

# Humboldt-Universität zu Berlin



Wirtschaftswissenschaftliche Fakultät  
Institut für Statistik und Ökonometrie  
Lehrstuhl Statistik

Bachelor-Abschlussarbeit  
Sommersemester 2005

## **Analyse der Arbeitslosigkeit mit Logit-Modellen**

Eine Anwendung auf ALBUS-Daten

Eingereicht von: Anzela Jotkute  
Eingereicht bei: Prof. Dr. Bernd Rönz  
Matrikelnummer: 3301685  
Studiengang: Statistik  
Adresse: Brüsseler Str.24, 13353 Berlin  
Abgabetermin: 03.08.2005

## Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit „Analyse der Arbeitslosigkeit mit Logit-Modellen“ selbstständig verfasst habe. Andere als die angegebenen Hilfsmittel und Quellen habe ich nicht benutzt. Die Arbeit hat keiner anderen Prüfungsbehörde vorgelegen.

Berlin, den 03.08.2005

Unterschrift:

## Inhaltsverzeichnis

<b>1 Einführung</b>	<b>6</b>
<b>2 Beschreibung der Daten</b>	<b>8</b>
2.1 Datenherkunft .....	8
2.2 Das Datenmaterial .....	8
2.3 Die Stichprobenziehung .....	9
2.4 Bildung der abhängigen Variable .....	9
2.5 Wahl der unabhängigen Variablen .....	10
<b>3 Voranalyse</b>	<b>14</b>
3.1 Theoretischer Hintergrund .....	14
3.2 Ergebnisse bivariater Analyse .....	14
3.2.1 Jahr 2002 .....	15
3.2.2 Jahr 2004 .....	16
<b>4 Logit-Modell</b>	<b>20</b>
4.1 Theoretischer Hintergrund .....	20
4.2 Praktische Durchführung und Ergebnisse .....	22
4.2.1 Jahr 2002 .....	23
4.2.2 Jahr 2004 .....	25
<b>5 Vergleich der beiden Jahre</b>	<b>30</b>
<b>6 Anhang</b>	<b>32</b>
<b>7 Literaturangaben</b>	<b>41</b>

## Tabellenverzeichnis

2.1	Erklärende Variablen für das Jahr 2002 .....	11
2.2	Erklärende Variablen für das Jahr 2004 .....	12
3.1	Chi-Quadrat-Unabhängigkeitstest für das Jahr 2002 .....	16
3.2	Chi-Quadrat-Unabhängigkeitstest für das Jahr 2004 .....	18
4.1	Koeffizienten des Logit-Modells für das Jahr 2002 .....	23
4.2	Koeffizienten des Logit-Modells für das Jahr 2004 .....	25
A.1	Stichprobenverteilung: Einflussgrößen und Variable arbeitslos für das Jahr 2002 .....	32
A.2	Stichprobenverteilung: Einflussgrößen und Variable arbeitslos für das Jahr 2004 .....	33
A.3	Stichprobenverteilung für getrennte Dateien für das Jahr 2004 .....	34
A.4	Testergebnisse der logistischen Regression für getrennte Dateien für das Jahr 2004 .....	
A.5	Effekte der logistischen Regression für Männer in 2004 .....	35
A.6	Effekte der logistischen Regression für Frauen in 2004 .....	36
A.7	Effekte der logistischen Regression für jüngere Leute (18-44) in 2004 .....	37
A.8	Effekte der logistischen Regression für ältere Leute (45-67) in 2004 .....	37
A.9	Effekte der logistischen Regression für alte Bundesländer in 2004 .....	38
A.10	Effekte der logistischen Regression für neue Bundesländer in 2004 .....	38
A.11	Effekte der logistischen Regression für jüngere Männer in 2004 .....	39
A.12	Effekte der logistischen Regression für ältere Männer in 2004 .....	39
A.13	Effekte der logistischen Regression für jüngere Frauen in 2004 .....	40
A.14	Effekte der logistischen Regression für ältere Frauen in 2004 .....	40

## Abbildungsverzeichnis

3.1 Arbeitslos und Familienstand .....	15
3.2 Arbeitslos und Schulausbildung .....	15
3.3 Arbeitslos und Gesundheitszustand .....	16
3.4 Arbeitslos und berufliche Stellung .....	17
3.5 Arbeitslos und Erhebungsgebiet getrennt nach Geschlechtszugehörigkeit .....	17
3.6 Arbeitslos und Rauchen getrennt nach Geschlechtszugehörigkeit .....	18

# 1 Einführung

Bei der Suche nach einem Thema für meine Bachelor-Arbeit hatte ich einige interessante Ideen. Auf jeden Fall sollte es keine rein theoretische Arbeit werden, sondern eine empirische Untersuchung mit echten Daten. Das Problem war aber, passende Daten zu finden. Es werden auch verschiedene Datensätze angeboten: kostenpflichtige und kostenfreie. Eine meiner Ideen war es, in irgendeiner Weise Arbeitslosigkeit zu untersuchen. Es ist ein sehr schmerzhaftes Problem, dass viele Menschen nicht nur in Deutschland betrifft. So bin ich auf die Daten der „Gesellschaft sozialwissenschaftlicher Infrastruktureinrichtungen“ (GESIS) gestoßen, die unter anderem auch Informationen für eine mögliche Analyse, weshalb eine Person arbeitslos wird und eine andere nicht, gesammelt hat.

## *Ziel der Arbeit*

Ziel der Arbeit ist es, zu untersuchen, ob eine Abhängigkeit bzw. ein Zusammenhang zwischen der Variable arbeitslos oder nicht arbeitslos und den anderen später genannten Variablen, die hier als abhängige (exogene) Variablen benutzt werden, besteht. Man stellt sich vor, dass es vielerlei solcher Abhängigkeiten gibt. Natürlich ist es nicht möglich und auch nicht Sinn der statistischen Untersuchung, alle Variablen, die die abhängige Variable beeinflussen, in die Analyse einzubeziehen. In dieser Arbeit ist es beabsichtigt, nur die Variablen einzubeziehen, die von der ALLBUS-Umfrage erfasst worden sind und die sinngemäß einen Einfluss bzw. Zusammenhang darstellen könnten. Aus diesem Grund wird durch die einbezogenen Variablen nur ein Teil der Variabilität der abhängigen Variablen erfasst. Das Ergebnis der statistischen Untersuchung wäre, festzustellen, ob sich zum Beispiel Vorurteile, wie Christen sind fleißiger (geben sich mehr Mühe bei der Arbeit oder bei der Arbeitssuche), bestätigen lassen oder nicht. Andererseits ist es auch interessant zu sehen, wie die sich leider üblichen bekannten Verhältnisse, dass Frauen, ältere Menschen oder Personen aus den neuen Bundesländern häufiger arbeitslos sind, in dieser Untersuchung niederschlagen.

Ziel dieser Arbeit ist also die Analyse der Arbeitslosigkeit mittels Logit-Modell.

Im Voraus wurden einige einführende Überprüfungen von Zusammenhängen von nominal- bzw. ordinalskalierten Variablen unternommen. Es bieten sich die Kontingenztabelle, der Chi-Quadrat-Unabhängigkeitstest nach Pearson und einige Zusammenhangsmaße an. In der bivariaten Kontingenztabelle sind neben den absoluten und relativen Zellhäufigkeiten die erwarteten Zellhäufigkeiten für die weitere Untersuchung von besonderem Interesse. Es ist diejenige absolute Häufigkeit in jeder Zelle, die unter der Annahme der stochastischen Unabhängigkeit zu erwarten ist. Im Ergebnis zeigten die benutzten Variablen eine gewisse Abhängigkeit. Es macht also durchaus Sinn, diese Zusammenhänge noch gründlicher und tiefer zu untersuchen. Einige repräsentable Ergebnisse der bivariaten Analyse wurden zum besseren Verständnis und zur Veranschaulichung auch direkt in diese Arbeit übernommen.

Hauptaufgabe dieser Arbeit ist es aber, die Arbeitslosigkeit aus relativ großen Stichproben aus zwei unterschiedlichen Jahren mittels Logit-Modellen möglichst umfangreich und interpretierbar zu untersuchen.

Zum Schluss werden die Ergebnisse der beiden Jahre miteinander verglichen.

## 2 Beschreibung der Daten

### 2.1 Datenherkunft

Die Daten, die ich benutze, habe ich aus den Internetseiten der „Gesellschaft sozialwissenschaftlicher Infrastruktureinrichtungen“ (GESIS). Die offizielle Information zu diesen Daten lautet:

„Die in diesem Beitrag benutzten Daten entstammen der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS 2002 und ALLBUS 2004). Das ALLBUS-Programm ist 1980-86, 1991 von der DFG gefördert worden. Die weiteren Erhebungen wurden von Bund und Ländern über die GESIS finanziert. ALLBUS wird von ZUMA (Zentrum für Umfragen, Methoden und Analysen e.V., Mannheim) und Zentralarchiv Empirische Sozialforschung (Köln) in Zusammenarbeit mit dem ALLBUS-Ausschuss realisiert. Die Daten sind beim Zentralarchiv für Empirische Sozialforschung (Köln) erhältlich. Die vorgenannten Institutionen und Personen tragen keine Verantwortung für die Verwendung der Daten in diesem Beitrag.“<sup>1</sup>

### 2.2 Das Datenmaterial

Für die Konzeption und Durchführung der ALLBUS-Umfragen ist die ALLBUS-Abteilung des Zentrums für Umfragen, Methoden und Analysen (ZUMA; Mannheim) in Abstimmung mit dem ALLBUS-Ausschuss zuständig.<sup>2</sup> Dabei handelt es sich um zyklische Umfragen zu schon „traditionellen“ Themen sowie um bestimmte inhaltliche Gebiete.

Für meine Bachelor-Arbeit habe ich aus den ALLBUS-Dateien 2002 und 2004 nur bestimmte einzelne Variablen ausgewählt. Für das Jahr 2002 hatte ich eine Kompakt-Version, für das Jahr 2004 eine vollständige Version. Außerdem werden für die ALLBUS-Umfragen für jedes Umfragejahr andere Schwerpunkte der Befragung gesetzt. So kommt es, dass sich die beiden Jahre auch innerhalb dieser Analyse, was die Variablenzusammensetzung betrifft, ein wenig unterscheiden.

---

<sup>1</sup> Gesellschaft sozialwissenschaftlicher Infrastruktureinrichtungen 2004, S.3

<sup>2</sup> GESIS



Folglich werden in dieser Arbeit zwei separate Dateien für das Jahr 2002 und für das Jahr 2004 benutzt. Die Tabelle für das Jahr 2002 umfasst insgesamt 1619 Personen, darunter 1432 hauptberuflich Erwerbstätige (ganztags und halbtags) und 187 zurzeit der Befragung Arbeitslose. Die Tabelle für das Jahr 2004 umfasst 1632 Personen, allerdings 1410 hauptberuflich Erwerbstätige (ganztags und halbtags) und 222 zurzeit der Befragung arbeitslose Personen. Die befragten Personen sind zwischen 18 und 67 Jahre alt, als Obergrenze wird also das genaue gesetzliche arbeitsfähige Alter gesetzt.

## 2.3 Die Stichprobenziehung

Für die ALLBUS-Jahre 2002 und 2004 wurde zur Stichprobenziehung ein personenbezogenes zweistufiges Ziehungsverfahren eingesetzt. In der ersten Auswahlstufe wurde eine Stichprobe von Gemeinden gezogen. In der zweiten Stufe wurden Personenadressen aus den Einwohnermelderegistern der betreffenden Gemeinden zufällig ausgewählt. Dieses Verfahren verspricht gegenüber der Haushaltsstichprobe des ADM-Verfahrens<sup>3</sup> einige Vorteile, sowohl was den Stichprobenplan und die Feldarbeit als auch was die Genauigkeit einiger Ergebnisse betrifft. Die Stichproben wurde für das Jahr 2002 in dem Zeitraum von Februar 2002 bis August 2002, für das Jahr 2004 von März 2004 bis Julie 2004 gezogen.

## 2.4 Bildung der abhängigen Variable

Da es in der Arbeit beabsichtigt ist, Einflüsse auf die Arbeitslosigkeit zu untersuchen, sind Variablen wie Berufstätigkeit und Status der Nichtberufstätigkeit von besonderem Interesse. Aus der Variable Berufstätigkeit wird die hauptberufliche Erwerbstätigkeit und aus der Variable Nichterwerbstätigkeit die Arbeitslosigkeit extrahiert. Daraus folgt, dass die abhängige Variable die Ausprägung null ( $Y=0$ ) hat, wenn die befragte Person halbtags oder ganztags hauptberuflich erwerbstätig ist, und die Ausprägung eins ( $Y=1$ ), wenn die befragte Person zurzeit der Befragung arbeitslos war. Infolgedessen sind Schüler und

---

<sup>3</sup> Arbeitskreis Deutscher Marktforschungsinstitute

Studenten, Rentner und Pensionäre, Hausfrauen und Hausmänner, Zivil- und Wehrdienstleistende sowie aus anderen Gründen nicht (hauptberuflich) erwerbstätige nicht in der zu untersuchenden Stichprobe enthalten.

## 2.5 Wahl der unabhängigen Variablen

In die Analyse werden neben dem Status der Arbeitslosigkeit noch andere Variablen berücksichtigt. Diese Variablen, so wird angenommen, beeinflussen die Variable Arbeitslosigkeit. So werden sie als Einflussgrößen oder exogene Variablen bezeichnet. Dazu gehören folgende Variablen:

- Geschlecht (männlich=1, weiblich=2),
- Alter (metrisch),
- Schulabschluss (kein Abschluss/ Volks-, Hauptschulabschluss/ mittlere Reife, Realschule/ Fachhochschule/ Abitur, Hochschulreife/ anderer Abschluss)
- Familienstand (verheiratet zusammenlebend/ verheiratet getrennt lebend/ verwitwet/ geschieden/ ledig)
- Konfession des Befragten (evangelisch ohne freikirch./ evangelisch freikirch./ römisch-katholisch/ andere christliche Religionen/ andere nicht christliche/ keiner Religion zugehörig)
- Erhebungsgebiet Ost-West
- Berufliche Stellung, Einordnung nach Terwey
- Gesundheitszustand des Befragten (Selbsteinstufung)
- Rauchen Sie? (ja / nein)

Die ALLBUS-Dateien haben außer den üblichen Fragen jedes Mal andere Schwerpunkte, so kommt es, dass sich auch in dieser Analyse für jedes Jahr die Zusammensetzung der Variablen unwesentlich ändert. Zum Beispiel die Variable: „Wie zufrieden sind Sie mit Ihrem Beruf“ wurde nur in einem der beiden Jahre erfasst, und trotz ihres grossen Einflusses wurde sie nicht in die Analyse hineingenommen, da sie häufig nicht beantwortet wurde.

Die Variable Alter wurden entsprechend der höchstmöglichen Signifikanz gruppiert, wobei nur Personen bis zu dem gesetzlich vorhergesehenen Erwerbsalter betrachtet werden. Trotzdem ist diese Variable nur für das Jahr 2002 signifikant. Die Variable

Schulbildung wurde aus den Variablen Allgemeine Schulbildung, Fachhochschulabschluss und Hochschulabschluss gebildet. Die Variable Familienstand unterscheidet sich marginal in ihrer Klassifizierung für jedes Jahr. Bei gleicher Klassifizierung wären die beiden Jahre zwar einwandfreier zu vergleichen, da aber diese Variable im Jahr 2004 nur schwach signifikant war, wurde für sich eine kleine Abweichung bei der Klassenbildung entschieden. In vielen Studien wurde bereits belegt, dass andauernde Arbeitslosigkeit bei einem der Eheleute das familiäre System labilisiert, dass sie entsprechend einen negativen Einfluss auf die Konfliktlösung hat<sup>4</sup>. Andererseits ist die Familie auch eine unersetzliche Stütze um bei der Arbeit erfolgreich zu sein, was wiederum bei geschiedenen Leuten schneller zu einem Arbeitsverlust führen kann. Die genaue Auswahl der Variablen und deren Kategorien für beide Jahre sind in folgenden Tabellen 2.1 und 2.2 dargestellt.

Variable	Kategorie	Label
Geschlechtszugehörigkeit	1	Mann
	2	<b>Frau</b>
Erhebungsgebiet	1	Alte Bundesländer
	2	<b>Neue Bundesländer</b>
Alter	1	18 bis 44 Jahre
	2	<b>45 bis 67 Jahre</b>
Berufliche Stellung	1	Arbeiter
	2	Angestellter
	3	<b>Sonstige Stellung</b>
Schulbildung	1	Ohne/anderer Schulabschluss
	2	Volks-, Hauptschulabschluss
	3	Mittlere Reife, Realschulabschluss
	4	Fachhochschul-, Hochschulreife
	5	<b>Fachhochschul-,Hochschulabschluss</b>
Familienstand	1	<b>Verheiratet</b>
	2	Verh.getr.leb./ verwitwet
	3	Geschieden
	4	Ledig

Tabelle 2.1: Erklärende Variablen für das Jahr 2002

Die fett gedruckten Wörter in beiden Tabellen sind die Referenzkategorien oder auch Bezugskategorien, die später bei der Analyse mit Logit-Modellen eine große Rolle spielen werden.

<sup>4</sup> Steffen Dauer (1999)

Die Variable berufliche Stellung ist aus der Variable berufliche Stellung, Einordnung nach N. Terwey in ALLBUS entstanden. Diese Variable enthielt bei der ALLBUS-Datei folgende Ausprägungen: Selbstständiger Landwirt, Akademischer freier Beruf (z. B. Arzt mit eigener Praxis, Rechtsanwalt), Selbstständiger in Gewerbe, Handel, Industrie, Dienstleistung u. a., Beamter, Richter, Berufssoldat, Angestellter, Arbeiter, in Ausbildung, mithelfende Familienangehörige, Genossenschaftsbauer und nicht bestimmbar. Die meisten Einträge fielen auf die Klassen Arbeiter und Angestellter, die anderen Klassen waren entweder sehr schwach oder relativ schwach vertreten, somit wurden die Modellbedingungen verletzt, dass in eine Zelle des Endmodells mindestens fünf Beobachtungen hineinfallen sollen. Somit war es auch nicht überraschend, dass nur die beiden genannten Klassen signifikante Testergebnisse vorwiesen. Aus diesem Grunde wurde aus der Variable berufliche Stellung, Einordnung nach N. Terwey die Variable berufliche Stellung mit nur drei Ausprägungen gebildet. Unter sonstige Stellung kommen dabei alle anderen Ausprägungen zusammen.

Variable	Kategorie	Label
Erhebungsgebiet	1	Alte Bundesländer
	2	<b>Neue Bundesländer</b>
Konfessionszugehörigkeit	0	Römisch-katholisch
	1	<b>Keine / andere</b>
Rauchen	0	Ja
	1	<b>Nein</b>
Berufliche Stellung	1	Arbeiter
	2	Angestellter
	3	<b>Sonstige Stellung</b>
Schulabschluss	1	Ohne/anderer Schulabschluss
	2	Volks-, Hauptschulabschluss
	3	Mittlere Reife
	4	Fachhochschul-, Hochschulreife
	5	<b>Fachhochschul-,Hochschulabschluss</b>
Gesundheitszustand	1	Sehr gut
	2	Gut
	3	Zufrieden stellend
	4	Weniger gut
	5	<b>Schlecht</b>
Familienstand	1	<b>Verheiratet</b>
	2	Verh. getr. leb./ verwitwet / ledig
	3	Geschieden

Tabelle 2.2: Erklärende Variablen für das Jahr 2004

Die Variable Gesundheitszustand stellt die Selbsteinstufung des Befragten dar. Bei dieser Variable stellt sich die Frage nach der Ursache und Wirkung. Bei Untersuchungen der Zusammenhänge zwischen der Arbeitslosigkeit und Gesundheit ist man in der Forschung schon längst zu den folgenden zwei grundlegenden Thesen gekommen: „Arbeitslosigkeit führt zu einem erhöhten Krankheitsrisiko (Kausalitätshypothese)“, und „Krankheit führt zu einem erhöhten Arbeitslosigkeitsrisiko (Selektionshypothese)“<sup>5</sup>. Es ist durch aktuelle Studien sowohl bewiesen worden, dass nach ihrer persönlichen Einschätzung arbeitslose Personen über einen deutlich schlechteren allgemeinen Gesundheitszustand verfügen, als auch, dass bei einer freien Marktwirtschaft bei einem gleichzeitig bestehenden Überangebot von Arbeitskräften kranke Arbeitnehmer häufig weniger wettbewerbsfähig sind, demnach wieder entlassen und seltener wieder eingestellt werden<sup>6</sup>. Dieses zeigt, dass die Variable Gesundheitszustand eine ernst zu nehmende Einflussvariable darstellt.

Für die Arbeit soll SPSS for Windows Release 10.0 verwendet werden.

---

<sup>5</sup> Gesundheitsberichterstattung des Bundes S. 5

<sup>6</sup> Gesundheitsberichterstattung des Bundes

## 3 Voranalyse

### 3.1 Theoretischer Hintergrund

Bevor man eine komplexere Untersuchung mittels Logit-Modellen durchführt, sollte man eine bivariate Untersuchung zwischen den Einflussvariablen und der abhängigen Variable vornehmen. Eine bivariate Voranalyse liefert einen ersten Eindruck über die Daten und hilft bei der weiteren Untersuchung. Alle Variablen, die in dieser Arbeit betrachtet werden, haben nominales oder ordinales Skalenniveau. Aus diesem Grund bietet sich ein Balkendiagramm für den besseren visuellen Einblick sowie in Zusammenhang mit einer Kontingenztafel der Chi-Quadrat-Unabhängigkeitstest nach Pearson.

Für ordinale und nominalskalierte Daten werden Balkendiagramme aufgestellt, dabei wird die Häufigkeit jeder Variablenausprägung der unabhängigen Variable X als separater Balken dargestellt. Die Unabhängigkeit wird hierbei mit Hilfe des Chi-Quadrat-Unabhängigkeitstests nach Pearson und der Somers`D durchgeführt. Somers`D ist eine Teststatistik, die den Zusammenhang zweier ordinalskalierten Variablen bewertet, wobei eine der Variablen als abhängig von der anderen betrachtet werden kann. In diesem Fall ist der Zustand, arbeitslos zu sein, die von den anderen soziodemografischen Variablen abhängige Größe. Der Chi-Quadrat-Unabhängigkeitstest kann in dieser Arbeit sowohl für ordinale als auch für dichotome Variablen eingesetzt werden. Die Nullhypothese „Die abhängige und die unabhängigen Zufallsvariablen sind stochastisch unabhängig“ soll nach Möglichkeit abgelehnt werden. Die Nullhypothese wird für zu große Werte abgelehnt. Für dichotome Variablen wird das Assoziationsmaß Phi untersucht. Das ist ein Maß für die Stärke des Zusammenhangs zwischen zwei nominalskalierten dichotomen Variablen.

### 3.2 Ergebnisse bivariater Analyse

Im Folgenden werden nur einige Abbildungen sowie einige repräsentative Testergebnisse zum besseren Verständnis des bivariaten Zusammenhangs gegeben. Die Balkendiagramme für die Variablen Schulausbildung und berufliche Stellung sind in den beiden Jahren äquivalent, so werden sie hier nur jeweils einmal aufgezeigt.

*Jahr 2002*

Bereits der erste Blick zeigt, dass zumindest im bivariaten Fall der vermutete Zusammenhang zwischen der abhängigen und den Einflussvariablen festzustellen ist. In der nächsten Abbildung sieht man, dass die Personen, die geschieden sind, öfter von der Arbeitslosigkeit betroffen sind als andere.

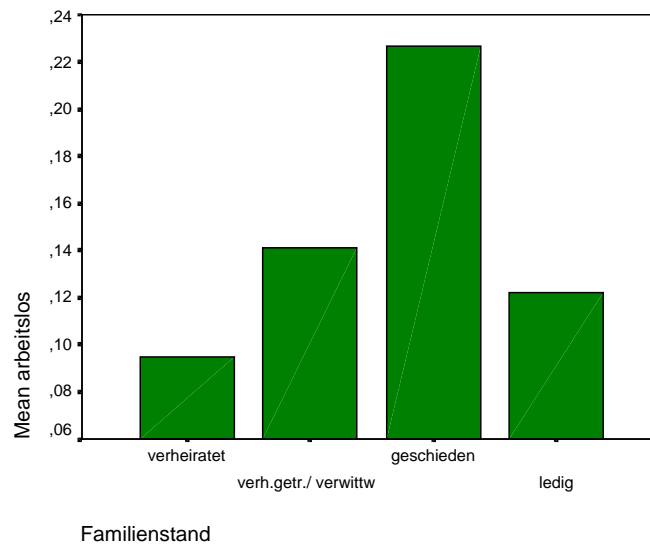


Abbildung 3.1: Arbeitslos und Familienstand

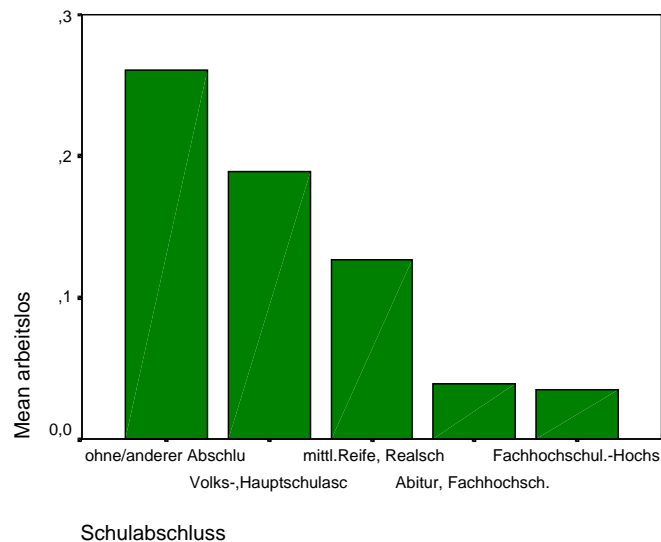


Abbildung 3.2: Arbeitslos und Schulausbildung

Auch die Variable Schulausbildung lässt einen deutlichen Zusammenhang vermuten.

Die Ergebnisse des Chi-Quadrat-Unabhängigkeitstests können der Tabelle 3.1 entnommen werden.

Wie man sieht, wird für die Variablen Erhebungsgebiet, berufliche Stellung, Schulausbildung und Familienstand die Nullhypothese auf Unabhängigkeit auf dem 5%-Niveau verworfen. Die Variablen Geschlechtszugehörigkeit und Alter sind allerdings nur schwach zum 10%-Niveau signifikant.

Einflussvariable	Chi-Quadrat	Freiheitsgrade	Sig. (asymptotisch)
Geschlechtszugehörigkeit	2,745	1	0,098
Erhebungsgebiet	85,675	1	0,000
Alter	3,360	1	0,067
Berufliche Stellung	86,393	2	0,000
Schulausbildung	63,311	4	0,000
Familienstand	21,902	3	0,012

Tabelle 3.1: Chi-Quadrat-Unabhängigkeitstest für das Jahr 2002

### *Jahr 2004*

Hinsichtlich des Gesundheitszustands ist zu erkennen, dass Personen, die sich gesundheitlich als sehr gut und gut einstufen, weniger von Arbeitslosigkeit betroffen sind, während Personen, die ihre Gesundheit als schlecht einstufen, öfter arbeitslos sind.

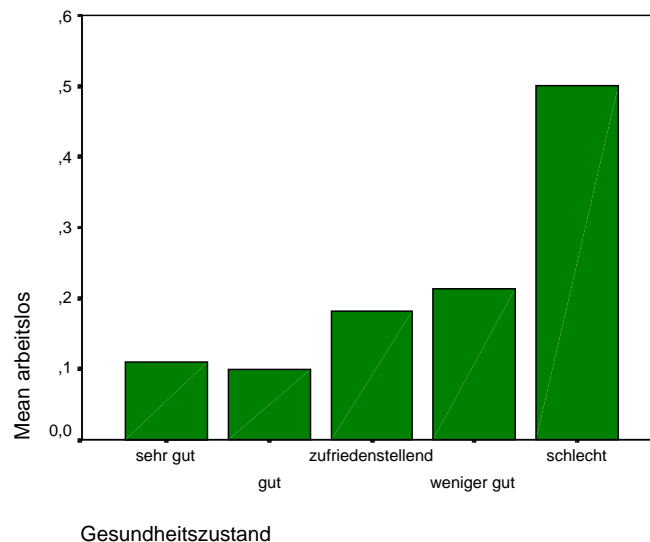


Abbildung 3.3: Arbeitslos und Gesundheitszustand



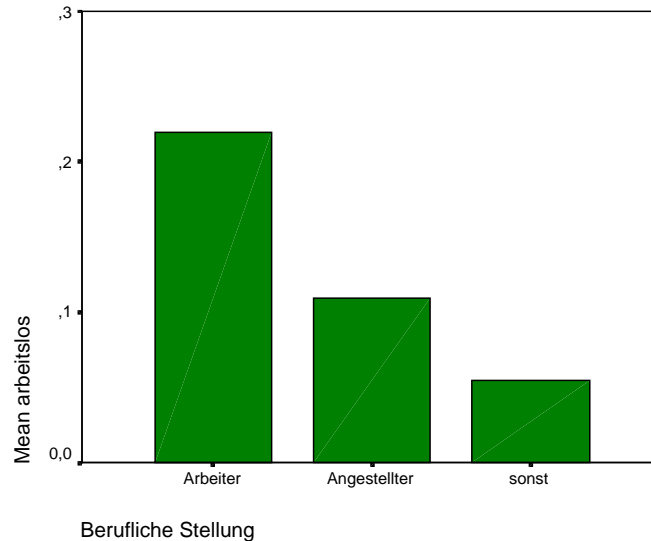


Abbildung 3.4: Arbeitslos und berufliche Stellung

Die Abbildung für die Berufliche Stellung zeigt, dass Arbeiter besonders häufig in der Stichprobe ohne Arbeit sind.

Im Folgenden wurde Balkendiagramme für die Variable Erhebungsgebiet erzeugt um den starken Zusammenhang anzudeuten. Dabei wurde noch nach Geschlecht getrennt, um mögliche Interaktionen leichter zu erkennen. Die Arbeitslosigkeit scheint in den neuen Bundesländern größer zu sein. Genauso ist zu sehen, dass in den alten Bundesländern weniger Frauen als Männer von der Arbeitslosigkeit betroffen sind, in den neuen Bundesländern aber umgekehrt, mehr Frauen als Männer arbeitslos sind.

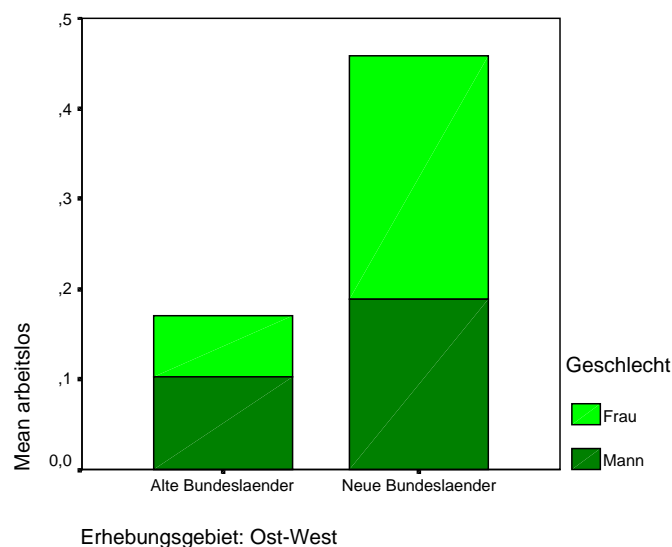


Abbildung 3.5: Arbeitslos und Erhebungsgebiet getrennt nach Geschlechtszugehörigkeit

Unerwartet deutlich zeigt sich der Zusammenhang der Variablen arbeitslos und Rauchverhalten in dem folgenden Fehlerbalkendiagramm. Anscheinend sind Personen, die rauchen, öfter von Arbeitslosigkeit betroffen als Personen, die nicht rauchen. Der Effekt der Geschlechtszugehörigkeit ist hier kaum zu erkennen.

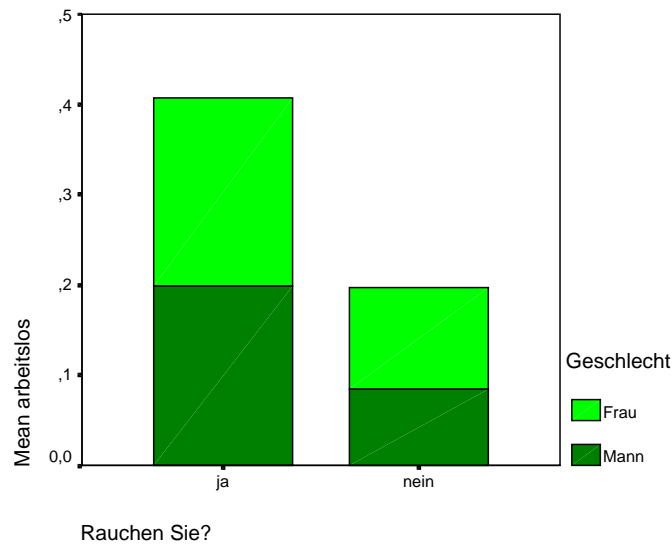


Abbildung 3.6: arbeitslos und Rauchen getrennt nach Geschlechtszugehörigkeit

Die Testergebnisse für die Unabhängigkeit der abhängigen Variable arbeitslos und die erklärenden Variablen für das Jahr 2004 können folgender Tabelle 3.2 entnommen werden.

Einflussvariable	Chi-Quadrat	Freiheitsgrade	Sig.(asymptotisch)
Erhebungsgebiet	59,880	1	0,000
Berufliche Stellung	58,142	2	0,000
Schulausbildung	67,083	4	0,000
Familienstand	18,379	2	0,000
Gesundheitszustand	60,446	4	0,000
Rauchen	36,030	1	0,000
Konfessionszugehörigkeit	30,543	1	0,000

Tabelle 3.2: Chi-Quadrat-Unabhängigkeitstest für das Jahr 2004

Hier wird für alle erklärenden Variablen die Nullhypothese auf Unabhängigkeit auf dem 5%-Niveau verworfen. Dieses bedeutet, dass alle Variablen in der bivariaten Analyse signifikanten Einfluss auf den Zustand arbeitslos zu sein ausüben. Für das Jahr 2004

könnte für die Variablen Geschlechtszugehörigkeit und Alter die Nullhypothese nicht verworfen werden. So wurden diese Variablen auch nicht mehr in die Analyse aufgenommen.

## 4 Logit-Modell

### 4.1 Theoretischer Hintergrund

Bei vielen praktischen Untersuchungen ist die interessierende, zu untersuchende Variable die nominalskalierte dichotome Zufallsvariable, sie hat also nur zwei mögliche Realisationen (Erfolg bzw. Misserfolg). Oft sind auch die beeinflussenden, abhängigen Variablen nominalskalierte dichotome oder mehrkategoriale Variablen, so kann keine lineare Regressionsanalyse zur Überprüfung von Zusammenhängen und Abhängigkeiten zwischen der zu erklärenden und den erklärenden Variablen durchgeführt werden. Eine andere Methode in der Statistik, die ein ähnliches Ziel hat ist die logistische Regression oder auch Logit-Modell genannt.

In dieser Arbeit sollen mögliche Zusammenhänge bzw. Abhängigkeiten von Variablen quantitativ nachgewiesen werden und mit Hilfe von Logit-Modellen untersucht werden. Dabei sollen die Stärke, die Form und die Richtung des Einflusses nach Möglichkeit ermittelt werden. Dies bedeutet, dass insbesondere das Ausmaß einer gemeinsamen Variation von Merkmalen bzw. der gemeinsamen Häufigkeitsverteilung der Variablen untersucht wird.

Ziel der statistischen Modellierung ist, möglichst viele Informationen aus den Daten zu gewinnen, so dass die Gesamtvariabilität der abhängigen Variablen so weit wie möglich durch das Modell bestimmt wird. Das heißt, dass die theoretischen Werte ( $\hat{\mu}_i$ ) möglichst nahe an den beobachteten Werten ( $y_i$ ) liegen sollen.

Die zu erklärende Variable hat die Ausprägung 1 bei Erfolg (wenn die betreffende Person arbeitslos ist) und 0 sonst. Dabei seien  $P(Y = 1) = \pi$  und  $P(Y = 0) = 1 - \pi$ . Die Erfolgswahrscheinlichkeiten  $\pi_k$  sind bedingte Wahrscheinlichkeiten, bedingt durch die jeweiligen Ausprägungen  $\chi_k^T = (\chi_{k0} \dots \chi_{kj} \dots \chi_{kJ})$  der erklärenden X-Variablen. Dabei sind  $k$  die Anzahl der Gruppen und  $j = 1, \dots, J$  die Anzahl der Variablen in dem Modell. Enthält das Modell eine Konstante, so sind  $\chi_{k0} = 1$  für alle  $k$ . Da die bedingten Wahrscheinlichkeiten  $\pi_k$  nur Werte im  $[0,1]$ -Intervall haben, die Linearkombination  $\chi_k^T \beta$  jedoch theoretisch Werte zwischen  $-\infty$  und  $+\infty$  annehmen kann, ist diese Linearkombination einer Transformation zu unterziehen. Die mathematische Form des Logit-Modells basiert auf der logistischen Verteilung, somit wird hier auch eine logistische Transformation benutzt.

Die Verteilungsfunktion der logistischen Verteilung ist gegeben mit:

$$\pi_k = F(\eta_k) = \frac{\exp(\eta_k)}{1 + \exp(\eta_k)}. \quad (1.1)$$

Nach Auflösung dieser Verteilungsfunktion nach  $(\eta_k)$  und nach einigen Umformungen resultiert:

$$\frac{\pi_k}{1 - \pi_k} = \exp(\eta_k). \quad (1.2)$$

Dabei wird das Verhältnis  $\pi_k / (1 - \pi_k)$  der Wahrscheinlichkeit eines Erfolges zur Wahrscheinlichkeit eines Misserfolges als odds (Chancen) eines Erfolges bezeichnet.

Logarithmieren von (1.2) führt zu:

$$\eta_k = \log\left(\frac{\pi_k}{1 - \pi_k}\right). \quad (1.3)$$

Diese logistische Transformation von  $\pi_k$  (ist die inverse Funktion für die logistische Verteilungsfunktion und) führt zu den logarithmierten odds eines Erfolges, die kurz als log odds oder logits bezeichnet werden.

Somit wurde erreicht, dass sich für jeden Wert  $\pi_k$  im Bereich (0,1) ein Wert von logit ( $\pi_k$ ) im Bereich  $(-\infty, +\infty)$  ergibt, zum Beispiel für  $\pi_k \rightarrow 0$  logit  $\pi_k \rightarrow -\infty$ , für  $\pi_k \rightarrow 1$  logit  $\pi_k \rightarrow \infty$ .

Die Beziehung zwischen der Erfolgswahrscheinlichkeit  $\pi_k$  und den erklärenden Variablen  $X_j$  ( $j = 1, \dots, J$ ) wird nunmehr in Argument der logistischen Verteilungsfunktion mittels

$$\eta_k = x_k^T \beta = \beta_0 \chi_{k0} + \beta_1 \chi_{k1} + \dots + \beta_J \chi_{kJ} \quad (1.4)$$

hergestellt, worin  $\beta = \beta(\beta_0, \beta_1, \dots, \beta_J)^T$  ein Vektor von unbekanntem zu schätzenden Parametern ist. Somit folgt:

$$\pi_k = F(\eta_k) = F(x_k^T \beta) = \frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} = \frac{\exp(\sum_j \chi_{kj} \beta_j)}{1 + \exp(\sum_j \chi_{kj} \beta_j)}. \quad (1.5)$$

Im Ergebnis der logistischen Transformation ergibt sich das lineare logistische Modell (Logit-Modell) für die Abhängigkeit der Erfolgswahrscheinlichkeit  $\pi_k$  von den Werten der erklärenden Variablen  $\chi_{k0}, \chi_{k1}, \dots, \chi_{kJ}$ :

$$\eta_k = g(\pi_k) = \log\left(\frac{\pi_k}{1 - \pi_k}\right) = x_k^T \beta = \sum_j \chi_{kj} \beta_j \quad (1.6)$$

Dabei ist  $g(\pi_k)$  die link (Verknüpfung) Funktion, die die Beziehung der abhängigen Variablen mit der linearen Komponente herstellt. Somit wird die Schätzung der  $k$  unbekanntem Wahrscheinlichkeiten  $\pi_k$  auf die Schätzung der geringeren Anzahl  $J$  von Parametern  $\beta_j$  verlagert. Nach deren Schätzung können auch die Wahrscheinlichkeiten  $\pi_k$  bestimmt werden.

Der Effekt einer erklärenden Variablen auf die geschätzten Wahrscheinlichkeiten eines Erfolges ist im Allgemeinen

$$\hat{\pi}_k = \frac{e^{b_0 + b_1 \chi_{k1} + \dots + b_J \chi_{kJ}}}{1 + e^{b_0 + b_1 \chi_{k1} + \dots + b_J \chi_{kJ}}} \quad (1.7)$$

wie folgt:

Ist der Parameter  $b_j$  positiv (negativ), dann erhöht sich (sinkt) die geschätzte link Funktion  $\hat{\eta}_k = \log[\hat{\pi}_k / (1 - \hat{\pi}_k)]$  (=log odds) bei Erhöhung der Variablen  $X_j$  um eine Einheit und Konstanz aller anderen X-Variablen um  $b_j$  und die odds  $\hat{\pi}_k / (1 - \hat{\pi}_k)$  erhöhen sich (sinken) multiplikativ um  $\exp(b_j)$ . Je höher (niedriger) jedoch  $\exp(\sum_j \chi_{kj} b_j)$  ist, desto höher (kleiner) ist die Wahrscheinlichkeit eines Erfolges.

Mit Odds ratio wird dabei das Verhältnis zweier odds bezeichnet. Dieses ist ein Maß der Assoziation, das direkt aus dem logistischen Modell geschätzt werden kann. Odds ratio gibt im Fall einer dichotomen Einflussvariable an, wie viel wahrscheinlicher es ist, dass das Ereignis eintritt ( $y=1$ ) unter den Fällen mit  $x=1$  im Gegensatz zu den Fällen mit  $x=0$ . Im Falle einer mehrkategorialen Variable ist es das Verhältnis zwischen zwei Kategorien.

Das Logit –Modell gehört zur Klasse der verallgemeinernden linearen Modelle.

## 4.2 Praktische Durchführung und Ergebnisse

In den folgenden Tabellen 4.1 und 4.2 sind die Ergebnisse der Koeffizienten des Logit-Modells zusammengefasst. Dabei wurden mit B die geschätzten Koeffizienten (logarithmierten odds ratios), mit S.E. die Standardfehler, mit Wald der Wald-Test für jeden Koeffizienten, mit Sig, deren Signifikanz und mit Exp(B) die odds ratio bezeichnet.

## Jahr 2002

Variable	B	S.E.	Wald	DF	Sig.	Exp(B)
Sex	,572	,187	9,475	1	,002	1,767
Erhebungsgebiet	1,697	,194	76,117	1	,000	5,442
Alter	,371	,198	3,519	1	,061	1,435
Berufliche Stellung			23,234	2	,000	
Arbeiter	2,089	,489	18,637	1	,000	8,089
Angestellte	1,460	,480	9,261	1	,002	4,279
Schulausbildung			39,409	4	,000	
1	2,049	,529	15,003	1	,000	7,635
2	1,734	,365	22,625	1	,000	5,635
3	,790	,344	5,285	1	,022	2,184
4	-,002	-,002	,000	1	,997	,994
Familienstand			20,027	3	,000	
Sonst	,205	,389	,276	1	,599	1,230
Geschieden	1,087	,258	17,813	1	,000	2,969
Ledig	,568	,219	6,697	1	,010	1,743
Konstante	-8,809	,762	133,595	1	,000	,000

Abbildung 4.1: Koeffizienten des Logit-Modells für das Jahr 2002

### *Die Effekte der exogenen Variablen für das Jahr 2002*

Arbeitslosigkeit ( $Y=1$ ) tritt in der Stichprobe rund fünfeinhalbmals häufiger bei Personen in den neuen Bundesländern ( $X=1$ ) auf im Vergleich zu Personen in den alten Bundesländern ( $X=0$ ). In diesem Jahr ist auch die Variable Geschlecht in den Kreis der signifikanten Variablen mit Signifikanz von  $p=0,002$  in die Analyse miteingeflossen worden. So tritt die Arbeitslosigkeit in der Stichprobe rund 1,8-mal häufiger bei Frauen ( $X=2$ ) auf im Vergleich zu Männern ( $X=1$ ) und die dichotomisierte Variable Alter, obwohl schwach signifikant ( $p=0,062$ ), besagt, dass ältere (45-67 Jahre alt) Menschen ( $X=2$ ) 1,4-mal höhere Chancen haben als jüngere (18-44 Jahre alt) Menschen ( $X=1$ ) arbeitslos zu werden.

Interpretation der mehrkategorialen Variablen:

Alle mehrkategorialen Variablen, die in diesem Modell geschätzt wurden, sind insgesamt laut Wald-Test zum vorgegebenen 5%-Niveau signifikant.

Zur Erstellung der Kontrast-Variablen wurde die Indikator-Kodierung benutzt. Diese Kodierung entspricht einer Kodierung in Scheinvariablen, auch Dummy-Kodierung

genannt. Die Kontrast-Variablen zeigen das Vorhandensein bzw. Nichtvorhandensein einer Kategorie an. Die Referenzkategorie nimmt den Wert null in allen Dummy-Variablen ein, die zutreffende Kategorie nimmt den Wert eins, alle anderen den Wert null ein.

Bei einer multiplen logistischen Regression, wie in diesem Fall, beinhalten die geschätzten Parameter  $b$  die (statistisch) bereinigten  $\log(\text{or})$   $\log$  odds ratio und  $\exp(b)$  die bereinigten odds ratio.

Odds ratio für die variable berufliche Stellung: Als Bezugskategorie wurde hier die letzte Kategorie „andere berufliche Stellung“ gewählt. In der Stichprobe tritt Arbeitslosigkeit achtmal häufiger bei Arbeitern ( $X=1$ ) und 4,3-mal häufiger bei Angestellten ( $X=2$ ) als bei Personen mit „anderer“ beruflicher Stellung ( $X=3$ ) auf.

Odds ratio für die Variable Familienstand: Variable Familienstand ist insgesamt stark signifikant, aber im Einzelnen nur die Klassen „geschieden“ ( $p=0,000$ ) und „ledig“ ( $p=0,002$ ). Die Klassen „verheiratet- getrennt lebend“ und „verwitwet“ sind getrennt oder auch zusammengefasst in einer Klasse nicht signifikant. So haben geschiedene Personen im Vergleich zu verheirateten Personen (hier Referenzkategorie) um drei mal und ledige Personen um 1,7-mal bessere Chancen, arbeitslos zu werden.

Odds ratio für die Variable Schulausbildung: Diese Variable ist hoch signifikant. Die Klasse „Fachhochschulabschluss oder Hochschulabschluss“ ist hier die Referenzkategorie. Die Arbeitslosigkeit tritt in der Stichprobe 7,7-mal häufiger bei Personen ohne einen Schulabschluss oder Personen, die einen anderen Schulabschluss haben, sowie 5,6-mal häufiger bei Personen mit einem Volks- oder Hauptschulabschluss, 2,2-mal häufiger bei Personen mit mittlerer Reife im Vergleich mit Personen mit Fachhochschul- oder Hochschulabschluss. Die Klasse „Abitur, Fachhochschul-,Hochschulreife“ ist hier nicht signifikant.

### *Modellanpassung für das Jahr 2002*

Für die Güte der Anpassung kann man den Wert der Differenz der Devianzen des Modells nur mit der Konstante ( $M_0$ ) und dem spezifizierten Modell mit allen exogenen Variablen ( $M_1$ ) anschauen. Der Wert beträgt 231,972 und ist für das vorgegebene Signifikanzniveau von 5 % hoch signifikant ( $p=0,000$ ). Somit wird die Hypothese, dass die exogenen Variablen insgesamt keinen Einfluss auf die endogene (abhängige) Variable haben, abgelehnt. Die gleiche Entscheidung trifft man, wenn man den Wert der Deviance mit dem Quantil der Chi-Quadrat-Verteilung vergleicht.



Das Nagelkerke mit dem Wert 0,261 sagt jedoch, dass nur 26,1 % der Variation durch dieses logistische Modell erklärt werden. Dieses bedeutet, dass es noch mehr Einflüsse auf die Tatsache, arbeitslos zu sein, gibt, die aber in der ALLBUS Studie nicht aufgenommen wurden oder auch nicht beobachtbar sind.

Im Ergebnis werden unter Verwendung des spezifizierten Modells 99,4 % der nicht arbeitslosen und 10,7 % der Arbeitslosen korrekt vorhergesagt, was insgesamt 89,1 % korrekt vorhergesagte Fälle ergibt.

## Jahr 2004

Variable	B	S.E.	Wald	DF	Sig.	Exp(B)
Konfession	,776	,246	9,918	1	,002	2,172
Erhebungsgebiet	1,153	,186	38,570	1	,000	3,166
Rauchen	-,560	,165	11,562	1	,001	,571
Berufliche Stellung			10,233	2	,006	
Arbeiter	,881	,279	9,962	1	,002	2,414
Angestellte	,798	,282	8,009	1	,005	2,222
Schulausbildung			42,818	4	,000	
1	3,306	,644	26,379	1	,000	27,278
2	2,479	,494	25,146	1	,000	11,926
3	1,707	,477	12,824	1	,000	5,512
4	1,384	,555	6,224	1	,013	3,991
Familienstand			6,402	2	,041	
Sonst	,253	,180	1,990	1	,158	1,288
Geschieden	,585	,242	5,847	1	,016	1,796
Gesundheitszustand			30,052	4	,000	
Sehr gut	-1,731	,451	14,754	1	,000	,177
Gut	-1,946	,419	21,534	1	,000	,143
Zufrieden stellend	-1,407	,425	10,956	1	,001	,245
Weniger gut	-1,042	,482	4,675	1	,031	,353
Konstante	-5,128	,847	36,691	1	,000	,006

Abbildung 4.2: Koeffizienten des Logit-Modells für das Jahr 2004

### *Die Effekte der exogenen Variablen für das Jahr 2004*

In der Tabelle 4.2 findet man die geschätzten Parameter des spezifizierten Modells.

Arbeitslosigkeit ( $Y=1$ ) tritt in der Stichprobe rund dreimal häufiger bei Personen in den neuen Bundesländern ( $X=1$ ) auf im Vergleich zu Personen in den alten Bundesländern ( $X=0$ ), zweimal häufiger bei Personen, die keiner Konfession oder einer anderen als der

römisch-katholischen ( $X=1$ ) angehören, im Vergleich zu Personen, die der römisch-katholischen Kirche ( $X=0$ ) angehören. Umgekehrt sinkt die Arbeitslosigkeit um die Hälfte bei Personen, die nicht rauchen ( $X=1$ ) im Vergleich zu Personen, die rauchen ( $X=0$ ).

Interpretation der mehrkategorialen Variablen:

Alle mehrkategorialen Variablen, die in diesem Modell geschätzt wurden, sind insgesamt laut des Wald-Tests zum vorgegebenen 5%-Niveau signifikant.

Zur Erstellung der Kontrast-Variablen wurde auch in diesem Jahr die Indikator-Kodierung benutzt.

Odds ratio für die Variable berufliche Stellung: Als Bezugskategorie wurde hier die letzte Kategorie „andere berufliche Stellung“ gewählt. In der Stichprobe tritt Arbeitslosigkeit 2,4-mal häufiger bei Arbeitern ( $X=1$ ) und 2,2-mal häufiger bei Angestellten ( $X=2$ ) als bei Personen mit „anderer beruflicher Stellung“ ( $X=3$ ) auf.

Odds ratio für die Variable Gesundheitszustand: Bezugskategorie ist hier die letzte Kategorie „schlecht“. Die Chancen arbeitslos zu werden betragen bei Personen, die sich gesundheitlich als „sehr gut“ einstufen fast zwei Zehntel, bei Personen, die sich gesundheitlich als „gut“ einstufen ein Siebtel, bei Personen, die sich gesundheitlich als „zufrieden stellend“ einstufen ein Viertel, und bei Personen, die sich gesundheitlich als „weniger gut“ einstufen ein Drittel, im Vergleich zu Personen, die sich gesundheitlich als „schlecht“ einstufen. Anders gesagt, die Chancen, nicht arbeitslos zu sein, steigen mit besserer Selbsteinstufung seines eigenen Gesundheitszustandes.

Odds ratio für die Variable Familienstand: Die Variable Familienstand ist insgesamt nur schwach signifikant ( $p=0,041$ ), so dass nur die Klasse „geschieden“ (mit  $p=0,016$ ) besagt, dass geschiedene Personen im Vergleich zu verheirateten Personen (hier Referenzkategorie) ein um 1,8-mal höheres Risiko haben, arbeitslos zu werden.

Odds ratio für die Variable Schulausbildung: Diese Variable ist hoch signifikant. Die Klasse „Fachhochschulabschluss oder Hochschulabschluss“ ist hier die Referenz-Kategorie. Die Arbeitslosigkeit tritt in der Stichprobe 27-mal häufiger bei Personen ohne einen Schulabschluss oder Personen, die einen anderen Schulabschluss haben auf, 12-mal häufiger bei Personen mit einem Volks- oder Hauptschulabschluss, fünfeinhalbmal häufiger bei Personen mit mittlerer Reife und viermal häufiger bei Personen mit Fachhochschul- oder Hochschulreife im Vergleich mit Personen mit Fachhochschul- oder Hochschulabschluss.

## *Modellanpassung für das Jahr 2004*

Der Wert der Differenz der Deviance der Modelle M0 (Modell ohne erklärende Variablen) und M1 (das spezifizierte Modell) von 239,029 wird mit dem Quantil der Chi-Quadrat-Verteilung verglichen. Für das vorgegebene Signifikanzniveau von 5 % ist dieser Wert signifikant ( $p=0,000$ ). Also kann die  $H_0$  auf Gleichheit der beiden Modelle abgelehnt werden.

Das Nagelkerke mit 0,248 gibt jedoch an, dass nur 25 % der Variation durch dieses Modell erklärt werden. Auch für dieses Jahr gilt, dass für diese Analyse nicht alle notwendigen Variablen erfasst worden sind oder sich erfassen lassen. Infolgedessen ist festzustellen, wie schon zu vermuten war, dass es sehr viele auch nicht erfassbare Ursachen dafür gibt, ob eine Person arbeitslos wird oder nicht.

Mit diesem geschätzten Modell werden 98,4 % der nicht Arbeitslosen und 15,3 % der Arbeitslosen, insgesamt also 87,1 % richtig vorhergesagt.

## *Wie könnte das Problem der Nichtsignifikanz der erklärenden Variablen Alter und Geschlechtszugehörigkeit angegangen werden?*

Die Nichtsignifikanz der erklärenden Variablen Alter und Geschlechtszugehörigkeit löst große Verwunderung aus. Meistens erklären diese beiden Variablen gut die Arbeitslosigkeit. Es könnte sein, dass die Stichproben nicht ausreichend gross für die gewählte Variablenzusammensetzung sind. Andererseits ist die Stichprobe des Jahres 2002 noch kleiner, und trotzdem liefert dieses Jahr vernünftige Ergebnisse auch für die oben genannten Variablen. Eine zweite Erklärung wäre, dass in diesem Jahr irgendwelche Interaktionen eine Rolle spielen.

Dieser Teil der Arbeit sollte als ein kleiner Exkurs verstanden werden.

Des Weiteren soll versucht werden durch Trennung des Datensatzes mögliche Interaktionseffekte anzudeuten. Alle Ergebnisse sind im Anhang (A.3-A.) entweder in tabellarischer Form zusammengefasst oder als SPSS-Output zu finden.

So wird der Datensatz erst nach Geschlechtszugehörigkeit getrennt untersucht. Bei der Analyse der Männer sind alle Variablen außer Alter hoch signifikant zum 5%-Niveau.

Auch alle folgenden Tests werden mit dem 5%-Signifikanzniveau verglichen. Im Gegensatz sind in dem gleichen Modell für Frauen nur die Variablen Erhebungsgebiet und Schulausbildung signifikant. Auch hier ist die Variable Alter nicht signifikant.

Bei Trennung des Datensatzes nach Alter in zwei Gruppen sind in dem Modell der jüngeren (18-44) Personen die Variablen Geschlechtszugehörigkeit und Familienstand nicht signifikant. Im Modell mit den älteren Personen sind die Variablen Geschlechtszugehörigkeit und Rauchen nicht signifikant. Die Variable berufliche Stellung ist nur zum 10%-Niveau ( $p=0.068$ ) signifikant.

Bei der Trennung nach Erhebungsgebiet sind bei der Analyse der alten Bundesländer die Variablen Geschlechtszugehörigkeit, Alter und Familienstand nicht signifikant. Die Variable Rauchen nur zum 10%-Niveau ( $p=0,083$ ) signifikant, aber die Variable Konfessionszugehörigkeit ist hier hoch signifikant mit einem odds ratio Wert, der besagt, dass die Personen, die der römisch-katholischen Kirche angehören 2.6-mal höhere Chancen haben, nicht arbeitslos zu sein, als Personen, die einer anderen oder keiner Konfession angehören. In den neuen Bundesländern sind dagegen die Variablen Geschlechtszugehörigkeit ( $p=0,063$ ), Familienstand ( $p=0,056$ ) und Gesundheitszustand ( $p=0,075$ ) zum 10%-Niveau signifikant und die Variablen Alter und Konfessionszugehörigkeit nicht signifikant.

Bei der Trennung der Datei nach Geschlecht und gleichzeitig in zwei Altersgruppen sind in der Tabelle der jüngeren Männer (18-44) alle Variablen außer Familienstand und Konfessionszugehörigkeit hoch signifikant. Bei den älteren Männer (45-67) sind die Variablen Erhebungsgebiet ( $p=0,064$ ) und berufliche Stellung ( $0,054$ ) schwach signifikant, alle anderen hoch signifikant. Bei jüngeren Frauen (18-44) sind nur die Variablen Erhebungsgebiet, Schulausbildung und Konfessionszugehörigkeit hoch signifikant und die Variable Rauchen ( $p=0,077$ ) ist zum 10%-Niveau signifikant. Bei älteren Frauen (45-67) ist dagegen nur die Variable Erhebungsgebiet hoch signifikant mit einem hohen Wert, der besagt, dass ältere Frauen in den neuen Bundesländern ein 7,4-mal höheres Risiko haben, arbeitslos zu werden, als Frauen des gleichen Alters in den alten Bundesländern. Es ist möglicherweise auch dadurch zu erklären, dass Frauen in den alten Bundesländern nach dem Großziehen ihrer Kinder sich nach erfolgloser Arbeitssuche, statt sich immer weiter um eine Arbeitsstelle zu bemühen und also in der Arbeitslosenstatistik zu bleiben, eher der Hausfrauentätigkeit widmen.

Die Ergebnisse der getrennten Tabellen, insbesondere der letzten, der doppelten Trennung, sind mit großer Vorsicht zu genießen. Diese Tabellen beinhalten aufgrund der

Teilung weniger Beobachtungen, insbesondere dadurch sind die Zahlen der Arbeitslosen niedriger (Siehe Tabelle A.3 im Anhang). Die Zahl der Einflussvariablen ist vergleichbarkeitshalber gleich geblieben, so kommt es, dass insgesamt mehr schwach besetzte oder leere Zellen anzutreffen sind, wodurch auch die Testergebnisse verzerrt werden können.

Im Ergebnis lässt sich sagen, dass möglicherweise durch ständigen Zuwachs der Arbeitslosigkeit die Analyse dessen schwieriger geworden ist, und um bessere Ergebnisse zu erzielen, sind größere Stichproben erforderlich. Insbesondere sieht man durch die Trennung, dass Frauen für die Analyse der Arbeitslosigkeit eine besondere Herausforderung darstellen. Bei Frauen ist das Erhebungsgebiet die einflussreichste gefundene Variable, bei jungen Frauen spielen noch Schulausbildung und Konfessionszugehörigkeit sowie das Rauchen eine Rolle. Bei Männern zeigte sich die Konfessionszugehörigkeit sowohl in jüngeren als auch in älteren Jahren als nicht wichtig. Genauso scheinen für ältere Männer das Erhebungsgebiet und die berufliche Stellung nicht mehr ausschlaggebend zu sein. Die Geschlechtszugehörigkeit ist nur noch bei Personen in den neuen Bundesländern für die Erklärung der Arbeitslosigkeit wichtig.

## 5 Zusammenfassender Vergleich der beiden Jahre

Ziel dieser Arbeit war unter anderem, herauszufinden, ob die Analyse derselben Fragestellung anhand zweier unterschiedlicher Jahre vergleichbare Ergebnisse liefert. In diesem Fall wurde die Fragestellung nach möglichen Einflussvariablen und Zusammenhängen auf den Zustand, arbeitslos zu werden, untersucht. Für diese Aufgabe wurden zwei aktuelle ALLBUS-Datensätze der Jahre 2002 und 2004 gewählt.

Auch wenn die beiden Datensätze aus einer Quelle stammen, gab es doch bei der Zusammensetzung der Variablen erhebliche Unterschiede. Einerseits sollte eine gute Vergleichbarkeit erreicht werden, indem gleichen Variablen auch eine gleiche Klasseneinteilung gegeben wurde, außer der Variable Familienstand, andererseits sollte jedes Jahr an sich auch möglichst viel Variabilität erklären können. So kam es, dass bei der Variable Familienstand marginal Unterschiede bei der Klassenunterteilung vorgenommen wurden, weil sie sonst im Jahr 2004 wenig signifikant gewesen wäre.

Die endogene Variable wurde in beiden Jahren aus den Variablen Befragte(r) berufstätig und Status der Nichterwerbstätigkeit gezogen, wobei Personen als nicht arbeitslos eingestuft wurden, die hauptberuflich ganztags oder halbtags erwerbstätig waren. Als arbeitslose galten Personen, die nicht erwerbstätig und zurzeit arbeitslos waren.

Die vorliegenden Untersuchungen, sowohl die binären als auch die multivariaten, zeigen, dass die Ergebnisse in der Tendenz in die gleiche Richtung weisen. Die Variablen Erhebungsgebiet, berufliche Stellung, Schulausbildung und Familienstand haben in beiden Jahren signifikante Effekte auf die Variable arbeitslos. Bei der Variable Schulausbildung zeigen sich in beiden Modellen Effekte in die gleiche Richtung, die jedoch in ihrer Höhe unterschiedlich sind. Außerdem ist im Jahr 2002 die Klasse Fachhochschul-, Hochschulreife nicht signifikant. Die Variable Familienstand ist im Jahr 2002 hoch signifikant ( $p=0,000$ ), im Jahr 2004 weniger deutlich ( $p=0,041$ ), jedoch auch signifikant zum 5%-Niveau. In beiden Jahren ist der Effekt der Klasse geschieden als bedeutsam hervorzuheben.

Die Variable berufliche Stellung hat in beiden Modellen hoch signifikante Effekte, die in eine Richtung weisen und eine sehr ähnliche Relation zeigen, die jedoch unterschiedlich hoch sind.

Unterschiede zeigen sich dadurch, dass die beiden Modelle, um eine gewisse Erklärbarkeit zu erreichen, doch teilweise unterschiedliche Variablen beinhalten. So

wurden für das Jahr 2004 die Variablen Geschlechtszugehörigkeit und Alter nicht hineingenommen, weil sie hier nicht signifikant waren. Andererseits wurden zwei neue hoch signifikante Variablen, Rauchverhalten und Gesundheitszustand, in dieses Jahr hineingenommen, die leider in ALLBUS 2002 nicht erfasst wurden.

Grundsätzlich lässt sich sagen, dass das Modell der beiden Jahre gleiche hoch signifikante Variablen wie Erhebungsgebiet, berufliche Stellung und Schulausbildung aufweist, die die Variable arbeitslos sehr deutlich beeinflussen und erklären. Meiner Einschätzung nach wären diese Variablen auch allgemein im globalen Sinne als erklärende Variablen für die Arbeitslosigkeit für andere Jahre geeignet. Sicherlich findet man für jedes Jahr noch andere signifikante Variablen, aber, wie schon oben erläutert wurde, die beiden Dateien stimmten in ihrer Variablenzusammensetzung nicht überein. Außerdem umfassen die Datensätze nicht sonderlich viele Beobachtungen, so dass nicht zu viele Variablen in das Modell aufgenommen werden können, so dass keine Modellvoraussetzungen verletzt werden.

## 6. Anhang

Variable	Arbeitslos				Total	
	Nein		Ja		Anzahl Anteil	
	Anzahl	Anteil	Anzahl	Anteil		
Geschlechtszugehörigkeit						
Mann	834	89,6	97	10,4	931	57,5
Frau	598	86,9	90	13,1	688	42,5
Erhebungsgebiet						
Alte Bundesländer	989	93,8	65	6,2	1054	65,1
Neue Bundesländer	443	78,4	122	21,6	565	34,9
Alter						
18 bis 44 Jahre	903	98,6	105	10,4	1008	62,3
45 bis 67 Jahre	529	86,6	82	13,4	611	37,7
Berufliche Stellung						
Arbeiter	393	78,1	110	21,9	503	31,1
Angestellter	753	91,3	72	8,7	825	51,0
Sonstige Stellung	286	98,3	5	1,7	291	18,0
Schulbildung						
Ohne/anderer Schulabschluss	32	74,4	11	25,6	43	2,7
Volks-, Hauptschulabschluss	343	81,5	78	18,5	421	26,0
Mittlere Reife	526	87,1	78	12,9	604	37,3
Fachhochschul-, Hochschulreife	198	96,1	8	3,9	203	12,5
Fachhochschul-, Hoch.-Abschluss	336	96,6	12	3,4	348	21,5
Familienstand						
Verheiratet	847	90,6	88	9,4	935	57,8
Verh. getr. leb./ verwitwet	61	85,9	10	14,1	71	4,4
Geschieden	109	77,3	32	22,7	141	8,7
Ledig	415	87,9	57	12,1	472	29,2
Gesamt (jeweils)	1432	88,4	187	11,6	1619	100

Tabelle A.1: Stichprobenverteilung: Einflussgrößen und Variable arbeitslos für das Jahr 2002



Variable	Arbeitslos				Gesamt	
	Nein		Ja		Anzahl	Anteil
	Anzahl	Anteil	Anzahl	Anteil		
<b>Erhebungsgebiet</b>						
Alte Bundesländer	972	91,2	94	8,8	1066	65,3
Neue Bundesländer	438	77,4	128	22,6	566	34,7
<b>Konfessionszugehörigkeit</b>						
Roemisch-katholisch	415	94,1	26	5,9	441	27,0
Keine / andere	995	83,5	196	16,5	1191	73,0
<b>Rauchen</b>						
Ja	480	97,7	122	20,3	602	36,9
Nein	930	90,3	100	9,7	1030	63,1
<b>Berufliche Stellung</b>						
Arbeiter	434	77,9	123	22,1	557	34,1
Angestellter	655	89,0	81	11,0	736	45,1
Sonstige Stellung	321	94,7	18	5,3	339	20,8
<b>Schulausbildung</b>						
Ohne/anderer Schulabschluss	20	66,7	10	33,3	30	1,8
Volks-, Hauptschulabschluss	357	79,7	91	20,3	448	27,5
Mittlere Reife	601	85,2	104	14,8	705	43,2
Fachhochschul-, Hochschulreife	161	93,1	12	6,9	173	10,6
Fachhochschul -,Hoch. Abschluss	271	98,2	5	1,8	276	16,9
<b>Gesundheitszustand</b>						
Sehr gut	259	89,0	32	11,0	291	17,8
Gut	745	90,1	82	9,9	827	50,7
Zufrieden stellend	311	81,6	70	18,4	381	23,3
Weniger gut	79	78,2	22	21,8	101	6,2
Schlecht	16	50,0	16	50,0	32	2,0
<b>Geschlechtszugehörigkeit</b>						
Mann	812	86,9	122	13,1	934	57,2
Frau	598	85,7	100	14,3	698	42,8
<b>Familienstand</b>						
Verheiratet	840	88,5	109	11,5	949	58,1
Verh.getr.leb./ verwitwet / ledig	458	85,6	77	14,4	535	32,3
Geschieden	112	75,7	36	24,3	148	9,1
Gesamt (jeweils)	1410	86,4	222	13,6	1632	100

Tabelle A.2: Stichprobenverteilung: Einflussgrößen und Variable arbeitslos für das Jahr 2004

Teilung der Datei nach	Arbeitslos		Insgesamt Anzahl
	Nein Anzahl	Ja Anzahl	
<b>Geschlechtszugehörigkeit</b>			
Mann	808	122	930
Frau	597	100	697
<b>Erhebungsgebiet</b>			
Alte Bundesländer	968	94	1062
Neue Bundesländer	437	128	568
<b>Alter</b>			
18 bis 44 Jahre	862	141	1003
45 bis 67 Jahre	543	81	624
<b>Geschlechtszugehörigkeit und Alter</b>			
Mann, 18 bis 44 Jahre	479	82	561
Mann, 45 bis 67 Jahre	329	40	369
Frau, 18 bis 44 Jahre	383	59	442
Frau, 45 bis 67 Jahre	214	41	255

Tabelle A.3: Stichprobenverteilung für getrennte Dateien für das Jahr 2004

Testergebnisse	Chi quadrat	df	Signf.	Nagelkerke
<b>Geschlechtszugehörigkeit</b>				
Mann	150,266	16	0,000	0,276
Frau	119,119	16	0,000	0,280
<b>Erhebungsgebiet</b>				
Alte Bundesländer	98,998	16	0,000	0,198
Neue Bundesländer	107,804	16	0,000	0,264
<b>Alter</b>				
18 bis 44 Jahre	138,055	16	0,000	0,231
45 bis 67 Jahre	116,691	16	0,000	0,317
<b>Geschlechtszugehörigkeit und Alter</b>				
Mann, 18 bis 44 Jahre	93,891	15	0,000	0,273
Mann, 45 bis 67 Jahre	69,847	15	0,000	0,347
Frau, 18 bis 44 Jahre	63,636	15	0,000	0,246
Frau, 45 bis 67 Jahre	68,006	15	0,000	0,399

Tabelle A.4: Testergebnisse der logistischen Regression für getrennte Dateien für das Jahr 2004

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step	GEBIET	1,002	,265	14,294	1	,000	2,723
1	GESUND			27,743	4	,000	
	GESUND(1)	-2,689	,615	19,122	1	,000	,068
	GESUND(2)	-2,842	,573	24,605	1	,000	,058
	GESUND(3)	-2,407	,581	17,146	1	,000	,090
	GESUND(4)	-1,895	,655	8,369	1	,004	,150
	RAUCHEN	-,638	,233	7,511	1	,006	,528
	ALTER	-,196	,263	,554	1	,457	,822
	BERUF			11,355	2	,003	
	BERUF(1)	1,284	,404	10,119	1	,001	3,611
	BERUF(2)	1,398	,438	10,181	1	,001	4,047
	FAMIL			11,912	2	,003	
	FAMIL(1)	,432	,256	2,842	1	,092	1,540
	FAMIL(2)	1,156	,344	11,295	1	,001	3,176
	SCHULE			29,950	4	,000	
	SCHULE(1)	3,770	,823	21,011	1	,000	43,396
	SCHULE(2)	2,312	,664	12,113	1	,001	10,092
	SCHULE(3)	1,432	,646	4,914	1	,027	4,186
	SCHULE(4)	1,596	,714	4,991	1	,025	4,934
	KONF	,638	,306	4,338	1	,037	1,893
	Constant	-3,820	1,174	10,583	1	,001	,022

Tabelle A.5: Effekte der logistischen Regression für Männer in 2004

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
GEBIET	1,373	,275	24,916	1	,000	3,949
GESUND			9,152	4	,057	
GESUND(1)	-,572	,706	,657	1	,418	,565
GESUND(2)	-,787	,655	1,443	1	,230	,455
GESUND(3)	-,139	,659	,045	1	,833	,870
GESUND(4)	,328	,752	,191	1	,662	1,389
RAUCHEN	-,441	,253	3,038	1	,081	,644
ALTER	-,126	,268	,222	1	,638	,881
BERUF			1,826	2	,401	
BERUF(1)	,408	,430	,902	1	,342	1,504
BERUF(2)	,051	,398	,016	1	,898	1,052
FAMIL			,059	2	,971	
FAMIL(1)	,014	,288	,002	1	,962	1,014
FAMIL(2)	,087	,359	,058	1	,809	1,090
SCHULE			17,541	4	,002	
SCHULE(1)	1,774	1,404	1,596	1	,207	5,893
SCHULE(2)	2,714	,778	12,159	1	,000	15,089
SCHULE(3)	2,101	,745	7,960	1	,005	8,174
SCHULE(4)	,907	,948	,915	1	,339	2,476
KONF	1,209	,504	5,754	1	,016	3,349
Constant	-6,948	1,524	20,772	1	,000	,001

Tabelle A.6: Effekte der logistischen Regression für Frauen in 2004

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
GEBIET	1,012	,232	18,979	1	,000	2,751
GESUND			18,194	4	,001	
GESUND(1)	-2,003	,603	11,026	1	,001	,135
GESUND(2)	-2,146	,578	13,787	1	,000	,117
GESUND(3)	-1,729	,594	8,467	1	,004	,178
GESUND(4)	-1,215	,692	3,082	1	,079	,297
SEX	,162	,224	,524	1	,469	1,176
RAUCHEN	-,736	,210	12,342	1	,000	,479
BERUF			6,703	2	,035	
BERUF(1)	,842	,336	6,285	1	,012	2,321
BERUF(2)	,510	,354	2,078	1	,149	1,665
FAMIL			,556	2	,757	
FAMIL(1)	,036	,213	,029	1	,865	1,037
FAMIL(2)	,263	,354	,554	1	,457	1,301
SCHULE			18,759	4	,001	
SCHULE(1)	2,672	,817	10,711	1	,001	14,476
SCHULE(2)	2,218	,647	11,748	1	,001	9,193
SCHULE(3)	1,461	,619	5,578	1	,018	4,311
SCHULE(4)	1,400	,676	4,296	1	,038	4,057
KONF(1)	-,832	,312	7,128	1	,008	,435
Constant	-2,543	1,030	6,101	1	,014	,079

Tabelle A.7: Effekte der logistischen Regression für jüngere Leute (18-44) in 2004

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
GEBIET	1,431	,331	18,679	1	,000	4,185
GESUND			11,824	4	,019	
GESUND(1)	-1,546	,783	3,902	1	,048	,213
GESUND(2)	-1,871	,672	7,751	1	,005	,154
GESUND(3)	-1,070	,659	2,633	1	,105	,343
GESUND(4)	-1,004	,728	1,899	1	,168	,367
SEX	,027	,294	,009	1	,925	1,028
RAUCHEN	-,263	,295	,794	1	,373	,769
BERUF			5,363	2	,068	
BERUF(1)	,899	,528	2,898	1	,089	2,458
BERUF(2)	1,194	,519	5,296	1	,021	3,299
FAMIL			8,876	2	,012	
FAMIL(1)	,582	,400	2,113	1	,146	1,789
FAMIL(2)	1,003	,353	8,100	1	,004	2,727
SCHULE			20,991	4	,000	
SCHULE(1)	4,171	1,120	13,874	1	,000	64,778
SCHULE(2)	2,832	,787	12,934	1	,000	16,979
SCHULE(3)	1,967	,762	6,660	1	,010	7,150
SCHULE(4)	-3,912	9,950	,155	1	,694	,020
KONF(1)	-,586	,426	1,887	1	,170	,557
Constant	-5,412	1,298	17,371	1	,000	,004

Tabelle A.8: Effekte der logistischen Regression für ältere Leute (45-67) in 2004

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
SEX	-,244	,267	,834	1	,361	,783
GESUND			27,569	4	,000	
GESUND(1)	-2,407	,607	15,742	1	,000	,090
GESUND(2)	-2,860	,573	24,884	1	,000	,057
GESUND(3)	-2,225	,576	14,935	1	,000	,108
GESUND(4)	-1,857	,653	8,088	1	,004	,156
RAUCHEN	-,429	,248	3,006	1	,083	,651
ALTER	-,324	,273	1,411	1	,235	,723
BERUF			6,525	2	,038	
BERUF(1)	1,066	,417	6,525	1	,011	2,904
BERUF(2)	,830	,425	3,809	1	,051	2,293
FAMIL			,745	2	,689	
FAMIL(1)	,091	,266	,117	1	,732	1,095
FAMIL(2)	,330	,388	,724	1	,395	1,392
SCHULE			16,094	4	,003	
SCHULE(1)	2,604	,783	11,066	1	,001	13,511
SCHULE(2)	1,706	,635	7,214	1	,007	5,510
SCHULE(3)	1,045	,634	2,721	1	,099	2,844
SCHULE(4)	1,113	,688	2,617	1	,106	3,044
KONF	,972	,272	12,768	1	,000	2,644
Constant	-2,297	1,183	3,768	1	,052	,101

Tabelle A.9: Effekte der logistischen Regression für alte Bundesländer in 2004

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
SEX	,464	,249	3,470	1	,063	1,590
GESUND			8,495	4	,075	
GESUND(1)	-,925	,700	1,748	1	,186	,397
GESUND(2)	-,925	,647	2,040	1	,153	,397
GESUND(3)	-,361	,657	,302	1	,583	,697
GESUND(4)	-,019	,745	,001	1	,980	,982
RAUCHEN	-,580	,233	6,215	1	,013	,560
ALTER	-,080	,263	,093	1	,760	,923
BERUF			3,206	2	,201	
BERUF(1)	,699	,394	3,146	1	,076	2,012
BERUF(2)	,617	,405	2,318	1	,128	1,853
FAMIL			5,750	2	,056	
FAMIL(1)	,356	,271	1,727	1	,189	1,427
FAMIL(2)	,772	,336	5,281	1	,022	2,164
SCHULE			29,429	4	,000	
SCHULE(1)	3,749	1,188	9,960	1	,002	42,480
SCHULE(2)	3,236	,776	17,368	1	,000	25,422
SCHULE(3)	2,228	,742	9,014	1	,003	9,282
SCHULE(4)	,985	,952	1,071	1	,301	2,678
KONF	-,822	,677	1,471	1	,225	,440
Constant	-1,593	1,754	,825	1	,364	,203

Tabelle A.10: Effekte der logistischen Regression für neue Bundesländer in 2004

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
GEBIET	,993	,317	9,809	1	,002	2,700
GESUND			16,551	4	,002	
GESUND(1)	-2,947	,827	12,705	1	,000	,052
GESUND(2)	-3,097	,801	14,953	1	,000	,045
GESUND(3)	-2,686	,819	10,762	1	,001	,068
GESUND(4)	-2,235	,977	5,231	1	,022	,107
RAUCHEN	-,924	,295	9,832	1	,002	,397
BERUF			9,700	2	,008	
BERUF(1)	1,566	,508	9,488	1	,002	4,786
BERUF(2)	1,203	,566	4,509	1	,034	3,329
FAMIL			2,856	2	,240	
FAMIL(1)	,204	,290	,495	1	,482	1,226
FAMIL(2)	,868	,518	2,811	1	,094	2,382
SCHULE			9,669	4	,046	
SCHULE(1)	2,979	1,214	6,023	1	,014	19,664
SCHULE(2)	2,046	1,087	3,543	1	,060	7,734
SCHULE(3)	1,439	1,063	1,835	1	,176	4,218
SCHULE(4)	1,850	1,091	2,875	1	,090	6,357
KONF	,548	,392	1,956	1	,162	1,730
Constant	-3,088	1,593	3,755	1	,053	,046

Tabelle A.11: Effekte der logistischen Regression für jüngere Männer in 2004

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
GEBIET	,917	,496	3,424	1	,064	2,502
GESUND			10,251	4	,036	
GESUND(1)	-2,636	1,140	5,349	1	,021	,072
GESUND(2)	-2,903	,935	9,645	1	,002	,055
GESUND(3)	-2,506	,934	7,203	1	,007	,082
GESUND(4)	-2,110	,992	4,519	1	,034	,121
RAUCHEN	-,437	,440	,989	1	,320	,646
BERUF			5,832	2	,054	
BERUF(1)	,741	,734	1,019	1	,313	2,098
BERUF(2)	1,631	,753	4,688	1	,030	5,110
FAMIL			12,035	2	,002	
FAMIL(1)	1,007	,583	2,981	1	,084	2,736
FAMIL(2)	1,668	,500	11,110	1	,001	5,299
SCHULE			19,849	4	,001	
SCHULE(1)	5,972	1,545	14,935	1	,000	392,277
SCHULE(2)	2,694	,882	9,326	1	,002	14,785
SCHULE(3)	1,484	,861	2,975	1	,085	4,413
SCHULE(4)	-4,739	22,553	,044	1	,834	,009
KONF	,831	,544	2,332	1	,127	2,294
Constant	-4,851	1,854	6,850	1	,009	,008

Tabelle A.12: Effekte der logistischen Regression für ältere Männer in 2004

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
GEBIET	1,096	,349	9,885	1	,002	2,994
GESUND			2,629	4	,622	
GESUND(1)	-,359	,991	,131	1	,718	,699
GESUND(2)	-,433	,947	,210	1	,647	,648
GESUND(3)	-,150	,972	,024	1	,878	,861
GESUND(4)	,516	1,087	,226	1	,635	1,676
RAUCHEN	-,564	,319	3,124	1	,077	,569
BERUF			,504	2	,777	
BERUF(1)	,020	,518	,001	1	,970	1,020
BERUF(2)	-,210	,474	,196	1	,658	,811
FAMIL			,208	2	,901	
FAMIL(1)	-,040	,334	,015	1	,904	,960
FAMIL(2)	-,232	,508	,208	1	,648	,793
SCHULE			13,897	4	,008	
SCHULE(1)	-2,345	14,570	,026	1	,872	,096
SCHULE(2)	2,612	,838	9,723	1	,002	13,628
SCHULE(3)	1,629	,771	4,461	1	,035	5,100
SCHULE(4)	,751	,962	,610	1	,435	2,119
KONF	1,412	,644	4,803	1	,028	4,103
Constant	-6,368	1,867	11,638	1	,001	,002

Tabelle A.13: Effekte der logistischen Regression für jüngere Frauen in 2004

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
GEBIET	2,001	,494	16,392	1	,000	7,393
GESUND			5,551	4	,235	
GESUND(1)	-,736	1,126	,427	1	,513	,479
GESUND(2)	-1,100	,975	1,272	1	,259	,333
GESUND(3)	-,048	,943	,003	1	,960	,953
GESUND(4)	,030	1,104	,001	1	,978	1,031
RAUCHEN	-,256	,450	,322	1	,570	,774
BERUF			2,732	2	,255	
BERUF(1)	1,391	,887	2,459	1	,117	4,020
BERUF(2)	,849	,841	1,019	1	,313	2,337
FAMIL			,956	2	,620	
FAMIL(1)	,025	,615	,002	1	,968	1,025
FAMIL(2)	,556	,581	,916	1	,338	1,744
SCHULE			,394	4	,983	
SCHULE(1)	7,828	24,535	,102	1	,750	2509,088
SCHULE(2)	8,467	24,501	,119	1	,730	4753,364
SCHULE(3)	8,359	24,498	,116	1	,733	4268,569
SCHULE(4)	-,083	46,452	,000	1	,999	,920
KONF	,545	,853	,408	1	,523	1,725
Constant	-14,019	24,594	,325	1	,569	,000

Tabelle A.14: Effekte der logistischen Regression für ältere Frauen in 2004



## Literaturverzeichnis

1. Andreß, H., Hagenars, J. A., Kühnel, S.: *Analyse von Tabellen und Kategorialen Daten: Log-lineare Modelle, latente Klassenanalyse, logistische Regression und GSK- Ansatz*  
Springer-Verlag Berlin Heidelberg 1997
2. Backhaus, K., Erichson, B., Plinke, W., Weiber, R.: *Multivariate Analysemethoden*  
Springer Verlag Berlin Heidelberg 2000
3. Dauer, S., Henning, H.: *Arbeitslosigkeit und Gesundheit*  
Mitteldeutscher Verlag GmbH Halle (Saale) 1999
4. Fahrmeier, L., Hamerle, A., Tutz, G.: *Multivariate statistische Verfahren*  
Walter de Gruyter, Berlin New York 1996
5. Gesundheitsberichterstattung des Bundes  
Robert-Koch-Institut in Zusammenarbeit mit dem Statistischen Bundesamt  
Heft 13: *Arbeitslosigkeit und Gesundheit*  
Februar 2003
6. Janssen, J., Laatz, W.: *Statistische Datenanalyse mit SPSS für Windows*  
Springer-Verlag Berlin Heidelberg 1999
7. Kleinbaum, D. G.: *Logistic Regression: A Self-Learning Text*  
Springer-Verlag New York, 1992
8. Ronning, G.: *Mikroökonomie*  
Springer-Verlag Berlin Heidelberg 1991
9. Rönz, B.: *Computergestützte Statistik II*  
Vorlesungsskript 2000
10. Rönz, B.: *Generalisierte lineare Modelle*  
Vorlesungsskript 2001
11. [www.emilea.de](http://www.emilea.de), Internetseite

