

**Analyse der ökonomischen und demografischen
Determinanten von Sportaktivitäten in Deutschland**

Martin Flohr

Arbeitspapier Nr. 27 (Juni 2004)

Institut für Statistik und Ökonometrie
Johannes Gutenberg-Universität Mainz
Fachbereich Rechts- und Wirtschafts-
wissenschaften
Haus Recht und Wirtschaft II

D 55099 Mainz

Herausgeber: Univ.-Prof. Dr. P.M. Schulze

© 2004 Institut für Statistik und Ökonometrie, Mainz
ISSN Nr. 1430 - 2136

Analyse der ökonomischen und demografischen Determinanten
von Sportaktivitäten in Deutschland

Martin Flohr

Gliederung

1 Einleitung	1
2 Bestimmung möglicher Einflussfaktoren der Sportaktivität	1
3 Binäres Logit-Modell	13
4 Auswertung und Interpretation der Schätzergebnisse	14
5 Fazit	21
Anhang	I
Literatur	III

Zusammenfassung

In der Studie wird mittels des binären Logit-Modells untersucht, welche demografischen und ökonomischen Determinanten die Wahrscheinlichkeit, in Deutschland Sport zu treiben, beeinflussen. Dabei zeigt sich für die befragten Personen, dass die Sportaktivität mit zunehmendem Alter abnimmt. Eine Rolle spielen auch das Geschlecht, die Zugehörigkeit zu einer "sozialen Schicht" sowie gesundheitliche Aspekte. Außerdem werden diesbezüglich signifikante Unterschiede zwischen Personen aus West- und Ostdeutschland sichtbar.

Summary

Using a binary logistic model this survey analyses the economic and demographic determinants of the probability to participate in sports in Germany. It turns out that the sporting participation decreases with age. Furthermore the influence of the sex, the belonging to a social class and aspects of health are proved. Referring to this there are significant differences between people from Western and Eastern Germany.

1 Einleitung

Der Sport hat sich zu einem gesellschaftlichen Massenphänomen mit bedeutsamen sozialen und individuellen Funktionen entwickelt. Damit einhergehend ist auch seine ökonomische Bedeutung gestiegen (Digel, 1990, S. 81). Exemplarisch kann dies anhand zweier ökonomischer Kennzahlen aus dem Jahr 1998 - aktuellere Zahlen liegen nicht vor - belegt werden. In diesem Jahr machte die sportbezogene Leistungserstellung mit knapp 53 Milliarden DM 1,4% des gesamten Bruttoinlandsproduktes aus. Außerdem übten 783.000 Personen eine bezahlte Tätigkeit aus, die im direkten oder indirekten Zusammenhang mit dem Sport stand, was in etwa der Anzahl der Arbeitnehmer im Kreditgewerbe entsprach (Meyer/Ahlert, 2000, S. 55 ff.). Das zeigt, dass die Analyse der ökonomischen und demografischen Faktoren, die die Nachfrage nach Sportaktivitäten determinieren, auch unter wirtschaftswissenschaftlichen Gesichtspunkten interessant ist.

Als Datenbasis der Untersuchung dient der Bundes-Gesundheitssurvey 1998 (BGS98) des Robert Koch-Instituts (Public Use File BGS98, Bundes-Gesundheitssurvey 1998, Robert Koch-Institut (RKI)). Die Beschreibung dieser Daten und der angewendeten Gewichtung erfolgt zu Beginn des zweiten Kapitels. Des Weiteren werden dort durch Anwendung verschiedener Testverfahren und deskriptiver Analysen mögliche Einflussgrößen der Sportaktivität bestimmt. Diese Einflussgrößen sollen in ein ökonometrisches Erklärungsmodell für einen qualitativen Regressanden eingehen. Da dieser Regressand die beiden Ausprägungen Sportaktivität und Nichtsportaktivität annehmen kann, bietet sich hierfür das binäre Logit-Modell an. Dessen ökonometrische Grundlagen werden in Kapitel 3 vorgestellt. Anschließend wird in Kapitel 4 die Schätzung ausgewertet und interpretiert. Mit der Zusammenfassung der Ergebnisse in Kapitel 5 wird diese Arbeit abgeschlossen.

2 Bestimmung möglicher Einflussfaktoren der Sportaktivität

In diesem Kapitel werden mit Hilfe deskriptiver Analysen und verschiedener Testverfahren demografische und ökonomische Variablen bestimmt, die die Sportaktivität der deutschen Bevölkerung später in einem Schätzmodell erklären sollen. Zunächst wird dabei auf die Datenbasis der durchgeführten Untersuchung eingegangen.

2.1 Datenbeschreibung

Die der Analyse zugrunde liegenden Daten gehen aus dem Fragebogen zu Leben und Gesundheit der Teilnehmer des BGS98 hervor, dessen Beantwortung neben einem Arztinterview und einer körperlichen Untersuchung zum Kernsurvey gehörte. Durchgeführt wurde die Erhebung vom RKI zwischen Oktober 1997 und März 1999. Da das menschliche Verhalten im Zeitverlauf nur langsamen Veränderungen unterliegt, ist auch knapp fünf Jahre später von einer immer noch vorhandenen Aktualität der hieraus gewonnenen Ergebnisse auszugehen. Zudem stehen derzeit keine aktuelleren seriösen Daten zur Verfügung.

Die Grundgesamtheit setzte sich aus allen während des Befragungs- und Untersuchungszeitraumes in Deutschland in Privathaushalten lebenden und in den Einwohnermelderegistern mit Hauptwohnsitz gemeldeten Erwachsenen im Alter von 18 bis 79 Jahren zusammen. Allerdings blieben Personen mit nicht ausreichenden Deutschkenntnissen außen vor (Stolzenberg, 2000, S. 3ff.).

Zur Bestimmung der Stichprobe wurde eine geschichtete Klumpenauswahl angewendet, wobei sich die 120 gezogenen Samplepoints mit 40 auf das Gebiet der ehemaligen DDR und 80 auf das frühere Bundesgebiet disproportional verteilen. Hinsichtlich der tatsächlichen Bevölkerungszahl ist in der Stichprobe die ostdeutsche Bevölkerung über- und die westdeutsche unterrepräsentiert.

Ein Verfahren, das versucht, diese aus dem Stichprobendesign resultierende Einschränkung der Repräsentativität in der gesamtdeutschen Betrachtung aufzuheben, stellt die Designgewichtung dar. Generell wird bei einer Gewichtung durch die Vergabe positiver reeller Zahlen, der Gewichtungsfaktoren, an die Befragten bzw. an die sie repräsentierenden Merkmalsvektoren deren relative Wichtigkeit verändert. Für die Designgewichtung sind diese Faktoren proportional zur inversen Auswahlwahrscheinlichkeit (Rösch, 1994, S. 7ff.). Sie lassen sich für die vorliegende Problemstellung dadurch berechnen, dass der jeweilige Anteil der west- bzw. ostdeutschen Bevölkerung an der Gesamtbevölkerung Deutschlands, die 1998 im Alter zwischen 18 und 79 Jahren war, ins Verhältnis zum jeweils entsprechenden Anteil in der Stichprobe gesetzt wird (Janssen/Laatz, 2003, S. 47):

$$(2.1) \quad G_i = \frac{\text{Soll}}{\text{Ist}},$$

Soll = jeweiliger Anteil der West- bzw. Ostdeutschen in der Grundgesamtheit

Ist = jeweiliger Anteil der West- bzw. Ostdeutschen in der Stichprobe.

Die aus den Berechnungen resultierenden Ergebnisse sind in folgender Übersicht tabellarisch dargestellt:

Tabelle 1: Gewichtungsfaktoren

	West	Ost
Soll	0,811	0,189
Ist	0,66	0,34
Gewichtungsfaktor G_i	1,22783	0,55686

Aufgrund der niedrigeren Auswahlwahrscheinlichkeiten erhalten die Merkmalsvektoren der Westdeutschen also mit 1,22783 ein höheres Gewicht als die der Ostdeutschen mit 0,55686. Die Fallzahl wird durch diese Gewichtung nicht verändert.

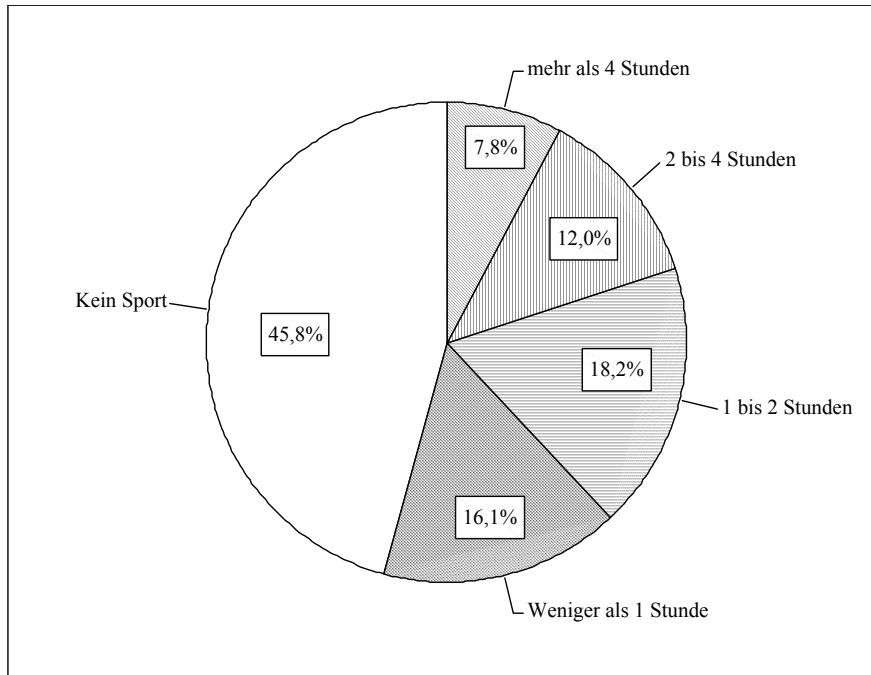
Die praktische Anwendung der Gewichtung ist umstritten. Während von manchen Befürwortern eine noch weitergehende Gewichtung gefordert wird, die auch den unterschiedlichen Nonresponse-Anteil in verschiedenen Bevölkerungsgruppen berücksichtigt, wird sie von anderer Seite grundsätzlich abgelehnt. Die ablehnende Haltung mancher Kritiker macht sich vor allem fest an der impliziten Annahme, dass die Abweichungen der bekannten und der unbekannt Variablen in Größenordnung und Variation gleich sind, und die Stichprobe homogene Subgruppen für alle vorhandenen Kategorienkombinationen der zu analysierenden Variablen liefert. Beides ist in der Praxis jedoch sehr unwahrscheinlich und führt zu Verzerrungen bei den Parameterschätzungen (Alt/Bien, 1994, S. 130).

2.2 Sportaktivität in Deutschland

Die nachfolgend vorgestellten Ergebnisse der Analyse der Sportaktivitäten in Deutschland basieren auf subjektiven Selbsteinschätzungen der befragten Personen. Besonders wichtig ist dies vor dem Hintergrund, dass der Sport in der jüngeren Vergangenheit einen starken Wachstums- und Differenzierungsprozess erlebt hat. Denn dieser hat zu einem Wandel der subjektiven Vorstellungen von Sport geführt, die sich zwischen verschiedenen Bevölkerungsgruppen kaum noch decken. So ist es zum Beispiel wahrscheinlich, dass ältere Personen Tätigkeiten wie Gymnastik oder Spaziergehen eher als sportliche Aktivitäten einstufen, als dies jüngere tun. Zusammen mit forschungsmethodischen Problemen liegen hierin wohl auch die Hauptursachen für große Unterschiede in den Befunden vorliegender Studien zur Sportbeteiligung. Hier ergeben sich Sportbeteiligungsquoten zwischen circa 20% und

75% (Nagel, 2002, S. 19). Von der gewichteten Stichprobe des BGS98 lässt sich auf eine Sportbeteiligungsquote von 54,2% für die in Deutschland lebende Bevölkerung im Alter von 18 bis 79 Jahren schließen. Relativierend ist jedoch zu sagen, dass hierzu auch 16,1% gehören, die weniger als eine Stunde pro Woche Sport treiben und mit der Gruppe der vollkommen Inaktiven zusammengenommen einen Anteil von über 60% der Gesamtbevölkerung ausmachen (siehe Abbildung 1).

Abbildung 1: Umfang der Sportaktivität in Deutschland 1998



2.3 Mögliche Determinanten der Sportaktivität in Deutschland

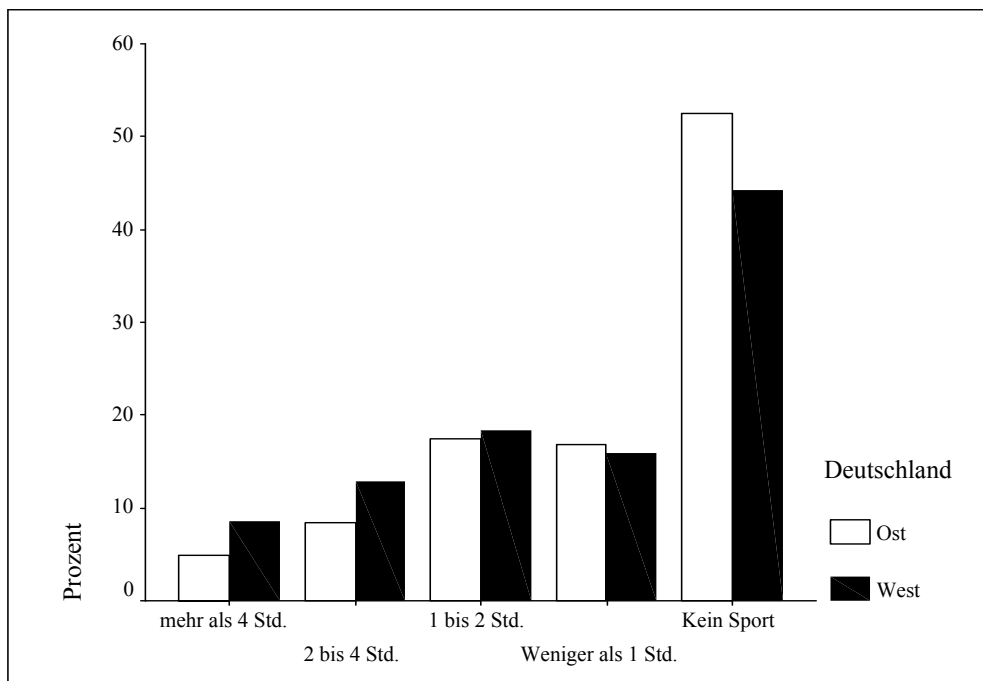
Häufig wird behauptet, dass Unterschiede in der Sportaktivität aufgrund der Expansion, Pluralisierung und Individualisierung des Sportes kaum noch sozialstrukturell begründet werden können (Nagel, 2002, S. 17). Ob sich dies auch durch die Stichprobenergebnisse des BGS98 bestätigt, soll die Untersuchung denkbarer demografischer und ökonomischer Determinanten zeigen. Diese orientiert sich größtenteils an den Determinanten, die in einer Studie von Farrell/Shields (2002) zur Sportaktivität in England analysiert wurden. Die folgende diesbezügliche Analyse für Deutschland basiert auf den Ergebnissen des parametrischen Anteilswertetests (AW-Test) (Bosch, 1996, S. 425) und des nicht-parametrischen χ^2 -Unabhängigkeitstests (Janssen/Laatz, 2003, S. 229). Zur Beurteilung der Stärke eines möglichen Einflusses unabhängig von der Stichprobengröße wird auf die Effektgröße Cramers V zurückgegriffen, die aus dem χ^2 -Wert ermittelt werden kann (Bühl/Zöfel, 2002, S. 247).

Alle nachfolgenden Entscheidungen der empirischen Analyse bezüglich Ablehnung oder Nicht-Ablehnung der Nullhypothese basieren auf einem Signifikanzniveau von 5%. Ist der p-Value (p) kleiner als 5%, wird die Nullhypothese abgelehnt. Während der AW-Test manuell berechnet wurde, sind die nachfolgenden Ergebnisse des χ^2 -Unabhängigkeitstests und von Cramers V mit SPSS 11.0 ermittelt worden. Die Ergebnisse der gesamtdeutschen Betrachtungen beruhen auf der gewichteten Stichprobe, während beim Vergleich der ost- und westdeutschen Verhältnisse auf eine Gewichtung verzichtet werden kann. Dies ergibt sich daraus, dass die gewählte Gewichtung nur die durch die disproportionale Schichtung entstandene Unterrepräsentierung der westdeutschen Bevölkerung in der Gesamtbetrachtung ausgleichen soll.

2.3.1 Ost-West

Beim Vergleich des Sportengagements zwischen Ostdeutschen und Westdeutschen zeigen sich signifikante Abweichungen ($\chi^2 = 42,58 / p = 0,000$; Z (AW-Test) = $-6,48 / p = 0,000$). Mit 52,5% Inaktiven betreibt im Osten mehr als die Hälfte der Bevölkerung gar keinen Sport, während diese Quote im Westen lediglich bei 44,3% liegt (siehe Abbildung 2).

Abbildung 2: Vergleich des Umfangs der Sportaktivität zwischen Ost und West



Diese Differenzen in der Sportbeteiligung sind im Prinzip die Fortsetzung der Verhältnisse, die auch schon vor der deutschen Einheit herrschten. Zwar kann die Aktivenrate in der DDR nur auf Basis der offiziellen Statistik geschätzt werden, die Größenordnung von rund 25%

für das Jahr 1988 wird jedoch in der Literatur (Praschek, 2000) als durchaus realistisch angesehen. Damit liegt sie deutlich unter der Aktivenrate in der Bundesrepublik Deutschland (BRD), die laut Sozio-ökonomischem Panel zum gleichen Zeitpunkt über 40% betrug (Erlinghagen, 2003, S. 9f.). Eine mögliche Ursache dafür, dass sich die Verhältnisse bisher noch nicht angeglichen haben, könnte im noch nicht so breiten Angebot an Sportstätten und Sportgeräten in den neuen Bundesländern liegen. Außerdem sorgt die Verkehrsinfrastruktur, die noch nicht auf dem Niveau des Westens ist, dafür, dass der Zugang zu Sportstätten für die Bevölkerung in ländlichen Regionen erschwert ist. Die Finanzschwäche der privaten Haushalte aufgrund der schwierigen Wirtschaftslage in Ostdeutschland kann ebenfalls ein Erklärungsansatz sein. Daneben wirkt wohl auch die langjährige Fokussierung in der DDR auf den Leistungssport nach, die dem eher gesundheits- und fitnessorientierten Sport unter Beteiligung breiterer Bevölkerungsgruppen in der BRD entgegenstand (Nagel, 2002, S. 21ff.)

2.3.2 Geschlecht

Die Abweichungen zwischen Ost- und Westdeutschland sind bei beiden Geschlechtern zu finden. Sowohl westdeutsche Frauen als auch westdeutsche Männer sind signifikant sportlich aktiver als Frauen bzw. Männer im Osten Deutschlands (siehe Tabelle 2 und Anhang II). Zwischen den Geschlechtern bestehen in der gesamtdeutschen Betrachtung der Sportbeteiligungsquoten signifikante Unterschiede. Wie Tabelle 2 zeigt, ist die Quote bei den Männern mit 56,7% höher als die der Frauen, die bei 51,7% liegt. Allerdings ergibt eine nähere Analyse, dass es sich hierbei um ein westdeutsches Phänomen handelt, das auf das gesamtdeutsche Ergebnis durchschlägt. Denn die sehr geringen, geschlechtsspezifischen Differenzen in der Sportbeteiligung in Ostdeutschland können als zufällig angenommen werden.

Tabelle 2: Sportbeteiligungsquoten, differenziert nach Geschlecht

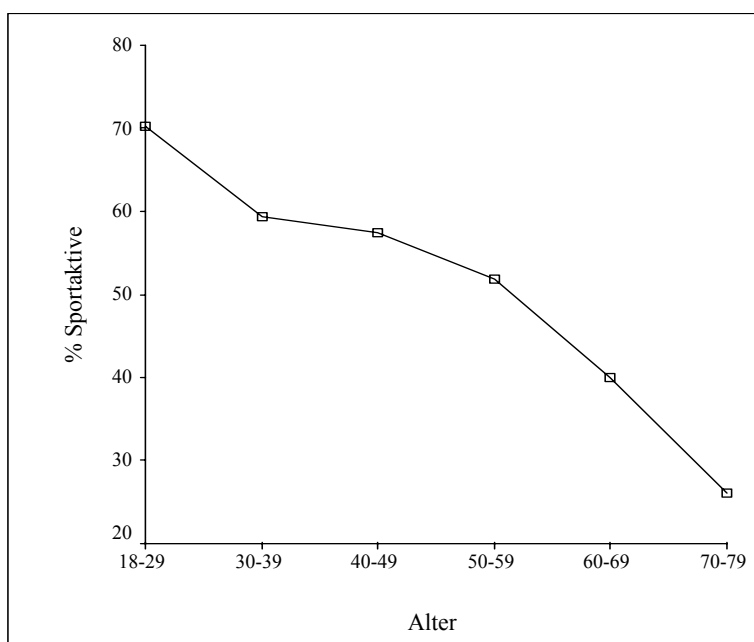
	Gesamtdeutschland	Westdeutschland	Ostdeutschland
Männer	56,7%	58,8%	47,9%
Frauen	51,7%	52,8%	47,1%
χ^2	17,57 (p = 0,000)	16,36 (p = 0,000)	0,18 (p = 0,000)
Cramers V	0,05 (p = 0,000)	0,06 (p = 0,000)	0,009 (p = 0,000)
Z (AW-Test)	4,17 (p = 0,000)	-5,78 (p = 0,000)	-1,36 (p = 0,000)

Das Ergebnis, dass im Westen Männer sportlich aktiver als Frauen sind, wird durch Nagels Studie aus dem Jahr 2002 nicht gestützt. Vielmehr geht er von einem vollständigen Abbau früher vorherrschender Unterschiede aus (Nagel, 2002, S. 29). Trotzdem soll hier kurz auf denkbare Gründe für geschlechtsspezifische Differenzen eingegangen werden. Ihre Wurzeln liegen möglicherweise in den Kindes- und Jugendjahren, in denen Jungen stärker zum Sporttreiben angeregt werden als Mädchen. Daraus kann sich dann ein geringeres Sportinteresse der Frauen im Erwachsenenalter entwickelt haben. Des Weiteren gilt die Sportkultur insgesamt als immer noch eher von Männern dominiert und zu wenig auf die Interessen der Frauen ausgerichtet. Als plausibler Grund erscheint zunächst auch die häufige Mehrfachbelastung der Frauen durch Haushalt, Kinder und gegebenenfalls Berufstätigkeit. Allerdings lässt die stärkere Erwerbsbeteiligung der Frauen in Ostdeutschland, wo es keine geschlechtsspezifischen Unterschiede in der Sportbeteiligung gibt, im Vergleich zu den westdeutschen Frauen gewisse Zweifel an diesem Argument zu (Baur/Braun, 2001, S. 71ff.).

2.3.3 Alter

Die Altersabhängigkeit der Sportaktivität wird durch Abbildung 3 ersichtlich. Mit steigendem Alter nimmt der Anteil der Personen, die sich sportlich engagieren, signifikant ab ($\chi^2 = 407,03 / p = 0,000$; Cramers V = 0,242 / $p = 0,000$). Die einzige Ausnahme bildet hier der Vergleich zwischen der Altersgruppe der 30- bis 39-Jährigen und den 40- bis 49-Jährigen, für den z. B. der AW-Test ($Z = 1,02 / p = 0,3576$) keine signifikanten Unterschiede ausweist.

Abbildung 3: Anteil der Sportaktiven in verschiedenen Altersgruppen



Für diesen Trend gibt es plausible Interpretationsmöglichkeiten. Eine geht etwa davon aus, dass der Rückzug aus dem Sport durch die nachlassende motorische Leistungsfähigkeit im zunehmenden Alter, bzw. durch die sozial vermittelte Vorstellung davon, gefördert wird. Außerdem sind bestehende Sportinfrastrukturen, z. B. in Sportvereinen, häufig primär auf die Interessen von Jugendlichen und jungen Erwachsenen ausgerichtet. Zu erwähnen sind jedoch auch gegenläufige Entwicklungen, wie ein stärker aufkommendes Gesundheitsbewusstsein oder höhere Freizeitdisponibilitäten, die sich positiv auf die Sportaktivitäten der älteren Generationen auswirken (Baur/Braun, 2001, S. 69f.).

Interessante Ergebnisse liefert der altersabhängige Vergleich zwischen Frauen und Männern (siehe Tabelle 3). Zwar liegt der berechnete Anteil der Sportaktiven bei den Männern mit Ausnahme der 50- bis 59-Jährigen in allen Altersgruppen höher als bei den Frauen, allerdings sind diese Unterschiede nach dem AW-Test nur in zwei Altersgruppen signifikant, nämlich bei den 18- bis 29-Jährigen und den 60- bis 69-Jährigen.

Tabelle 3: Anteil der Sportaktiven in verschiedenen Altersgruppen, differenziert nach Geschlecht

	19-29	30-39	40-49	50-59	60-69	70-79
Männer	74,8%	61%	58,4%	50,7%	46%	27,9%
Frauen	65,5%	57,6%	56,5%	52,9%	34,6%	24,7%
Z(AW-Test)	3,62	1,35	0,69	-0,80	3,65	0,84
p	0,000	0,177	0,490	0,424	0,000	0,401

2.3.4 Haushaltskonstellation

Über die Auswirkungen, die das Zusammenleben mit einem Partner auf die Sportaktivität hat, gibt es verschiedene, gegensätzliche Annahmen. Auf der einen Seite liegt die Vermutung nahe, dass Alleinlebenden ein größeres Zeitbudget für Freizeitaktivitäten, also auch für Sport, zur Verfügung steht als in Partnerschaft lebenden Personen. Andererseits ist auch denkbar, dass sich dieses Zeitproblem durch gemeinsame Sportausübung in der Partnerschaft nicht stellt oder sogar eine gegenseitige Motivation zum Sport stattfindet (Baur/Braun, 2001, S. 75). Da im Rahmen des BGS98 mit 2468 Personen nur gut ein Drittel aller Personen die Frage nach dem Zusammenleben mit einem Partner beantwortet haben, sollten die nachfolgenden Ergebnisse nicht überbewertet werden. Trotzdem sind die in Tabelle 4 dargestellten empirischen Befunde bemerkenswert. In der Gesamtbetrachtung aller antwortenden Personen stellen sich zwar weder in der Sportbeteiligungsquote noch im Umfang der Sportaktivi-

tät signifikante Abweichungen zwischen Alleinlebenden und in Partnerschaft Lebenden heraus. Getrennte Analysen für beide Geschlechter ergeben jedoch jeweils signifikante Unterschiede, die allerdings in entgegengesetzte Richtungen gehen. Während alleinlebende Männer signifikant sportaktiver sind als in Partnerschaft lebende Männer, ist dies bei den Frauen genau umgekehrt. Sie ziehen sich nicht in die Partnerschaft zurück, sondern lassen sich wohl häufig durch den, zumindest im Westen, sportlich aktiveren männlichen Partner zum Sport animieren.

Tabelle 4: Sportbeteiligungsquoten von Alleinlebenden und in Partnerschaft Lebenden, differenziert nach Geschlecht

	Gesamt	Männer	Frauen
Alleinlebende	59,3%	69,7%	50,9%
In Partnerschaft	58,9%	58,4%	59,4%
χ^2	0,04 (p = 0,848)	14,01 (p = 0,000)	7,57 (p = 0,006)
Cramers V	0,004 (p = 0,848)	0,111 (p = 0,000)	0,076 (p = 0,006)

Über den Einfluss von Kindern auf die sportliche Aktivität der Eltern gibt es ebenfalls verschiedene Hypothesen. Vor allem bei kleinen Kindern wird erwartet, dass aufgrund des hohen Betreuungsaufwands weniger Zeit zum Sport verbleibt (Baur/Braun, 2001, S. 75). Dies kann jedoch durch die vorliegende Stichprobe nicht gestützt werden, da zwischen Personen mit Kindern bis zu 5 Jahren und Personen ohne Kinder bis zu 5 Jahren keine signifikanten Differenzen angezeigt werden (Z (AW-Test) = 1,76 / p = 0,078). Dagegen sind Eltern von Kindern im Alter zwischen 6 und 14 Jahren mit 58,1% signifikant sportaktiver (Z (AW-Test) = -3,35 / p = 0,000; Cramers V = 0,04 / p = 0,001) als diejenigen ohne Kinder in diesem Alter, von denen sich 53,2% am Sport beteiligen. Daraus könnte man nun den Schluss ziehen, dass die Eltern durch das starke Sportengagement ihrer Kinder selbst zum vermehrten Sporttreiben veranlasst werden. Aber ein Blick auf die Altersstruktur der beiden betrachteten Gruppen zeigt, dass dieser Effekt nicht überschätzt werden sollte, da das Durchschnittsalter der Personen mit Kindern in diesem Alter mit etwa 38 Jahren deutlich niedriger ist als die 45,5 Jahre derer ohne Kinder in diesem Alter. Auch die Messung dieses Zusammenhangs durch Cramers V (= 0,421) spricht dafür, dass der Effekt zu einem Teil durch den nachgewiesenen altersbedingten Rückgang der Sportaktivität verursacht ist.

2.3.5 Gesundheit

Die Untersuchung, inwieweit das subjektive gesundheitliche Wohlbefinden der Bevölkerung die Sportaktivität beeinflusst, kommt zu einem eindeutigen, intuitiv leicht nachvollziehbaren Resultat. Je besser der eigene Gesundheitszustand eingeschätzt wird, desto größer ist die Beteiligung am Sport ($\chi^2 = 319,17 / p = 0,000$; Cramers V = 0,215 / $p = 0,000$). Die Abstände in den Sportbeteiligungsquoten zwischen den Kategorien betragen immer mehr als 10 Prozentpunkte, so dass sich eine Spanne von 71,6% für Personen mit ausgezeichnetem bzw. sehr gutem Gesundheitszustand bis zu 26% für Personen mit schlechter gesundheitlicher Befindlichkeit ergibt. Allerdings ist anzunehmen, dass umgekehrt die sportliche Aktivität auch einen positiven Einfluss auf den Gesundheitszustand hat.

Unter dem Gesundheitsaspekt ist auch von Interesse, ob das Sportengagement der Ausdruck eines auch in anderen Bereichen gesundheitsbewussten Lebensstils ist. Treiben beispielsweise Nichtraucher mehr Sport als Raucher? Diese Frage ist für Gesamtdeutschland aufgrund der Testergebnisse zunächst zu bejahen. Eine differenzierte Betrachtung spricht aber dafür, dass dieser Zusammenhang nur für Westdeutschland vorliegt, nicht jedoch für Ostdeutschland (siehe Tabelle 5).

Tabelle 5: Sportbeteiligungsquoten, differenziert nach Rauchgewohnheit

	Gesamtdeutschland	Westdeutschland	Ostdeutschland
Raucher	50,4%	51,2%	47,2%
Nichtraucher	56,0%	57,9%	47,6%
χ^2	19,55 ($p = 0,000$)	18,65 ($p = 0,000$)	0,04 ($p = 0,847$)
Cramers V	0,053 ($p = 0,000$)	0,064 ($p = 0,000$)	0,004 ($p = 0,847$)

2.3.6 Schulbildung

Die empirischen Ergebnisse deuten auf einen relativ starken Zusammenhang zwischen dem Bildungsniveau und der Sportaktivität in der Bevölkerung hin. Dieser Zusammenhang äußert sich derart, dass das Sportengagement mit zunehmender Schulbildung steigt. Während sich von den Personen mit niedrigem Schulabschluss, also unterhalb der 10. Klasse, lediglich 42,4% aktiv am Sport beteiligen, steigt die Quote bei Personen mit mittlerem Schulabschluss nach der 10. Klasse zunächst auf 59,3% und bei Personen mit hohem Schulabschluss nochmals auf 72% ($\chi^2 = 400,34 / p = 0,000$; Cramers V = 0,243 / $p = 0,000$).

Die Erklärungsmöglichkeiten, die bisher von der Sportsoziologie für dieses Phänomen gefunden wurden, sind ziemlich beschränkt. Ein u.a. von Lehnertz (1979) propagierter Ansatz,

der eine größere Sportbindung als Resultat einer längeren Schulkarriere mit Sportunterricht sieht, wurde bereits durch die empirischen Untersuchungen von Schlagenhaut (1977) in Frage gestellt (Baur/Braun, 2001, S. 83).

2.3.7 Erwerbsstatus

Aus Tabelle 6 lässt sich das Vorliegen eines signifikanten Zusammenhangs zwischen Erwerbsstatus und Sportbeteiligung ablesen. Dabei bilden die Auszubildenden einerseits und die Vorruheständler, Rentner bzw. Pensionäre andererseits die beiden Gruppen, die sich am deutlichsten von den anderen unterscheiden. Dass erstere dabei am meisten und letztere am wenigsten Sport treiben, ist wohl in erster Linie mit der jeweiligen Altersstruktur in den beiden Gruppen zu erklären. Durch den Wert, den Cramers V bei der Messung des Zusammenhangs zwischen Alter und Erwerbsstatus annimmt (0,409), wird diese These gestützt. Die Annahme, dass Teilzeit- oder stundenweise Erwerbstätige aufgrund der größeren verfügbaren Freizeit mit höherer Wahrscheinlichkeit sportaktiv sind als Vollzeit-Erwerbstätige, kann durch die Stichprobenergebnisse nicht bestätigt werden (AW-Test: $Z = 0,16 / p = 0,873$ (Gesamt); $Z = 0,74 / p = 0,459$ (West); $Z = -0,72 / p = 0,472$ (Ost)).

Tabelle 6: Sportbeteiligungsquoten, differenziert nach Erwerbsstatus

	Gesamtdeutschland	Westdeutschland	Ostdeutschland
Vollzeit-erwerbstätig	60%	62,1%	51,5%
Teilzeit- oder stundenweise erwerbstätig	59,7%	60,5%	54,2%
arbeitslos	46,2%	42,4%	51,7%
Auszubildender	77,1%	78,5%	72,3%
Rentner, Pensionär, im Vorruhestand	35%	36,6%	29,9%
Sonstige Nichterwerbstätige	56,9%	56,6%	61,3%
χ^2	308,1 (p = 0,000)	203,9 (p = 0,000)	112,5 (p = 0,000)
Cramers V	0,214 (p = 0,000)	0,214 (p = 0,000)	0,222 (p = 0,000)

Als möglichen Grund für die geringere Sportaktivität von Arbeitslosen gegenüber Vollzeit-Erwerbstätigen wird in der Literatur auf ihre eingeschränkten Handlungsspielräume und den häufig stattfindenden Rückzug aus dem öffentlichen Leben verwiesen (Nagel, 2002, S. 51). Die differenzierte Auswertung der Stichprobe zeigt diesen Effekt jedoch nur für die Arbeitslosen im Westen (Z (AW-Test) = 5,05 / $p = 0,000$ (West)).

2.3.8 Haushaltseinkommen und soziale Schichtung

Sport treiben ist oft auch mit Kosten, etwa für Ausrüstungsgegenstände oder Mitgliedsbeiträge, verbunden. Daher erscheint es plausibel, dass zwischen dem verfügbaren Einkommen und der Sportaktivität ein Zusammenhang besteht (Nagel, 2002, S. 42). Das im Fragebogen des BGS98 abgefragte Haushaltsnettoeinkommen ist als Indikator für die ökonomische Stellung einer Person nicht unproblematisch, weil es nicht beachtet, wie viele Personen davon leben. Zur Berechnung eines sinnvollen Äquivalenzeinkommens, das auch die Haushaltgröße berücksichtigt, müsste das genaue Haushaltsnettoeinkommen aller Teilnehmer am BGS98 vorliegen. Da dieses jedoch im Fragebogen des BGS98 lediglich in vorgegebenen Intervallen abgefragt wurde, wird hier darauf verzichtet und auf das Haushaltsnettoeinkommen abgestellt.

In Tabelle 7 ist eine signifikante Einkommensabhängigkeit der Sportaktivität erkennbar ($\chi^2 = 204,93 / p = 0,000$; Cramers V = 0,192 / p = 0,000). Mit Ausnahme der letzten steigt mit jeder höheren Einkommensstufe die Sportbeteiligungsquote.

Tabelle 7: Sportbeteiligungsquoten, differenziert nach Haushaltsnettoeinkommen

Unter 2000 DM	2000 bis 3000 DM	3000 bis 4000 DM	4000 bis 5000 DM	5000 bis 6000 DM	6000 bis 8000 DM	Über 8000DM
43,4%	49,1%	54,9%	60,9%	69,3%	73,3%	71,5%

Die empirischen Ergebnisse über den Einfluss von Haushaltsnettoeinkommen und Schulabschluss deuten darauf hin, dass im Sport, wie auch in anderen sozialwissenschaftlichen Bereichen, eine soziale Schichtung vorliegt. D. h., je höher der soziale und ökonomische Status einer Person ist, desto eher ist sie sportlich aktiv. Zwar hat sich der Sport in der modernen Industriegesellschaft zu einem für alle gesellschaftlichen Schichten attraktiven Massenphänomen entwickelt, doch eine von der sozialen Lage unabhängige Gleichstellung hat noch nicht stattgefunden (Schwier, 2003, S. 4). Der BGS98 bietet hierfür den Schichtindex nach Winkler¹ an, der aus den beiden oben genannten Variablen "Haushaltsnettoeinkommen" und "Schulabschluss" sowie der Variablen "Berufliche Stellung" gebildet wird. Er nimmt eine Einteilung in Unter-, Mittel- und Oberschicht vor. Die Sportbeteiligungsquoten steigern sich in dieser Reihenfolge von 39,2% über 54,6% auf 69,6% ($\chi^2 = 286,82 / p = 0,000$; Cramers V = 0,204 / p = 0,000).

¹ Wie dieser Index berechnet wird, ist in Stolzenberg, 2000, S. 27ff., beschrieben.

3 Binäres Logit-Modell

Für die abhängige Variable "SPORTAK" soll ein geeignetes Erklärungsmodell gefunden werden. Da sie nur zwei Ausprägungen annehmen kann, also qualitativer Natur ist, ist eine Übertragung der Vorgehensweise des klassischen Regressionsmodells auf die Problemstellung nicht geeignet. Denn dieses setzt das Vorliegen einer metrisch skalierten, abhängigen Variablen voraus, die frei variieren kann (Tutz, 2000, S. 30). In einem linearen Wahrscheinlichkeitsmodell etwa wäre die Schätzung der dichotomen Variablen "SPORTAK" mit mehreren unerwünschten Problemen, z. B. Annahmeverletzungen der Kleinstquadratmethode, verbunden. Der entscheidende Nachteil wäre aber die Tatsache, dass die systematische Komponente nicht unbedingt zwischen 0 und 1 liegt (Gujarati, 2003, S. 583ff.). Somit wäre deren Interpretation als Wahrscheinlichkeit für ein bestimmtes Ereignis nicht mehr ohne weiteres möglich (Eckey/Kosfeld/Dreger, 1995, S. 171). Ein adäquates Verfahren für qualitative Regressanden stellen dagegen u. a. die Logit-Modelle dar. Mit dem binären Logit-Modell lässt sich die Abhängigkeitsstruktur der aus zwei Zuständen bestehenden Variable "SPORTAK" analysieren.

Die abhängige Variable kann in der Regel (i. d. R.) als ein Ereignis (hier: Sportaktivität) ausgelegt werden, welches eintritt ($Y = 1$) oder nicht eintritt ($Y = 0$). Im binären Logit-Modell wird für jedes Individuum i die Wahrscheinlichkeit für den Eintritt dieses Ereignisses mit Hilfe der logistischen Verteilungsfunktion, wie folgt, dargestellt:

$$(3.1) \quad P_i = \frac{1}{1 + e^{-Z_i}},$$

$$\text{mit } Z_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_j X_{ij}, \quad i = 1, \dots, n.$$

Ist $P_i \geq 0,5$, geht man von einer sportaktiven Person aus. Andernfalls, also wenn $P_i < 0,5$ ist, nimmt man an, dass die Person keinen Sport treibt (Bühl/Zöfel, 2002, S. 351). Für letzteres lässt sich die Wahrscheinlichkeit als Gegenwahrscheinlichkeit von P_i ermitteln:

$$(3.2) \quad 1 - P_i = \frac{1}{1 + e^{Z_i}}.$$

Setzt man (3.1) und (3.2) nun ins Verhältnis, ergibt sich folgendes, auch *odds ratio* genanntes, Wahrscheinlichkeits- bzw. Chancenverhältnis:

$$(3.3) \quad \frac{P_i}{1-P_i} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = e^{Z_i}.$$

Demnach stünden bei $P_i = 0,8$ die Chancen 4 zu 1, dass die Person Sport treibt. Durch Logarithmieren von (3.3) ergibt sich die Modellgleichung bzw. der Logit:

$$(3.4) \quad L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_j X_{ij}.$$

Während der Logit L_i linear in den Regressoren ist, zeigen die Verknüpfungen über die logistische Verteilungsfunktion in (3.1) und (3.2), dass dies auf die Wahrscheinlichkeiten nicht zutrifft (Gujarati, 2003, S. 596). Durch den S-förmigen Verlauf der logistischen Verteilungsfunktion weist die abhängige Variable keine konstanten Veränderungsraten auf. Das Ausmaß der Veränderung wird durch die Steilheit dieser Funktion im aktuellen Niveau bestimmt (Eckey/Kosfeld/Dreger, 1995, S. 175). Die Interpretation der Koeffizienten bezüglich der Wahrscheinlichkeiten wird erschwert, da sich aus den Koeffizienten der Regressoren allein lediglich eine Aussage darüber treffen lässt, in welche Richtung sich die Wahrscheinlichkeit durch bestimmte Veränderungen der Einflussgrößen bewegt (Eckey/Kosfeld/Dreger, 1995, S. 175). Wenn $\beta_j < 0$ ist, dann wird die Erhöhung eines Regressors X_{ij} die Wahrscheinlichkeit für das Individuum i , sportlich aktiv zu sein, senken. Für $\beta_j > 0$ dagegen steigt P_i mit wachsendem X_{ij} .

Am häufigsten werden Logit-Modelle mit der gewichteten Kleinst-Quadrate-Methode (WLS-Schätzung) und der Maximum-Likelihood-Methode (ML-Schätzung) geschätzt. Im hier vorliegenden Fall, dass nur kategoriale Variablen und eine große Stichprobe vorliegen, kann man für beide Methoden ähnliche Schätzergebnisse erwarten (Agresti, 1990, S. 462). Die ML-Schätzung ist in vielen Belangen überlegen (Urban, 1993, S. 53), weshalb sie hier angewendet wird.

4 Auswertung und Interpretation der Schätzergebnisse

In diesem Kapitel werden die Schätzergebnisse, die aus der Anwendung des binären Logit-Modells resultieren, dargestellt. Dabei werden im Prinzip alle Determinanten, die sich bei der Analyse in Kapitel 2 als signifikant erwiesen haben, zunächst in das zu schätzende Modell aufgenommen - auch diejenigen, die etwa nur in einem Teil Deutschlands oder für ein Geschlecht Signifikanz aufwiesen. Ausgenommen wird allerdings die Variable Partner, da

die hohe Anzahl von fehlenden Werten (vgl. Kapitel 2.3.4) die Aussagekraft der Schätzung deutlich einschränken würde. Um dem Prinzip der sparsamen Modellierung gerecht zu werden, wird außerdem anstelle der beiden Determinanten Haushaltsnettoeinkommen und Schulabschluss der mit ihnen relativ hoch korrelierte Schichtindex ($r_s = 0,658$ bzw. $0,48$) aus Kapitel 2.3.8 verwendet.

Bei der Aufstellung der zu schätzenden Modellgleichungen für das binäre Logit-Modell ist zu beachten, dass die erklärenden Variablen hier alle qualitativ bzw. kategorial sind. Würde man sie wie metrische Variablen ins Modell aufnehmen, wären die geschätzten Parameter nicht interpretierbar. Deshalb wird eine Darstellung mit 0-1-kodierten Dummy-Variablen gewählt (Tutz, 2000, S. 45). Für Variablen mit I Kategorien sind dabei nur $I-1$ Dummy-Variablen zu bilden, da sich die Darstellung einer Kategorie, der sogenannten Referenzkategorie, dadurch ergibt, dass alle $I-1$ Dummy-Variablen den Wert 0 annehmen (Tutz, 2000, S. 19). Die entsprechende Kodierung der vorliegenden Variablen ist dem Anhang I zu entnehmen.

Bevor mit der Auswertung der Schätzergebnisse begonnen wird, sollte ein Blick darauf geworfen werden, ob die Anwendungsvoraussetzungen für die ML-Schätzung erfüllt sind. Dazu gehören eine möglichst große Stichprobe, geringe lineare Beziehungen zwischen den unabhängigen Variablen bzw. keine Multikollinearität und für jedes Ereignis zumindest eine kleine Personengruppe der Stichprobe, die dieses gewählt hat (Urban, 1993, S. 53ff.). Eindeutig ist, dass eine ausreichend große Stichprobe, in der jede Handlungsalternative von einer ausreichenden Anzahl von Personen gewählt wurde, vorliegt. Wie die Analysen in Kapitel 2 gezeigt haben, gibt es zwischen den unabhängigen Variablen teilweise nennenswerte Zusammenhänge, z. B. zwischen Alter und Gesundheitszustand, allerdings nicht in Größenordnungen, in denen man Multikollinearität vermuten könnte.

Zu Beginn der Auswertung der Schätzergebnisse wurde mit Hilfe des Wald-Tests überprüft, welche der erklärenden Variablen im Schätzmodell einen signifikanten Einfluss auf die Sportaktivität ausüben. Ähnlich wie der t-Test in der linearen Regressionsanalyse testet der Wald-Test die Nullhypothese, dass ein bestimmter, geschätzter Koeffizient b_j gleich Null ist und die zugehörige Variable somit keinen Einfluss hat. Berechnet wird die Prüfgröße wie folgt:

$$(4.1) \quad t_{\text{wald}} = \left(\frac{b_j}{s_{b_j}} \right)^2, \text{ mit } s_{b_j} = \text{Standardfehler von } b_j.$$

Sie ist für jede einzelne Dummy-Variable χ^2 -verteilt mit einem Freiheitsgrad (Backhaus, 2000, S. 119f.).

Ausgehend von einem 5%-igen Signifikanzniveau erweist sich nur die Variable "KIND6_14" als nicht signifikant und wird aus dem Modell eliminiert. Allerdings sind von den fünf Dummy-Variablen, die die Variable "ERWERBST" darstellen, drei nicht signifikant und eine nur knapp signifikant (siehe Anhang III). Daher wird diese ebenfalls aus dem endgültigen Modell herausgenommen.

Mehrere Maße für die Anpassungsgüte des Regressionsmodells basieren auf der Likelihood-Funktion L bzw. auf deren negativen doppelten Wert des Logarithmus $-2LL$. Je kleiner $-2LL$ wird, desto höher ist die Anpassungsgüte (Bühl/Zöfel, 2002, S. 354). Beim Likelihood-Ratio-(LR-)Test wird der entsprechende Wert für obiges Schätzmodell (unrestringiertes Modell) mit dem Wert des Regressionsmodells (restringiertes Modell), das nur die Konstante als erklärende Variable beinhaltet, verglichen.

Die Prüfgröße

$$(4.2) \quad LR = -2(LL_r - LL_u), \quad \text{mit } r = \text{restringiert und } u = \text{unrestringiert,}$$

ist asymptotisch χ^2 -verteilt mit r Freiheitsgraden, wobei r die Anzahl der Regressionskoeffizienten ohne die Konstante darstellt (Eckey/Kosfeld/Dreger, 1995, S. 178). Hier nimmt sie einen Wert von 896,4 an. Damit wird die Nullhypothese, dass alle r Koeffizienten gleich Null sind, abgelehnt. Das unrestringierte Modell bietet also einen höheren Erklärungsgehalt.

Analog zum R^2 in der linearen Regression existieren auch in Logit-Modellen Bestimmtheitsmaße, die den Anteil der durch das Modell erklärten Varianz angeben (Bühl/Zöfel, 2002, S. 354). Dazu zählen u. a. die Maße von Cox/Snell und Nagelkerke. Der Unterschied zwischen diesen beiden Maßen liegt darin, dass letzteres im Gegensatz zu ersterem auch im optimalen Fall tatsächlich den Wert Eins annehmen kann (Backhaus, 2000, S. 133). Nach Nagelkerkes R^2 werden 16,4% der Varianz der abhängigen Variable durch die Varianz der im Modell verbliebenen unabhängigen Variablen erklärt (Cox/Snells $R^2 = 0,123$). Da solche Bestimmtheitsmaße im allgemeinen vergleichsweise niedrige Werte besitzen

(Gujarati, 2003, S. 586), kann man insgesamt zumindest von einem befriedigenden Modell sprechen. Zusätzlich ist zu beachten, dass es sich hierbei um modell-relative Gütemaße handelt, die nicht prüfen, ob die geschätzten den tatsächlich beobachteten Y-Werten entsprechen, sondern allein die Ergebnisse zweier Logit-Schätzungen vergleichen (Urban, 1993, S. 63). Abhilfe für diesen Mangel kann die Klassifikationsstabelle (siehe Tabelle 9) leisten, in der die vorhergesagten Wahlhandlungen bezüglich der Sportaktivität den beobachteten gegenübergestellt werden.

Tabelle 9: Klassifikationstabelle für die Sportaktivität

		Vorhergesagte Sportaktivität		Prozentsatz der Richtigen
		inaktiv	aktiv	
Beobachtete Sportaktivität	inaktiv	1738	1386	55,6
	Aktiv	983	2736	73,6
Gesamtprozentsatz				65,4

In der Gesamtbetrachtung liegt die Prognose in 65,4% der Fälle richtig. Vergleicht man dies mit dem Fall, in dem keine Einflussfaktoren ins Modell aufgenommen werden und alle Personen der anteilmäßig größeren Kategorie zugeordnet werden (54,3%), liefert das Modell mit erklärenden Variablen eine bessere Prognose. Dabei scheint die Vorhersagefähigkeit des Modells für Aktive, von denen 73,6% richtig eingeschätzt wurden, besser zu sein als für Inaktive (55,6%). Die dieser Klassifikation zugrunde liegende Regel, dass für $P_i \geq 0,5$ Sportaktivität und für $P_i < 0,5$ Inaktivität prognostiziert wird, ist aber durchaus kritisch zu sehen. Denn die Tatsache, dass beispielsweise die geschätzten Wahrscheinlichkeiten von 98% und 51% zur gleichen Prognose führen, lässt Zweifel an der Genauigkeit der Klassifikation zu (Urban, 1993, S. 68).

Die Interpretation der Regressionskoeffizienten stützt sich auf die in Tabelle 8 wiedergegebenen Schätzergebnisse.

Tabelle 8: Schätzung der Sportaktivität mit dem binären Logit-Modell

Variable	Regressionskoeffizient b	Standardfehler s_b	t_{Wald}	FG	Sig.	Exp(b)
ALTER			285,152	5	0,000	
ALTER2	-0,629	0,085	54,188	1	0,000	0,533
ALTER3	-0,697	0,090	60,300	1	0,000	0,498
ALTER4	-0,911	0,091	100,149	1	0,000	0,402
ALTER5	-1,236	0,099	157,107	1	0,000	0,291
ALTER6	-1,908	0,125	232,610	1	0,000	0,148
RAUCHEN1	-0,473	0,058	67,081	1	0,000	0,623
GESUND			91,343	3	0,000	
GESUND2	-0,484	0,071	46,490	1	0,000	0,616
GESUND3	-0,813	0,094	74,656	1	0,000	0,443
GESUND4	-1,280	0,222	33,291	1	0,000	0,278
GESCHL1	0,138	0,052	6,947	1	0,008	1,148
OSTWEST1	0,293	0,066	19,989	1	0,000	1,341
SCHICHT			196,955	2	0,000	
SCHICHT2	0,536	0,066	66,956	1	0,000	1,710
SCHICHT3	1,162	0,083	196,920	1	0,000	3,195
Konstante	0,237	0,111	4,519	1	0,034	1,267

Der Aussagegehalt der, nur knapp signifikanten, Konstante wird deutlich, wenn der Fall betrachtet wird, dass alle Dummy-Variablen im Modell den Wert Null annehmen. Dies trifft genau dann zu, wenn für alle Merkmale die Referenzkategorie angenommen wird (Tutz, 2000, S. 61). Hier wird also von einer weiblichen, ostdeutschen Person im Alter von 18 bis 29 Jahren, die einen sehr guten Gesundheitszustand hat, nicht raucht, der Unterschicht zugeordnet ist und kein Kind im Alter von 6 bis 14 Jahren hat, ausgegangen. Das positive Vorzeichen der Konstante lässt darauf schließen, dass die Wahrscheinlichkeit P_i dieser Person, sportlich aktiv zu sein, größer ist als die, inaktiv zu sein - und zwar genau 1,267 mal so groß. Dieser Wert ergibt sich, wenn man den Wert der Konstante in die Gleichung der *Odds Ratio* (3.3) einsetzt, also $e^{0,237}$. Gemäß (3.1) beträgt P_i hier 56,6%, was für die Sportbeteiligung der Person spricht. Der Wert der Konstante (0,237), isoliert betrachtet, steht hier für den wenig anschaulichen Wert des Logits aus (3.4).

Um die anderen Schätzkoeffizienten interpretieren zu können, muss man aufgrund der Darstellung durch Dummy-Variablen immer den Bezug zur jeweiligen Referenzkategorie herstellen. Die Veränderung der *Odds Ratio* (3.3) respektive des Logits (3.4), die aus dem

Übergang von einer Referenzkategorie I zu einer anderen Kategorie J der gleichen Einflussgröße A, ceteris paribus (c. p.), resultiert, kann wie folgt bestimmt werden:

$$(4.3) \quad \frac{P_i(A = J)/(1 - P_i(A = J))}{P_i(A = I)/(1 - P_i(A = I))} = e^{\beta_{A(J)}}$$

$$(4.4) \quad \ln\left(\frac{P_i(A = J)/(1 - P_i(A = J))}{P_i(A = I)/(1 - P_i(A = I))}\right) = \beta_{A(J)}.$$

Die Wirkung in (4.3) ist multiplikativ, d. h., die Odds Ratio erhöht sich um diesen Faktor, wohingegen die Veränderung des Logits in (4.4) additiver Natur ist (Tutz, 2000, S. 61f.). Letzteres wird wegen der schon angesprochenen mangelnden Anschaulichkeit bei den folgenden Interpretationen nicht mehr explizit berücksichtigt.

Aus den negativen Vorzeichen der Regressionskoeffizienten der Dummy-Variablen, die die Variable "ALTER" darstellen, ist ersichtlich, dass, ausgehend von der Referenzkategorie der 18- bis 29-Jährigen, die Wahrscheinlichkeit, sportlich aktiv zu sein, in den anderen Altersklassen sinkt. Der Vergleich der Koeffizientenwerte zeigt, dass dieser Rückgang umso stärker ausfällt, je höher das Alter bzw. die Altersklasse ist. Exemplarisch sei der Koeffizient für die Gruppe der 70- bis 79-Jährigen (-1,916) herausgehoben. Multipliziert man das Chancenverhältnis bzw. die *Odds Ratio* in der Referenzkategorie mit $e^{-1,916} = 0,147$, dann ergibt sich, c. p., das Chancenverhältnis dieser Gruppe. Für den Fall, dass sonst alle Merkmale die Referenzkategorie annehmen, betrüge das Chancenverhältnis dann $e^{0,237} \cdot e^{-1,916} = 0,187$. Die Wahrscheinlichkeit, keinen Sport zu treiben, ist also mit 0,843 mehr als 5 mal so groß wie die Wahrscheinlichkeit, sportlich aktiv zu sein ($P_i = 0,157$). Daher kann davon ausgegangen werden, dass sich die Person nicht am Sport beteiligt. Durch die Veränderung der Dummy-Variablen verringert sich P_i also um $0,566 - 0,157 = 0,409$.

Die bei metrischen Variablen übliche Berechnung des Einflusses, den eine marginale Veränderung einer erklärenden Variable auf die Wahrscheinlichkeit ausübt, über die partielle Ableitung ($\partial P_i / \partial X_k = P_i (1 - P_i) b_k$), ist im Prinzip für Dummy-Variablen nicht geeignet. Denn Ableitungen einer Funktion sind nur für kontinuierliche, nicht für diskrete Variablen definiert. Somit wird der Effekt bei Veränderung einer Dummy-Variablen i. d. R. durch die partielle Ableitung überschätzt, bietet aber zumindest eine grobe Approximation (Liao, 1994, S. 18ff.). Mit $0,566 \cdot (1 - 0,566) \cdot (-1,916) = -0,471$ wird dies auch für obiges Beispiel bestätigt.

Analog lassen sich auch die Koeffizienten der anderen Variablen aus Tabelle 8 ablesen und interpretieren. Mit folgender Übersicht soll lediglich kurz deren jeweilige Wirkungsrichtung beschrieben werden:

- Rauchen hat eine negative Wirkung auf die Wahrscheinlichkeit, Sport zu treiben.
- Je besser der eigene Gesundheitszustand eingeschätzt wird, umso größer wird die Wahrscheinlichkeit, Sport zu treiben.
- Ein Mann, der ansonsten die gleichen Merkmalsausprägungen wie eine Frau besitzt, ist mit höherer Wahrscheinlichkeit sportaktiv.
- In den neuen Bundesländern zu leben, wirkt sich im Vergleich zum Wohnen in den alten Bundesländern negativ auf die Wahrscheinlichkeit für Sportaktivitäten aus.
- Mit der Zugehörigkeit zu einer "höheren" Schicht steigt die Wahrscheinlichkeit, sich am Sport zu beteiligen.

Zur Verdeutlichung der Wirkungsweise des geschätzten Modells soll exemplarisch die Sportaktivität einer Person, die bei mehr als einer Variablen nicht zur Referenzkategorie gehört (18- bis 29-Jahre, Nichtraucher, guter Gesundheitszustand, männlich, Westdeutscher, zur Mittelschicht gehörig), analysiert werden. Aus (3.1) in Verbindung mit (3.4) berechnet sich die Wahrscheinlichkeit für diese Person, sportaktiv zu sein:

$$L_i = 0,247 - 0,484 \cdot 1 + 0,138 \cdot 1 + 0,293 \cdot 1 + 0,536 \cdot 1 = 0,73$$

$$P_i = \frac{1}{1 + e^{-L}} = \frac{1}{1 + e^{-0,73}} = 0,675.$$

Da diese größer als 0,5 ist, kann man davon ausgehen, dass die Person Sport treibt. Die zugehörige *Odds Ratio* ist $e^{0,73} = 2,075$. Im Falle eines "Aufstiegs in die Oberschicht" würde sie sich, c. p., auf $e^{0,73} \cdot e^{(1,162 - 0,536)} = 3,881$ ($b_{1,\text{Schicht}} = 0,536$, $b_{2,\text{Schicht}} = 1,162$) erhöhen. Daraus folgt eine gestiegene Wahrscheinlichkeit P_i von 0,795, die noch stärker für ein sportliches Engagement der Person spricht.

Eine Möglichkeit auch den Umfang der Sportaktivität, d.h. wie viel Zeit pro Woche für Sport verwendet wird, in einem Modell zu erklären, ist das ordinale Logit-Modell. Da jedoch eine entsprechende Schätzung keinen zusätzlichen Erkenntnisgewinn lieferte, wird hier auf die Darstellung dieses Modells verzichtet.

5 Fazit

Die vorangegangene Analyse der Daten des BGS98 zeigt, dass entgegen anderslautender Vermutungen viele sozialstrukturelle Merkmale einen signifikanten Einfluss auf die Sportaktivität der deutschen Bevölkerung haben bzw. im Jahr 1998 hatten. Mit folgender stichpunktartiger Aufstellung sollen die wichtigsten Ergebnisse zusammengefasst werden:

- Die Gesamt-Sportbeteiligungsquote in der gewichteten Stichprobe beträgt 54,2%.
- Das Sportengagement der Westdeutschen ist stärker als das der Ostdeutschen.
- Die gesamtdeutsche Sportaktivität ist abhängig vom Geschlecht. Während 56,7% der Männer Sport treiben, sind dies bei den Frauen lediglich 51,7%. Allerdings handelt es sich dabei um ein rein westdeutsches Phänomen, das sich im Gesamtergebnis niederschlägt.
- Mit zunehmendem Alter reduziert sich die Beteiligung der Bevölkerung am Sport. Im geschätzten Logit-Modell erweist es sich als stärkste Einflussgröße auf die Sportaktivität.
- Während in Partnerschaft lebende Frauen stärker sportlich engagiert sind als alleinlebende, ist dies bei Männern genau umgekehrt.
- Das Vorhandensein von Kindern unter 5 Jahren hat offensichtlich keinen Einfluss auf die Sportaktivität der Individuen, wohingegen ältere Kinder (6-14 Jahre) das Sportengagement ihrer Eltern erhöhen.
- Ein guter gesundheitlicher Zustand wirkt sich positiv auf die sportliche Aktivität aus. Zumindest im Westen treiben Nichtraucher mehr Sport als Raucher.
- Determiniert wird die Sportaktivität auch durch den Erwerbsstatus. Die höchste sportliche Beteiligung weisen Auszubildende, gefolgt von Voll- und Teilzeitbeschäftigten, auf. Am wenigsten Sport treiben die Arbeitslosen.
- Das Sportengagement steigt mit zunehmender Schulbildung.
- Es zeigt sich eine signifikante Einkommensabhängigkeit der Sportbeteiligung.
- Die "soziale Schicht" einer Person, die durch das Haushaltsnettoeinkommen, die Schulbildung und die berufliche Stellung definiert wird, spielt für deren sportliches Engagement eine Rolle. Personen aus der Oberschicht sind mit einer höheren Wahrscheinlichkeit sportlich aktiv als Personen aus der Mittelschicht und diese sind wiederum häufiger am Sport beteiligt als Personen aus der Unterschicht.

Anhand der Schätzergebnisse lässt sich erkennen, dass durch die Aufnahme der genannten Merkmale als unabhängige Variablen in das binäre Logit-Modell etwa 15% der Variation der abhängigen Variablen erklärt werden kann. Dadurch wird deutlich, dass trotz der heutzutage vorherrschenden starken Individualisierung, Pluralisierung und Differenzierung des Sports auch relevante objektive Determinanten einwirken (Baur/Braun, 2001, S. 84).

Anhang I: Variablenübersicht

Variablenname	Variablenlabel	Kategorien
SPORTAK	Sportaktivität	SPORTAK0: inaktiv SPORTAK1: aktiv
GESCHL	Geschlecht	GESCHL0: weiblich GESCHL1: männlich
OSTWEST	Teil Deutschlands	OSTWEST0: Ost OSTWEST1: West
SCHICHT	Soziale Schicht nach Winkler	SCHICHT1: Unterschicht SCHICHT2: Mittelschicht SCHICHT3: Oberschicht
RAUCHEN	Rauchgewohnheit	RAUCHEN0: Nichtraucher RAUCHEN1: Raucher
ALTER	Altersklassen	ALTER1: 18-29 Jahre ALTER2: 30-39 Jahre ALTER3: 40-49 Jahre ALTER4: 50-59 Jahre ALTER5: 60-69 Jahre ALTER6: 70-79 Jahre
GESUND	Allgemeiner Gesundheitszustand	GESUND1: ausgezeichnet, sehr gut GESUND2: gut GESUND3: weniger gut GESUND4: schlecht
ERWERBST	Erwerbsstatus	ERWERBST1: vollzeit-erwerbstätig ERWERBST2: teilzeit- oder stundenweise erwerbstätig ERWERBST3: arbeitslos ERWERBST4: Auszubildender ERWERBST5: Rentner, Pensionär, im Vorruhestand ERWERBST6: sonstige
KIND0_5	Kind(er) zwischen 0 und 5 Jahren	KIND0_50: nein KIND0_51: ja
KIND6_14	Kind(er) zwischen 6 und 14 Jahren	KIND6_140: nein KIND6_141: ja
PARTNER	Zusammenleben mit Partner	PARTNER0: nein PARTNER1: ja
SCHULAB	Schulabschluss	SCHULAB1: niedrig (bis 9. Klasse) SCHULAB2: mittel (bis 10. Klasse) SCHULAB3: hoch (ab 12. Klasse)
EINK	Haushaltseinkommen	EINK1: unter 2000 DM EINK2: 2000-3000 DM EINK3: 3000-4000 DM EINK4: 4000-5000 DM EINK5: 5000-6000 DM EINK6: 6000-8000 DM EINK7: mehr als 8000 DM

Anhang II: Signifikanztests für den Vergleich zwischen Ost und West, differenziert nach Geschlecht

	Männer	Frauen
χ^2	24,52 (p = 0,000)	7,32 (p = 0,007)
Cramers V	0,085 (p = 0,000)	0,041 (p = 0,007)
Z (M.W.-U-Test)	-7,12 (p = 0,000)	-4,55 (p = 0,000)

Anhang III: Vorläufige Schätzung der Sportaktivität mit dem binären Logit-Modell

Variablen in der Gleichung

Variablen	Regressions- koeffizient b	Standard- fehler s _b	t _{wald}	FG	Sig.	Exp(b)
ALTER			132,868	5	0,000	
ALTER2	-0,543	0,091	35,666	1	0,000	0,581
ALTER3	-0,622	0,095	42,594	1	0,000	0,537
ALTER4	-0,799	0,097	67,417	1	0,000	0,450
ALTER5	-1,120	0,134	70,320	1	0,000	0,326
ALTER6	-1,815	0,169	115,867	1	0,000	0,163
RAUCHEN1	0,493	0,059	70,494	1	0,000	1,637
ERWERBST			19,537	5	0,002	
ERWERBST2	0,180	0,087	4,340	1	0,037	1,198
ERWERBST3	-0,075	0,122	0,378	1	0,539	0,928
ERWERBST4	0,714	0,184	15,110	1	0,000	2,041
ERWERBST5	0,012	0,121	0,010	1	0,920	1,012
ERWERBST6	0,093	0,093	1,011	1	0,315	1,098
GESUND			80,121	3	0,000	
GESUND2	-0,455	0,072	39,997	1	0,000	0,635
GESUND3	-0,789	0,096	67,560	1	0,000	0,454
GESUND4	-1,194	0,231	26,628	1	0,000	0,303
GESCHL1	0,185	0,060	9,671	1	0,002	1,204
OSTWEST1	0,275	0,068	16,294	1	0,000	1,316
SCHICHT			203,640	2	0,000	
SCHICHT2	0,573	0,068	70,409	1	0,000	1,774
SCHICHT3	1,225	0,086	203,426	1	0,000	3,403
Konstante	0,027	0,128	0,044	1	0,834	1,027

Variablen nicht in der Gleichung

Variablen	Wert	FG	Sig.
KIND6_141	0,853	1	0,356
Gesamtstatistik	0,853	1	0,356

Literatur

Agresti, A., *Categorical Data Analysis*, New York, 1990.

Alt, C., Bien, W., Gewichtung, ein sinnvolles Verfahren in den Sozialwissenschaften? Fragen Probleme und Schlussfolgerungen, Gabler, Hofmeyer-Zlotnik, in: Krebs (Hrsg.), *Gewichtung in der Umfragepraxis*, Opladen, 1994, S. 124-140.

Backhaus, K., *Multivariate Analysemethoden*, 10. Auflage, Berlin, 2002.

Baur, J., Braun, S., *Der vereinsorganisierte Sport in Ostdeutschland*, Bundesinstitut für Sportwissenschaft (Hrsg.), *Wissenschaftliche Berichte und Materialien*, Band 12, Köln, 2001.

Bosch, K., *Großes Lehrbuch der Statistik*, München/Wien, 1996.

Bühl, A., Zöfel, P., *SPSS 11 - Einführung in die moderne Datenanalyse unter Windows*, 8. Auflage, München, 2002.

Digel, H., Die Versportlichung unserer Kultur und deren Folgen für den Sport - ein Beitrag zur Uneigentlichkeit des Sports, in: Gabler, H., Göhner, U. (Hrsg.), *Für einen besseren Sport...: Themen, Entwicklungen und Perspektiven aus Sport und Sportwissenschaft*, Schorndorf, 1990, S. 73-96.

Eckey, H.-F., Kosfeld, R., Dreger, C., *Ökonometrie*, Wiesbaden, 1995.

Erlinghagen, M., Wer treibt Sport im geteilten und vereinten Deutschland? - Eine quantitative Analyse sozio-ökonomischer Determinanten des Breitensports, <http://www.iatge.de/aktuell/veroeff/2003/gr2003-04.pdf>, Stand: 2003, Abruf 10.11.2003.

Farrell, L., Shields, M. A., Investigating the economic and demographic determinants of sporting participation in England, in: *Journal of the Royal Statistical Society, Series A*, No.165, Part 2, 2002, S. 335-348.

Gujarati, D.N., *Basic Econometrics*, 4. Auflage, Boston, 2003.

Janssen, J., Laatz, W., *Statistische Datenanalyse mit SPSS für Windows*, 4. Auflage, Berlin/Heidelberg, 2003.

Liao, T. F., *Interpreting Probability Models : Logit, Probit, and Other Generalized Linear Models*, Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-101, Thousand Oaks, 1994.

Meyer, B., Ahlert, G., *Die ökonomischen Perspektiven des Sports*, Schriftenreihe des Bundesinstituts für Sportwissenschaft, Band 100, Köln, 2000.

Nagel, M., Sportengagements im soziostrukturellen Kontext – Eine repräsentative Untersuchung über soziale Ungleichheit im Sport (Kapitel 4,5 und 6), <http://www.uni-potsdam.de/u/asug/Lehre/WS2002-2003/Sozial-strukturanalyse/nagel.pdf>, Stand 2002, Abruf 02.11.2003.

Rösch, G., Kriterien der Gewichtung einer nationalen Bevölkerungsstichprobe, in: Gabler, Hoffmeyer-Zlotnik, Krebs (Hrsg.), Gewichtung in der Umfragepraxis, Opladen, 1994, S. 7-26.

Schwier, J., Sport und soziale Ungleichheit, <http://uni-giessen.de/~g51039/> VorlesungVI.htm, Stand 28.05.2003, Abruf 26.10.2003.

Stolzenberg, H., Bundesgesundheitsurvey, Public Use File BGS98, Dokumentation des Datensatzes, Berlin, 2000.

Tutz, G., Die Analyse kategorialer Daten, München/Wien, 2000.

Urban, D., Logit-Analyse, Stuttgart, 1993.

Autor:

Dipl.-Kfm. Martin Flohr, Projektbearbeiter

Bisher erschienene Arbeitspapiere:

1. Peter M. Schulze, Prognoseverfahren wissenschaftlicher Institute in der Bundesrepublik Deutschland. Überblick über eine Umfrage (Dezember 1993)
2. Martina Nold / Peter M. Schulze, Möglichkeiten und Grenzen der Quantifizierung der Schattenwirtschaft (April 1994)
3. Armin Seher, Einfluß der Integrationsordnung bei Zeitreihen auf die Spezifikation von Fehlerkorrekturmodellen (Juni 1994)
4. Lars Berg / Armin Gemünden / Frank Hubert / Ralf Leonhardt / Michael Leroudier, Die Situation der Studentenschaft in den Wirtschaftswissenschaften an der Universität Mainz im Frühjahr 1994. Ergebnisse einer Umfrage (August 1994)
5. Christoph Balz, Ein Fehlerkorrekturmodell zur Entwicklung des Kapitalmarktzinses in der Bundesrepublik Deutschland (Oktober 1994)
6. Reinhard Elkmann / Nora Lauterbach / Stephan Wind, Tertiärisierung regionaler Wirtschaftsstrukturen. Eine empirische Analyse kreisfreier Städte und Landkreise in Hessen, Rheinland-Pfalz und dem Saarland (Dezember 1994)
7. Peter M. Schulze / Uwe Spieker, Deutsche Aktienindizes. Statistische Konzepte und Beispiele (Dezember 1994)
8. Armin Seher / Peter M. Schulze, Fehlerkorrekturmodelle und die Bewertung von Aktienkursindizes. Empirische Analyse zur Eignung des Konzepts (Januar 1995)
9. Reinhard Elkmann / Annette Klostermann / Kerstin Lieder, Zur intertemporalen Konstanz der Struktur regionaler Lohn- und Gehaltsniveaus in der Bundesrepublik Deutschland (Mai 1995)
10. Christoph Fischer, Ein Fehlerkorrekturmodell zur Kaufkraftparitätentheorie (März 1996)
11. Ralf Becker / Claudia Müller, Zur Schätzung regionaler Konsumfunktionen (Oktober 1996)
12. Frank Hubert, Klassifizierung der Arbeitsmärkte in den OECD-Ländern mittels Cluster- und Diskriminanzanalyse (April 1997)
13. Frank Hubert, Das Okun'sche Gesetz: Eine empirische Überprüfung für ausgewählte OECD-Länder unter besonderer Berücksichtigung der nationalen Arbeitsmarktordnungen (September 1997)
14. Christoph Balz/ Peter M. Schulze, Die Rolle nationaler, regionaler und sektoraler Faktoren für die Variation von Output, Beschäftigung und Produktivität in der Bundesrepublik Deutschland (Dezember 1997)

15. Peter M. Schulze, Steigende Skalenerträge und regionales Wachstum:
Eine quantitative Analyse mit kleinräumigen Daten (März 1998)
16. Ralf Becker, Die Verallgemeinerte Momentenmethode (Generalized Method of Moments - GMM).
Darstellung und Anwendung (Juni 1998)
17. Peter M. Schulze, Regionales Wachstum: Sind die Dienstleistungen der Motor?
(August 1998)
18. Ke Ma, Absatzanalyse für den chinesischen Pkw-Markt (Oktober 1998)
19. Christoph Balz/Peter M. Schulze, Die sektorale Dimension der Konvergenz.
Eine empirische Untersuchung für die Bundesrepublik Deutschland (Januar 1999)
- 20.* Robert Skarupke, Quantifizierung des Heimvorteils im deutschen Profifußball:
Eine empirische Untersuchung für die 1. Fußball-Bundesliga (August 2000)
- 21.* Peter M. Schulze, Regionalwirtschaftlicher Datenkatalog für die Bundesrepublik Deutschland
(September 2000)
- 22.* Yvonne Lange, Ein logistisches Regressionsmodell zur Analyse der Verkehrsmittelwahl im Raum
Mainz (Oktober 2000)
- 23.* Verena Dexheimer, Zählmodellen (Count Data Models). Ansätze und Anwendungen (Mai
2002)
- 24.* Andreas Handel, Die Entwicklung des Geldvermögens der privaten Haushalte in Deutschland
(September 2003)
- 25.* Christina Bastian/Yvonne Lange/Peter M. Schulze, Hedonische Preisindizes - Überblick und An-
wendung auf Personalcomputer (Mai 2004)
- 26.* Alexander Prinz/Peter M. Schulze, Zur Entwicklung von Containerschiffsflotten - Eine Paneldat-
enanalyse - (Mai 2004)
- 27.* Martin Flohr, Analyse der ökonomischen und demografischen Determinanten von Sportaktivitä-
ten in Deutschland (Juni 2004)

* Im Internet unter <http://www.statoek.vwl.uni-mainz.de/scpu.htm> verfügbar.