Friedrich-Alexander-Universität Erlangen-Nürnberg

Wirtschafts-und Sozialwissenschaftliche Fakultät

Diskussionspapier
28 / 1999

Rangordnungsstatistiken als Verteilungsmaßzahlen für ordinalskalierte Merkmale

I. Schiefemessung

Ingo Klein

Lehrstuhl für Statistik und Ökonometrie
Lehrstuhl für Statistik und empirische Wirtschaftsforschung
Lange Gasse 20 · D-90403 Nürnberg

RANGORDNUNGSSTATISTIKEN ALS VERTEILUNGSMASSZAHLEN FÜR ORDINALSKALIERTE MERKMALE

II. Schiefemessung

Ingo Klein

Lehrstuhl für Statistik und Ökonometrie

Universität Erlangen–Nürnberg

Lange Gasse 20

D–90403 Nürnberg

Germany

E-mail: ingo.klein@wiso.uni-erlangen.de

Abstract

We introduce several measures of skewness for ordinal variables. These measures are based on an axiomatic approach for the phenomen of skewness and the functional form stems from a rank–statistic with special scores–functions. We show the favorite properties of one measure of skewness that can be interpreted by exchanging frequencies to get extremely skewed distributions. If the maximum or the minimum of the ordinal variable have no positive frequency it is shown that these extreme values can influence the amount of skewness heavily.

Keywords: descriptive statistics, skewness, ordinal variables, rank statistics

1 Einleitung

Schiefemaße werden in der Literatur ebenso wie Streuungsmaße zumeist für quantitative Merkmale angegeben. Eine Ausnahme bildet die Arbeit von Klein (1999a), in der auf der Basis von Häufigkeitsdifferenzen ein Schiefemaß auch für ordinalskalierte Merkmale eingeführt wird.

Wir wollen im folgenden analog zu dem Vorgehen von Klein (1999b) für Streuungsmaße zeigen, wie sich mittels linearer Rangordnungsstatistiken eine Klasse von Schiefemaßen konstruieren läßt, deren Elemente sich nur durch die Wahl der jeweiligen Scoresfunktionen unterscheiden. Für das von Klein (1999a) betrachtete Schiefemaß wird eine anschauliche Interpretation mittels der aus Klein (1999b) bekannten Operation der Häufigkeitsumschichtung vorgenommen. [1]

2 Formaler Rahmen

Der formale Rahmen entspricht dem in Klein (1999b) eingeführten und soll kurz wiederholt werden.

Es seien E eine endliche Grundgesamtheit des Umfangs n und U ein ordinalskaliertes Merkmal mit endlich vielen geordneten reellwertigen Ausprägungen $u_i < u_{i+1}$ für $i \in \mathbb{N}$. Es wird davon ausgegangen, daß die Definition des Merkmals und die Festlegung der Merkmalsausprägungen im Vorfeld statistischer Erhebungen festgelegt werden. So wird zum Beispiel entschieden, daß die Links-Selbsteinschätzung auf einer Zehnpunkteskala erfolgt, ohne daß zunächst klar, ob alle Skalenwerte in einer konkreten Erhebung auch angekreuzt werden. Es soll deshalb im folgenden von "möglichen" Merkmalsausprägungen gesprochen werden, die datenunabhängig festgelegt sind und deren Festlegung sehr wohl das Ausmaß der Schiefe einer Häufigkeitsverteilung beeinflussen kann.

n_i bezeichne die absolute Häufigkeit mit der die i–te Merkmalsausprägung in der Grundgesamtheit E vorkommt für $i \in \mathbb{N}$. Es sei $n = \sum_i n_i$. Dann sind $f_i = n_i/n$ die relativen Häufigkeiten, $R_i = \sum_{j \leq i} n_j$ die kumulierten absoluten und $F_i = R_i/n$ die kumulierten

[1]Für Korrekturen und Hinweise danke ich Herrn Diplom–Mathematiker Hans Kiesl, Lehrstuhl für Statistik, Universität Bamberg.

relativen Häufigkeiten für $i \in \mathbb{N}$. Ist $k \in \mathbb{N}$ die Anzahl der möglichen Merkmalsausprägungen, so sind $R_k = n$ und $F_k = 1$.

In der nichtparametrischen Statistik wird

$$\sum_{i=1}^{k} c_i a^*(R_i) = \sum_{i=1}^{k} c_i a(F_i)$$

mit $a(F_i) = a^*(R_i)$, $i = 1, 2, \ldots, k$ als lineare Rangordnungsstatistik bezeichnet. $c_1, \ldots, c_k$ heißen Regressionskoeffizienten und $a^*(.)$ bzw. $a(.)$ Scorefunktion (vgl. z.B. Hajek & Sidak (1967), S. 57ff.). Wir werden im folgenden nicht–lineare Rangordnungsstatistiken der Form

$$\sum_{i=1}^{k} c_i a(F_i, F_{k-i})$$

betrachten, wobei die Scorefunktion $a(.)$ von zwei Argumenten abhängt.

3 Schiefemessung für ordinalskalierte Merkmale

Die elementaren Anforderungen an ein Schiefemaß betreffen zunächst die Extrempunkte, wann keine und wann eine minimale oder maximale Schiefe vorliegen soll. Diese Situationen werden in Klein (1999a) für den Fall der Rechtsschiefe (Linkssteilheit) diskutiert.

3.1 Symmetrie

Keine Schiefe liegt offenbar dann vor, wenn die Verteilung symmetrisch ist, d.h. es gelten

$$f_i = f_{k+1-i} \text{ bzw. } F_i = 1 - F_{k-i}$$

für $i = 1, 2, \ldots, [k/2]$, wobei $[.]$ den ganzzahligen Anteil des Arguments bezeichnet.

3.2 Extreme Schiefe bei fester Anzahl der Merkmalsausprägungen

Sei $k \in \mathbb{N}$ die Anzahl möglicher Merkmalsausprägungen, dann tritt der Fall der maximalen Rechtsschiefe auf, wenn die kleinste Merkmalsausprägungen u_1 die gesamte Häufigkeitsmasse auf sich vereint, d.h. es gilt $F_i = 1$ für $i = 1, 2, \ldots, k$. Wir bezeichnen das zugehörige k–dimensionale Tupel kumulierter relativer Häufigkeiten als $\iota_{\mathbf{1}} = (1, 1, \ldots, 1, 1)$.

Umgekehrt liegt eine minimale Rechtsschiefe (d.h. maximale Linksschiefe) vor, wenn lediglich die maximale Ausprägung u_k eine positive Häufigkeit besitzt. Das zugehörige k–dimensionale Tupel kumulierter relativer Häufigkeiten wird mit $\iota_{\mathbf{k}} = (0, 0, \ldots, 0, 1)$ bezeichnet.

3.3 Funktionale Abhängigkeit

Auch Schiefemaße können mit derselben Argumentation wie für Streuungsmaße nur von dem Vektor der kumulierten relativen Häufigkeiten funktional abhängen, da das Schiefephänomen von Translationen nicht berührt wird. Damit greift für ordinalskalierte Merkmale die Argumentation aus Klein (1994): Wenn eine bezüglich aller streng monoton zunehmenden Transformationen vergleichsinvariante Maßzahl absolutinvariant bezüglich einer streng monoton zunehmenden Transformation ist, dann ist sie auch absolutinvariant bezüglich aller streng monoton zunehmenden Transformationen und damit nur funktional abhängig von kumulierten Häufigkeiten (siehe auch Klein (1999a), S. 6f.).

3.4 Partialordnung der Schiefe bei endlicher Anzahl von Ausprägungen

Ausgangspunkt für jede Schiefemessung ist die Definition einer geeigneten Ordnung von Verteilungen unterschiedlicher Schiefe. Wir betrachten dazu die Menge $\mathcal{F}_k$ aller k–dimensionalen Tupel kumulierter relativer Häufigkeiten. Es liegt nun nahe, zwei k–dimensionale Vektoren kumulierter relativer Häufigkeiten $\mathbf{F} = (F_1, \ldots, F_{k-1}, 1)$ und $\mathbf{F}' = (F'_1, \ldots, F'_{k-1}, 1)$ entsprechend der folgenden Relation zu ordnen: $\mathbf{F}$ ist nicht rechtsschiefer als $\mathbf{F}'$ (kurz: $\mathbf{F} \preceq_{SO} \mathbf{F}'$), wenn

$$F_i - (1 - F_{k-i}) \leq F'_i - (1 - F'_{k-i})$$

für $i = 1, 2, \ldots, k-1$ gilt. Diese Bedingung läßt sich alternativ als

$$H_i = (F_i + F_{k-i})/2 \leq H'_i = (F'_i + F'_{k-i})/2$$

für $i = 1, 2, \ldots, k-1$ formulieren. Es handelt sich lediglich um eine Partialordnung auf $\mathcal{F}_k$, da nicht alle Verteilungen entsprechend $\preceq_{SO}$ geordnet werden können.

Von Schiefemaßen SO für ordinalskalierte Merkmale ist zumindest zu verlangen, daß sie diese sehr spezielle Ordnung erhalten.

Betrachtet man als Referenz eine symmetrische Verteilung $\mathbf{F^s} = (F_1^s, \ldots, F_{k-1}^s, 1)$, für die offensichtlich $F_i^s - (1 - F_{k-i}^s) = 0$ für $i = 1, 2, \ldots, k-1$ gelten muß, dann ist eine Verteilung $\mathbf{F} \in \mathcal{F}_k$ nicht rechtsschief, wenn $\mathbf{F} \preceq_{SO} \mathbf{F^s}$ ist. Analog folgt, daß $\mathbf{F}$ nicht linksschief ist, wenn $\mathbf{F^s} \preceq_{SO} \mathbf{F}$ ist.

3.5 Spiegelung der Verteilung

Permutationen der Komponenten des Vektors $\mathbf{f} = (f_1, \ldots, f_k)$ der relativen Häufigkeiten werden im allgemeinen die Schiefe der Verteilung beeinflussen. Dies gilt auch, wenn die Spiegelung

$$\overline{\mathbf{F}} = (1 - F_{k-1}, \ldots, 1 - F_1, 1)$$

von $\mathbf{F} = (F_1, \ldots, F_{k-1}, 1)$ betrachtet wird. Ist $\mathbf{F}$ rechtsschief, so muß $\overline{\mathbf{F}}$ linksschief sein und umgekehrt. Von einem Schiefemaß ist deshalb zu verlangen, daß die Spiegelung zwar absolutbetragsmäßig den Wert eines Schiefemaßes nicht beeinflußt aber sehr wohl das Vorzeichen.

3.6 Ein allgemeines Konstruktionsprinzip

Es ist zunächst zu klären, was unter einem (additiven) Schiefemaß für ordinalskalierte Merkmale zu verstehen ist.

Definition 3.1 *Seien $\mathcal{F}_k$ die Menge aller Vektoren von kumulierten relativen Häufigkeiten für $k \in \mathbb{N}$ Merkmalsausprägungen. $SO : \mathcal{F}_k \to \mathbb{R}$ heißt Schiefemaß, wenn*

1. *$SO(\mathbf{F}) = 0$, falls $\mathbf{F} \in \mathcal{F}_k$ symmetrisch ist,*
2. *$\mathbf{F} \preceq_{SO} \mathbf{F}' \Longrightarrow SO(\mathbf{F}) \leq SO(\mathbf{F}')$ für $\mathbf{F}, \mathbf{F}' \in \mathcal{F}_k$,*
3. *$SO(F_1, \ldots, F_{k-1}, 1) = -SO(1 - F_{k-1}, \ldots, 1 - F_1, 1)$ für $\mathbf{F} = (F_1, \ldots, F_k) \in \mathcal{F}_k$*

gelten.

Als unmittelbare Konsequenz dieser Definition ergibt sich, daß

$$SO(\iota_{\mathbf{k}}) \leq SO(\mathbf{F}) \leq SO(\iota_{\mathbf{1}})$$

für $F \in \mathcal{F}_k$ ist.

Wir betrachten hier ausschließlich additive Maßzahlen der Form

$$SO(\mathbf{F}) = \sum_{i=1}^{k-1} b(H_i)$$

mit einer sog. Scorefunktion $b(.)$. $SO(\mathbf{F})$ ist offenbar ein Spezialfall der eingangs eingeführten Rangordnungsstatistik $\sum_{i=1}^{k} c_i a(F_i, F_{k-i})$.

Im weiteren sollen für die Scorefunktion $b(.)$ die folgenden Annahmen getroffen werden:

(1) $b : [0,1] \rightarrow \mathbb{R}$,

(2) b ist stetig und beschränkt im Intervall $[0,1]$,

(3) b ist antisymmetrisch, d.h. $b(p) = -b(1-p)$ für $p \in [0,1]$,

(4) b ist streng monoton zunehmend auf dem Intervall $[0,1]$.

Insbesondere ist dann $b(1/2) = 0$. b nimmt für $u = 0$ das Minimum und für $u = 1$ das Maximum an.

Ein Beispiel, wie $b(.)$ gewählt werden kann, liefert das in Klein (1999a) diskutierte Schiefemaß

$$SO_{1,2,\ldots,k-1} = -(k-1) + 2\sum_{i=1}^{k-1} F_i.$$

Setze $b(p) = 2(p - (k-1)/2)$ für $p \in [0,1]$.

Wenn $b(.)$ die Eigenschaften (1) bis (4) besitzt, dann kann gezeigt werden, daß die zugehörige additive Maßzahl SO ein Schiefemaß ist.

Theorem 3.1 *Sei $SO : \mathcal{F}_k \rightarrow \mathbb{R}$ eine (additive) Maßzahl mit Scorefunktion $b(.)$, die die Eigenschaften (1) bis (4) erfüllt, dann ist SO ein Schiefemaß für ordinalskalierte Merkmale mit k Merkmalsausprägungen.*

Beweis:

1. Sei $\mathbf{F} = (F_1, \ldots, F_k) \in \mathcal{F}_k$ symmetrisch. Dann ist $F_i = 1 - F_{k-i}$ und damit $H_i = 1/2$, so daß $b(H_i) = 0$ für $i = 1, 2, \ldots, k$ ist. Damit ist auch $SO(\mathbf{F}) = 0$.

2. Seien $\mathbf{F}, \mathbf{F}' \in \mathcal{F}_k$ mit $\mathbf{F} \preceq_{SO} \mathbf{F}'$, dann ist $H_i \leq H_i'$ und wegen der strengen Monotonie von $b(.)$ auch $b(H_i) \leq b(H_i')$ für $i = 1, 2, \ldots, k$. Somit folgt $SO(\mathbf{F}) \leq SO(\mathbf{F}')$.

3. Sei $\mathbf{F} = (F_1, \ldots, F_{k-1}, 1) \in \mathcal{F}_k$ und

$$\overline{\mathbf{F}} = (\overline{F}_1, \ldots, \overline{F}_{k-1}, 1) = (1 - F_{k-1}, \ldots, 1 - F_1, 1).$$

Dann ist

$$(\overline{F}_i + \overline{F}_{k-i})/2 = 1 - (F_i + F_{k-i})/2$$

für $i = 1, 2, \ldots, k-1$. Wegen der Antisymmetrie folgt dann

$$SO(\overline{\mathbf{F}}) = \sum_{i=1}^{k-1} b(1 - (F_i + F_{k-i})/2) = -\sum_{i=1}^{k-1} b((F_i + F_{k-i})/2) = -SO(\mathbf{F}).$$

□

Es wurde bereits erwähnt, daß $SO(\iota_{\mathbf{k}}) \leq SO(\mathbf{F}) \leq SO(\iota_{\mathbf{1}})$ ist. Die Unter– und Obergrenze lassen sich für die betrachteten additiven Schiefemaßzahlen mit Scorefunktion $b(.)$ angeben. Es ist

$$(k-1)b(0) \leq SO(\mathbf{F}) \leq (k-1)b(1).$$

Desweiteren gilt wegen $H_i = H_{k-i}$ für $i = 1, 2, \ldots, k-1$

$$SO(\mathbf{F}) = 2\sum_{i=1}^{[k/2]} b(F_i + F_{k-i})/2),$$

falls k ungerade ist. Für gerades k ergibt sich

$$SO(\mathbf{F}) = 2\sum_{i=1}^{k/2-1} b(F_i + F_{k-i})/2) + b(F_{k/2}).$$

Schließlich können die Schiefemaße entsprechend

$$SO(\mathbf{F}) = \sum_{H_i > 1/2} b(H_i) - \sum_{H_i \leq 1/2} b(H_i)$$

als Differenz zweier nicht-negativer Summanden dargestellt werden.

Der vorstehende Satz läßt eine große Freiheit der Konstruktion von Schiefemaßen für ordinalskalierte Merkmale. Insbesondere nennt er keine Anweisung, wie die Scorefunktion außerhalb der Eigenschaften (1) bis (4) festzulegen. Dies ist auch mit rein deskriptiven Anforderungen nicht möglich. Mittels einer Scorefunktion ist eine Gewichtung insbesondere der Ränder der Häufigkeitsverteilung möglich. Verwendet man aber auf das Intervall $[0, 1]$ normierte Schiefemaße entsprechend,

$$SO^*(\mathbf{F}) = SO(\mathbf{F})/((k-1)b(1))$$

so wird dieser Gewichtungseffekt automatisch nivelliert, und die Schiefewerte werden deutlich weniger stark durch die Wahl der Scorefunktion beeinflußt.

Es sollen nun im folgenden ausschließlich Scorefunktionen betrachtet werden, die aus Scorefunktion hergeleitet werden können, die auch bei der Streuungsmessung eine Rolle spielen.

3.7 Aus Streuungsmaßen abgeleitete Schiefemaße

Spezielle Scorefunktionen $b(.)$ mit den gewünschten Eigenschaften (1) bis (4) lassen sich aus den Scorefunktionen $a(.)$ für Streuungsmaße herleiten. Diese sollen nach Klein (1999b) die folgenden Eigenschaften besitzen:

(1) $a : [0, 1] \to [0, \infty)$,

(2) a ist stetig im Intervall $[0, 1]$,

(3) a ist symmetrisch, d.h. $a(p) = a(1-p)$ für $p \in [0, 1]$,

(4) a ist streng monoton zunehmend auf dem Intervall $[0, 1/2]$,

(5) $a(0) = 0$.

Wähle

$$b(p) = \begin{cases} a(p) - a(1/2) & \text{für } 0 \le p \le 1/2 \\ a(1/2) - a(p) & \text{für } 1/2 \le p \le 1, \end{cases}$$

dann besitzt $b(.)$ die für die Schiefemessung geforderten Eigenschaften. Wegen

$$\begin{aligned} SO(\mathbf{F}) &= \sum_{H_i>1/2} b(H_i) - \sum_{H_i\leq 1/2} b(H_i) \\ &= \sum_{H_i>1/2} (a(H_i) - a(1/2)) + \sum_{H_i\leq 1/2} (a(H_i) - a(1/2)) \end{aligned}$$

kann $SO(\mathbf{F}$ auch als Summe zweier Streuungsmaßwerte interpretiert werden, die für zwei Hälften der Häufigkeitsverteilung separat berechnet werden. Die Messung der Schiefe durch Streuungsmaße auf Teilbereichen der Verteilung ist aber wohlbekannt.

In Klein (1999b) werden in Anlehnung an und Erweiterung von Vogel (1991) insgesamt acht verschiedene Scorefunktionen betrachtet, die zur Erzeugung von Scorefunktionen für Schiefemaße benutzt werden können. Wir wollen hier lediglich drei herausgreifen, die zum einen das Schiefemaß $SO_{1,2,\ldots,k-1}$ liefern und zum anderen zum Gini–Streuungsmaß und zu einem Streuungsmaß auf der Basis der Entropie gehören, das bereits von Vogel & Dobbener (1982) betrachtet wird. Die Scorefunktionen sind in der Numerierung von Klein (1999b), die sich an Vogel (1991), anlehnt:

1. $a_7(p) = 1/2 - |p - 1/2|$,
2. $a_1(p) = -p\mathrm{ld}p - (1-p)\mathrm{ld}(1-p)$,
3. $a_2(p) = p(1-p)$.

Die korrespondierenden Scorefunktionen für Schiefemaße sind

1. $b_7(p) = p - 1/2$
2. $$b_1(p) = \begin{cases} -p\mathrm{ld}p - (1-p)\mathrm{ld}(1-p) - 1 & \text{für } p \leq 1/2 \\ 1 + p\mathrm{ld}p + (1-p)\mathrm{ld}(1-p) & \text{für } p > 1/2 \end{cases}$$
3. $$b_2(p) = \begin{cases} p(1-p) - 1/4 & \text{für } p \leq 1/2 \\ 1/4 - p(1-p) & \text{für } p > 1/2 \end{cases}$$

Betrachte $b_7(.)$, dann ist

$$SO_7(\mathbf{F}) = \sum_{i=1}^{k-1} b_7(H_i) = \sum_{i=1}^{k-1}(F_i - 1/2) = SO_{1,2,\ldots,k-1}(\mathbf{F})/2.$$

D.h. es gibt einen engen Bezug dieses Schiefemaßes zu dem in Klein (1999a) vorgeschlagenen Maß $SO_{1,2,\ldots,k-1}$, das Werte zwischen $-(k-1)$ und $k-1$ annimmt.

Eine weitere naheliegende Formulierung eines Schiefemaßes ist

$$\begin{aligned} SO_2(\mathbf{F}) &= \sum_{i=1}^{k-1} b_2(H_i) \\ &= \sum_{H_i > 1/2} (H_i(1-H_i) - 1/4) - \sum_{H_i \le 1/2} (H_i(1-H_i) - 1/4), \end{aligned}$$

für das $-(k-1)/4 \le SO_2(\mathbf{F}) \le (k-1)/4$ gilt.

Vogel & Dobbener (1981) haben ein Streuungsmaß für ordinalskalierte Merkmale auf der Basis der Entropie vorgeschlagen. Das zugehörige Schiefemaß lautet lautet

$$\begin{aligned} SO_1(\mathbf{F}) &= \sum_{i=1}^{k-1} b_1(H_i) \\ &= \sum_{H_i \le 1/2} (-H_i \mathrm{ld} H_i - (1-H_i)\mathrm{ld}(1-H_i) - 1)) \\ &\quad - \sum_{H_i > 1/2} (-H_i \mathrm{ld} H_i - (1-H_i)\mathrm{ld}(1-H_i) - 1)\,. \end{aligned}$$

Der Wertebereich ist $-(k-1) \le SO_1(\mathbf{F}) \le k-1$.

4 Schiefemaße auf der Basis einer Austauschoperation

4.1 Häufigkeitsaustausch

Wir betrachten in Anlehnung an Klein (1999b) die folgende Austauschoperation von Häufigkeiten: Es sei ein Vektor $\mathbf{n} = (n_1, \ldots, n_k)$ absoluter Häufigkeiten gegeben. Unter einer Häufigkeitsumschichtung verstehen wir eine Operation, die $n_i > 0$ für ein geeignetes $i \in \{1, 2, \ldots, k\}$ um Eins reduziert und die benachbarte Häufigkeit n_{i+1} oder n_{i-1} um

Eins erhöht. Durch sukzessive Anwendung der Häufigkeitsumschichtung kann der Vektor **n** z.B. derart verändert werden, daß der Fall einer Einpunktverteilung vorliegt.

Betrachtet man

$$U_{ij} = |i - j| f_j,$$

dann mißt

$$nU_{ij} = |i - j| n_j$$

gerade die Anzahl der Häufigkeitsumschichtungen, die nötig sind, um n_j auf Null zu reduzieren und n_i um n_j zu steigern. Dann gibt

$$nU_i = n \sum_{j=1}^{k} U_{ij} = \sum_{j=1}^{k} |i - j| n_j$$

die Anzahl von Häufigkeitsumschichtungen an, die benötigt werden, um eine Einpunktverteilung zu erzeugen mit $F_j = 0$, $j = 1, 2, \ldots, i - 1$ und $F_j = 1$, $j = i, i + 1, \ldots, k$.

Als mögliches Schiefemaß läßt sich die Differenz $U_k - U_1$ zwischen der Anzahl der Umschichtungen betrachten, die nötig sind, um die extremen Einpunktverteilungen $\iota_{\mathbf{1}}$ und $\iota_{\mathbf{k}}$ zu erzeugen. Der folgende Satz zeigt, daß diese Differenz mit dem Maß S_7 identisch ist.

Theorem 4.1 *Sei* $\mathbf{F} = (F_1, \ldots, F_{k-1}, 1)$ *ein Vektor kumulierter relativer Häufigkeiten mit* $F_k = 1$*, dann gilt:*

$$SO_7(\mathbf{F}) = \sum_{i=1}^{k-1} (F_i - 1/2) = \frac{1}{2}(U_k - U_1).$$

Beweis: Es sind

$$\begin{aligned} U_1 &= \sum_{i=1}^{k} (i - 1) f_i = \sum_{i=1}^{k} i f_i - 1, \\ U_k &= \sum_{i=1}^{k} (k - i) f_i = k - \sum_{i=1}^{k} i f_i. \end{aligned}$$

Weiterhin ist

$$\begin{aligned} \sum_{i=1}^{k} i f_i &= f_1 + 2 f_2 + \ldots + k f_k \\ &= \sum_{i=1}^{k} f_i + \sum_{i=2}^{k} f_i + \ldots + f_k \end{aligned}$$

$$\begin{aligned} &= 1 + (1 - F_1) + \ldots + (1 - F_{k-1}) \\ &= 1 + \sum_{i=1}^{k-1}(1 - F_i) = k - \sum_{i=1}^{k-1} F_i. \end{aligned}$$

Dann ist

$$\begin{aligned} U_k - U_1 &= k + 1 - 2\sum_{i=1}^{k} i f_i \\ &= k + 1 - 2k + 2\sum_{i=1}^{k-1} F_i = 2SO_7(\mathbf{F}). \end{aligned}$$

□

Im Gegensatz zur Streuungsmessung ist eine ähnlich anschauliche Interpretation der auf dem Gini–Maß bzw. der Entropie basierenden Schiefemaße nicht einsichtig.

5 Anwendungsbeispiel

Das Zentrum für Umfragen, Meinungen und Analysen (ZUMA) führt regelmäßig mit dem ALLBUS eine Befragung der deutschen Bevölkerung durch, wobei auch die Links-Rechtseinstufung der Befragten erhoben wird. Vorgelegt wird eine zehnstufige Rating-Skala mit den diskreten Ausprägungen "1" ("Links") und "10" ("Rechts"), auf die sich die Befragten selbst einordnen sollen. Im ALLBUS des Jahres 1996 haben diese Einordnung 2344 Westdeutsche und 1098 Ostdeutsche vorgenommen. Die Tabelle 1 enthält die u.a. die Häufigkeitsverteilung des ordinalskalierten Merkmals "Links-Rechts-Selbsteinstufung". Im einzelnen bezeichnen $\mathbf{f^w}$ und $\mathbf{f^o}$ die relativen Häufigkeiten für West- und Ostdeutschland und $\mathbf{F^w}$ bzw. $\mathbf{F^o}$ die zugehörigen kumulierten relativen Häufigkeiten. Die Differenzen $D_i^w = F_i^w - (1 - F_{10-i}^w)$ bzw. $D_i^o = F_i^o - (1 - F_{10-i}^o)$ für $i = 1, 2, \ldots, 9$ sind die bekannten Grundbausteine der Schiefemessung.

Tabelle 1: Relative Häufigkeiten der Links-Rechts-Selbsteinstufung im ALLBUS 1996 für West- und Ostdeutschland

i	1	2	3	4	5	6	7	8	9	10
West										
f_i^w	0.016	0.038	0.102	0.122	0.254	0.242	0.115	0.066	0.021	0.024
F_i^w	0.016	0.054	0.156	0.278	0.532	0.774	0.889	0.955	0.976	1.000
D_i^w	-0.008	0.009	0.045	0.052	0.064	0.052	0.045	0.009	-0.008	
Ost										
f_i^o	0.034	0.047	0.135	0.118	0.352	0.197	0.054	0.044	0.009	0.010
F_i^o	0.034	0.088	0.216	0.334	0.686	0.883	0.917	0.981	0.990	1.000
D_i^o	0.024	0.062	0.153	0.217	0.372	0.217	0.153	0.062	0.024	

Da $D_i^w = F_i^o - (1 - F_{10-i}^o) > 0$ für $i = 1, 2, \ldots, 9$ ist die Häufigkeitsverteilung der Links-Rechts-Selbsteinschätzung für Ostdeutschland eindeutig rechtsschief, wobei über das Ausmaß der Rechtsschiefe zunächst nichts gesagt werden kann. In Westdeutschland ist ein solche eindeutige Schiefeaussage generell nicht möglich, da $D_1^w = F_1^w - (1 - F_9^w) < 0$ und $D_i^w = F_i^w - (1 - F_{10-i}^w) > 0$ für $i = 2, 3, \ldots, 9$ sind. Die Tatsache, daß jedoch $F_1^w - (1 - F_9^w) = -0.008$ nur leicht negativ ist, läßt ebenfalls eine rechtsschiefe Verteilung für Westdeutschland vermuten. D.h. in beiden Teilen Deutschlands dominiert die Linkseinschätzung die Rechtseinschätzung, wobei offen bleiben muß, ob jeweils die Links-Rechts-Einschätzungen inhaltich identisch interpretiert werden.

Da $D_i^w = F_i^w - (1 - F_{10-i}^w) < F^o - (1 - F_{10-i}^o)$ für $i = 1, 2, \ldots, 9$ ist, muß die Verteilung in Ostdeutschland nach der partiellen Ordnung $\preceq_{SO}$ linkssteiler als in Westdeutschland sein. Die Linkseinschätzung scheint im Osten stärker zu sein als im Westen. Diese Ordnung führt dazu, daß sämtliche Schiefemaße mit einer die Bedingungen (1) bis (4) (siehe Seite 5) erfüllenden Scorefunktion die Rechtsschiefe in Ostdeutschland höher beurteilen als die Linksschiefe. Dies zeigen auch die folgenden nicht-normierten Maßzahlwerte:

Tabelle 2: Ausgewählte Schiefemaße für das Merkmal "Links-Rechts-Selbsteinstufung" in West- und Ostdeutschland

Schiefemaß	nicht normiert		normiert	
	West	Ost	West	Ost
SO_1	0.1996	1.3174	0.0222	0.1464
SO_2	0.0602	0.3428	0.0268	0.1523
SO_7	0.1300	0.6428	0.0289	0.1427

Die Normierung durch das jeweilige Maximum der Schiefemaßzahl nivelliert die Schiefeunterschiede, da insbesondere die starke Gewichting der Ränder der Häufigkeitsverteilung durch die Scorefunktion b_1 zurückgenommen wird. Das qualitative Ergebnis des Schiefevergleiches wird dadurch nicht berührt.

Häufig wird ein ordinalskaliertes Merkmals als ordinalskaliert behandelt, wenn die Anzahl der möglichen Merkmalsausprägungen hinreichend groß ist. In diesem Falle läßt sich das dritte standardisierte Moment als Schiefemaß berechnen, was im vorliegenden Beispiel zu den Schiefewerten 0.1161 für Westdeutschland und 0.0423 für Ostdeutschland führt. Damit wird die Schiefe in Ostdeutschland als geringer eingestuft, was im Widerspruch zur Schiefeordnung $\preceq_{SO}$ steht, die von dem dritten standardisierten Moment somit offensichtlich nicht eingehalten wird.

6 Zusammenfassung

Wir haben eine Reihe von Schiefemaßen für ordinalskalierte Merkmale eingeführt, die allesamt wichtige axiomatische Anforderungen an die Schiefemessung erfüllen. Von einem Maß wird zusätzlich eine anschauliche Interpretierbarkeit nachgewiesen. Die betrachtete Schiefedefinition rekurriert auf eine feste Anzahl von Merkmalsausprägungen. Im Regelfall können somit die Streuungen von Merkmalen mit unterschiedlicher Anzahl von Ausprägungen nicht verglichen werden. Betrachtet man jedoch den Fall, daß extreme Ausprägungen eine Nullhäufigkeit aufweisen, so zeigen die Maßzahlen wie gewünscht eine Verstärkung der jeweiligen Schiefe an.

Eine weitergehende Festlegung eines Schiefemaßes verlangt eine zusätzliche Axiomatisierung des Krümmungsverhaltens der Scoresfunktion $b(.)$ oder aber deren Rechtfertigung aus testtheoretischen Überlegungen, wenn der Fall quantitativer Merkmale betrachtet wird.

7 Literatur

1. HAJEK, J., SIDAK, Z. (1968). *Theory of rank tests.* Academic Press, New York.

2. KLEIN, I. (1994). *Mögliche Skalentypen, invariante Relationen und wissenschaftliche Gesetze.* Vandenhoeck & Ruprecht, Göttingen.

3. KLEIN, I. (1999a). Systematik der Schiefemessung auch für ordinalskalierte Merkmale. *Diskussionspapier der Lehrstühle für Statistik der Universität Erlangen–Nürnberg* Nr. **26/1999**.

4. KLEIN, I. (1999b). Lineare Rangordnungsstatistiken als Verteilungsmaßzahlen für ordinalskalierte Merkmale. I. Streuungsmessung. Diskussionspapiere der Lehrstühle für Statistik der Universität Erlangen–Nürnberg Nr. **27/1999**.

5. VOGEL, F. (1991). Streuungsmessung ordinalskalierter Merkmale. *Jahrbücher für Nationalökonomie und Statistik* **208** 299–318.

6. VOGEL, F., DOBBENER, R. (1982). Ein Streuungsmaß für komparative Merkmale. *Jahrbücher für Nationalökonomie und Statistik* **197** 145-157.