den Spalten 17-24 abgespeichert sind und die Werte der Gruppierungsvariablen X in den Spalten 25-32 zu finden sind. Eingeleitet werden diese Angaben mit einem „Slash (/)", dem dann die 1 für Zeile 1 der Fälle im Rohdatensatz folgt. In der Variable anzahl sind die Zellenhäufigkeiten der Kreuztabelle eingetragen, die im Datensatz in den Spalten 9 bis 16 zu finden sind. Der jeweils eingelesene Datenfile sieht in SPSS für die fiktiven Verteilungen in den obigen Tabellen 1a und 2a dann wie folgt aus:

Gleichverteilung

| anzahl | y | x |
|---|---|---|
| 50 | 1 | 1 |
| 50 | 2 | 1 |
| 50 | 3 | 1 |
| 50 | 1 | 2 |
| 50 | 2 | 2 |
| 50 | 3 | 2 |

maximale Konzentration

| anzahl | y | x |
|---|---|---|
| 150 | 1 | 1 |
| 0 | 2 | 1 |
| 0 | 3 | 1 |
| 0 | 1 | 2 |
| 0 | 2 | 2 |
| 150 | 3 | 2 |

Es wird anschließend pro Ausprägung von X eine Spalte mit dem Variablennamen XKON(n) eingerichtet, in die die absoluten Häufigkeiten von X für Y eingetragen werden. Das erreicht man mit dem Kommando DO REPEAT.

```
DO REPEAT KON=1 to 2 / XKON=XKON1 TO XKON2.
IF (X=KON) XKON=ANZAHL.
END REPEAT.
```

Mit KON=1 to 2 wird ein Stellvertreter eingeführt, der die Ausprägungen der Kontrollvariablen X aufführt. Die Anweisung XKON=XKON1 TO XKON2 richtet die Variablen xkon1 (für die absoluten Häufigkeiten der Kontrollvariable X=1) und xkon2 (für die absoluten Häufigkeiten der Kontrollvariable X=2) ein, in denen dann die Werte eingetragen werden, die durch den IF-Befehl zugewiesen werden. Der Datensatz hat nach dieser Operation folgende Struktur:

Gleichverteilung

| anzahl | y | x | xkon1 | xkon2 |
|---|---|---|---|---|
| 50 | 1 | 1 | 50 | |
| 50 | 2 | 1 | 50 | |
| 50 | 3 | 1 | 50 | |
| 50 | 1 | 2 | | 50 |
| 50 | 2 | 2 | | 50 |
| 50 | 3 | 2 | | 50 |

maximale Konzentration

| anzahl | y | x | xkon1 | xkon2 |
|---|---|---|---|---|
| 150 | 1 | 1 | 150 | |
| 0 | 2 | 1 | .00 | |
| 0 | 3 | 1 | .00 | |
| 0 | 1 | 2 | | .00 |
| 0 | 2 | 2 | | .00 |
| 150 | 3 | 2 | | 150 |

Der folgende Schritt besteht nun darin, die Werte von X, eingetragen in xkon1 und xkon2, so anzuordnen, daß mit SPSS die Prozentwerte sowie besprochenen Indizes für die Verteilung von X über die Kategorien von Y berechnet werden können. Y ist somit die Indikator-Variable, anhand der der Datensatz neu geordnet werden muß bzw. die Fälle im Datensatz gruppiert werden. Dies geschieht in SPSS mit dem AGGREGATE-Kommando. Zuvor müssen jedoch die Missing Values von xkon1 und xkon2 zu Null umkodiert werden, da diese für die gleichen Kategorien von Y aufsummiert werden. Vorhandene Missing Values würden in diesem Vorgang zu einem Ausfall aller Werte führen. Die Summen von xkon1 und xkon2 werden in neue Variablen eingetragen, die hier mit x1 und x2 benannt sind.

```
RECODE XKON1 TO XKON2 (SYSMIS=0).

AGGREGATE OUTFILE=*
        /BREAK=Y
        /X1 TO X2=SUM(XKON1 XKON2).
```

Der mit diesem Befehl neu gebildete Datensatz beinhaltet die Variablen Y, X1 mit den absoluten Werten von X=1 und X2 mit den absoluten Werten von X=2.

Gleichverteilung

| y | x1 | x2 |
|---|---|---|
| 1 | 50 | 50 |
| 2 | 50 | 50 |
| 3 | 50 | 50 |

maximale Konzentration

| y | x1 | x2 |
|---|---|---|
| 1 | 150 | .00 |
| 2 | .00 | .00 |
| 3 | .00 | 150 |

Da SPSS Werte nur zeilenweise berechnen kann, muß die Datenmatrix gedreht werden, um dann mit compute-Befehlen weiterarbeiten zu können. Das Drehen einer Datenmatrix um 90 Grad ist mit dem Kommando FLIP möglich und führt zu folgender Datenmatrix:

Gleichverteilung

| case_lbl | var001 | var002 | var003 |
|---|---|---|---|
| y | 1 | 2 | 3 |
| x1 | 50 | 50 | 50 |
| x2 | 50 | 50 | 50 |

maximale Konzentration

| case_lbl | var001 | var002 | var003 |
|---|---|---|---|
| y | 1 | 2 | 3 |
| x1 | 150 | .00 | .00 |
| x2 | .00 | .00 | 150 |

Zur Berechnung der Prozentanteile für die Variable X wird im folgenden Schritt die Summenvariable total mit Hilfe eines COMPUTE-Befehls gebildet:

```
COMPUTE TOTAL=SUM(VAR001 TO VAR003).
```

SPSS fügt die Variable total hinzu, in der die Spaltenhäufigkeiten von X=1 und X=2 enthalten sind. Die Variable total wird herangezogen, um die Anteilswerte für X zu berechnen. Dies geschieht durch folgenden Befehl:

```
VECTOR Y (3).
VECTOR VARY=VAR001 TO VAR003.
LOOP #i=1 TO 3.
COMPUTE Y(#i)=VARY(#i)/TOTAL.
END LOOP.
```

Mit VECTOR Y (3) werden die Variablen Y1 bis Y3 in der Datenmatrix eingerichtet, die zudem Elemente eines Vektors sind. Die Anweisung VECTOR VARY=VAR001 TO VAR003 richtet den Vektor vary mit den Elemente var001 bis var003 ein. Damit ist die Voraussetzung für das LOOP-Kommando geschaffen. Mit LOOP #i=1 TO 3 wird dem Stellvertreter #i der Wert 1 bis 3 zugeteilt. Das COMPUTE-Kommando bildet die Variablen Y1 bis Y3, indem die Ergebnisse aus der Division von var001 bis var003 durch total in den Variablen Y1 bis Y3 eingetragen werden.

Die Zeilenvariable Y beinhaltet nur die codierten Ausprägungen der Y-Variablen. Da SPSS für die Berechnung der Heterogenitätsmaße zeilenweise vorgeht, wird auch für diese erste Zeile ein Wert berechnet. Dieser Wert kann unbeachtet bleiben.

### 4.3 Umsetzung der Maße mit SPSS[11,12]

*Berechnung der Entropie nach **Shannon**:*
Nach der oben angegebenen Formel von ***Shannon*** müssen die Zellenanteile mit ihrem Logarithmus auf Basis 2 gewichtet werden. Da der Logarithmus für nichtbesetzte Kategorien nicht berechnet werden kann, setzt SPSS hier einen System-Missing-Wert ein.[13] System-Missings wird durch eine RECODE Anweisung der Wert Null zugeordnet, um einen fehlerlosen Ablauf des Programms zu ermöglichen. Weiterhin ist der Logarithmus aus der Anzahl der Kategorien von Y - hier sind es drei Kategorien - notwendig, um eine Standardisierung zwischen 0 und 1 zu erreichen. Die ersten Schritte sehen also folgendermaßen aus:

```
***LOGARITHMEN (auf Basis 2) DER ANTEILE VON X***.
DO REPEAT A=Y1 TO Y3 / B=LGY1 TO LGY3.
COMPUTE B=LN(A)/LN(2).
END REPEAT.
RECODE LGY1 TO LGY3 (SYSMIS=0).
EXECUTE.
***LOGARITHMUS AUS DER ANZAHL DER KATEGORIEN VON Y (=3 KATEGORIEN)***.
COMPUTE LGK3=LN(3)/LN(2).
EXECUTE.
```

In einem weiteren Schritt wird nun das unstandardisierte und das standardisierte Entropie-Maß mit folgenden COMPUTE-Befehlen berechnet:

```
***BERECHNUNG DES ENTROPIE-MASSES***.
COMPUTE ENTRO=(Y1*(-LGY1) )
        + (Y2*(-LGY2) )
        + (Y3*(-LGY3) ).
EXECUTE.
```

11 Dr. ***Christof Wolf*** vom Forschungsinstitut für Soziologie der Universität zu Köln war immer ein sehr kompetenter Ansprechpartner, der uns mit hilfreichen Hinweisen zur Seite stand.

12 Auf die Darstellung der Umsetzung des Dissimilaritätsindexes mit SPSS wird verzichtet, da man diesen Index gut per Hand ausrechnen kann.

13 In diesen Fällen erhält man im SPSS-Output eine Warnmeldung (Warning#602), die aber nicht beachtet werden muß.

```
***BERECHNUNG DES STANDARDISIERTEN ENTROPIE-MASSES***.
COMPUTE ENTROSTD=ENTRO/LGK3.
EXECUTE.
```

*Berechnung des Diversity-Index:*
Zur Berechnung des Diversity-Index müssen alle Zellenanteile quadriert, aufsummiert und von 1 subtrahiert werden. Entsprechend sieht die compute-Anweisung folgendermaßen aus:

```
***BERECHNUNG DES UNSTANDARDISIERTEN DIVERSITY-INDEX***.
COMPUTE D=1 - ( (Y1*Y1) + (Y2*Y2) + (Y3*Y3) ).

***OBERGRENZE DES DIVERSITY-INDEX***.
COMPUTE DMAX=( 1 - (1/3) ).

***BERECHNUNG DES STANDARDISIERTEN DIVERSITY-INDEX***.
COMPUTE DSTAND=D/DMAX.
```

## 5 Ein Anwendungsbeispiel

Abschließend soll die Anwendung der Heterogenitätsmaße im Rahmen unserer Ausgangsfragestellung erläutert werden. Ziel ist es also in diesem Abschnitt zu prüfen, ob eine Pluralisierung der Lebensformen gemessen werden kann. Mit Hilfe der im kumulierten ALLBUS 1980-1996[14] erfaßten Haushaltsfeinklassifikation bildeten wir eine Haushaltsvariable mit acht Kategorien, welche sich grob in die drei Subgruppen Haushalte ohne Kinder, Haushalte mit Kinder und Mehrgenerationenhaushalte einordnen lassen. Die Kategorien sind folgende: ‘Ehepaare ohne Kinder’, ‘nichteheliche Lebensgemeinschaften ohne Kinder’, ‘Ein-Personen-Haushalte’, ‘Ehepaare mit Kinder’, ‘nichteheliche Lebensgemeinschaften mit Kinder’, ‘Alleinerziehende’, ‘Haushalte mit nicht mehr ledigen Kindern’ und schließlich ‘Drei- und Mehrgenerationenhaushalte’.

Im Rahmen dieses Beitrags wird beispielhaft das Merkmal Konfessionszugehörigkeit gewählt, um zu testen, inwieweit sich die Heterogenität der Haushaltsformen zwischen der Gruppe von Personen, die einer Konfession angehören, von der Gruppe der konfessionslosen Personen unterscheidet. Es wird vermutet, daß sich mit der Zugehörigkeit zu einer Konfessionsgruppe bestimmte Wertvorstellungen bezüglich der Wahl einer Lebensform verbinden, die die Haushaltsformen der Ehepaare ohne oder mit Kindern bevorzugen.

---

14 Von 1980 bis 1992 wurde die Stichprobe des ALLBUS mit Hilfe des ADM-Stichprobendesigns als Haushaltsstichprobe erhoben. Seit 1994 wird der ALLBUS als Personenstichprobe erhoben. Der Veränderung des Erhebungsdesigns trägt das Transformationsgewicht insbesondere dann Rechnung, wenn die Haushaltsvariable in die Datenanalyse einbezogen wird. Die bis 1992 erhobenen Daten auf Haushaltsebene wurden mit Hilfe des Transformationsgewichts so gewichtet, daß sie sich wie die Datensätze von 1994 und 1996 auf die Individualebene beziehen.

Trifft das zu, dann weist die Gruppe der Personen mit einer Konfessionszugehörigkeit eine geringere Heterogenität auf als die Vergleichsgruppe der konfessionslosen Personen. Dabei werden drei verschiedene Verteilungen betrachtet. Zunächst betrachten wir die Verteilung über alle 8 Haushaltskategorien, anschließend werden die Heterogenitätsmaße für die Haushalts-Subgruppen ‘Haushalte ohne Kinder’ und ‘Haushalte mit Kinder’ getrennt berechnet.

**Tab.5):** Haushalte nach der Zugehörigkeit zu einer Konfession[15]

| Haushalte | | Einer Konfession angehörig | konfessionslos | Summe |
|---|---|---|---|---|
| Haushalte ohne Kinder (eine Generation) | Ehepaare | 5 847 (25,7%) | 593 (26,4%) | 6 440 |
| | nichteheliche Lebensgemeinschaften | 625 (2,7%) | 172 (7,7%) | 797 |
| | Ein-Personen-Haushalte | 2 973 (13,1%) | 439 (19,6%) | 3 412 |
| Haushalte mit Kindern (zwei Generationen) | Ehepaare | 10 576 (46,4%) | 830 (37,0%) | 11 406 |
| | nichteheliche Lebensgemeinschaften | 204 (0,9%) | 37 (1,6%) | 241 |
| | Alleinerziehende | 1 007 (4,4%) | 107 (4,8%) | 1 114 |
| | HH mit nicht mehr led. Kindern | 450 (2,0%) | 24 (1,1%) | 474 |
| Mehrgeneration | Drei- und Mehrgenerationenhaushalte | 1 090 (4,8%) | 42 (1,9%) | 1 132 |
| Summe | | 22 772 (100%) | 2 244 (100%) | 25 016 |

Quelle: Kumulierter ALLBUS 1980-1996; eigene Berechnungen mit gewichteten Daten

**Tab. 6):** Haushalte ohne Kinder nach der Zugehörigkeit zu einer Konfession

| Haushalte | | Einer Konfession angehörig | konfessionslos | Summe |
|---|---|---|---|---|
| Haushalte ohne Kinder (eine Generation) | Ehepaare | 5 847 (61,9%) | 593 (49,3%) | 6 440 |
| | nichteheliche Lebensgemeinschaften | 625 (6,6%) | 172 (14,3%) | 797 |
| | Ein-Personen-Haushalte | 2 973 (31,5%) | 439 (36,5%) | 3 412 |
| Summe | | 9 445 (100%) | 1 204 (100%) | 10 649 |

Quelle: Kumulierter ALLBUS 1980-1996; eigene Berechnungen mit gewichteten Daten

**Tab. 7):** Haushalte mit Kindern nach der Zugehörigkeit zu einer Konfession

| Haushalte | | Einer Konfession angehörig | konfessionslos | Summe |
|---|---|---|---|---|
| Haushalte mit Kindern (zwei Generationen) | Ehepaare | 10 576 (86,4%) | 830 (83,2%) | 11 406 |
| | nichteheliche Lebensgemeinschaften | 204 (1,7%) | 37 (3,7%) | 241 |
| | Alleinerziehende | 1 007 (8,2%) | 107 (10,7%) | 1 114 |
| | HH mit nicht mehr led. Kindern | 450 (3,7%) | 24 (2,4%) | 474 |
| Summe | | 12 237 (100%) | 998 (100%) | 13 235 |

Quelle: Kumulierter ALLBUS 1980-1996; eigene Berechnungen mit gewichteten Daten

15 In die Berechnung wurden nur Personen aus den alten Bundesländern und mit deutscher Staatsangehörigkeit mit einbezogen.

**Tab. 8):** Heterogenitätsindizes zu den Verteilungen in den Tabellen 5 bis 7[16]

| Haushalte | Einer Konfession angehörig | | konfessionslos | | Gruppenvergleich |
|---|---|---|---|---|---|
| | Entropie | Diversity | Entropie | Diversity | Dissimilarity |
| Haushalte insgesamt [1] | 0.71 | 0.80 | 0.76 | 0.85 | 13,3 % |
| Haushalte ohne Kinder [2] | 0.77 | 0.77 | 0.91 | 0.91 | 12,5 % |
| Haushalte mit Kindern [3] | 0.38 | 0.33 | 0.44 | 0.39 | 5,5 % |

1 Alle acht Haushaltsformen werden in die Berechnung einbezogen (Tab. 5).
2 Die Basis der Prozentuierung - und damit die Berechnung der Indizes - bezieht sich nur auf die Eingenerationenhaushalte (Tab. 6).
3 Die Basis der Prozentuierung - und damit die Berechnung der Indizes - bezieht sich nur auf die Zweigenerationenhaushalte (Tab. 7).

Aus den prozentualen Verteilungen (Tab.5) ist zu erkennen, daß konfessionslose Personen eher in Haushaltsformen ohne Kinder sowie in nichtehelichen Lebensgemeinschaften mit Kindern anzutreffen sind. Der traditionelle Familienhaushalt (Ehepaare mit Kind) und Mehrgenerationenhaushalte werden vorzugsweise von Personen gewählt, die einer Konfession angehören. Die Entropiemaße und Diversity-Indizes (Tab. 8) für die Haushalte insgesamt zeigen eine höhere Heterogenität für die Personengruppe ohne Konfessionszugehörigkeit an. Betrachtet man die Untergruppe der Haushalte ohne Kinder gesondert, wird die Differenz hinsichtlich der Pluralität noch deutlicher. Die Gruppe der konfessionslosen Personen hat eine sehr hohe Entropie von 0.91, während die Personen, die einer Konfession angehören, eine Entropie von 0.77 aufweisen. Die Werte des Diversity-Index stimmen mit denen des Entropie-Maßes überein. Haushaltsformen mit Kindern zeigen generell eine geringere Pluralität im Vergleich zu den Haushalten ohne Kinder. Aber auch hier wirkt sich die Zugehörigkeit zu einer Kirchengemeinschaft homogenisierend aus. Zieht man den Dissimilaritäts-Index heran, wird die Tatsache unterstrichen, daß sich die Personengruppen in bezug auf die Eingenerationenhaushalte sehr viel stärker unterscheiden als hinsichtlich der Zweigenerationenhaushalte. Während bei den Eingenerationenhaushalten 12,5% einer Gruppe eine andere Haushaltsform wählen müßten, um eine identische Verteilung zwischen konfessionslosen und einer Kirche angehörigen Personen zu erreichen, sind es bei den Zweigenerationenhaushalten nur 5,5%. Die Konfession hängt mit der Pluralität der Haushaltsformen zusammen, was bei den Eingenerationenhaushalten sehr deutlich wird. Ob die Konfessionslosigkeit deshalb mit einer höheren Pluralität der Lebensformen verknüpft ist, weil sie mit einer normativen Heterogenität einhergeht, kann nur vermutet, aber hier nicht empirisch belegt werden. Die Pluralität der Lebensformen kann möglicherweise auch durch andere Faktoren erklärt werden, die mit der Konfessionslosigkeit assoziiert sind.

16 Wir haben die gleichen Berechnungen mit den ungewichteten Daten des kumulierten ALLBUS durchgeführt. Auch hier sind wir zu dem Ergebnis gekommen, daß bei den Haushalten, die von der Personengruppe ohne Religionszugehörigkeit gebildet werden, eine höhere Heterogenität vorliegt.

## 6 Diskussion

Ziel dieses Beitrags war es, Heterogenitätsmaße vorzustellen und ihre Eigenschaften anhand eines Beispiels aus der Familiensoziologie zu diskutieren. Diesen Maßen liegen unterschiedliche mathematische Konzepte zugrunde, was sich auf ihre Meßeigenschaften auswirkt. So berücksichtigt das Entropiemaß stärker die schwach besetzten Merkmalskategorien, während der Diversity-Index als Konzentrationsmaß mehr Gewicht auf die stark besetzten Kategorien legt. Beides sind Maße für die Intragruppenheterogenität. Mit dem Dissimilaritäts-Index vergleicht man hingegen die Heterogenität zweier Gruppen bezüglich eines Merkmals. Die Wahl eines Heterogenitätsmaßes sollte daher theoretisch begründet werden. Ferner ist es für eine gehaltvolle Interpretation von Heterogenitätsmaßen unabdingbar, Häufigkeitsverteilungen in die Betrachtung einzubeziehen.

Inwieweit Heterogenitätsmaße in Kausal- oder Regressionsmodellen zur Anwendung kommen, war nicht Gegenstand dieses Beitrags. Wir wollen lediglich darauf hinweisen, daß diese Maße Aussagen über ein Aggregat treffen, so daß sie auch in Mehrebenen-Analysen berücksichtigt werden können.

In statistischen Programmpaketen sind die Heterogenitätsmaße, so wie sie in unserem Beitrag besprochen wurden, unserer Kenntnis nach nicht implementiert. Mit dem Entropie- und dem Konzentrationsmaß, welche von SPSS im Rahmen der Analyse kategorialer Daten mit der Prozedur GENLOG berechnet werden, können gemäß des Varianzzerlegungsansatzes von ***Magidson*** (1981) und ***Theil*** (1970) Aussagen über den Anteil erklärter Varianz in einem statistischen Modell getroffen werden. Die Heterogenität innerhalb von Gruppen kann auf diese Art jedoch nicht bestimmt werden. Gerade darum ging es uns aber in diesem Beitrag.

Die Messung der Ungleichheit bei qualitativen Variablen sollte in der Soziologie häufiger vorgenommen werden, zumal sich viele Konzepte in der Soziologie - wie zum Beispiel die Pluralisierung der Lebensformen - durch derartige statistische Maßzahlen präzise ausdrükken lassen.

### Anhang: SPSS-Anweisungen im Macro qvarianz

```
define qvarianz     (y =          !charend ( '/' )
                    /x =          !charend ( '/' )
                    /predcat =    !charend ( '/' )
                    /depcat =     !charend ( '/' ) ).

    procedure output outfile='c:\crosstab.dat'.

    Crosstabs variables=!y (1, !depcat)  !x (1, !predcat)
                    /tables=!y by !x
                    /format=avalue tables
                    /cells=count
                    /write=all.

    Data list file='c:\crosstab.dat' fixed records=1 / 1 anzahl 9-16 y 17-24 x 25-32.
    exec.
```

```
    do repeat kon=1 to !predcat/ xkon=xkon1 to !concat (xkon, !predcat).
    if (x=kon) xkon=anzahl.
    end repeat.
    exec.

    recode xkon1 to !concat (xkon, !predcat) (sysmis=0).

    aggregate outfile=*
    /break=y
    /x1 to !concat (x, !predcat)=sum(xkon1 to !concat (xkon, !predcat)).
    flip.

    compute total=sum(var001 to !concat(var00, !depcat)).
    ++vector y (!depcat).
    ++vector vary=var001 to !concat(var00, !depcat).
    loop #i=1 to !depcat.
    ++compute y(#i)=vary(#i)/total.
    end loop.
    exec.

    formats y1 to !concat (y, !depcat) (f8.4).
    rename variables case_lbl=predict.

    do repeat a=y1 to !concat (y, !depcat)/ b=lgy1 to !concat (lgy, !depcat).
    compute b=ln(a)/ln(2).
    end repeat.
    recode lgy1 to !concat (lgy, !depcat) (sysmis=0).
    exec.

    compute entromax=ln(!depcat)/ln(2).
    exec.

    ++vector p (!depcat).
    ++vector yv=y1 to !concat(y, !depcat)
            /lgv=lgy1 to !concat(lgy, !depcat).
    loop #i=1 to !depcat.
    ++compute p(#i) = yv(#i) * (-lgv(#i)).
    end loop.
    exec.
    compute entro=sum(p1 to !concat(p, !depcat)).
    exec.
    compute entrostd=entro/entromax.

    compute dmax=(1-(1/!depcat)).
    exec.
    ++vector square (!depcat).
    ++vector yv=y1 to !concat(y, !depcat).
    loop #i=1 to !depcat.
    ++compute square(#i)=(yv(#i)*yv(#i)).
    end loop.
    exec.
    compute d=1-(sum(square1 to !concat (square, !depcat))).
    exec.
    compute dstand=d/dmax.
    exec.


    list var predict entro entromax entrostd/ cases from 2.
    list var predict d dmax dstand/ cases from 2.
    exec.

!enddefine.
```

## Literatur

***Agresti, Alan/Barbara F. Agresti*** (1978): Statistical Analysis of Qualitative Variation. In: ***Karl F. Schuessler*** (ed.): Sociological Methodology 1978. London: Jossey-Bass Publishers, 204-237.

***Beck, Ulrich*** (1989): Risikogesellschaft. Auf dem Weg in eine andere Moderne. Frankfurt/Main: Suhrkamp.

***Coulter, Philip B.*** (1989): Measuring Inequality: A Methodological Handbook. London: Westview Press.

***Duncan, Dudley/Beverly Duncan*** (1955a): A Methodological Analysis of Segregation Indexes. In: American Sociological Review 20, 210-217.

***Duncan, Dudley/Beverly Duncan*** (1955b): Residential Distribution and Occupational Stratification. In: American Sociological Review 60, 493-505.

***Friedrichs, Jürgen*** (1983): Stadtanalyse. Soziale und räumliche Organisation der Gesellschaft. Opladen: Westdeutscher Verlag.

***Friedrichs, Jürgen*** (1998): Die Individualisierungs-These. Eine Explikation im Rahmen der Rational-Choice-Theorie. In: ***Jürgen Friedrichs*** (Hrsg.): Die Individualisierungs-These. Opladen: Leske und Budrich, 33-47.

***Gibbs, Jack/Walter T. Martin*** (1962): Urbanization, Technology, and the Division of Labor: International Patterns. In: American Sociological Review 27, 667-677.

***Herfindahl, Orris C.*** (1950): Concentration in the Steel Industry. Columbia University: Diss.

***Huinink, Johannes/Michael Wagner*** (1998): Individualisierung und die Pluralisierung von Lebensformen. In: ***Jürgen Friedrichs*** (Hrsg.): Die Individualisierungs-These. Opladen: Leske und Budrich, 85-106.

***Jagodzinski, Wolfgang/Markus Quandt*** (1997): Wahlverhalten und Religion im Lichte der Individualisierungsthese. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 49, 761-782.

***Kaufmann, Franz-Xaver*** (1988): Familie und Modernität. In: ***Kurt Lüscher*** et al.(Hrsg.): Die postmoderne Familie. Familiale Strategien und Familienpolitik in einer Übergangszeit. Konstanz: Universitätsverlag Konstanz, 391-415.

***Kaufmann, Franz-Xaver*** (1995): Zukunft der Familie im vereinten Deutschland. Gesellschaftliche und politische Bedingungen. München: Beck.

***Lieberson, Stanley*** (1969): Measuring Population Diversity. In: American Sociological Review 34, 850-862.

***Magidson, Jay*** (1981): Qualitative Variance, Entropy, and Correlation Ratios for Nominal Dependent Variables. In: Social Science Research 10, 177-194.

***Nave-Herz, Rosemarie*** (1997): Pluralisierung familialer Lebensformen - ein Konstrukt der Wissenschaft? In: ***Laszlo A. Vaskovics*** (Hrsg.), Familienleitbilder und Familienwelten. Opladen: Leske und Budrich, 36-49.

***Peet, Robert K.*** (1974): The Measurement of Species Diversity. In: ***Richard F. Johnston*** et al. (eds.): Annual Review of Ecology and Systematics, Vol. 5. Palo Alto, California, 285-307.

***Shannon, Claude E./Warren Weaver*** (1959): The Mathematical Theory of Communication. Urbana: University of Illinois Press.

***Theil, Henri*** (1969): The Desired Political Entropy. In: The American Political Science Review 63: 521-525.

***Theil, Henri*** (1970): On the Estimation of Relationships Involving Qualitative Variables. In: American Journal of Sociology 76, 103-154.

***Tyrell, Hartmann*** (1979): Familie und gesellschaftliche Differenzierung. In: ***Helge Pross*** (Hrsg.): Familie - wohin? Leistungen, Leistungsdefizite und Leistungswandlungen der Familie in hochindustrialisierten Gesellschaften. Reinbek bei Hamburg: Rowohlt, 13-77.

***Wehrspaun, Michael*** (1988): Alternative Lebensformen und postmoderne Identitätskonstruktion. In: ***Kurt Lüscher*** et al. (Hrsg.): Die postmoderne Familie. Familiale Strategien und Familienpolitik in einer Übergangszeit. Konstanz: Universitätsverlag Konstanz, 157-168.

***Zapf, Wolfgang*** et al. (1987): Individualisierung und Sicherheit. Untersuchungen zur Lebensqualität in der Bundesrepublik Deutschland. München: Beck.