

# Anhang

## A Ridit-Werte und Ridit-Test

Die Ridit-Analyse ist eine recht unbekannte Analyse ordinaler Daten, die von Bross (1958) eingeführt wurde.<sup>1</sup> Sie wird in der vorliegenden Arbeit insbesondere im ersten Anwendungsbeispiel im Abschnitt 6.1 verwendet. Zusätzlich zu den Erläuterungen im Abschnitt 6.1.3.2 soll an dieser Stelle das mathematische Fundament der Ridit-Analyse kurz vorgestellt werden.

Es werden Verteilungen mit einer Referenzverteilung verglichen, die aus theoretischen Überlegungen heraus bestimmt werden muss.<sup>2</sup> Seien  $x_1, \dots, x_p$  ( $p \in \mathbb{N}$ ) die geordneten Kategorien einer Skala und  $\{p_j : j = 1, \dots, p\}$  die Häufigkeiten für jede Antwortkategorie. Die Ridits  $w_1, \dots, w_p$  der Referenzgruppe werden wie folgt bestimmt:<sup>3</sup>

$$w_1 = \frac{(1/2)p_1}{n}$$
$$w_j = \frac{1/2p_j + \sum_{k < j} p_k}{n},$$

mit  $j = 2, \dots, p$  und  $n = \sum_{l=1}^p p_l$ .

Die Ridit-Werte für die Vergleichsgruppen  $v = 1, \dots, m$  sind die Summe der Ridits  $w_k$  ( $k = 1, \dots, p$ ) der Referenzgruppe multipliziert mit den absoluten Häufigkeiten  $\pi_{vk}$  der Vergleichsgruppe  $v$  in der Antwortkategorie  $k$ , dividiert durch die Anzahl  $\pi_v = (\sum_{m=1}^p \pi_{vm})$  der Fälle der Vergleichsgruppe  $v$  in allen Antwortkategorien:<sup>4</sup>

$$r_v = \sum_{k=1}^p \frac{w_k \pi_{vk}}{\pi_v}. \quad (\text{A.1})$$

Die Ridit-Werte der Referenzgruppe ergeben sich durch Einsetzen in (A.1) und sind definitionsgemäß gleich 0,5.

---

<sup>1</sup> Vgl. [SD96], S. 354.

<sup>2</sup> Vgl. [Bro58], S. 21f.

<sup>3</sup> Vgl. [PQS97], S. 420.

<sup>4</sup> Vgl. [Hol03], S. 46.

Zusätzlich bietet die Ridit-Analyse einen Hypothesentest auf systematische oder zufällige Unterscheidungen in den untersuchten Gruppen.<sup>5</sup> Getestet wird dabei, ob sich der Ridit-Wert  $r_v$  einer Vergleichsgruppe  $v$  systematisch vom Ridit-Wert 0,5 der Referenzgruppe unterscheidet. Es werden also die folgenden Hypothesen unter der Kruskal-Wallis-Statistik  $W$  betrachtet:<sup>6</sup>

$$H_0 : \forall v, r_v = 0,5$$

$$H_1 : \exists v, r_v \neq 0,5$$

$$W = 12 \sum_{v=1}^m \pi_v (r_v - 0,5)^2.$$

$W$  folgt einer  $\chi^2$ -Verteilung mit  $(m - 1)$  Freiheitsgraden.<sup>7</sup>

## B Einstellungen zur deutschen Vereinigung von Arbeitslosen und ihren statistischen Zwillingen

In diesem Abschnitt werden die Einstellungen von arbeitslosen Ost- und Westdeutschen und ihrer statistischen Zwillinge aus der Gruppe der Erwerbstätigen dargestellt, wie sie in Abschnitt 6.1.4 ermittelt wurden. Es werden sowohl die neun Variablen V491 bis V499 aus der ALLBUS 2006 vorgestellt als auch die drei Faktoren (Faktor1 „Verhalten und Beurteilung des Ostens“, Faktor2 „Verhalten und Beurteilung des Westens“ und Faktor3 „Sozialismus und Stasi“), die mit Hilfe der Faktorenanalyse gewonnen wurden.<sup>8</sup> Für die neun Variablen V491 bis V499 werden die Ridits ausgewiesen, d. h. die Lage relativ zur Referenzgruppe der Arbeitslosen.<sup>9</sup> Die Faktoren werden durch Mittelwertbildung der Antwortkategorien bestimmt, die in der ALLBUS 2006 mit 1 bis 4 codiert sind. Signifikante Unterschiede in den Einstellungen der Erwerbstätigen gegenüber den Arbeitslosen bzw. der statistischen Zwillinge aus der Gruppe der Erwerbstätigen gegenüber den Arbeitslosen werden bei einem Signifikanzniveau von  $\alpha = 0,05$  mittels \* gekennzeichnet.

---

<sup>5</sup> Vgl. [Hol03], S. 47.

<sup>6</sup> Vgl. [Wu07], S. 679f.

<sup>7</sup> Vgl. [Wu07], S. 680.

<sup>8</sup> Vgl. Abschnitt 6.1.4.

<sup>9</sup> Siehe Abschnitt 6.1.3.2 und Anhang A.

In Tabelle B.1 werden zunächst die ostdeutschen Arbeitslosen allen ostdeutschen Erwerbstätigen gegenübergestellt:

Tabelle B.1: Einstellungen zur deutschen Vereinigung: Ostdeutsche Arbeitslose ( $n = 127$ ) und Erwerbstätige ( $n = 534$ )

	V491	V492	V493	V494	V495	V496
Cases	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000
Controls	0,6070*	0,4560	0,5124	0,3425*	0,4189*	0,6215*
	V497	V498	V499	Faktor1	Faktor2	Faktor3
Cases	0,5000	0,5000	0,5000	2,84	2,29	1,80
Controls	0,5265	0,5366	0,5393	2,62*	2,46*	1,97*

Zur Durchführung der Hypothesentests wurde bei den quantitativen Variablen (Faktoren) bei Annahme der Normalverteilung nach Kolmogorov-Smirnov und bei Varianzhomogenität der gewöhnliche Studentsche T-Test verwendet, bei Varianzheterogenität der T-Test nach Welch. Konnte Normalverteilung nach Kolmogorov-Smirnov nicht angenommen werden, wurde der Mann-Whitney-U-Test verwendet. Für die ordinal skalierten Variablen wurde der Ridit-Test gerechnet. Die folgenden Tabellen B.2 und B.3 enthalten die Ridits der statistischen Zwillinge der ostdeutschen Arbeitslosen nach Durchführung des constrained und unconstrained Fuzzy-Matchings.

Tabelle B.2: Einstellungen zur deutschen Vereinigung: Statistische Zwillinge der ostdeutschen Arbeitslosen, constrained Fuzzy-Matching ( $n = 127$ )

Verknüpfungstyp	V491	V492	V493	V494	V495	V496	V497	V498	V499	Faktor1	Faktor2	Faktor3
und-Verkn., abs.	0,6002*	0,4309	0,5182	0,53623*	0,4142*	0,5857*	0,5045	0,5190	0,4936	2,58*	2,45	1,88
und-Verkn., eukl.	0,5943*	0,4391	0,5101	0,5373*	0,4243*	0,5934*	0,5111	0,5190	0,4993	2,61*	2,43	1,89
oder-Verkn., abs.	0,5602	0,4543	0,5291	0,3822*	0,4212*	0,5950*	0,5048	0,4958	0,4635	2,64*	2,41	1,78
oder-Verkn., eukl.	0,5389	0,4366	0,5149	0,3929*	0,4132*	0,5702	0,5149	0,4931	0,4708	2,63*	2,37	1,80
MinMas-Op., $\gamma = 0,1$ , abs.	0,5552	0,4571	0,5316	0,3822*	0,4212*	0,5924*	0,5074	0,4907	0,4635	2,64*	2,40	1,77
MinMas-Op., $\gamma = 0,1$ , eukl.	0,5343	0,4334	0,5188	0,3892*	0,4157*	0,5702	0,5149	0,4931	0,4730	2,63*	2,35	1,80
MinMas-Op., $\gamma = 0,3$ , abs.	0,5580	0,4561	0,5267	0,3799*	0,4236*	0,5947*	0,5152	0,4888	0,4614	2,65*	2,40	1,76
MinMas-Op., $\gamma = 0,3$ , eukl.	0,5389	0,4366	0,5241	0,3929*	0,4132*	0,5702	0,5149	0,4708	0,4708	2,63*	2,37	1,80
MinMas-Op., $\gamma = 0,5$ , abs.	0,5602	0,4543	0,5291	0,3822*	0,4212*	0,5950*	0,5048	0,4931	0,4635	2,64*	2,41	1,78
MinMas-Op., $\gamma = 0,5$ , eukl.	0,5389	0,4366	0,5241	0,3929*	0,4132*	0,5702	0,5149	0,4931	0,4708	2,63*	2,37	1,80
MinMas-Op., $\gamma = 0,7$ , abs.	0,5602	0,4592	0,5291	0,3822*	0,4254*	0,5924*	0,5100	0,4907	0,4635	2,65*	2,41	1,77
MinMas-Op., $\gamma = 0,7$ , eukl.	0,5389	0,4255*	0,5306	0,3841*	0,4067*	0,5789*	0,5123	0,4981	0,4801	2,61*	2,38	1,82
MinMas-Op., $\gamma = 0,9$ , abs.	0,5531	0,4561	0,5231	0,3859*	0,4110*	0,5851*	0,4956	0,4964	0,4678	2,62*	2,39	1,80
MinMas-Op., $\gamma = 0,9$ , eukl.	0,5303	0,4313	0,5223	0,3864*	0,4072*	0,5843*	0,5228	0,4857	0,4778	2,62*	2,35	1,79
Alg.-PS-Op., $\gamma = 0,1$ , abs.	0,5254	0,4479	0,5063	0,3649*	0,4154*	0,5638	0,5086	0,4994	0,4701	2,60*	2,31	1,82
Alg.-PS-Op., $\gamma = 0,1$ , eukl.	0,5273	0,4391	0,5015	0,3755*	0,4172*	0,5671	0,5072	0,4952	0,4723	2,61*	2,31	1,81
Alg.-PS-Op., $\gamma = 0,3$ , abs.	0,5254	0,4479	0,5063	0,3649*	0,4154*	0,5638	0,5086	0,4994	0,4701	2,60*	2,31	1,82
Alg.-PS-Op., $\gamma = 0,3$ , eukl.	0,5273	0,4391	0,5015	0,3755*	0,4172*	0,5671	0,5072	0,4952	0,4723	2,61*	2,31	1,81
Alg.-PS-Op., $\gamma = 0,5$ , abs.	0,5254	0,4479	0,5063	0,3649*	0,4154*	0,5638	0,5086	0,4994	0,4701	2,60*	2,31	1,82
Alg.-PS-Op., $\gamma = 0,5$ , eukl.	0,5273	0,4391	0,5015	0,3755*	0,4172*	0,5671	0,5072	0,4952	0,4723	2,61*	2,31	1,81
Alg.-PS-Op., $\gamma = 0,7$ , abs.	0,5254	0,4479	0,5063	0,3649*	0,4154*	0,5638	0,5086	0,4994	0,4701	2,60*	2,31	1,82
Alg.-PS-Op., $\gamma = 0,7$ , eukl.	0,5415	0,4362	0,5053	0,3774*	0,4288	0,5698	0,5072	0,4912	0,4795	2,62*	2,34	1,81
Alg.-PS-Op., $\gamma = 0,9$ , abs.	0,5254	0,4398	0,5039	0,3659*	0,4132*	0,5638	0,5112	0,5078	0,4784	2,60*	2,31	1,84
Alg.-PS-Op., $\gamma = 0,9$ , eukl.	0,5624	0,4342	0,5295	0,3782*	0,4166*	0,5827*	0,4833	0,4935	0,4804	2,58*	2,42	1,82
fcm-oder-Verkn., abs.	0,6060*	0,4530	0,5170	0,3180*	0,4090*	0,6500*	0,4920	0,4860	0,5590	2,55*	2,45*	1,91
fcm-oder-Verkn., eukl.	0,6050*	0,4550	0,5110	0,3190*	0,4140*	0,6470*	0,4980	0,4890	0,5560	2,56*	2,45*	1,91
fcm-und-Verkn., abs.	0,5870*	0,4350	0,5090	0,3250*	0,4040*	0,5980*	0,4730	0,4800	0,5080	2,52*	2,40	1,81
fcm-und-Verkn., eukl.	0,6010*	0,4380	0,5140	0,3180*	0,3990*	0,6070*	0,4990	0,4840	0,5190	2,53*	2,44	1,85

Tabelle B.3: Einstellungen zur deutschen Vereinigung: Statistische Zwillinge der ostdeutschen Arbeitslosen, unconstrained Fuzzy-Matching ( $n = 127$ )

Verknüpfungstyp	V491	V492	V493	V494	V495	V496	V497	V498	V499	Faktor1	Faktor2	Faktor3
und. abs.	0.5691	0.4373	0.5323	0.3832*	0.3987*	0.6001*	0.4949	0.4778	0.5021	2.58*	2.44	1.81
und. eukl.	0.5614	0.4454	0.5218	0.3920*	0.4093*	0.5836*	0.4949	0.4834	0.4969	2.60*	2.41	1.81
oder. abs.	0.5159	0.4147*	0.5283	0.4153*	0.4012*	0.5917*	0.4798	0.4636	0.4666	2.59*	2.34	1.73
oder. eukl.	0.5126	0.4260*	0.5166	0.4164*	0.4097*	0.5619	0.4713	0.4710	0.4596	2.60*	2.32	1.74
MinMax-Op., $\gamma = 0.1$ , abs.	0.5109	0.4225*	0.5255	0.4176*	0.4012*	0.5917*	0.4798	0.4619	0.4696	2.60*	2.33	1.73
MinMax-Op., $\gamma = 0.1$ , eukl.	0.5029	0.4199*	0.5113	0.4128*	0.4057*	0.5592	0.4740	0.4760	0.4617	2.59*	2.30	1.75
MinMax-Op., $\gamma = 0.3$ , abs.	0.5137	0.4165*	0.5259	0.4130*	0.4037*	0.5914*	0.4902	0.4570	0.4645	2.60*	2.34	1.71
MinMax-Op., $\gamma = 0.3$ , eukl.	0.5076	0.4231*	0.5166	0.4164*	0.4032*	0.5592	0.4740	0.4760	0.4596	2.59*	2.32	1.75
MinMax-Op., $\gamma = 0.5$ , abs.	0.5159	0.4147*	0.5283	0.4153*	0.4012*	0.5917*	0.4798	0.4636	0.4666	2.59*	2.34	1.73
MinMax-Op., $\gamma = 0.5$ , eukl.	0.5076	0.4231*	0.5166	0.4164*	0.4032*	0.5592	0.4740	0.4760	0.4596	2.59*	2.32	1.75
MinMax-Op., $\gamma = 0.7$ , abs.	0.5159	0.4196*	0.5283	0.4153*	0.4054*	0.5890*	0.4850	0.4586	0.4666	2.60*	2.34	1.72
MinMax-Op., $\gamma = 0.7$ , eukl.	0.5076	0.4181*	0.5234	0.4128*	0.3984*	0.5656	0.4713	0.4811	0.4658	2.58*	2.33	1.77
MinMax-Op., $\gamma = 0.9$ , abs.	0.5137	0.4186*	0.5202	0.4140*	0.3970*	0.5971*	0.4759	0.4676	0.4728	2.58*	2.33	1.75
MinMax-Op., $\gamma = 0.9$ , eukl.	0.5095	0.4181*	0.5184	0.4091*	0.3855*	0.5606	0.4647	0.4734	0.4628	2.56*	2.32	1.76
Allg. PS-Op., $\gamma = 0.1$ , abs.	0.5338	0.4353	0.5079	0.3787*	0.4187*	0.5641	0.4876	0.4695	0.4517	2.59*	2.34	1.74
Allg. PS-Op., $\gamma = 0.1$ , eukl.	0.5305	0.4215*	0.5099	0.3852*	0.4051*	0.5610	0.4706	0.4903	0.4462	2.56*	2.33	1.74
Allg. PS-Op., $\gamma = 0.3$ , abs.	0.5338	0.4353	0.5079	0.3787*	0.4187*	0.5641	0.4876	0.4695	0.4517	2.59*	2.34	1.74
Allg. PS-Op., $\gamma = 0.3$ , eukl.	0.5305	0.4215*	0.5099	0.3852*	0.4051*	0.5610	0.4706	0.4903	0.4462	2.56*	2.33	1.74
Allg. PS-Op., $\gamma = 0.5$ , abs.	0.5243	0.4424	0.5050	0.3833*	0.4204*	0.5730*	0.4850	0.4728	0.4600	2.60*	2.32	1.75
Allg. PS-Op., $\gamma = 0.5$ , eukl.	0.5195	0.4215*	0.5079	0.3898*	0.4053*	0.5661	0.4693	0.4886	0.4535	2.56*	2.31	1.76
Allg. PS-Op., $\gamma = 0.7$ , abs.	0.5243	0.4424	0.5050	0.3833*	0.4204*	0.5730*	0.4850	0.4728	0.4600	2.60*	2.32	1.75
Allg. PS-Op., $\gamma = 0.7$ , eukl.	0.5381	0.4152*	0.4903	0.3873*	0.4040*	0.5585	0.4587	0.4903	0.4609	2.54*	2.31	1.78
Allg. PS-Op., $\gamma = 0.9$ , abs.	0.5307	0.4321	0.4946	0.3807*	0.4174*	0.5644	0.4810	0.4852	0.4569	2.57*	2.32	1.78
Allg. PS-Op., $\gamma = 0.9$ , eukl.	0.5460	0.4133*	0.4987	0.3687*	0.3959*	0.5648	0.4304	0.4809	0.4612	2.48*	2.35	1.76
fcm-oder-Verkn., abs.	0.5949*	0.4272*	0.5029	0.3109*	0.3881*	0.6193*	0.4839	0.4558	0.5550	2.48*	2.42	1.83
fcm-oder-Verkn., eukl.	0.6025*	0.4282*	0.4954	0.3098*	0.3850*	0.6288*	0.4839	0.4649	0.5562	2.48*	2.42	1.86
fcm-und-Verkn., abs.	0.5584	0.4129*	0.4795	0.3504*	0.4190*	0.5516	0.4953	0.5185	0.5660	2.56*	2.36	2.00*
fcm-und-Verkn., eukl.	0.5074	0.4820	0.4425	0.4111*	0.3404*	0.5008	0.4739	0.4806	0.5971*	2.56*	2.15	2.07*

Da beim statistischen Fuzzy-Matching nur 127 Paare statistischer Zwillinge gefunden wurden, beim Distanzmatching aber zu allen Arbeitslosen ein statistischer Zwilling gefunden wird, enthält die folgende Tabelle B.4 die Gegenüberstellung aller 128 ostdeutschen Arbeitslosen mit allen Erwerbstätigen.

Tabelle B.4: Einstellungen zur deutschen Vereinigung: Ostdeutsche Arbeitslose ( $n = 128$ ) und Erwerbstätige ( $n = 534$ )

	V491	V492	V493	V494	V495	V496
Cases	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000
Controls	0,6070*	0,4586	0,5098	0,3402*	0,4178*	0,6215*
	V497	V498	V499	Faktor1	Faktor2	Faktor3
Cases	0,5000	0,5000	0,5000	2,84	2,29	1,80
Controls	0,5264	0,5354	0,5388	2,62*	2,46*	1,97*

Die Tabellen B.5 und B.6 stellen den Vergleich der nach constrained und unconstrained Distanzmatching gefundenen statistischen Zwillinge mit den ostdeutschen Arbeitslosen dar.

Tabelle B.5: Einstellungen zur deutschen Vereinigung: Statistische Zwillinge der ostdeutschen Arbeitslosen, constrained Distanzmatching ( $n = 128$ )

Distanz	V491	V492	V493	V494	V495	V496
Abs. Distanz	0,5604	0,4465	0,4905	0,3544*	0,4198*	0,5829*
Eukl. Distanz	0,5726*	0,4524	0,4825	0,3593*	0,4295	0,5872*
Distanz	V497	V498	V499	Faktor1	Faktor2	Faktor3
Abs. Distanz	0,4718	0,4916	0,5116	2,57*	2,36	1,86
Eukl. Distanz	0,5019	0,5006	0,5073	2,61*	2,37	1,87

Tabelle B.6: Einstellungen zur deutschen Vereinigung: Statistische Zwillinge der ostdeutschen Arbeitslosen, unconstrained Distanzmatching ( $n = 128$ )

Distanz	V491	V492	V493	V494	V495	V496
Abs. Distanz	0,5689	0,4428	0,4446	0,3752*	0,4328	0,6068*
Eukl. Distanz	0,5671	0,4432	0,4349	0,3724*	0,4239*	0,5991*
Distanz	V497	V498	V499	Faktor1	Faktor2	Faktor3
Abs. Distanz	0,4652	0,4804	0,4922	2,58*	2,31	1,81
Eukl. Distanz	0,4691	0,4887	0,5001	2,58*	2,29	1,84

Zum Vergleich der westdeutschen Arbeitslosen mit den Erwerbstätigen werden in der folgenden Tabelle B.7 die Ridits und Mittelwerte der westdeutschen Erwerbstätigen im Vergleich zu den Arbeitslosen dargestellt.

Tabelle B.7: Einstellungen zur deutschen Vereinigung: Westdeutsche Arbeitslose ( $n = 128$ ) und Erwerbstätige ( $n = 1116$ )

	V491	V492	V493	V494	V495	V496
Cases	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000
Controls	0,5032	0,4885	0,5230	0,4533	0,5274	0,4755
	V497	V498	V499	Faktor1	Faktor2	Faktor3
Cases	0,5000	0,5000	0,5000	2,11	3,02	2,21
Controls	0,5037	0,5430	0,5721*	2,06	3,11	2,40*

Die beiden folgenden Tabellen B.8 und B.9 enthalten die Ridits bzw. Mittelwerte der Faktoren der nach statistischem constrained und unconstrained Fuzzy-Matching gefundenen statistischen Zwillinge im Vergleich zur Referenzgruppe der Arbeitslosen.

Tabelle B.8: Einstellungen zur deutschen Vereinigung: Statistische Zwillinge der westdeutschen Arbeitslosen, constrained Fuzzy-Matching ( $n = 128$ )

Verknüpfungstyp	V491	V492	V493	V494	V495	V496	V497	V498	V499	Faktor1	Faktor2	Faktor3
und, abs.	0,4842	0,4774	0,5361	0,4177*	0,5154	0,4319	0,4661	0,5472	0,5560	1,99	3,08	2,35
und, eukl.	0,5044	0,5028	0,5412	0,4140*	0,5239	0,4342	0,4888	0,5263	0,5524	2,03	3,13	2,31
oder, abs.	0,4717	0,4859	0,5720	0,4377	0,5329	0,4446	0,4936	0,5324	0,5150	2,04	3,12	2,24
oder, eukl.	0,4960	0,5068	0,5604	0,4377	0,5223	0,4514	0,4941	0,5198	0,5184	2,05	3,15	2,22
MinMax-Op., $\gamma = 0,1$ , abs.	0,4746	0,4881	0,5684	0,4427	0,5348	0,4497	0,4891	0,5310	0,5184	2,05	3,12	2,25
MinMax-Op., $\gamma = 0,1$ , eukl.	0,4953	0,5072	0,5596	0,4407	0,5249	0,4505	0,4953	0,5232	0,5208	2,06	3,15	2,23
MinMax-Op., $\gamma = 0,3$ , abs.	0,4775	0,4911	0,5684	0,4427	0,5347	0,4480	0,4912	0,5339	0,5234	2,05	3,12	2,26
MinMax-Op., $\gamma = 0,3$ , eukl.	0,5049	0,5057	0,5640	0,4364	0,5247	0,4575	0,4941	0,5166	0,5312	2,05	3,17	2,23
MinMax-Op., $\gamma = 0,5$ , abs.	0,4717	0,4859	0,5720	0,4377	0,5282	0,4446	0,4936	0,5324	0,5150	2,04	3,12	2,24
MinMax-Op., $\gamma = 0,5$ , eukl.	0,4989	0,5098	0,5604	0,4406	0,5223	0,4514	0,4941	0,5242	0,5193	2,06	3,15	2,22
MinMax-Op., $\gamma = 0,7$ , abs.	0,4746	0,4859	0,5720	0,4377	0,5282	0,4446	0,4936	0,5353	0,5220	2,04	3,13	2,25
MinMax-Op., $\gamma = 0,7$ , eukl.	0,5047	0,5068	0,5633	0,4377	0,5192	0,4486	0,4897	0,5286	0,5262	2,05	3,16	2,25
MinMax-Op., $\gamma = 0,9$ , abs.	0,4793	0,4798	0,5747*	0,4377	0,5239	0,4402	0,4835	0,5315	0,5141	2,03	3,14	2,23
MinMax-Op., $\gamma = 0,9$ , eukl.	0,4930	0,5046	0,5617	0,4511	0,5251	0,4385	0,4922	0,5352	0,5225	2,07	3,14	2,28
Alg. PS-Op., $\gamma = 0,1$ , abs.	0,4916	0,4874	0,5888*	0,4403	0,5291	0,4366	0,5000	0,5370	0,5409	2,05	3,19	2,29
Alg. PS-Op., $\gamma = 0,1$ , eukl.	0,4941	0,4947	0,5799*	0,4436	0,5189	0,4396	0,5104	0,5495	0,5491	2,06	3,17	2,35
Alg. PS-Op., $\gamma = 0,3$ , abs.	0,4916	0,4874	0,5888*	0,4403	0,5291	0,4366	0,5000	0,5370	0,5409	2,05	3,19	2,29
Alg. PS-Op., $\gamma = 0,3$ , eukl.	0,4941	0,4947	0,5799*	0,4436	0,5291	0,4366	0,5000	0,5370	0,5409	2,05	3,17	2,30
Alg. PS-Op., $\gamma = 0,5$ , abs.	0,4886	0,4907	0,5822*	0,4403	0,5225	0,4340	0,4934	0,5311	0,5446	2,04	3,17	2,35
Alg. PS-Op., $\gamma = 0,5$ , eukl.	0,5030	0,5040	0,5762*	0,4449	0,5205	0,5370	0,5125	0,5387	0,5418	2,07	3,18	2,31
Alg. PS-Op., $\gamma = 0,7$ , abs.	0,4886	0,4928	0,5822*	0,4341	0,5225	0,4373	0,4911	0,5283	0,5427	2,03	3,17	2,27
Alg. PS-Op., $\gamma = 0,7$ , eukl.	0,5107	0,4985	0,5755*	0,4394	0,5353	0,4564	0,5137	0,5325	0,5370	2,08	3,20	2,28
Alg. PS-Op., $\gamma = 0,9$ , abs.	0,4973	0,4982	0,5815*	0,4353	0,5241	0,4387	0,4888	0,5176	0,5395	2,04	3,19	2,25
Alg. PS-Op., $\gamma = 0,9$ , eukl.	0,5116	0,5042	0,5549	0,4461	0,5332	0,4538	0,5193	0,5297	0,5124	2,10	3,15	2,24
fem-oder-Verkn., abs.	0,4860	0,5060	0,5510	0,4530	0,4940	0,4460	0,5160	0,5270	0,5000	2,05	3,11	2,29
fem-oder-Verkn., eukl.	0,4710	0,4992	0,5308	0,4570	0,5040	0,4320	0,4870	0,5120	0,5350	2,03	3,05	2,23
fem-und-Verkn., abs.	0,5400	0,4900	0,5770*	0,4290	0,5030	0,4520	0,4790	0,5260*	0,5860*	2,00	3,24*	2,36
fem-und-Verkn., eukl.	0,5470	0,4960	0,5650	0,4490	0,5330	0,4620	0,4990	0,5070	0,5560	2,06	3,25*	2,28



Tabelle B.9: Einstellungen zur deutschen Vereinigung: Statistische Zwillinge der westdeutschen Arbeitslosen, unconstrained Fuzzy-Matching ( $n = 128$ )

Verknüpfungsart	V491	V492	V493	V494	V495	V496	V497	V498	V499	Faktor1	Faktor2	Faktor3
und. abs.	0,4971	0,4401	0,6007*	0,3858*	0,4800	0,4319	0,4431	0,5371	0,5824*	1,87*	3,21	2,39
und. eukl.	0,5206	0,4786	0,5962*	0,3895*	0,5112	0,4474	0,4632	0,5298	0,5712	1,95	3,25*	2,35
oder. eukl.	0,4897	0,4853	0,5749*	0,4332	0,5252	0,4257	0,4913	0,5249	0,5169	2,02	3,15	2,23
MinMax-Op., $\gamma = 0,1$ , abs.	0,5049	0,5062	0,5683	0,4237*	0,5279	0,4518	0,4915	0,5252	0,5301	2,04	3,18	2,25
MinMax-Op., $\gamma = 0,1$ , eukl.	0,4878	0,4823	0,5779*	0,4332	0,5288	0,4420	0,4889	0,5282	0,5204	2,02	3,15	2,25
MinMax-Op., $\gamma = 0,3$ , abs.	0,5041	0,5066	0,5675	0,4269	0,5305	0,4509	0,4987	0,5287	0,5314	2,05	3,17	2,26
MinMax-Op., $\gamma = 0,3$ , eukl.	0,4878	0,4853	0,5779*	0,4332	0,5286	0,4402	0,4911	0,5282	0,5184	2,03	3,15	2,25
MinMax-Op., $\gamma = 0,5$ , abs.	0,5109	0,5062	0,5683	0,4237*	0,5279	0,4570	0,4997	0,5235	0,5320	2,04	3,19	2,25
MinMax-Op., $\gamma = 0,5$ , eukl.	0,4897	0,4853	0,5749*	0,4332	0,5252	0,4454	0,4913	0,5249	0,5169	2,02	3,15	2,23
MinMax-Op., $\gamma = 0,7$ , abs.	0,5078	0,5093	0,5683	0,4267	0,5279	0,4518	0,4975	0,5298	0,5301	2,04	3,18	2,26
MinMax-Op., $\gamma = 0,7$ , eukl.	0,4897	0,4853	0,5749*	0,4332	0,5252	0,4454	0,4913	0,5249	0,5169	2,02	3,15	2,23
MinMax-Op., $\gamma = 0,9$ , abs.	0,5107	0,5062	0,5713	0,4237*	0,5246	0,4490	0,4929	0,5314	0,5301	2,03	3,19	2,26
MinMax-Op., $\gamma = 0,9$ , eukl.	0,4957	0,4853	0,5770*	0,4332	0,5241	0,4420	0,4880	0,5303	0,5231	2,02	3,17	2,25
Alg. PS-Op., $\gamma = 0,9$ , abs.	0,5020	0,5035	0,5754*	0,4432	0,5284	0,4429	0,4872	0,5386	0,5448	2,05	3,19	2,31
Alg. PS-Op., $\gamma = 0,9$ , eukl.	0,4975	0,4802	0,5930*	0,4470	0,5286	0,4268	0,4934	0,5254	0,5504	2,03	3,21	2,28
Alg. PS-Op., $\gamma = 0,1$ , abs.	0,4914	0,4843	0,5854*	0,4482	0,5237	0,4334	0,5183	0,5467	0,5459	2,06	3,17	2,31
Alg. PS-Op., $\gamma = 0,1$ , eukl.	0,4975	0,4802	0,5930*	0,4470	0,5286	0,4268	0,4934	0,5254	0,5504	2,03	3,21	2,28
Alg. PS-Op., $\gamma = 0,3$ , abs.	0,4914	0,4843	0,5854*	0,4482	0,5237	0,4334	0,5183	0,5467	0,5459	2,06	3,17	2,31
Alg. PS-Op., $\gamma = 0,3$ , eukl.	0,4975	0,4802	0,5930*	0,4470	0,5286	0,4268	0,4934	0,5254	0,5504	2,03	3,21	2,28
Alg. PS-Op., $\gamma = 0,5$ , abs.	0,4975	0,4802	0,5930*	0,4470	0,5286	0,4268	0,4934	0,5254	0,5504	2,03	3,21	2,28
Alg. PS-Op., $\gamma = 0,5$ , eukl.	0,4914	0,4873	0,5884*	0,4461	0,5262	0,4334	0,5183	0,5405	0,5437	2,06	3,17	2,30
Alg. PS-Op., $\gamma = 0,7$ , abs.	0,5022	0,4802	0,5937*	0,4408	0,5262	0,4330	0,4843	0,5222	0,5466	2,02	3,22	2,26
Alg. PS-Op., $\gamma = 0,7$ , eukl.	0,5030	0,4891	0,5894*	0,4440	0,5381	0,4488	0,4980	0,5246	0,5198	2,06	3,20	2,21
Alg. PS-Op., $\gamma = 0,9$ , abs.	0,5050	0,4802	0,5906*	0,4358	0,5222	0,4336	0,4798	0,5130	0,5396	2,01	3,23	2,23
Alg. PS-Op., $\gamma = 0,9$ , eukl.	0,5032	0,4867	0,5822*	0,4432	0,5381	0,4584	0,4790	0,5400	0,5177	2,08	3,19	2,25
fem-oder-Verkn., abs.	0,4945	0,4865	0,5581	0,4637	0,5222	0,4324	0,4778	0,5227	0,5699	2,02	3,13	2,35
fem-oder-Verkn., eukl.	0,4851	0,4981	0,5416	0,4615	0,5314	0,4256	0,4991	0,5369	0,5661	2,07	3,08	2,37
fem-und-Verkn., abs.	0,4926	0,5042	0,5372	0,4629	0,5282	0,4205*	0,4933	0,5369	0,5727	2,07	3,08	2,38
fem-und-Verkn., eukl.	0,4950	0,5082	0,5338	0,4692	0,5269	0,4285	0,4862	0,5282	0,5550	2,07	3,09	2,32

In den abschließenden beiden Tabellen B.10 und B.11 werden die nach statistischem constrained und unconstrained Distanzmatching gefundenen statistischen Zwillinge mit der Referenzgruppe der Arbeitslosen verglichen.

Tabelle B.10: Einstellungen zur deutschen Vereinigung: Statistische Zwillinge der west-deutschen Arbeitslosen, constrained Distanzmatching ( $n = 128$ )

Distanz	V491	V492	V493	V494	V495	V496
Abs. Distanz	0,5161	0,4817	0,5727	0,4620	0,4838	0,4544
Eukl. Distanz	0,5270	0,4880	0,5700	0,4640	0,4900	0,4700
Distanz	V497	V498	V499	Faktor1	Faktor2	Faktor3
Abs. Distanz	0,4835	0,5235	0,5671	2,00	3,19	2,32
Eukl. Distanz	0,4900	0,5170	0,5560	2,03	3,18	2,29

Tabelle B.11: Einstellungen zur deutschen Vereinigung: Statistische Zwillinge der west-deutschen Arbeitslosen, unconstrained Distanzmatching ( $n = 128$ )

Distanz	V491	V492	V493	V494	V495	V496
Abs. Distanz	0,5304	0,4828	0,5842*	0,4774	0,4975	0,4474
Eukl. Distanz	0,5154	0,4828	0,5648	0,4779	0,5059	0,4605
Distanz	V497	V498	V499	Faktor1	Faktor2	Faktor3
Abs. Distanz	0,5066	0,5110	0,5754	2,03	3,26*	2,29
Eukl. Distanz	0,5267	0,5032	0,5642	2,06	3,20	2,27

## C Quellcode des Programms zum statistischen Fuzzy-Matching

In diesem Abschnitt wird der Quellcode der wichtigsten Teile des Programms zum statistischen Fuzzy-Matching vorgestellt. Der gesamte Quellcode ist auf der beiliegenden CD-ROM enthalten. Kommentare innerhalb des Quellcodes sind mit einem vorangestellten %-Zeichen gekennzeichnet.

## Statistisches Matching mit Fuzzy-Logic

```

% Ablauf des ganzen Programms:

% Definition aller notwendigen Variablen:
definition_variablen;
% Laden der Daten und Abfragen aller notwendigen Parameter:
load_data_gui;
% Warten, bis das GUI mit "OK" geschlossen wird, damit mit den
% erzeugten Variablen und Daten weiter gerechnet werden kann:
uiwait(load_data_gui);
% Abfangen des Falls, dass der Benutzer das Programm abbricht.
% Kein Abbrechen durch den Benutzer bei den Angaben zu den Daten:
if abbruch==0;
    % Bestimmen der ling. Terme, berechnen der Zugehörigkeits-
    % funktionen und fuzzyfizieren der Ausgangsdaten.
    zugehoerigkeitsfunktionen;
    % Kein Abbrechen durch den Benutzer bei der Eingabe der
    % ling. Terme:
    if abbruch==0;
        % Mitteilung, dass die Berechnung gestartet wurde
        h=msgbox(['Das statistische Fuzzy-Matching berechnet die...
        ...statistischen Zwillinge. Je nach Größe der Ausgangsdaten...
        ...kann dies etwas Zeit in Anspruch nehmen!']);
        % Berechnen der Regelbasis
        regelbasis;
        % Bestimmen der Distanzmatrix
        distanzen;
        % Ermitteln der stat. Zwillinge
        matching;
        % Ausgeben der Datensätze der stat. Zwillinge
        stat_zwillinge;
        h=msgbox(['Die statistischen Zwillinge sind ermittelt...
        ...worden!']);
    else
        % Abbruch des Programms
        msgboxText = (['Abbruch des Programms!']);
        % Das Programm wartet bis die MessageBox geschlossen wird.
        uiwait(msgbox(msgboxText));
    end;
else % Abbruch des Programms
    msgboxText = (['Abbruch des Programms!']);
    %Das Programm wartet bis die MessageBox geschlossen wird.

```

```

    uiwait(msgbox(msgboxText));
end;

```

### Bestimmen der linguistischen Terme und berechnen der Zugehörigkeitsfunktionen

Es werden die Bestimmung der linguistischen Terme und die Berechnung der Zugehörigkeitsfunktionen für bis zu drei linguistische Terme pro Matchingvariable dargestellt. Für mehr als drei linguistische Terme verlaufen die Berechnungen analog. Darüber hinaus wird die automatische Bestimmung von Zugehörigkeitsfunktionen durch die Fuzzy c-Means Clusteranalyse gezeigt.

```

% Festlegen der ling. Terme und berechnen der Zugehörigkeits-
% funktionen.
% Bestimmen der Anzahlen Cases und Controls
Anz_Cases=length(Case_Daten(:,1));
Anz_Controls=length(Control_Daten(:,1));
% Die Matrizen, die später die fuzzyfizzierten Ausgangsdaten in den
% Matchingvariablen aufnehmen sollen, werden zunächst pre-allokiert.
mu_Cases=zeros(Anz_Cases,max(Anz_LingTerme),Anz_Matchingvar);
mu_Controls=zeros(Anz_Controls,max(Anz_LingTerme),Anz_Matchingvar);

% Sortieren der Case- und Controldaten in eine zufällige Reihenfolge:
r=randperm(Anz_Cases)';
% Erzeugt einen Spaltenvektor mit zufälliger Anordnung der Zahlen
% von 1 bis Anz_Cases.
Case_Daten=Case_Daten(r,:);
% Sortiert die Case_Daten entsprechend der Reihenfolge der Zeilen
% in r.
r=randperm(Anz_Controls)';
% Erzeugt einen Spaltenvektor mit zufälliger Anordnung der Zahlen
% von 1 bis Anz_Controls.
Control_Daten=Control_Daten(r,:);
% Sortiert die Control_Daten entsprechend der Reihenfolge der
% Zeilen in r.
if fcmund_verkn==0 & fcmoder_verkn==0
    global form1; global form2; global form3;global form4;global
    form5; global form6;global form7; global form8; global form9;
    global ecken1_1;global ecken1_2;global ecken1_3;global ecken1_4;
    global ecken2_1;global ecken2_2;global ecken2_3;global ecken2_4;
    global ecken3_1;global ecken3_2;global ecken3_3;global ecken3_4;
    global ecken4_1;global ecken4_2;global ecken4_3;global ecken4_4;
    global ecken5_1;global ecken5_2;global ecken5_3;global ecken5_4;

```

```

global ecken6_1;global ecken6_2;global ecken6_3;global ecken6_4;
global ecken7_1;global ecken7_2;global ecken7_3;global ecken7_4;
global ecken8_1;global ecken8_2;global ecken8_3;global ecken8_4;
global ecken9_1;global ecken9_2;global ecken9_3;global ecken9_4;

% Abfragen der ling. Terme für alle Matchingvariablen:
for index1=1:Anz_Matchingvar
    if Anz_LingTerme(index1,1)==1
        % Ausgeben der Meldung zur Eingabe der linguistischen Terme
        % für die index1-te Matchingvariable
        msgboxText = ([ 'Bitte geben Sie die Parameter der...
        ...linguistischen Terme für die ' num2str(index1) ' . ...
        ...Matchingvariable ein. ']);
        % Das Programm wartet mit der weiteren Ausführung so lange,
        % bis die Messagebox geschlossen wird:
        uiwait(msgbox(msgboxText,'Eingabe ling. Terme'));
        % Aufruf des entsprechenden Eingabefensters:
        lingterme_1;
        % Der Ablauf des Programms wird an dieser Stelle so lange
        % gestoppt, bis der Benutzer seine Eingabe gemacht hat.
        uiwait(lingterme_1);
        % Entsprechend der Form des ling. Terms werden die Arrays
        % mit den Informationen über die Ecken geschrieben.
        if form1==1
            ecken1=[ecken1_1 ecken1_2 ecken1_3 ecken1_4];
            % Trapezfunktion
            mu_1=@(x)trapmf(x,[ecken1]);
        elseif form1==2
            ecken1=[ecken1_1 ecken1_2 ecken1_3];
            mu_1=@(x)trimf(x,[ecken1]);
        elseif form1==3
            ecken1=[ecken1_1 ecken1_1 ecken1_1];
            mu_1=@(x)trimf(x,[ecken1]);
        end;
        % Fuzzyfizieren der Ausgangsdaten
        % Die fuzzyfizierten Daten der Cases und Controls werden in
        % zwei unterschiedlichen Matrizen gespeichert. Für jede
        % Matchingvariable werden die fuzzyfizierten Daten in eine
        % neue "dritte" Dimension der Matrix geschrieben.
        mu_Cases(:,1:Anz_LingTerme(index1,1),index1)=...
        ... [mu_1(Case_Daten(:,Matchingvar_Cases(index1)))];
        mu_Controls(:,1:Anz_LingTerme(index1,1),index1)=...

```

```

...[mu_1(Control_Daten(:,Matchingvar_Controls(index1)))];
elseif Anz_LingTerme(index1,1)==2
% Ausgeben der Meldung zur Eingabe der linguistischen Terme
% für die index1-te Matchingvariable:
msgboxText = (['Bitte geben Sie die Parameter ...
...der linguistischen Terme für die ' num2str(index1) '. ...
...Matchingvariable ein.']);
% Das Programm wartet mit der weiteren Ausführung so lange,
% bis die MessageBox geschlossen wird:
uiwait(msgbox(msgboxText,'Eingabe ling. Terme'));
lingterme_2;
uiwait(lingterme_2);
% Entsprechend der Form des ling. Terms werden die Arrays mit
% den Informationen über die Ecken geschrieben.
if form1==1
    ecken1=[ecken1_1 ecken1_2 ecken1_3 ecken1_4];
    % Trapezfunktion
    mu_1=@(x)trapmf(x,[ecken1]);
elseif form1==2
    ecken1=[ecken1_1 ecken1_2 ecken1_3];
    mu_1=@(x)trimf(x,[ecken1]);
elseif form1==3
    ecken1=[ecken1_1 ecken1_1 ecken1_1];
    mu_1=@(x)trimf(x,[ecken1]);
end;
%Fuzzifizieren der Ausgangsdaten
if form2==1
    ecken2=[ecken2_1 ecken2_2 ecken2_3 ecken2_4];
    mu_2=@(x)trapmf(x,[ecken2]);
elseif form2==2
    ecken2=[ecken2_1 ecken2_2 ecken2_3];
    mu_2=@(x)trimf(x,[ecken2]);
elseif form2==3
    ecken2=[ecken2_1 ecken2_1 ecken2_1];
    mu_2=@(x)trimf(x,[ecken2]);
end;
%Fuzzifizieren der Ausgangsdaten
mu_Cases(:,1:Anz_LingTerme(index1,1),index1)=...
...[mu_1(Case_Daten(:,Matchingvar_Cases(index1))) ...
...mu_2(Case_Daten(:,Matchingvar_Cases(index1)))];
mu_Controls(:,1:Anz_LingTerme(index1,1),index1)=...
...[mu_1(Control_Daten(:,Matchingvar_Controls(index1))) ...

```

```

        ...mu_2(Control_Daten(:,Matchingvar_Controls(index1))))];
elseif Anz_LingTerme(index1,1)==3
    % Ausgeben der Meldung zur Eingabe der linguistischen Terme
    % für die index1-te Matchingvariable:
    msgboxText = ([ 'Bitte geben Sie die Parameter ...
...der linguistischen Terme für die ' num2str(index1) '. ...
...Matchingvariable ein.']);
    % Das Programm wartet mit der weiteren Ausführung so lange,
    % bis die MessageBox geschlossen wird:
    uiwait(msgbox(msgboxText,'Eingabe ling. Terme'));
    lingterme_3;
    uiwait(lingterme_3);
    % Entsprechend der Form des ling. Terms werden die Arrays
    % mit den Informationen über die Ecken geschrieben.
    if form1==1
        ecken1=[ecken1_1 ecken1_2 ecken1_3 ecken1_4];
        % Trapezfunktion
        mu_1=@(x)trapmf(x,[ecken1]);
    elseif form1==2
        ecken1=[ecken1_1 ecken1_2 ecken1_3];
        mu_1=@(x)trimf(x,[ecken1]);
    elseif form1==3
        ecken1=[ecken1_1 ecken1_1 ecken1_1];
        mu_1=@(x)trimf(x,[ecken1]);
    end;
    %Fuzzyfizieren der Ausgangsdaten
    if form2==1
        ecken2=[ecken2_1 ecken2_2 ecken2_3 ecken2_4];
        mu_2=@(x)trapmf(x,[ecken2]);
    elseif form2==2
        ecken2=[ecken2_1 ecken2_2 ecken2_3];
        mu_2=@(x)trimf(x,[ecken2]);
    elseif form2==3
        ecken2=[ecken2_1 ecken2_1 ecken2_1];
        mu_2=@(x)trimf(x,[ecken2]);
    end;
    % Fuzzyfizieren der Ausgangsdaten
    if form3==1
        ecken3=[ecken3_1 ecken3_2 ecken3_3 ecken3_4];
        mu_3=@(x)trapmf(x,[ecken3]);
    elseif form3==2
        ecken3=[ecken3_1 ecken3_2 ecken3_3];

```

```

        mu_3=@(x)trimf(x,[ecken3]);
    elseif form3==3
        ecken3=[ecken3_1 ecken3_1 ecken3_1];
        mu_3=@(x)trimf(x,[ecken3]);
    end;
    %Fuzzyfizieren der Ausgangsdaten
    mu_Cases(:,1:Anz_LingTerme(index1,1),index1)=...
    ...[mu_1(Case_Daten(:,Matchingvar_Cases(index1)))
        ...mu_2(Case_Daten(:,Matchingvar_Cases(index1))) ...
        ...mu_3(Case_Daten(:,Matchingvar_Cases(index1)))];
    mu_Controls(:,1:Anz_LingTerme(index1,1),index1)=...
    ...[mu_1(Control_Daten(:,Matchingvar_Controls(index1)))
        ...mu_2(Control_Daten(:,Matchingvar_Controls(index1))) ...
        ...mu_3(Control_Daten(:,Matchingvar_Controls(index1)))];
    end;%if
end;%for

else % fcmund_verkn==1 | fcmoder_verkn==1
    % Fuzzy Clustering: FCM
    Anzahl_Cluster = Anz_LingTerme;
    for index=1:Anz_Matchingvar
        data = [Case_Daten(:,Matchingvar_Cases(index));...
            ...Control_Daten(:,Matchingvar_Controls(index))];
        [center,U,obj_fcn] = fcm(data, Anzahl_Cluster(index));
        U=U';
        U_Cases{1,index}=[U(1:Anz_Cases,1:Anzahl_Cluster(index))];
        U_Controls{1,index}=U((Anz_Cases+1):...
            ...(Anz_Cases+Anz_Controls),1:Anzahl_Cluster(index));
    end;%for
end;%if

```

## Berechnen der Zugehörigkeitsgrade der Datensätze zur Regelbasis

```

% Grundlage bilden die Matrizen mit den fuzzyfizierten Ausgangsdaten.
% Es werden die Zugehörigkeitsgrade eines bestimmten Cases oder eines
% bestimmten Controls zu den jeweiligen ling. Variablen resp. Termen
% betrachtet.

```

```

%Berechnen der Anzahl Regeln und initialisieren der Variablen:
Anz_Regeln=1;
for index=1:Anz_Matchingvar
    % Die Anzahl Regeln ergibt sich aus dem Produkt der Anzahlen

```



```

% ling. Terme:
Anz_Regeln=Anz_Regeln*Anz_LingTerme(index);
end;
C={};% Hilfsvariable C

% Falls beim prod-sum-Operator keine Gewichte vom Nutzer angegeben
% werden, wird die Gleichgewichtung aller Variablen verwendet.
% Mittels der Funktion repmat wird ein Zeilenvektor erzeugt, der
% das Gewicht (1/n) so oft repliziert, wie Matchingvariablen
% vorhanden sind:
if isempty(gewichte) % keine Gewichte angegeben
    delta=repmat(1/Anz_Matchingvar,1,Anz_Matchingvar);
    delta=delta';
else
    delta=gewichte;
end;
mu_Regelbasis_Cases_ges=zeros(Anz_Cases,Anz_Regeln);% Preallocation
if fcmund_verkn==0 & fcmoder_verkn==0
    for index1=1:Anz_Cases
        for index2=1:Anz_Matchingvar
            % Es werden die Zeilen der Matrix mu_Cases verwendet, die
            % den fuzzyfizierten Ausgangsdaten jedem Datensatz der
            % Cases entsprechen. Die Matrix C ist ein CellArray und
            % speichert die Datensätze aus mu_Cases entsprechend der
            % Anzahl ling. Terme für die jeweilige Matchingvariable in
            % der ersten Zeile:
            C(1,index2)={mu_Cases(index1,1:Anz_LingTerme(index2),...
                ...index2)};
        end;
        % Die Funktion allcomb berechnet alle möglichen Kombinationen
        % zwischen den Vektoren von C. Zur Bestimmung der "und"-
        % Verknüpfung wird das Minimum jeder möglichen Kombination
        % gebildet. Da nur die erste Zeile des Cellarrays C besetzt
        % ist, muss nur diese bei der Ermittlung aller möglichen
        % Kombinationen berücksichtigt werden.
        if und_verkn==1
            % Bestimmung des Minimums
            mu_Regelbasis_Cases_ges(index1,:) = min(allcomb(C{1,:}));
        elseif oder_verkn==1
            % Bestimmung des Maximums
            mu_Regelbasis_Cases_ges(index1,:) = max(allcomb(C{1,:}));
        elseif minmax_verkn==1

```

```

% Bestimmung der Linearkombination aus Minimum und
% Maximum zur Berechnung der kompensatorischen
% "minmax"-Verknüpfung
mu_Regelbasis_Cases_ges(index1,:)=...
...Kompensationsgrad*min(allcomb(C{1,:})')+...
...((1-Kompensationsgrad)*max(allcomb(C{1,:})'));
elseif prodsum_verkn==1
    dummy1=[];% Hilfsvariable dummy1
    dummy1=allcomb(C{1,:})';
    % Bestimmung der Linearkombination aus alg. Produkt und
    % alg. Summe zur Berechnung der kompensatorischen
    % "prodsum"-Verknüpfung. Zunächst wird eine Matrix mit
    % den Gewichten delta mittels der Funktion repmat
    % erzeugt. Dabei wird der Spaltenvektor der Gewichte so
    % oft repliziert, wie Anzahlen ling. Terme vorhanden
    % sind. Dann werden die Zugehörigkeitsgrade zu den mit
    % allcomb erzeugten möglichen Kombinationen der ling.
    % Terme mit den Gewichten potenziert, das Produkt
    % und abschließend die Linearkombination gebildet.
    dummy2=[];
    dummy2=repmat(delta,1,Anz_Regeln);
    mu_Regelbasis_Cases_ges(index1,:)=...
    ...Kompensationsgrad*prod(dummy1.^dummy2)+...
    ...((1-Kompensationsgrad)*...
    ...((1-(prod((1-dummy1).^dummy2)))));
end;
end;
elseif fcmund_verkn==1
    for index1=1:Anz_Cases
        % Bestimmung des Minimums
        for index2=1:Anz_Matchingvar
            dummy2=U_Cases{1,index2};
            C(1,index2)={dummy2(index1,:)};
        end;
        mu_Regelbasis_Cases_ges(index1,:) = min(allcomb(C{1,:})');
    end;
else % fcmoder_verkn=1
    for index1=1:Anz_Cases
        % Bestimmung des Maximums
        for index2=1:Anz_Matchingvar
            dummy2=U_Cases{1,index2};
            C(1,index2)={dummy2(index1,:)};
        end;
    end;
end;

```

```

    end;
    mu_Regelbasis_Cases_ges(index1,:) = max(allcomb(C{1,:})');
end;
end;

% Und das gleiche noch einmal für die Controls...
%Preallocation
mu_Regelbasis_Controls_ges=zeros(Anz_Controls,Anz_Regeln);
if fcmund_verkn==0 & fcmoder_verkn==0
    for index1=1:Anz_Controls
        for index2=1:Anz_Matchingvar
            % Es werden die Zeilen der Matrix mu_Controls verwendet,
            % die den fuzzyfizzierten Ausgangsdaten jedem Datensatz der
            % Controls entsprechen.
            C(1,index2)={mu_Controls(index1,1:Anz_LingTerme(index2),...
                ...index2)};
        end;
    end;
    if und_verkn==1
        %Bestimmung des Minimums
        mu_Regelbasis_Controls_ges(index1,:) = ...
            ... min(allcomb(C{1,:})');
    elseif oder_verkn==1
        % Bestimmung des Maximums
        mu_Regelbasis_Controls_ges(index1,:) = ...
            ... max(allcomb(C{1,:})');
    elseif minmax_verkn==1
        % Bestimmung der Linearkombination aus Minimum und Maximum
        % zur Berechnung der kompensatorischen "minmax"-Verknüpfung
        mu_Regelbasis_Controls_ges(index1,:)=...
            ...Kompensationsgrad*min(allcomb(C{1,:})')+...
            ... (1-Kompensationsgrad)*max(allcomb(C{1,:})');
    elseif prodsum_verkn==1
        dummy1=[]; % Hilfsvariable dummy1
        dummy1=allcomb(C{1,:})';
        dummy2=[];
        dummy2=repmat(delta,1,Anz_Regeln);
        mu_Regelbasis_Controls_ges(index1,:)=...
            ...Kompensationsgrad*prod(dummy1.~dummy2)+...
            ...((1-Kompensationsgrad)*...
            ... (1-(prod((1-dummy1).~dummy2)))));
    end;
end;
end;

```

```

elseif fcmund_verkn==1
    for index1=1:Anz_Controls
        % Bestimmung des Minimums
        dummy2=[];
        for index2=1:Anz_Matchingvar
            dummy2=U_Controls{1,index2};
            C(1,index2)={dummy2(index1,:)};
        end;
        mu_Regelbasis_Controls_ges(index1,:) = min(allcomb(C{1,:})');
    end;
else %fcmoder_verkn=1
    for index1=1:Anz_Controls
        % Bestimmung des Maximums
        for index2=1:Anz_Matchingvar
            dummy2=U_Controls{1,index2};
            C(1,index2)={dummy2(index1,:)};
        end;
        mu_Regelbasis_Controls_ges(index1,:) = max(allcomb(C{1,:})');
    end;
end;
end;

```

## Bestimmen der Distanzen

% In jeder Spalte stehen die Distanzen zwischen einem Case und  
 % allen Controls. D.h. in Spalte 1 stehen die Distanzen zwischen  
 % Case 1 und allen Controls. Diese Art der Speicherung ist  
 % notwendig für die folgende Bestimmung der Distanzen.

```

dist = zeros(Anz_Controls,Anz_Cases);%Preallocation
% Bestimmung der absoluten Distanz:
if abs_dist==1
    for index=1:Anz_Cases
        % Ausschneiden einer Zeile der Cases
        zeile=mu_Regelbasis_Cases_ges(index,:);
        % Bestimmen der Distanz indem die ausgeschnittene Zeile auf
        % die Anzahl Controls repliziert wird und anschließend die
        % absolute Differenz zwischen den Zeilen der so erzeugten
        % Matrizen bestimmt werden. Ergebnis ist ein Zeilenvektor,
        % der die Distanz von einem Case zu allen Controls
        % beinhaltet.
        dist(:,index)=nansum(abs(zeile(ones(1,Anz_Controls),:)-...
        ...mu_Regelbasis_Controls_ges),2);
    end;
end;

```

```

% Bestimmung der euklidischen Distanz:
elseif eukl_dist==1
    for index=1:Anz_Cases
        zeile=mu_Regelbasis_Cases_ges(index,:);
        dist(:,index)=sqrt(nansum((zeile(ones(1,Anz_Controls),:)-...
            ..mu_Regelbasis_Controls_ges).^2,2));
    end;
end;

```

## Identifizieren der statistischen Zwillinge

```

% Einlesen der IDs der Cases und Controls
Case_IDS=Case_Daten(:,Case_ID);
Control_IDS=Control_Daten(:,Control_ID);

% Vorbereitungen, falls Perfect Matches vorkommen:
if length(perfect_matches)>0
    for index1=1:length(perfect_matches)
        for index=1:Anz_Cases
            % Bestimmen der Indizes der kritischen Variablen in der
            % Menge der Matchingvariablen der Cases.
            index2(index1)=...
                ...find(Matchingvar_Cases==perfect_matches(index1));
            % Aufbau der Matrix mit den Daten der kritischen Variablen
            % der Cases.
            PMCases=Case_Daten(:,Matchingvar_Cases(index2(index1)));
            % Aufbau der Matrix mit den Daten der kritischen Variablen
            % der Controls.
            PMControls=...
                ...Control_Daten(:,Matchingvar_Controls(index2(index1)));
            % Der Wert der Cases in der kritischen Matchingvariablen
            % wird mittels repmat so oft dupliziert, wie Controls vor-
            % handen sind und davon dann die Ausprägungen der Controls
            % in der kritischen Variablen subtrahiert. In der Ergebnis-
            % matrix stehen also an den Stellen Nullen, wo Case und
            % Control in der kritischen Variablen übereinstimmen. Die
            % Cases sind in den Spalten untergebracht, die Controls
            % in den Zeilen. In Spalte 1 stehen also die Differenzen
            % in der kritischen Variablen von Case 1 zu allen Controls.
            PMMatrix(:,index)=...
                ...abs(repmat(PMCases(index,1),Anz_Controls,1)-PMControls);
            % Mit Hilfe des Befehls "find" werden die Positionen der

```

```

% Nullen in der Matrix ausgemacht. Diese stellen die Basis
% für die Suche nach einem stat. Zwilling dar. Controls,
% die keine Übereinstimmung mit dem jeweiligen Case in der
% krit. Variablen haben, werden aus der weiteren Untersu-
% chung ausgeschlossen, indem ihr Eintrag in der Distanz-
% Matrix als "inf" gekennzeichnet wird.
i=find(PMMatrix(:,index)~=0);
dist([i],index)=inf;
end;
end;
% Ein Else-Pfad ist an dieser Stelle nicht notwendig.
end;
Stat_Zwillinge=zeros(Anz_Cases,2);%Preallocation
if constrained==1% Constrained Matching
    if minimierung==1 %Minimierung der Gesamtdistanz
        % Falls NaN in dist vorkommen, werden diese zu inf gemacht:
        i=0;
        i=find(isnan(dist)~=0);
        dist(i)=inf;
        % Aufruf der Funktion assignmentoptimal mit der Distanzmatrix
        % zur Bestimmung der optimalen Zuweisung nach dem
        % Kuhn-Munkres-Algorithmus:
        [assignment, cost] = assignmentoptimal(dist);
        zaehler=0;
        for index=1:Anz_Controls
            % Die Funktion assignmentoptimal weist jedem Control einen
            % Case oder 0 zu. Betrachtet werden nur die Fälle, in denen
            % ein Case zugewiesen wurde:
            if assignment(index)>0
                zaehler=zaehler+1;
                % In der ersten Spalte von Stat_Zwillinge steht die
                % Case-ID, in der zweiten Spalte die Control-ID.
                Stat_Zwillinge(zaehler,1:2)=...
                    ...[Case_IDs(assignment(index)) Control_IDs(index)];
            end;%drittes if
        end;%for
        % Die Liste der stat. Zwillinge wird nach den Case-IDs sortiert.
        Stat_Zwillinge=sortrows(Stat_Zwillinge,1);
    else %Keine Minimierung der Gesamtdistanz
        % Dürfen Controls nicht mehrfach als stat. Zwillinge verwendet
        % werden und wird die Gesamtdistanz nicht minimiert, dann
        % werden die Distanzen zwischen Cases und Controls zunächst

```

```
% aufsteigend sortiert und jeweils die geringste Distanz zuerst
% betrachtet.
for index=1:Anz_Cases
    % A ist ein Zeilenvektor und speichert die Werte der minimalen
    % Distanzen zwischen einem Case und allen Controls.
    % I1 speichert die zugehörigen Indizes.
    [A,I1]=min(dist);
    % A wird aufsteigend sortiert, d.h. die kleinste Distanz
    % zwischen einem Case und einem Control steht vorne, während
    % die größte Distanz am Ende steht. IX enthält die Indizes,
    % die der ursprünglichen Stelle in A entsprechen.
    [A1,IX]=sort(A);
    % Die tatsächlichen IDs der Cases und Controls müssen
    % entsprechend neu sortiert werden und es muss der Fall
    % abgefangen werden, dass keine Übereinstimmung in den krit.
    % Variablen vorkommt.
    if isnan(A1(1)) | isinf(A1(1)) % A1(1) enthält keinen Wert
        Case_IDs_neu(index,1)=Case_IDs(IX(1),1);
        Case_IDs(IX(1),:)=[];
        % Es wird kein Control zu einem Case gefunden, wenn es
        % keine Übereinstimmung in den kritischen Variablen gibt.
        Control_IDs_neu(index,1)=NaN;
    else
        Case_IDs_neu(index,1)=Case_IDs(IX(1),1);
        Case_IDs(IX(1),:)=[];
        % Hier werden die Controls zu den Cases bestimmt anhand
        % der Reihenfolge der geringsten Distanzen.
        [A2(index),I2(index)]=min(dist(:,IX(1)));
        % Anschließend werden der gematchte Case und Control aus
        % der Matrix der Distanzen gelöscht:
        dist(I2(index),:)=[];
        dist(:,IX(1))=[];
        % Control_IDs_neu enthält die IDs der verwendeten Controls
        Control_IDs_neu(index,1)=Control_IDs(I2(index),1);
        % Die gematchte ID wird aus der Liste entfernt:
        Control_IDs(I2(index),:)=[];
    end;
    % Die Variablen A, I1, A1 und IX werden im nächsten
    % Schleifen-Durchlauf neu bestimmt.
    clear A;clear I1;clear A1;clear IX;
end; % for
Stat_Zwillinge=[Case_IDs_neu Control_IDs_neu];
```

```

        Stat_Zwillinge=sortrows(Stat_Zwillinge,1);
    end;% zweites if
else % Unconstrained Matching
    for index=1:Anz_Cases
        [A,I1]=min(dist);
        [A1,IX]=sort(A);
        if isnan(A1(1)) | isinf(A1(1))
            Case_IDs_neu(index,1)=Case_IDs(IX(1),1);
            Case_IDs(IX(1),:)=[];
            Control_IDs_neu(index,1)=NaN;
        else
            Case_IDs_neu(index,1)=Case_IDs(IX(1),1);
            Case_IDs(IX(1),:)=[];
            % Hier werden die Controls zu den Cases bestimmt anhand der
            % Reihenfolge der geringsten Distanzen:
            [A2(index),I2(index)]=min(dist(:,IX(1)));
            % Anschließend wird der Case aus der Matrix der
            % Distanzen gelöscht:
            dist(:,IX(1))=[];
            % Control_IDs_neu enthält die ID der verwendeten Controls
            Control_IDs_neu(index,1)=Control_IDs(I2(index),1);
        end;%if
        clear A;clear I1;clear A1;clear IX;
    end;%for
    Stat_Zwillinge=[Case_IDs_neu Control_IDs_neu];
    Stat_Zwillinge=sortrows(Stat_Zwillinge,1);
end; % if

```

## Ausgabe der Ergebnisse

```

% Aufbau der Datenmatrix der Statistischen Zwillinge.
% Die Matrix der Cases entspricht der nach Case-IDs aufsteigend
% sortierten Ausgangsmatrix der Case_Daten, da jeder Case zum
% Matching verwendet wird. Die Matrix der Controls ist ein
% Ausschnitt aus der ursprünglichen Matrix der Control_Daten,
% da nicht alle Controls zum Matching verwendet werden bzw.
% manche sogar mehrfach.
Stat_Zwillinge_Cases=sortrows(Case_Daten,Case_ID);
Stat_Zwillinge_Cases_IDs=Stat_Zwillinge_Cases(:,Case_ID);
Stat_Zwillinge_Controls=...
...zeros(size(Stat_Zwillinge,1),size(Control_Daten,2));
for i=1:size(Stat_Zwillinge,1)

```



```
for j=1:size(Control_Daten,1)
    if Stat_Zwillinge(i,2)==Control_Daten(j,Control_ID)
        Stat_Zwillinge_Controls(i,:) = Control_Daten(j,:);
    elseif isnan(Stat_Zwillinge(i,2))
        Stat_Zwillinge_Controls(i,:) = NaN;
    end;
end;
end;
Stat_Zwillinge_Controls_IDs=Stat_Zwillinge_Controls(:,Control_ID);
csvwrite('Stat_Zwillinge_Cases.csv',Stat_Zwillinge_Cases);
csvwrite('Stat_Zwillinge_Controls.csv',Stat_Zwillinge_Controls);
xlswrite('Stat_Zwillinge_Cases_IDs.xls',Stat_Zwillinge_Cases_IDs);
xlswrite('Stat_Zwillinge_Controls_IDs.xls',...
...Stat_Zwillinge_Controls_IDs);
```

# Literaturverzeichnis

- [AB84] ALDENDERFER, M. S. und R. K. BLASHFIELD: *Cluster Analysis. Quantitative Applications in the Social Sciences*. Sage Publications, Inc., Newbury Park, 1984.
- [AB08] ALPAR, P. und S. BLASCHKE: *Einleitung*. In: ALPAR, P. und S. BLASCHKE (Herausgeber): *Web 2.0 - Eine empirische Bestandsaufnahme*, Seiten 3–14. Vieweg + Teubner, Wiesbaden, 2008.
- [ABK07] ALPAR, P., S. BLASCHKE und S. KESSLER: *Web 2.0: Neue erfolgreiche Kommunikationsstrategien für kleine und mittlere Unternehmen*. Hessen-Media Band 57. Hessen Agentur GmbH, Wiesbaden, 2007.
- [ABRW07] ANGERMANN, A., M. BEUSCHEL, M. RAU und U. WOHLFARTH: *Matlab - Simulink - Stateflow*. Oldenbourg, München, 5. Auflage, 2007.
- [ACA93] ABDULGHAFOUR, M., T. CHANDRA und M. A. ABIDI: *Data Fusion Through Fuzzy Logic Applied to Feature Extraction from Multi-Sensory Images*. In: *IEEE International Conference on Robotics and Automation*, Band 2, Seiten 359–366, 1993.
- [AH05] AGARWAL, S. und P. HITZLER: *Modeling Fuzzy Rules with Description Logics*. In: *Proceedings of Workshop on OWL Experiences and Directions*, Seiten 1–10, Galway, 2005.
- [Alb05] ALBRECHT, C.: *Automatisierte Daten-Transformation im GALA-Projekt*. Interner Bericht FZJ-ZAM-IB-2005-13, Forschungszentrum Jülich, 2005. Online verfügbar unter <http://www.fz-juelich.de/jsc/math/RD/projects/gruenenthal/Claudia-Albrecht-Autom-Datentransformation.pdf>, abgerufen am 18.10.2008.
- [Alp04] ALPAR, P.: *What Data Is Necessary to Data Mine for Knowledge?* In: *ICEB International Conference on Electronic Business*, Seiten 1219–1223, 2004.

- [AN00] ALPAR, P. und J. NIEDEREICHHOLZ: *Einführung zu Data Mining*. In: ALPAR, P. und J. NIEDEREICHHOLZ (Herausgeber): *Data Mining im praktischen Einsatz*, Seiten 1–28. Vieweg, Braunschweig, 2000.
- [AP02] ALMUS, M. und S. PRANTL: *Die Auswirkungen öffentlicher Gründungsförderung auf das Überleben und Wachstum junger Unternehmen*. Jahrbücher für Nationalökonomie und Statistik, 222(2):161–185, 2002.
- [AR05] ANTON, H. und C. RORRES: *Elementary Linear Algebra*. John Wiley & Sons, Inc., Hoboken, 9. Auflage, 2005.
- [Arb08] ARBEITSGEMEINSCHAFT ONLINE-FORSCHUNG E.V.: *Berichtsband Teil 1 zur internet facts 2008-I*, 2008. Online verfügbar unter <http://www.agof.de/studie.353.html>, abgerufen am 21.09.2008.
- [Att06] ATTESLANDER, P.: *Methoden der empirischen Sozialforschung*. Erich Schmidt Verlag GmbH & Co., Berlin, 11. Auflage, 2006.
- [Bö93] BÖHME, G.: *Fuzzy Logik: Einführung in die algebraischen und logischen Grundlagen*. Springer, Berlin u.a., 1993.
- [BA94] BLAND, J. M. und D. G. ALTMAN: *Statistic Notes: Matching*. BMJ, 309:1128, 1994.
- [Bac02] BACHER, J.: *Statistisches Matching: Anwendungsmöglichkeiten, Verfahren und ihre praktische Umsetzung in SPSS*. ZA-Informationen, 51:38–66, 2002.
- [Bar07] BARTEL, R.: *Blogs für alle: Das Weblog-Kompendium*. Smart Books Publishing AG, Baar, 2007.
- [Bau02] BAUNE, A. K.: *Dynamische Clusteranalyse-Verfahren, ihre Bewertung und deren Anwendung im medizinischen Umfeld*. Dissertation, Fakultät für Informatik der Universität Ulm, 2002.
- [BB07] BOJADZIEV, G. und M. BOJADZIEV: *Fuzzy Logic for Business, Finance, and Management*. World Scientific Publishing Co. Pte. Ltd., Singapur, 2. Auflage, 2007.

- [BC64] BOX, G. E. P. und D. R. COX: *An Analysis of Transformation*. Journal of the Royal Statistical Society, Series B(26):211–252, 1964.
- [BD86] BONISSONE, P. P. und K. S. DECKER: *Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-off Precision and Complexity*. In: KANAL, L. N. und J. F. LEMMER (Herausgeber): *Uncertainty in Artificial Intelligence*, Seiten 217–247. North-Holland, Amsterdam, 1986.
- [BD02] BORTZ, J. und N. DÖRING: *Forschungsmethoden und Evaluation*. Springer, Heidelberg, 2002.
- [BE07] BONTSCHEV, G. und M. ELING: *Wo investieren Distressed-Securities-Hedgfonds? Ein Asset-based Style-Faktorenmodell*. Working Papers on Risk Management and Insurance No. 47, Institute of Insurance Economics, Universität St. Gallen, St. Gallen, 2007.
- [Bed90] BEDER, J. H.: *On the Use of RIDIT Analysis*. Psychometrika, 55(4):603–616, 1990.
- [Ben01] BENKER, H.: *Statistik mit MATHCAT und MATLAB*. Addison-Wesley, Berlin, 2001.
- [Ben05] BENNINGHAUS, H.: *Einführung in die sozialwissenschaftliche Datenanalyse*. Oldenbourg, München, 2005.
- [BEPW06] BACKHAUS, K., B. ERICHSON, W. PLINKE und R. WEIBER: *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Springer, Berlin u.a., 11. Auflage, 2006.
- [Bez73] BEZDEK, J. C.: *Fuzzy Mathematics in Pattern Classification*. Dissertation, Applied Math. Center, Cornell University Ithaca, 1973.
- [BG93] BANDEMER, H. und S. GOTTWALD: *Einführung in Fuzzy-Methoden*. Akademie Verlag, Berlin, 4. Auflage, 1993.
- [Bie97] BIEWER, B.: *Fuzzy-Methoden: Praxisrelevante Rechenmodelle und Fuzzy-Programmiersprachen*. Springer, Berlin, 1997.
- [BIT08] BITKOM: *Deutschland holt bei Breitband auf*, 2008. Online verfügbar unter [http://www.bitkom.org/de/presse/30739\\_52242.aspx](http://www.bitkom.org/de/presse/30739_52242.aspx), abgerufen am 15.09.2008.

- [BKl00] BEIERLE, C. und G. KERN-ISBERNER: *Methoden wissensbasierter Systeme*. Vieweg, Wiesbaden, 2000.
- [BKKN03] BORGELT, C., F. KLAWONN, R. KRUSE und D. NAUCK: *Neuro-Fuzzy-Systeme*. Vieweg, Wiesbaden, 2003.
- [BL71] BOURGEOIS, F. und J.-C. LASSALLE: *An Extension of the Munkres Algorithm for the Assignment Problem to Rectangular Matrices*. Communications of the ACM, 14(12):802–804, 1971.
- [Bla01] BLASIUS, J.: *Korrespondenzanalyse*. Oldenbourg, München, 2001.
- [Ble04] BLEIHOLDER, J.: *Techniken des Data Merging in Integrationssystemen*. In: SAMIA, M. und S. CONRAD (Herausgeber): *Tagungsband zum 16. GI-Workshop über Grundlagen von Datenbanken*, Seiten 23–27. Institute of Computer Science, Heinrich-Heine Universität Düsseldorf, 2004.
- [Blo05] BLOHM, M.: *Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)*. In: GRÖZINGER, G. und W. MATIASKE (Herausgeber): *Deutschland regional. Sozialwissenschaftliche Daten im Forschungsverbund*, Seiten 43–55. Hampp Verlag, München, 2005.
- [Boc74] BOCK, H. H.: *Automatische Klassifikation: Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten*. Vandenhoeck & Ruprecht, Göttingen, 1974.
- [Bon95] BONITZ, H.: *Aristoteles: Philosophische Schriften*, Band 5. Meiner Verlag, Hamburg, 1995.
- [Bos98] BOSCH, K.: *Statistik Taschenbuch*. Oldenbourg, München, 3. Auflage, 1998.
- [Bos07] BOSCH, K.: *Basiswissen Statistik*. Oldenbourg, München, 2007.
- [Bro58] BROSS, I. D. J.: *How to use Redit Analysis*. Biometrics, 14:18–38, 1958.
- [Bro06] BRODA, S.: *Marktforschungspraxis: Konzepte, Methoden, Erfahrungen*. Gabler, Wiesbaden, 2006.

- [BT99] BILGIÇ, T. und I. B. TÜRKSEN: *Measurement of Membership Functions: Theoretical and Empirical Work*. In: DUBOIS, D. und H. PRADE (Herausgeber): *Handbook of Fuzzy Sets and Systems*, Band 1, Seiten 195–232. Kluwer, Boston, London, Dordrecht, 1999.
- [BTK03] BORGELT, C., H. TIMM und R. KRUSE: *Unsicheres und vages Wissen*. In: GÖRZ, G., C.-R. ROLLINGER und J. SCHNEEBERGER (Herausgeber): *Handbuch der künstlichen Intelligenz*. Oldenbourg, München, 2003.
- [BVW04] BUHL, H. U., S. VOLKERT und V. WINKLER: *Individualisierte Anlageberatung: Axiomatische Fundierung von Zielfunktionen zur Bewertung von Anlagealternativen*. Technischer Bericht Diskussionspapier WI-150, Lehrstuhl für Betriebswirtschaftslehre, Wirtschaftsinformatik & Financial Engineering Kernkompetenzzentrum IT & Finanzdienstleistungen, Augsburg, 2004.
- [CCK<sup>+</sup>00] CHAPMAN, P., J. CLINTON, R. KERBER, T. KHABAZA, T. REINARTZ, C. SHEARER und R. WIRTH: *CRISP-DM 1.0: Step-by-Step Data Mining Guide*. The CRISP-DM Consortium, Seiten 1–78, 2000. Online verfügbar unter <http://www.crisp-dm.org/CRISPWP-0800.pdf>, abgerufen am 28.11.2008.
- [Cle08] CLEFF, T.: *Deskriptive Statistik und moderne Datenanalyse*. Gabler, Wiesbaden, 2008.
- [Coc53] COCHRAN, W. G.: *Matching in Analytical Studies*. Journal of the American Public Health Association, 43(6):684–691, 1953.
- [Cod70] CODD, E. F.: *A Relational Model of Data for Large Shared Data Banks*. Communications of the ACM, 13(6):377–387, 1970.
- [Cox94] COX, E.: *The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems*. AP Professional, Boston, 1994.
- [Cro51] CRONBACH, L. J.: *Coefficient Alpha and the Internal Structure of Tests*. Psychometrika, 16:297–334, 1951.
- [Cun06] CUNNINGHAM, J.: *Determining an Optimal Membership Function Based on Community Consensus in a Fuzzy Database System*. In: *Proceedings of the 44th Annual Southeast Regional Conference*,

- Seiten 632–637, New York, 2006. ACM Press. Session: Database Systems and Computer Vision.
- [D'A98] D'AGOSTINO, R. B.: *Tutorial in Biostatistics: Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group*. *Statistics in Medicine*, 17:2265–2281, 1998.
- [DKV<sup>+</sup>04] DRIGAS, A., S. KOUREMENOS, S. VRETTOS, J. VRETTAROS und D. KOUREMENOS: *An Expert System for Job Matching of the Unemployed*. *Expert Systems with Applications*, 26:217–224, 2004.
- [DL97] DEZA, M. M. und M. LAURENT: *Geometry of Cuts and Metrics*. Springer, Berlin u.a., 1997.
- [DP80] DUBOIS, D. und H. PRADE: *Fuzzy Sets and Systems: Theory and Applications*, Band 144 der Reihe *Mathematics in Science and Engineering*. Academic Press, New York, 1980.
- [DP85] DUBOIS, D. und H. PRADE: *A Review of Fuzzy Set Aggregation Connectives*. *Information Science*, 36:85–121, 1985.
- [dTC06] DE CARVALHO DE A.T., F., C. P. TENÓRIO und N. L. CAVALCANTI: *Partitional Fuzzy Clustering Methods Based on Adaptive Quadratic Distances*. *Fuzzy Sets and Systems*, 157:2833–2857, 2006.
- [Dun73] DUNN, J.C.: *A Fuzzy Relative of the ISODATA Process and its use in Detecting Compact Well-Separated Clusters*. *Journal of Cybernetics*, 3:32–57, 1973.
- [DW02] DEHEJIA, R. H. und S. WAHBA: *Propensity Score-Matching Methods For Nonexperimental Causal Studies*. *The Review of Economics and Statistics*, 84(1):151–161, 2002.
- [DZS01] D'ORAZIO, M., M. DI ZIO und M. SCANU: *Statistical Matching: a Tool for Integrating Data in National Statistical Institutes*. In: *Second International Seminar of Exchange of Technology and Know-How / Fourth New Techniques and Technologies for Statistics Seminar*, Seiten 433–440, Crete, 2001.
- [DZS06] D'ORAZIO, M., M. DI ZIO und M. SCANU: *Statistical Matching: Theory and Practice*. John Wiley & Sons, Ltd., West Sussex, 2006.

- [Eic95] EICHLER, M.: *Methoden zur automatischen Klassifikation gekoppelter Meßgrößen*. Seminararbeit am Institut für Phonetik, Johann Wolfgang Goethe Universität Frankfurt, 1995. Online verfügbar unter <http://www.informatik.uni-frankfurt.de/~ifb/exphon/ss95/martin01.html>, abgerufen am 07.07.2007.
- [EKR02] ECKEY, H.-F., R. KOSFELD und M. RENGERS: *Multivariate Statistik*. Gabler, Wiesbaden, 2002.
- [ELL01] EVERITT, B. S., S. LANDAU und M. LEESE: *Cluster Analysis*. Arnold, London, 2001.
- [Eng01] ENGEL, D.: *Höheres Beschäftigungswachstum durch Venture Capital?* Discussion Paper No. 01-34, Zentrum für Europäische Wirtschaftsforschung GmbH (ZEW), Mannheim, 2001.
- [Fai97] FAIR, M. E.: *Record Linkage in an Information Age Society*. In: ALVEY, W. und B. JAMERSON (Herausgeber): *Record Linkage Techniques*, Seiten 427–441, Washington D.C., 1997.
- [Fel06] FELDBUSCH, F.: *Intelligente Datenanalyse*. Vorlesungsskript, Universität Karlsruhe, Sommersemester 2006. Online verfügbar unter [http://ces.univ-karlsruhe.de/teaching/IDA\\_s05/folien/ida-3a-fuzzy.pdf](http://ces.univ-karlsruhe.de/teaching/IDA_s05/folien/ida-3a-fuzzy.pdf), abgerufen am 24.01.2007.
- [FG08] FISCH, M. und C. GSCHIEDLE: *Mitmachnetz 2.0: Rege Beteiligung nur in Communities*. Media Perspektiven, 7:356–364, 2008.
- [FM03] FRANKE, K.-H. und T. MACHLEIDT: *Farbpixelklassifikation mittels partitionierender Clusteralgorithmen und Vergleich mit neuronalen Ansätzen*. Wissenschaftlicher Beitrag, Technische Universität Ilmenau, 2003.
- [FMV08] FUEGLISTALLER, U., C. A. MÜLLER und T. VOLERY: *Entrepreneurship*. Gabler, Wiesbaden, 2. Auflage, 2008.
- [For93] FORSTER, O.: *Analysis 2*. Vieweg, Braunschweig, 1993.
- [FPSS96a] FAYYAD, U., G. PIATETSKY-SHAPIRO und P. SMYTH: *From Data Mining to Knowledge Discovery in Databases*. American Association for Artificial Intelligence, Fall 1996:37–54, 1996.



- [FPSS96b] FAYYAD, U., G. PIATETSKY-SHAPIRO und P. SMYTH: *From Data Mining to Knowledge Discovery in Databases: An Overview*. In: FAYYAD, U., G. PIATETSKY-SHAPIRO, P. SMYTH und R. UTHURUSAMY (Herausgeber): *Advances in Knowledge Discovery and Data Mining*, Seiten 1–30. 1996.
- [Fra02] FRANK, H.: *Fuzzy-Methoden in der Wirtschaftsmathematik*. Vieweg, Wiesbaden, 2002.
- [Fre04] FREISTÄTTER, J.: *Wissensmanagement als Basis für wissensbasiertes Management*. Vielfalt in Uniform, 4:1–70, 2004.
- [Fro07] FROESE, N.: *Aristoteles: Logik und Methodik in der Antike*, 2007. Online verfügbar unter <http://www.antike-griechische.de/Aristoteles.pdf>, abgerufen am 10.03.2008.
- [GBG04] GUO, S., R. BARTH und C. GIBBONS: *Introduction to Propensity Score Matching: A New Device for Program Evaluation*. Workshop Presented at the Annual Conference of the Society for Social Work Research, New Orleans, 2004. Online verfügbar unter [http://ssw.unc.edu/VRC/Lectures/PSM\\_SSWR\\_2004.pdf](http://ssw.unc.edu/VRC/Lectures/PSM_SSWR_2004.pdf), abgerufen am 28.02.2007.
- [GD04] GANGL, M. und T. DiPRETE: *Kausalanalyse durch Matchingverfahren*. DIW- Diskussionspapiere 401, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin, 2004.
- [Ger96] GERNERT, D.: *Fuzzy Logic*. Betriebswirtschaftliches Seminar in Automation, Robotik und betriebliche Auswirkungen, Technische Universität München, 1996.
- [Ges06] GESELLSCHAFT SOZIALWISSENSCHAFTLICHER INFRASTRUKTUREINRICHTUNGEN: *Datenhandbuch 2006*. Technischer Bericht, Zentralarchiv für empirische Sozialforschung an der Universität zu Köln und Zentrum für Umfragen, Methoden und Analysen ZUMA in Mannheim, 2006. Online verfügbar unter <http://www.gesis.org/Datenservice/ALLBUS/index.htm>, abgerufen am 24.05.2008.
- [Ges07] GESELLSCHAFT SOZIALWISSENSCHAFTLICHER INFRASTRUKTUREINRICHTUNGEN: *Die allgemeine Bevölkerungsumfrage der*

- Sozialwissenschaften* - ALLBUS, 2007. Online verfügbar unter <http://www.gesis.org/Dauerbeobachtung/Allbus>, abgerufen am 24.05.2008.
- [GF07] GSCHIEDLE, C. und M. FISCH: *Onliner 2007: Das "Mitmach-Netz" im Breitbandzeitalter*. *Media Perspektiven*, 8:393–405, 2007.
- [Gol06] GOLIN, M. J.: *Bipartite Matching & the Hungarian Method*. Course Notes, Hong Kong University of Science and Technology, Hong Kong, 2006.
- [Gro02] GROSSMANN, S.: *Fuzzy Clusteranalyse zur automatischen Erzeugung von Fuzzy-Modellen in praktischen Anwendungen*. Diplomarbeit, Institut für innovative Informatik-Anwendungen, Fachhochschule Trier, November 2002.
- [Gru02] GRUNDMANN, W.: *Operations Research*. Vieweg Verlag, Wiesbaden, 2002.
- [Grz08] GRZEGORZEWSKI, P.: *Trapezoidal Approximations of Fuzzy Numbers Preserving the Expected Interval - Algorithms and Properties*. *Fuzzy Sets and Systems*, 159:1354–1364, 2008.
- [GSB05] GENSLER, S., B. SKIERA und M. BÖHM: *Einsatzmöglichkeiten der Matching Methode zur Berücksichtigung von Selbstselektion*. *Journal für Betriebswirtschaft*, 55(1):37–62, 2005.
- [Göt07] GÖTHLICH, S. E.: *Zum Umgang mit fehlenden Daten in großzahligen empirischen Erhebungen*. In: ALBERS, S., D. KLAPPER, U. KONRADT, A. WALTER und J. WOLF (Herausgeber): *Methodik der empirischen Forschung*, Seiten 119–134. Gabler, Wiesbaden, 2. Auflage, 2007.
- [Gut54] GUTTMAN, L.: *Some Necessary Conditions for Common Factor Analysis*. *Psychometrika*, 19:149–161, 1954.
- [Hae03] HAENDEL, L.: *Clusterverfahren zur datenbasierten Generierung interpretierbarer Regeln unter Verwendung lokaler Entscheidungskriterien*. Dissertation, Fakultät für Elektrotechnik und Informatik der Universität Dortmund, 2003.
- [Haf00] HAFNER, R.: *Statistik für Sozial- und Wirtschaftswissenschaftler*. Band 1. Springer, Wien, 2000.

- [Hai06] HAIR, J. F.: *Multivariate Data Analysis*. Prentice Hall, Upper Saddle River, 6. Auflage, 2006.
- [Häd06] HÄDER, M.: *Empirische Sozialforschung: Eine Einführung*. CRRS Publications, Toronto, 2006.
- [Höf04] HÖFLER, M.: *Statistik in der Epidemiologie psychischer Störungen*. Springer, Berlin, 2004.
- [HH07] HARTMANN, P. H. und I. HÖHNE: *MNT 2.0 - Zur Weiterentwicklung der MedienNutzerTypologie*. Media Perspektiven, 5:235–241, 2007.
- [Hil94] HILLMANN, K.-H.: *Wörterbuch der Soziologie*. Alfred Kröner Verlag, Stuttgart, 4. Auflage, 1994.
- [Hir01] HIRJI, K. K.: *Exploring Data Mining Implementation*. Communications of the ACM, 44(7):87–93, 2001.
- [HIT97] HECKMAN, J. J., H. ICHIMURA und P. TODD: *Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program*. Review of Economic Studies, 64:605–654, 1997.
- [HIT98] HECKMAN, J. J., H. ICHIMURA und P. TODD: *Matching as an Econometric Evaluation Estimator*. Review of Economic Studies, 65:261–294, 1998.
- [HK00] HÖPPNER, F. und F. KLAWONN: *Obtaining Interpretable Fuzzy Models from Fuzzy Clustering and Fuzzy Regression*. In: *Proceedings of the 4th Int. Conf. on Knowledge-Based Intelligent Engineering Systems & Allied Technologies*, Seiten 162–165. Innovation in Knowledge-Based & Intelligent Engineering Systems, 2000.
- [HK06] HAN, J. und M. KAMBER: *Data Mining*. Morgan Kaufmann Publishers, San Francisco, 2006.
- [HKK97] HÖPPNER, F., F. KLAWONN und R. KRUSE: *Fuzzy-Clusteranalyse: Verfahren für die Bilderkennung, Klassifikation und Datenanalyse*. Vieweg, Braunschweig, Wiesbaden, 1997.
- [HKK04] HARTMANN, W., R. KREUTZER und H. KUHFUSS: *Kundenclubs & More*. Gabler, Wiesbaden, 2004.

- [Hof04] HOFFMANN, O.: *Fuzzy Logik und neuronale Netze*. Vorlesungsskript, Fachhochschule Gießen-Friedberg, 2004. Online verfügbar unter [http://homepages.fh-giessen.de/~hg14032/Skripte/S5\\_NNFuzzyLogic\\_Skript.pdf](http://homepages.fh-giessen.de/~hg14032/Skripte/S5_NNFuzzyLogic_Skript.pdf), abgerufen am 16.06.2008.
- [Hol98] HOLZ, R.: *Rating, Ranking, Scoring und Fuzzy Sets: Eine Methoden (Stilelement)-Zusammenführung am Beispiel von LV-Produktatings*. Blätter der DGVM, 23(4):363–384, 1998.
- [Hol03] HOLM, K.: *Almo Statistik-System Handbuch Teil 3b*. Linz, 2003.
- [HRL05] HAM, J. C., P. B. REAGAN und X. LI: *Propensity Score Matching, a Distance-Based Measure of Migration, and the Wage Growth of Young Men*. IEPR Working Paper 05.13, University of Southern California, Institute of Economic Policy Research, Los Angeles, 2005.
- [HTGK07] HAAS, S., T. TRUMP, M. GERHARDS und W. KLINGLER: *Web 2.0: Nutzer und Nutzertypen*. Media Perspektiven, 4:215–222, 2007.
- [HW06] HUPPERT, B. und W. WILLEMS: *Lineare Algebra*. Teubner, Wiesbaden, 2006.
- [IOST00] INGRAM, D. D., J. O’HARE, F. SCHEUREN und J. TUREK: *Statistical Matching: A New Validation Case Study*. In: *Proceedings of the Survey Research Methods Section*, Seiten 746–751. American Statistical Association, 2000.
- [Iwe00] IWE, H.: *Einführung in die Fuzzy-Technologie*. Vorlesungsskript, Hochschule für Technik und Wirtschaft Dresden, 2000. Online verfügbar unter <http://www.informatik.htw-dresden.de/~iwe/lvfuzzy/FuzzySkript.pdf>, abgerufen am 24.01.2007.
- [Jan98] JANTZEN, J.: *Neuro-Fuzzy Modelling*. Technical Report 98-H-874, Technical University of Denmark, Lyngby, 1998.
- [JM96] JAANINEH, G. und M. MAIJOHANN: *Fuzzy-Logik und Fuzzy-Control*. Vogel Buchverlag, Würzburg, 1996.
- [Kad78] KADANE, J. B.: *Some Statistical Problems in Merging Data Files*. Compendium of Tax Research, Washington, D.C., Seiten 159–179, 1978.

- [KD59] KAISER, H. F. und K. DICKMAN: *Analytic Determination of Common Factors*. American Psychological Reports, 14:425–437, 1959.
- [KF88] KLIR, G. J. und T. A. FOLGER: *Fuzzy Sets, Uncertainty, and Information*. Prentice Hall, Englewood Cliffs, 1988.
- [KGG08] KHAN, S., A. R. GANGULY und A. GUPTA: *Data Mining and Data Fusion for Enhanced Decision Support*. In: BURSTEIN, F. und C. W. HOLSAPPLE (Herausgeber): *Handbook on Decision Support Systems*, Seiten 581–608. Springer, Berlin, 2008.
- [KKG93] KRUSE, R., J. GEBHARDT und F. KLAWONN: *Fuzzy-Systeme*. B. G. Teubner, Stuttgart, 1993.
- [KH03a] KLAWONN, F. und F. HÖPPNER: *An Alternative Approach to the Fuzzifier in Fuzzy Clustering to Obtain Better Clustering Results*. In: *Proceedings of 3rd EUSFLAT*, Seiten 730–734, Zittau, 2003.
- [KH03b] KLAWONN, F. und F. HÖPPNER: *What is Fuzzy About Fuzzy Clustering? - Understanding and Improving the Concept of the Fuzzifier*. In: BERTHOLD, M. R., H.-J. LENZ, E. BRADLEY, R. KRUSE und C. BORGELT (Herausgeber): *Advances in Intelligent Data Analysis V*, Seiten 254–264. Springer, Berlin, 2003.
- [KI92] KACPRZYK, J. und C. IWANSKI: *Fuzzy Logic with Linguistic Quantifiers in Inductive Learning*. In: ZADEH, L. und J. KACPRZYK (Herausgeber): *Fuzzy Logic for the Management of Uncertainty*, Seiten 465–478. John Wiley & Sons, Inc., New York, 1992.
- [KL94] KOVACEVIC, M. S. und T.-P. LIU: *Statistical Matching of Survey Datafiles: A Simulation Study*. In: *Proceedings of the Survey Research Methods Section, American Statistical Association*, Seiten 479–484, 1994.
- [Kla04] KLAWONN, F.: *Fuzzy-Clusteranalyse*. CS Research Seminar Data and Knowledge Engineering, Otto-von-Guericke-Universität, Magdeburg, Mai 2004.
- [KM93] KIRCHNER, F. und C. MICHAELIS: *Wörterbuch der Philosophischen Grundbegriffe*. Verlag der Dörr'schen Buchhandlung, Leipzig, 5. Auflage, 1993.

- [KMMW06] KUHLENKAMP, A., S. MANOUCHEHRI, I. MERGEL und U. WINAND: *Privatsphäre versus Erreichbarkeit bei der Nutzung von Social Software*. HMD - Praxis der Wirtschaftsinformatik, 252:27–35, 2006.
- [KMN<sup>+</sup>02] KANUNGO, T., D. M. MOUNT, N. S. NETANYAHU, C. D. PIATKO, R. SILVERMAN und A. Y. WU: *An Efficient k-Means Clustering Algorithm: Analysis and Implementation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):881–892, 2002.
- [KNK97] KRUSE, R., D. NAUCK und F. KLAWONN: *Neuronale Fuzzy-Systeme*. Spektrum der Wissenschaften: Kopf oder Computer, 4:92–99, 1997.
- [Kno96] KNORZ, G.: *Datenbank-Entwurfsmethoden*. In: BUDER, M., W. REHFELD und T. SEEGER (Herausgeber): *Grundlagen der praktischen Information und Dokumentation*, Band 2, Seiten 664–687. K.G. Saur, München, 4. Auflage, 1996.
- [Koh03] KOHLER, T. C.: *Wirkungen des Produktdesigns: Analyse und Messung am Beispiel Automobildesign*. Deutscher Universitätsverlag, Wiesbaden, 2003.
- [Koz82] KOZELKA, R. M.: *How to Work Through a Clustering Problem*. In: H. C. HUDSON (Herausgeber): *Classifying Social Data*, Seiten 1–12. Jossey-Bass, San Francisco, 1982.
- [KR05] KLEINSCHMIDT, P. und C. RANK: *Relationale Datenbanksysteme: Eine praktische Einführung*. Springer, Berlin, 2005.
- [KR06] KIESL, H. und S. RÄSSLER: *How Valid Can Data Fusion Be?* Technischer Bericht Discussionpaper No. 15/2006, Bundesagentur für Arbeit, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg, 2006.
- [KSM<sup>+</sup>07] KUMAR, B. P., J. SELVAM, V. S. MEENAKSHI, K. KANTHI, A. L. SUSEELA und V. L. KUMAR: *Business Decision Making, Management and Information Technology*. Ubiquity, 8(8):1–20, 2007.
- [Kuh55] KUHN, H. W.: *The Hungarian Method for the Assignment Problem*. Naval Research Logistics Quarterly, 2:83–97, 1955.

- [Kun00] KUNCHEVA, L. I.: *How Good are Fuzzy If-Then Classifiers*. IEEE transactions on Systems, MAN and Cybernetics - Part B: Cybernetics, 30(4):501–509, 2000.
- [KW69] KANTOR, S. und W. WINKELSTEIN: *The Rationale and Use of Riddit Analysis in Epidemiologic Studies on Blood Pressure*. American Journal of Epidemiology, 90(3):201–213, 1969.
- [KWHS01] KOCH, A., M. WASMER, J. HARKNESS und E. SCHOLZ: *Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" ALLBUS 2000*. ZUMA-Methodenbericht 2001-05, ZUMA, Mannheim, 2001.
- [Lec98] LECHNER, M.: *Mikroökonomische Evaluationsstudien: Anmerkungen zu Theorie und Praxis*. In: PFEIFFER, F. und W. POHLMAYER (Herausgeber): *Qualifikation, Weiterbildung und Arbeitsmarkterfolg*, ZEW-Wirtschaftsanalysen, Seiten 13–38. Nomos-Verlag, Baden-Baden, 1998.
- [Lit00] LITZ, H. P.: *Multivariate statistische Methoden*. Oldenbourg, München, 2000.
- [LSC05] LUELLEN, J. K., W. R. SHADISH und M. H. CLARK: *Propensity Scores: An Introduction and Experimental Test*. Evaluation Review, 29(6):530–558, 2005.
- [LW67] LANCE, G. N. und W. T. WILLIAMS: *Mixed-Data Classificatory Programs I - Agglomerative Systems*. Australian Computer Journal, 1(1):15–20, 1967.
- [Mah36] MAHALANOBIS, P. C.: *On the Generalised Distance in Statistics*. In: *Proceedings of the National Institute of Science of India*, Seiten 49–55, 1936.
- [MAN08] MAURER, T., P. ALPAR und P. NOLL: *Nutzertypen junger Erwachsener in sozialen Online-Netzwerken in Deutschland*. In: ALPAR, P. und S. BLASCHKE (Herausgeber): *Web 2.0 - Eine empirische Bestandsaufnahme*, Seiten 207–232. Vieweg + Teubner, Wiesbaden, 2008.
- [Mar03] MARTENS, J.: *Statistische Datenanalyse mit SPSS für Windows*. Oldenbourg, München, 2. Auflage, 2003.

- [Mau07] MAURER, T.: *Mehrwert sozialer Online-Netzwerke aus Benutzersicht: Eine empirische Untersuchung*. Diplomarbeit, Institut für Wirtschaftsinformatik, Philipps-Universität Marburg, Juli 2007.
- [MB01] MIELKE, P. W. und K. J. BERRY: *Permutation Methods: A Distance Function Approach*. Springer, New York u.a., 2001.
- [Men51] MENGER, K.: *Ensembles flous et fonctions aléatoires*. Comptes Rendus Académie des Sciences, 37:2001–2003, 1951.
- [MH01] MILLER, H. J. und J. HAN: *Geographic Data Mining and Knowledge Discovery: An Overview*. In: MILLER, H. J. und J. HAN (Herausgeber): *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, London, New York, 2001.
- [Miz89] MIZUMOTO, M.: *Pictorial Representations of Fuzzy Connectives, Part I: Cases of  $t$ -Norms,  $t$ -Conorms and Averaging Operators*. Fuzzy Sets and Systems, 31:217–242, 1989.
- [MKB79] MARDIA, K. V., J. T. KENT und J. M. BIBBY: *Multivariate Analysis*. Academic Press, London, 1979.
- [MS00] MENDES, M. E. S. und L. SACKS: *Assessment of the Performance of Fuzzy Cluster Analysis in the Classification of RFC Documents*. In: *Proceedings of the London Communications Symposium*, London, 2000.
- [MS01] MORIARITY, C. und F. SCHEUREN: *Statistical matching: A Paradigm for Assessing the Uncertainty in the Procedure*. Journal of Official Statistics, 17:407–422, 2001.
- [MS03] MORIARITY, C. und F. SCHEUREN: *Statistical Matching with Assessment of Uncertainty in the Procedure: New Findings*. In: *Proceedings of the Survey Research Methods Section, American Statistical Association*, Seiten 2904–2909, 2003.
- [MS04] MÜLLER-SCHOLZ, W. K.: *Die stille Transformation*. Gabler, Wiesbaden, 2004.
- [MW03] MERKL, R. und S. WAACK: *Bioinformatik interaktiv*. Wiley-VCH, Weinheim, 2003.



- [NA07] NOLL, P. und P. ALPAR: *A Methodology for Statistical Matching with Fuzzy Logic*. In: *Proceedings of NAFIPS 2007, Fuzzy Information Processing Society*, Seiten 73–78, San Diego, 2007.
- [NKAJ59] NEWCOMBE, H., J. KENNEDY, S. AXFORD und A. JAMES: *Automatic linkage of vital records*. *Science*, 130:954–959, 1959.
- [Oeh07] OEHMICHEN, E.: *Die neue MedienNutzerTypologie MNT 2.0*. *Media Perspektiven*, 5:226–234, 2007.
- [OS04] OEHMICHEN, E. und C. SCHRÖTER: *Die OnlineNutzerTypologie (ONT)*. *Media Perspektiven*, 8:386–393, 2004.
- [OS07] OEHMICHEN, E. und C. SCHRÖTER: *Zur typologischen Struktur medienübergreifender Nutzungsmuster*. *Media Perspektiven*, 8:406–421, 2007.
- [Paa85] PAASS, G.: *Disclosure Risk and Disclosure Avoidance for Microdata*. *Journal of Business and Economic Statistics*, 6:487–500, 1985.
- [Pd00] PAJARES, G. und J. M. DE LA CRUZ: *A new Learning Strategy for Stereo Matching Derived from a Fuzzy Clustering Method*. *Fuzzy Sets and Systems*, 110:413–427, 2000.
- [Pet05] PETERSOHN, H.: *Data Mining*. Oldenbourg, München, 2005.
- [Pil08] PILGRIM, R. A.: *Munkres' Assignment Algorithm. Modified for Rectangular Matrices*. Course Notes, Murray State University, Murray, 2008. Online verfügbar unter <http://cscslab.murraystate.edu/bob.pilgrim/445/munkres.html>, abgerufen am 06.10.2008.
- [Pow00] POWER, D. J.: *Decision Support Systems Hyperbook*, 2000. Online verfügbar unter <http://dssresources.com/dssbook/ch1sbdm.pdf>, abgerufen am 16.11.2008.
- [PQS97] POUPLARD, N., E. M. QANNARI und S. SIMON: *Use of RIDITS to Analyse Categorical Data in Preference Studies*. *Food Quality and Preference*, 8(5/6):419–422, 1997.
- [PU03] PERNUL, G. und R. UNLAND: *Datenbanken im Unternehmen: Analyse, Modellbildung und Einsatz*. Oldenbourg, München, 2. Auflage, 2003.

- [RÖ2] RÄSSLER, S.: *Statistical Matching: A Frequentist Theory, Practical Applications and Alternative Bayesian Approaches*. Springer, New York u.a., 2002.
- [RÖ4a] RÄSSLER, S.: *Data Fusion*. Präsentation University of Ljubljana 21./22.05.2004, Ljubljana, 2004.
- [RÖ4b] RÄSSLER, S.: *Data Fusion: Identification Problems, Validity, and Multiple Imputation*. Australian Journal of Statistics, 33:153–171, 2004.
- [RD81] RODGERS, W. L. und E. B. DEVOL: *An Evaluation of Statistical Matching*. In: *Proceedings of the Survey Research Methods Section, American Statistical Association*, Seiten 128–132, 1981.
- [Red03] REDWAY, H.: *Data Fusion by Statistical Matching*. In: *International Microsimulation Conference on Population, Ageing and Health: Modelling our Future*, Seiten 1–18, Canberra, 2003.
- [RF98] RÄSSLER, S. und K. FLEISCHER: *Aspects Concerning Data Fusion Techniques*. ZUMA-Nachrichten, Spezial No. 4:317–333, 1998.
- [Rön01] RÖNZ, B.: *Computergestützte Statistik 1*, 2001. Online verfügbar unter <http://www.quantlet.com/mdstat/scripts/cs1/cs1-html/index.html>, abgerufen am 18.10.2008.
- [Rod84] RODGERS, W. L.: *An Evaluation of Statistical Matching*. Journal of Business and Economic Statistics, 2(1):91–102, 1984.
- [RP02] ROHWER, G. und U. PÖTTER: *Methoden sozialwissenschaftlicher Datenkonstruktion*. Juventa Verlag, Weinheim, 2002.
- [RR83] ROSENBAUM, P. R. und D. B. RUBIN: *The Central Role of the Propensity Score in Observational Studies for Causal Effects*. Biometrika, 70(1):41–55, 1983.
- [RS06] REINOWSKI, E. und B. SCHULTZ: *Microeconomic Evaluation of Selected ESF-Funded ALMP-Programmes*. IWH Discussion Paper 17, Institut für Wirtschaftsforschung, Halle, 2006.
- [Rub74] RUBIN, D. B.: *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*. Journal of Educational Psychology, 66:688–701, 1974.

- [Rub76] RUBIN, D. B.: *Inference and Missing Data*. Biometrika, 63:581–592, 1976.
- [Rub78] RUBIN, D. B.: *Bayesian Inference for Causal Effects*. Annals of Statistics, 6:34–58, 1978.
- [Rub79] RUBIN, D. B.: *Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies*. Journal of the American Statistical Association, 74:318–328, 1979.
- [Rus23] RUSSEL, B.: *Vagueness*. Australian Journal of Psychology and Philosophy, 1:84–92, 1923.
- [San06] SANGHI, S.: *Determining the Membership Values to Optimize Retrieval in a Fuzzy Relational Database*. In: *Proceedings of the 44th Annual Southeast Regional Conference*, Seiten 537–542, New York, 2006. ACM Press. Session: Database Systems II.
- [Sap00] SAPORTA, G.: *Data Fusion and Data Grafting*. In: *NMDM2000 International Meeting on Nonlinear Methods and Data Mining*, Rom, 2000.
- [Sau04] SAUER, J.: *Neuronale Netze und Fuzzy Control-Systeme*. Vorlesungsskript, Fachhochschule Regensburg, März 2004. Online verfügbar unter <http://fbim.fh-regensburg.de/~saj39122/vhb/NN-Script/script/gen/title.html#link1>, abgerufen am 24.01.2007.
- [Sau07] SAUER, M.: *Weblogs, Podcasting & Online-Journalismus*. O'Reilly, Köln, 2007.
- [SCC02] SHADISH, W. R., T. D. COOK und D. T. CAMPBELL: *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company, Boston, New York, 2002.
- [Sch74] SCHMID, H. J.: *Eine geometrische Deutung der Ungarischen Methode*. Mathematische Zeitschrift, 138:213–218, 1974.
- [Sch93] SCHULTE, U.: *Einführung in die Fuzzy-Logik: Fortschritte durch Unschärfe*. Franzis-Verlag GmbH, München, 1993.
- [Sch00] SCHLITGEN, R.: *Einführung in die Statistik: Analyse und Modellierung von Daten*. Oldenbourg, München, 9. Auflage, 2000.

- [Sch03] SCHWENKER, F.: *Data Mining*. Vorlesungsskript, Universität Ulm, 2003. Online verfügbar unter <http://www.informatik.uni-ulm.de/ni/Lehre/WS03/DMM/dmm.html>, abgerufen am 11.05.2007.
- [Sch04] SCHENDERA, C. F. G.: *Datenmanagement und Datenanalyse mit dem SAS-System*. Oldenbourg, München, 2004.
- [Sch06] SCHULTZ, B.: *Möglichkeiten und Grenzen des Matching-Ansatzes: Am Beispiel der betrieblichen Mitbestimmung*. IWH-Diskussionspapiere 15, Institut für Wirtschaftsforschung Halle, Juli 2006.
- [Sch07a] SCHMIDT, J.: *Umfrage "Wie ich blogge?!"*, 2007. Online verfügbar unter <http://www.schmidtmitdete.de/lebenslauf-aktivitaten-publikationen/umfrage-wie-ich-blogge>, abgerufen am 22.07.2008.
- [Sch07b] SCHMIDT, J.: *WIB 2005-Daten werden freigegeben*, 2007. Online verfügbar unter <http://www.bamberg-gewinnt.de/wordpress/archives/791>, abgerufen am 22.07.2008.
- [Sch07c] SCHMIDTHALS, J.: *Technologiekoperationen in radikalen Innovationsvorhaben*. DUV, Wiesbaden, 2007.
- [Sch08] SCHMIDT, J.: *Geschlechterunterschiede in der deutschsprachigen Blogosphäre*. In: ALPAR, P. und S. BLASCHKE (Herausgeber): *Web 2.0 - Eine empirische Bestandsaufnahme*, Seiten 73–86. Vieweg + Teubner, Wiesbaden, 2008.
- [SD96] SERMEUS, W. und L. DELESIE: *Ridit Analysis on Ordinal Data*. *Western Journal of Nursing Research*, 18:351–359, 1996.
- [SG02] SCHAFFER, J. L. und J. W. GRAHAM: *Missing Data: Our View of the State of the Art*. *Psychological Methods*, 7(2):147–177, 2002.
- [Sia01] SIANESI, B.: *Implementing Propensity Score Matching Estimators with STATA*. Presentation for UK Stata Users Group, VII Meeting, London, 2001. Online verfügbar unter <http://fmwww.bc.edu/RePEc/usug2001/psmatch.pdf>, abgerufen am 28.02.2007.
- [Sim74] SIMS, C. A.: *Comment*. *Annals of Economic and Social Measurement*, 3:395–397, 1974.

- [Smi97] SMITH, H. L.: *Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies*. *Sociological Methodology*, 27:325–353, 1997.
- [SN91] SAXENA, P. C. und K. NAVANEETHAM: *The Effect of Cluster Size, Dimensionality, and Number of Clusters on Recovery of True Cluster Structure Through Chernoff-Type Faces*. *The Statistician*, 40(4):415–425, 1991.
- [SS07] SCHIRMER, T. und K. SIEMER: *Excel 2007 für Späteinsteiger*. Franzis Verlag, Poing, 2007.
- [Sud03] SUDARYANTO: *A Fuzzy Multi-Attribute Decision Making Approach for the Identification of the Key Sectors of an Economy: The Case of Indonesia*. Dissertation, Rheinisch-Westfälische Technische Hochschule Aachen, 2003. Online verfügbar unter [http://darwin.bth.rwth-aachen.de/opus3/volltexte/2003/591/pdf/03\\_121.pdf](http://darwin.bth.rwth-aachen.de/opus3/volltexte/2003/591/pdf/03_121.pdf), abgerufen am 16.06.2008.
- [SW06] SCHMIDT, J. und M. WILBERS: *Wie ich blogge?! Erste Ergebnisse der Weblogbefragung 2005*. Bericht 06-01, Forschungsstelle “Neue Kommunikationsmedien”, Bamberg, 2006. Online verfügbar unter <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-9874>, abgerufen am 23.07.2008.
- [TH08] TOUTENBURG, H. und C. HEUMANN: *Deskriptive Statistik*. Springer, Berlin, 6. Auflage, 2008.
- [Tim02] TIMM, H.: *Fuzzy Clusteranalyse: Methoden zur Exploration von Daten mit fehlenden Werten sowie klassifizierten Daten*. Dissertation, Fakultät für Informatik der Otto-von-Guericke-Universität Magdeburg, 2002.
- [TNS06] TNS INFRATEST: *Dokumentation Fragebogen ALLBUS 2006*, 2006. Online verfügbar unter <http://www.za.uni-koeln.de/data/allbus/fragebogen/za4500fb1.pdf>, abgerufen am 26.05.2008.
- [TPT01] TOLIAS, Y. A., S. M. PANAS und L. H. TSOUKALAS: *Generalized Fuzzy Indices for Similarity Matching*. *Fuzzy Sets and Systems*, 120:255–270, 2001.

- [Tra94] TRAEGER, D. H.: *Einführung in die Fuzzy-Logik*. B.G. Teubner, Stuttgart, 1994.
- [TS04] THOMAS, O. und A.-W. SCHEER: *Referenzmodellbasiertes Customizing unter Berücksichtigung unscharfer Daten*. Technischer Bericht Heft 177, Institut für Wirtschaftsinformatik im Deutschen Forschungszentrum für Künstliche Intelligenz, Saarbrücken, 2004. Online verfügbar unter [http://www.iwi.uni-sb.de/Download/iwihefte/Thomas\\_Scheer\\_IWi-Heft\\_177.pdf](http://www.iwi.uni-sb.de/Download/iwihefte/Thomas_Scheer_IWi-Heft_177.pdf), abgerufen am 16.06.2008.
- [Unb08] UNBEHAUEN, H.: *Regelungstechnik I*. Vieweg Verlag, Wiesbaden, 2008.
- [Utz08] UTZ, S.: *(Selbst-)marketing auf Hyves*. In: ALPAR, P. und S. BLASCHKE (Herausgeber): *Web 2.0 - Eine empirische Bestandsaufnahme*, Seiten 233–258. Vieweg + Teubner, Wiesbaden, 2008.
- [van00] VAN DER PUTTEN, P.: *Data Fusion for Data Mining*. Coil Seminar 2000, Chios, Juni 2000.
- [vF05] VAN EIMEREN, B. und B. FREES: *Nach dem Boom: Größter Zuwachs in internetfernen Gruppen*. Media Perspektiven, 8:362–379, 2005.
- [vF07] VAN EIMEREN, B. und B. FREES: *Internetnutzung zwischen Pragmatismus und YouTube-Euphorie*. Media Perspektiven, 8:362–378, 2007.
- [VH06] VIERTL, R. und D. HARETER: *Beschreibung und Analyse unscharfer Informationen: Statistische Methoden für unscharfe Daten*. Springer, Wien, 2006.
- [Vie02] VIERTL, R.: *Allgemeine Informationstheorie und Statistik*. In: DUTTER, R. (Herausgeber): *Festschrift 50 Jahre österreichische statistische Gesellschaft*, Seiten 105–114. Österreichische Statistische Gesellschaft, Wien, 2002.
- [vKG02] VAN DER PUTTEN, P., J. N. KOK und A. GUPTA: *Data Fusion through Statistical Matching*. MIT Sloan Working Paper 185, Cambridge, MA, 2002.

- [Voß04] VOSS, W.: *Taschenbuch der Statistik*. Carl Hanser Verlag, München, 2. Auflage, 2004.
- [von04] VON HOYNINGEN-HOENE, J.: *Integration nach Unternehmensakquisitionen*. DUV, Wiesbaden, 2004.
- [Wal08] WALDFORST, S.: *Die Wirkung von Zielen auf die Arbeitsleistung von Akteuren*. Gabler, Wiesbaden, 2008.
- [WB05] WEBER, S. und A. BRAKE: *Internetbasierte Befragung*. In: KÜHL, S., P. STRODTOLZ und A. TAFFERTSHOFER (Herausgeber): *Quantitative Methoden der Organisationsforschung*. VS Verlag für Sozialwissenschaften, Wiesbaden, 2005.
- [Web99] WEBER, K.: *Industriebetriebslehre*. Springer, Berlin, 1999.
- [WG99] WILLIAMSON, M. und K. J. GASTON: *A Simple Transformation for Sets of Range Sizes*. *Ecography*, 22:674–680, 1999.
- [WHSW03] WENZLAU, A., U. HOFER, M. SIEGERT und S. WOHLRAB: *Kundenprofiling*. Publicis Corporate Publishing, Erlangen, 2003.
- [Wie98] WIEDEN, E.-A.: *Prototypische Realisierung von Fuzzy-Zugriffen auf relationale Datenbanken im PIA*. Vortrag im STS Oberseminar, Technische Universität Hamburg-Harburg, November 1998. Online verfügbar unter <http://www.sts.tu-harburg.de/slides/1998/11-98-Wied-0S-DA.pdf>, abgerufen am 24.01.2007.
- [Win95] WINKLER, W. E.: *Matching and Record Linkage*. In: COX, B. G. (Herausgeber): *Business Survey Methods*, Seiten 374–403. John Wiley & Sons, Inc., 1995.
- [Wu07] WU, C.-H.: *On the Application of Grey Relational Analysis and RIDIT Analysis to Likert Scale Surveys*. *International Mathematical Forum*, 2(14):675–687, 2007.
- [WZ01] WIEDENBECK, M. und C. ZÜLL: *Klassifikation mit Clusteranalyse: Grundlegende Techniken hierarchischer und K-means-Verfahren*. *ZUMA How-To*, 10:1–18, 2001.
- [YA99] YOSHIZOE, Y. und M. ARAKI: *Statistical Matching of Household Survey Files*. Working Paper No. 10, Aoyama Gakuin University, Tokyo, 1999.

- [Yan98] YANG, Y.: *Rechnergestützte Östrusüberwachung bei Milchkühen unter Anwendung der Fuzzy-Logic-Methode*. Herbert Utz Verlag, München, 1998.
- [Zad65a] ZADEH, L. A.: *Fuzzy Sets*. *Information and Control*, 8(3):338–353, 1965.
- [Zad65b] ZADEH, L. A.: *Fuzzy Sets and Systems*. In: FOX, J. (Herausgeber): *System Theory*, Seiten 29–39. Polytechnic Press, New York, 1965.
- [Zad66a] ZADEH, L. A.: *Abstraction and Pattern Classification*. *Journal of Mathematical Analysis and Applications*, 13(1):1–7, 1966.
- [Zad66b] ZADEH, L. A.: *Shadows of Fuzzy Sets*. *Problems in Transmission of Information*, 2:37–44, 1966.
- [Zad71] ZADEH, L. A.: *Towards a Theory of Fuzzy Systems*. In: KALMAN, R. E. und R. N. DECLAIRIS (Herausgeber): *Aspects of Networks and System Theory*, Seiten 469–490. Holt, Rinehart & Winston, New York, 1971.
- [Zad73] ZADEH, L. A.: *The Concept of a Linguistic Variable and its Application to Approximate Reasoning*. Memorandum ERL-M 411 Berkeley, 1973.
- [Zad76] ZADEH, L. A.: *The Linguistic Approach and its Application to Decision Analysis*. In: HO, Y. C. und S. K. MITTER (Herausgeber): *Directions in Large Scale Systems*, Seiten 339–370. Plenum Press, New York, 1976.
- [Zad81] ZADEH, L. A.: *PRUF-A Meaning Representation Language for Natural Languages*. In: MAMDANI, E. H. und G. R. GAINES (Herausgeber): *Fuzzy Reasoning and its Applications*, Seiten 1–66. Academic Press, London, New York, Toronto, 1981.
- [Zen04] ZENTRUM FÜR EUROPÄISCHE WIRTSCHAFTSFORSCHUNG GMBH: *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS*, 2004. Online verfügbar unter <http://www.zew.de/de/publikationen/dfgflex/allbus.html>, abgerufen am 24.05.2008.



- [ZG91] ZIMMERMANN, H.-J. und L. GUTSCHE: *Multi-Criteria Analyse: Einführung in die Theorie der Entscheidungen bei Mehrfachzielsetzungen*. Springer, Berlin, 1991.
- [Zim93] ZIMMERMANN, H.-J.: *Fuzzy Sets, Decision Making, and Expert Systems*. Kluwer Academic Publishers, Boston, 4. Auflage, 1993.
- [Zim94] ZIMMERMANN, H.-J.: *Fuzzy Set Theory: And its Applications*. Kluwer Academic Publishers, Boston, 2. Auflage, 1994.
- [ZL96] ZHANG, T. und R. RAMAKRISHNAN M. LIVNY: *BIRCH: An Efficient Data Clustering Method for Very Large Databases*. In: *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, Seiten 103–114, Montreal, 1996.
- [ZZ80] ZIMMERMANN, H.-J. und P. ZYSNO: *Latent Connectives in Human Decision Making*. *Fuzzy Sets and Systems*, 4:37–51, 1980.

# Sachverzeichnis

- $\gamma$ -Operator, 35, 66
- Abstand, 70
- ALLBUS, 94
- Aussagenlogik, 25
- Blogroll, 144
- C-Means Algorithmus, 44
- Canberra-Metrik, 72
- Case, 13
- Charakterisierende Funktion, 29
- Charakteristische Funktion, 26, 27
- City-Block-Distanz, 73
- Clusteranalyse, 43
  - Agglomerative Verfahren, 44
  - Deterministische Zuordnung, 43
  - Divisive Verfahren, 44
  - Fuzzy, 44, 58, 81, 85
  - Hierarchisch, 44
  - Partitionierend, 44
  - Possibilistische Zuordnung, 43
  - Probabilistische Zuordnung, 43
- Clusterschwerpunkt, 47
- Cold Deck Imputation, 3
- Complete-Linkage, 44
- Conditional Independence Assumption, 15, 24
- Control, 13
- CRISP-DM, 2
- Cronbachs-*alpha*, 130
- Data Merging, 10
- Data Mining, 1
  - Prozess, 2, 3
- Datenanreicherung, 2
- Datenauswahl, 2
- Datenfusion, 13
- Datenschutz, 183
- Direct Rating, 41
- Distanz, 69, 71, 88
  - Absolute, 71, 72
  - Euklidische, 71, 73
- Distanzfunktion, 13, 71
- Donor Unit, 13
- Elementaroperationen, 31
- Faktoranalyse, 130
  - Eigenwert, 131
  - Erklärte Gesamtvarianz, 131
  - Faktorladung, 131
  - Kaiser-Guttman-Kriterium, 131
  - Rotation, 131
  - Schiefwinklige Rotation, 131
- Fuzzy
  - $\alpha$ -Schnitt, 29
  - C-Means Algorithmus, 45, 47, 85
  - Gleichheit, 28
  - Menge, 28
  - Regeln, 49, 51, 64
  - Regelungstechnik, 62
  - Teilmenge, 28
- Fuzzy Logic, 25, 38

- Fuzzyfier, 47
- Gesamtdistanz
  - Minimierung, 21, 75, 89, 90
- Heterogenität in Vergleichsgruppen,
  - 12, 93, 97, 130
- Homogenitätstest, 106
  - $\chi^2$ , 106, 153
- Hot Deck Imputation, 3
- Imputation fehlender Werte, 12
- Indikatorfunktion, 28
- Intervallschätzung, 41
- Knowledge Discovery in Databases (KDD), 1
- Kolmogorov-Smirnov-Test, 132
- Kompensationsgrad, 36, 37, 81, 112
- Kompensatorische Operatoren, 34, 37,
  - 64, 66, 68
  - Nicht-parametrisiert, 35
  - Parametrisiert, 35
- Kontrollgruppe, 12
- Kuhn-Munkres-Algorithmus, 75
- Linguistische Variable, 38, 56, 98, 149
- Linguistischer Ausdruck, 37
- Linguistischer Term, 38, 39, 56, 62,
  - 98, 149
- Mahalanobis-Distanz, 72
- Manhattan-Distanz, 73
- Matching-Güte, 105, 139, 152
- Matching-Prozess, 13
- Matchingvariablen
  - Gewichtung, 67, 161, 165
- MATLAB, 79
- Metrik, 70
- Minkowski-Metrik, 71
- MinMax-Operator, 36, 68
- Modellierte Operationen, 31
- Multivariate Analysen, 12
- Onlineanwendungen, 146
- Paarweiser Vergleich, 42
- Perfect Match, 82, 89, 108
- Polling, 40
- ProdSum-Operator, 37, 68
- Propensity Score, 17
  - Matching, 16, 17
- Prototyp, 45, 47
- PRUF, 49
- Recipient Unit, 13
- Record Linkage, 10
- Regelbasis, 55, 61, 62
- Reliabilität, 130
- Reverse Rating, 41
- Ridit
  - Analyse, 107, 187
  - Test, 130, 187
  - Wert, 187
- Ridits, 189
- s-Norm, 33, 35, 65
- Sample Percent Reduction in Bias,
  - 109, 112, 153, 163
- Scharfe Menge, 26
- Single-Linkage, 44
- Soziale Online-Netzwerke, 142, 143,
  - 145
- Statistische Zwillinge, 1, 10, 16, 22,
  - 60, 69, 88
- Statistisches Fuzzy-Matching, 4, 53,
  - 133, 138
  - Constrained, 74, 82, 88, 92, 108,
    - 120, 153

- Unconstrained, 74, 82, 88, 92, 114, 123, 156
- Statistisches Matching, 1, 2, 9, 10, 53
  - Ausgangssituation, 13
  - Constrained, 19, 21
  - Distanzmatching, 117, 126, 158
  - Kritik, 22
  - Unconstrained, 19
- StudiVZ, 145, 167, 177
- t-Conorm, 33
- t-Norm, 32, 35, 51, 65
- T-Test, 106, 132, 189
  - nach Welch, 132, 189
- Trackback, 176
- Transformationsfunktion, 75
- Ungarische Methode, 75
- Unschärfe, 25
- Unscharfe Menge, 26, 28
- Unscharfe Zahl, 28
- Variable
  - Metrisch skaliert, 6, 60, 106
  - Nominal skaliert, 5, 53, 60, 106
  - Ordinal skaliert, 6, 60, 106, 107
- Varianzheterogenität, 132, 189
- Varianzhomogenität, 132, 189
- Verknüpfung linguistischer Terme, 61, 62, 64, 80
- Vorverarbeitung, 3
- Ward's Methode, 44
- Web 2.0, 142, 146, 148
- Weblog, 142–144
  - Autoren, 148, 170, 171
  - Leser, 148, 171, 174
- Xing, 145, 167, 177
- Zimmermann-Operator, 66
- Zugehörigkeit, 26
- Zugehörigkeitsfunktion, 27–29, 57
  - Bestimmung, 58, 83
  - Objektive Ermittlung, 42
  - Schätzung, 42
  - Subjektive Ermittlung, 40
- Zugehörigkeitsgrad, 4, 6, 27, 32, 46, 48, 51
  - zur Regelbasis, 4, 55, 64, 86