

Universität Trier

**Zentrum für Informations-, Medien-
und Kommunikationstechnologie
(ZIMK)**



Trier, den 25.06.2012

B. Baltes-Götz

Logistische Regressionsanalyse mit SPSS

Inhaltsübersicht

VORWORT	5
1 EINLEITUNG	6
2 DIE BINÄRE LOGISTISCHE REGRESSION	9
2.1 Modell	9
2.1.1 Populationsmodell	9
2.1.2 Stichprobenmodell	10
2.1.3 Äquivalente Modellformulierungen	10
2.1.4 Ein möglicher Entstehungshintergrund	12
2.1.5 Vergleich mit der Probit-Analyse	14
2.2 Anwendungsbeispiel	15
2.3 Schätzung der Parameter	17
2.3.1 Die Maximum-Likelihood-Methode	17
2.3.2 Alternative Verfahren	19
2.4 Beurteilung der Modellgültigkeit	20
2.4.1 Globale Modellgültigkeitstests	20
2.4.1.1 Pearson- χ^2 -Statistik	20
2.4.1.2 Devianz-Statistik	23
2.4.1.3 Hosmer-Lemeshow - Statistik	24
2.4.2 Untersuchung von Residuen und Einflussindikatoren	26
2.4.2.1 Residuendiagnostik	26
2.4.2.2 Einflussreiche Fälle	34
2.5 Beurteilung der Modellrelevanz	36
2.5.1 Der Likelihood-Quotiententest zur globalen Nullhypothese	36
2.5.2 Pseudo- R^2 -Statistiken	37
2.5.3 Prädiktive Effizienz	39
2.5.3.1 Die Klassifikationstabelle	39
2.5.3.2 Klassifikationsdiagramm	40
2.6 Beurteilung der einzelnen Regressoren	42
2.6.1 Regressionskoeffizienten und Effektgrößen	42
2.6.2 Signifikanz	43
2.6.3 Fehlende bzw. irrelevante Prädiktoren	45
2.7 Nominalskalierte Regressoren mit mehr als zwei Kategorien	46
2.8 Interaktionen	51
2.8.1 Interaktionen zwischen nominalskalierten Regressoren	52
2.8.1.1 Bedeutung der Regressionsgewichte bei Indikatorcodierung	52
2.8.1.2 Bedeutung der Regressionsgewichte bei Abweichungskodierung	54
2.8.2 „Haupteffekte“ in Modellen mit Wechselwirkung	56
2.9 Strategien zur Modellbildung	58
2.9.1 Signifikanztests zu Prädiktorblöcken	58
2.9.2 Automatische Modellsuche	58
2.9.3 Empfehlungen zur Modellbildung	61

3	DIE MULTINOMIALE LOGISTISCHE REGRESSION	62
3.1	Populationsmodell	62
3.2	Stichprobenmodell	63
3.3	Anwendungsbeispiel	64
3.4	Parameterschätzung	66
3.5	Modellgültigkeit	67
3.6	Beurteilung der Modellrelevanz	68
3.7	Beurteilung der einzelnen Regressoren	69
3.8	Log-Likelihood - Varianten	70
4	DIE ORDINALE LOGISTISCHE REGRESSION	72
4.1	Das kumulative Logit-Modell	72
4.2	Anwendungsbeispiel	75
4.3	Parameterschätzung	76
4.4	Modellgültigkeit	78
4.4.1	Parallelität	78
4.4.2	Globale Modellgültigkeit	78
4.4.3	Lokale Modellanalyse	79
4.5	Beurteilung der Modellrelevanz	80
4.6	Beurteilung der einzelnen Regressoren	81
4.7	Vergleiche mit alternativen Auswertungsverfahren	82
4.7.1	Multinomiale logistische Regression	82
4.7.2	Lineare Regressions- bzw. Varianzanalyse	83
4.8	Lokations-Skalen - Modell	84
5	NUMERISCHE SCHÄTZPROBLEME	89
5.1	Multikollinearität	89
5.2	Quasi-vollständige Trennung	89
5.3	Vollständige Trennung	92

6	ANHANG	94
6.1	Symbolverzeichnis	94
6.2	SPSS-Programme zu den Beispielen	94
6.2.1	SPSS-Syntaxdatei zum DBS-Beispiel	95
6.2.2	SPSS-Syntaxdatei zum Beispiel für die multinomiale Regression	96
	LITERATUR	97
	STICHWORTVERZEICHNIS	99

Herausgeber: Universität Trier
Zentrum für Informations-, Medien und Kommunikationstechnologie (ZIMK)
Universitätsring 15
D-54286 Trier
Tel.: (0651) 201-3417, Fax.: (0651) 3921

Autor: Bernhard Baltes-Götz (E-Mail: baltes@uni-trier.de)

Copyright © 2012; ZIMK

Vorwort

In diesem Manuskript wird die logistische Regressionsanalyse für Kriteriumsvariablen mit folgender Struktur behandelt:

- nominalskaliert mit zwei oder mehr Kategorien
- ordinalskaliert

Als Software kommt SPSS Statistics 20.0 für Windows zum Einsatz, jedoch können praktisch alle vorgestellten Verfahren auch mit anderen SPSS-Versionen unter Linux, MacOS oder Windows realisiert werden.

Die aktuelle Version des Manuskripts ist als PDF-Dokument zusammen mit allen im Kurs benutzen Daten und SPSS-Programmen auf dem Webserver der Universität Trier von der Startseite (<http://www.uni-trier.de/>) ausgehend folgendermaßen zu finden:

[Rechenzentrum](#) > [Studierende](#) > [EDV-Dokumentationen](#) >
[Statistik](#) > [Logistische Regressionsanalyse mit SPSS](#)

Leider sind in diesem Manuskript einige Teile mit heißer Nadel gestrickt, so dass Unzulänglichkeiten zu befürchten und entsprechende Hinweise der Leser(innen) zu erhoffen sind (z.B. an die Mail-Adresse baltes@uni-trier.de).

Trier, im Juni 2012

Bernhard Baltes-Götz

1 Einleitung

In der statistischen Forschungspraxis sind oft nominal- oder ordinalskalierte Kriterien zu untersuchen, z.B.:

- Kaufentscheidung für ein Produkt (nominales Kriterium mit zwei Kategorien):
 - ja
 - nein
- Wahl eines Verkehrsmittels für den Weg zur Uni (nominales Kriterium mit drei Kategorien):
 - per Pedes oder Pedal (Fahrrad)
 - ÖPNV
 - PKW
- Durchblutungsstörung (ordinales Kriterium):
 - keine
 - periphere
 - koronare
- Stellungnahme (ordinales Kriterium):
 - entschieden dagegen
 - eher dagegen
 - neutral
 - eher dafür
 - entschieden dafür

Auf der Suche nach einem Modell zur Erklärung und oder Vorhersage solcher Kriterien sollen in der Regel mehrere Regressoren mit metrischer oder nominaler Skalenqualität einbezogen werden, wobei eventuell auch Interaktionen zwischen zwei oder mehreren Regressoren unterstellt werden sollen.

In dieser Situation kann die vertraute *lineare Regressionsanalyse* nicht eingesetzt werden:

- Ihre Voraussetzung normalverteilter und varianzhomogener Residuen ist offensichtlich verletzt.
- Ihre prognostizierten Werte können außerhalb des plausiblen Bereichs liegen.
Z.B. sind bei einer dichotomen Kriteriumsvariablen nur Prognosewerte von 0 bis 1 sinnvoll, damit diese als Wahrscheinlichkeiten der Zugehörigkeit zur ersten Gruppe interpretiert werden können.

Die *lineare Diskriminanzanalyse* bietet zwar eine Prognose der Gruppenzugehörigkeit, ist aber vielfach wegen ihrer Voraussetzungen bzgl. der Prädiktoren nicht anwendbar:

- Intervallskalenqualität
- multivariate Normalverteilung innerhalb der Populationen zu den Kriteriumsausprägungen
- Homogenität der Kovarianzmatrizen
- Keine Wechselwirkungen zwischen den Prädiktoren in Bezug auf das Kriterium

Sofern nur ein einzelner Prädiktor in Frage kommt, der zudem nominales Messniveau besitzt, kann zur Analyse eines nominalen Kriteriums die Kreuztabellenanalyse verwendet werden (siehe z.B. Baltes-Götz 2012, Kap. 11).

Die Beschränkung auf *einen* Prädiktor entfällt bei den zur Analyse von multivariaten Kontingenztabellen oft vorgeschlagenen *log-lineare Modellen* (siehe z.B. Bühl 2012, Kap. 22). Zwar werden hier zunächst nur *Assoziationen* modelliert (ohne Unterscheidung zwischen abhängigen und unabhängigen Variablen), doch lassen sich auch asymmetrische Modelle für die Erklärung kategorialer Kriteriumsvariablen aus *kategorialen* Prädiktorvariablen formulieren, die oft als *logit-lineare Modelle* bezeichnet werden. In SPSS ermöglicht die Prozedur GENLOG sowohl symmetrische als auch asymmetrische Modelle. Als Probleme bei der Anwendung log- bzw. logit-linearer Modelle sind zu nennen:

- Metrische Prädiktoren können nur durch eine durch künstliche (und willkürliche) Kategorisierung einbezogen werden. Dabei verliert man sowohl Information (durch Vergröberung) als auch statistische Effizienz (durch eine erhöhte Anzahl von Parametern).
- Viele Forschungspraktiker(innen) mit einem regressionsanalytischen Denkansatz empfinden die primär assoziativen log-linearen Modelle, die oft verwirrend viele Parameter besitzen, und deren asymmetrische Reformulierung als relativ unhandlich.

Trotz einer deutlichen Verwandtschaft¹ mit dem log-linearen Ansatz ist die *logistische Regressionsanalyse* zur Untersuchung von kategorialen oder ordinalen Kriterien oft besser geeignet. Bei dieser multivariaten Methode werden die Wahrscheinlichkeiten der Zugehörigkeit zu den Kriteriumsgruppen aufgrund von *intervall- oder nominalskalierten* Prädiktoren modelliert, wobei ein *verallgemeinertes lineares Modell* zum Einsatz kommt (vgl. McCullagh & Nelder 1989). Viele Schätzer und Tests im Rahmen der logistischen Regressionsanalyse haben direkte Entsprechungen bei der linearen Regressionsanalyse, z.B.:

- der Test zur globalen Nullhypothese, dass alle Parameter außer dem konstanten Term gleich 0 sind
- die Tests zu den Nullhypothesen zu den einzelnen Parametern
- Bestimmtheitsmaße zur Beurteilung der Modellrelevanz

Es sind Modellgültigkeitstests verfügbar, die auch bei *Individualdaten* (also lediglich einfach besetzten Prädiktorwertkombinationen) anwendbar sind.

Weil die logistische Regression *Wahrscheinlichkeiten* für die Zugehörigkeit zu Kriteriumsgruppen zu modellieren hat, besitzt ihre Modellgleichung einige Besonderheiten im Vergleich zur linearen Regression, die aber nach der Lektüre dieses Manuskripts keine Schwierigkeiten mehr machen sollten. Lohn für diese Bemühungen ist ein für nahezu beliebige abhängige Variablen (ohne Bauchschmerzen) anwendbares Analyseverfahren. Auf der *unabhängigen* Seite bestehen dieselben Möglichkeiten und Einschränkungen wie bei einer linearen Regressionsanalyse. Man kann metrische und kategoriale Regressoren verwenden, muss ordinale Variablen also entweder als kategorial oder als metrisch behandeln.

In SPSS stehen für die logistische Regressionsanalyse u.a. die *drei* folgenden Prozeduren bereit:²

- **LOGISTIC REGRESSION**
Diese über den Menübefehl

Analysieren > Regression > Binär Logistisch

ansprechbare Prozedur analysiert *dichotome* Kriterien unter Verwendung von *Individualdaten* und ist damit besonders geeignet für Modelle, die neben kategorialen auch *metrische*, in zahlreichen Ausprägungen realisierte Regressoren enthalten. LOGISTIC REGRESSION bietet u.a. den für Modelle mit überwiegend *einfach* besetzten Prädiktorwertkombinationen (Zellen) geeigneten Hosmer-Lemeshow-Modellgültigkeitstest, kann zahlreiche diagnostische Informationen zum ge-

¹ Zu jedem logistischem Regressionsmodell mit ausschließlich kategorialen Regressoren existiert ein äquivalentes log-lineares Modell.

² Die drei aufgelisteten und im Manuskript behandelten Prozeduren sind schon in der SPSS-Version 10 vorhanden. Aktuelle SPSS-Versionen bieten weitere Optionen:

- **Analysieren > Verallgemeinerte Lineare Modelle > Verallgemeinerte Schätzgleichungen**
Für Daten mit Abhängigkeitsstrukturen (durch Clusterbildung oder Messwiederholung) sind Modelle mit dichotomen Kriterien möglich, wobei die Abhängigkeiten durch die GEE-Methodologie nach Liang & Zeger (1986) neutralisiert werden.
- **Analysieren > Gemischte Modelle > Verallgemeinert Linear**
Für Daten mit Abhängigkeitsstrukturen (durch Clusterbildung oder Messwiederholung) sind Mehrebenenmodelle für kategoriale Kriterien (mit zwei oder mehr Ausprägungen) möglich.
- **Analysieren > Komplexe Stichproben**
Für komplexe Stichproben, die nicht durch einfache Zufallsauswahl zustande gekommen sind, können logistische Regressionen mit einem kategorialen oder ordinalen Kriterium gerechnet werden.

schätzten Modell (z.B. Residuen, Cook-Distanzen) als neue Variablen abspeichern und erlaubt bei nominalskalierten Prädiktoren eine flexible Wahl der Kontrastkodierung.

- **NOMREG**

Diese über den Menübefehl

Analysieren > Regression > Multinomial Logistisch

erreichbare Prozedur unterstützt auch nominalskalierte Kriterien mit mehr als zwei Kategorien. Obwohl NOMREG auch für den binären Spezialfall verwendbar ist, wird LOGISTIC REGRESSION *nicht* komplett ersetzt, weil beide Prozeduren mit teilweise unterschiedlichen Algorithmen arbeiten. Während die rein binäre Variante mit *Individualdaten* rechnet, fasst NOMREG alle Fälle mit einer gemeinsamen Prädiktorwertekombinationen zu einer Gruppe zusammen. Bei zentralen Ergebnissen einer binären logistischen Regressionsanalyse (z.B. bei der Parameterschätzung) wirken sich die Algorithmus-Unterschiede *nicht* aus, so dass die Entscheidung zwischen den beiden SPSS-Prozeduren irrelevant ist. Speziell zur Beschreibung und Testung der Anpassungsgüte eines Modells sind jedoch Statistiken vorgeschlagen worden, die eher für aggregierte Daten oder eher für Individualdaten geeignet sind. Dementsprechend werden sie nur von NOMREG (z.B. Modellgültigkeitstest über Pearsons Goodness of Fit – Statistik) oder nur von LOGISTIC REGRESSION (z.B. Hosmer-Lemeshow - Modellgültigkeitstest) berechnet.

- **PLUM**

Diese über den Menübefehl

Analysieren > Regression > Ordinal

erreichbare Prozedur verwendet die PLUM-Technologie (*PoLytomous Universal Model*) zur Analyse von *ordinalen* Kriterien. Neben der logistischen Linkfunktion (siehe unten) werden auch etliche Alternativen unterstützt. Wie NOMREG arbeitet auch PLUM intern mit aggregierten Daten.

Je nach SPSS-Version sind die drei Prozeduren unterschiedlich auf die Module *Base*, *Regression* und *Advanced* verteilt. In den meisten SPSS-Installationen dürften jedoch alle genannten Module und damit auch alle im Manuskript behandelten Optionen zur logistischen Regressionsanalyse enthalten sein.

Wir werden in diesem Manuskript die wichtigsten statistischen Grundlagen der logistischen Regression in einiger Ausführlichkeit besprechen und natürlich auch die Verwendung der SPSS-Prozeduren behandeln. Im Abschnitt 2 wird der besonders wichtige und angenehm einfache Spezialfall der *binären* logistischen Regression (mit einer *dichotomen* Kriteriumsvariablen) vorgestellt. Im Abschnitt 3 folgt mit der *multinomialen* logistischen Regression die Generalisierung auf nominalskalierte Kriterien mit mehr als zwei Kategorien, und im Abschnitt 4 werden *ordinale* Kriteriumsvariablen behandelt. Schließlich kommen im Abschnitt 4.8 noch kritische Datenverhältnisse zur Sprache, die zu irregulären Ergebnissen führen können.

2 Die binäre logistische Regression

Der Bequemlichkeit halber wird im Manuskript gelegentlich die Abkürzung *BLR* für die binäre logistische Regression verwendet.

2.1 Modell

2.1.1 Populationsmodell

Für eine (0,1) - kodierte Kriteriumsvariable Y und die Prädiktorvariablen X_1 bis X_M (intervallskaliert oder durch Kodierung von kategorialen Variablen entstanden) erklärt das logistische Regressionsmodell die Wahrscheinlichkeit für das Ereignis $\{Y = 1\}$ folgendermaßen:

$$P(Y = 1) = \frac{e^{\beta X}}{1 + e^{\beta X}} = \frac{1}{1 + e^{-\beta X}}, \quad \text{mit } \boldsymbol{\beta} := [\beta_0, \beta_1, \beta_2, \dots, \beta_M] \quad \text{und} \quad \mathbf{X} := \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} \quad (1)$$

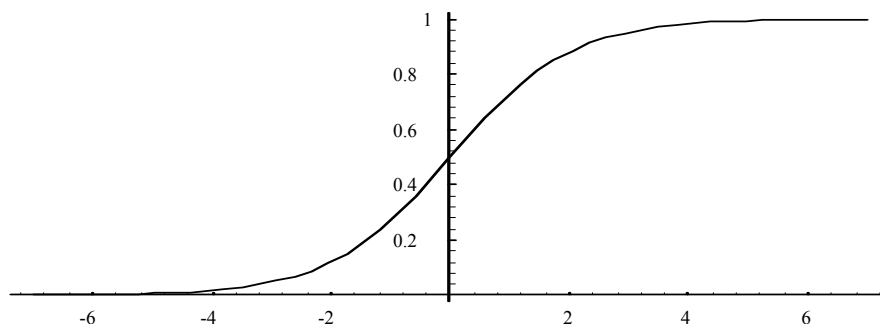
Anmerkungen:

- Zur Vereinfachung der Formeln werden im Manuskript gelegentlich elementare Vektorausdrücke verwendet. Zum Verständnis ist aber lediglich die folgende Multiplikationsregel erforderlich (am Beispiel $\boldsymbol{\beta X}$):

$$\boldsymbol{\beta X} = [\beta_0, \beta_1, \beta_2, \dots, \beta_M] \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} = \beta_0 + \sum_{m=1}^M \beta_m X_m$$

- Mit e ist die Eulersche Zahl gemeint, also die Basis zum natürlichen Logarithmus.

In der Modellgleichung ist hinter die lineare Funktion $\boldsymbol{\beta X}$ der Prädiktoren die **logistische Verteilungsfunktion** geschaltet. Sie sorgt dafür, dass alle Modellprognosen im Intervall von 0 bis 1 liegen und daher als Wahrscheinlichkeiten interpretiert werden können. Wie die folgende Abbildung zeigt, liefert die logistische Verteilungsfunktion für beliebige Argumente (von $-\infty$ bis ∞) einen Funktionswert im Intervall (0,1):



Im Vergleich zur Standardnormalverteilungsfunktion besitzt die logistische Verteilungsfunktion einerseits einen sehr ähnlichen Verlauf und andererseits eine mathematisch einfachere Beschreibung (siehe Abschnitt 2.1.4).

2.1.2 Stichprobenmodell

Während sich das eben vorgestellte *Populationsmodell* auf theoretischer Ebene bewegt, beschreibt das *Stichprobenmodell*, wie die Daten einer empirischen Studie zustande gekommen sind. Beobachtet man bei N Fällen mit den Prädiktorwertekombinationen \mathbf{x}_i ($i = 1, \dots, N$) jeweils die Kriteriumskategorie, kommt folgendes BLR-Stichprobenmodell zum Einsatz:

- Es sind N unabhängige (0, 1) - wertige Zufallsvariablen Y_i vorhanden.
- Für die Wahrscheinlichkeit zum Einserereignis gilt bei der i -ten Beobachtung:

$$P(Y_i = 1) = \frac{e^{\beta \mathbf{x}_i}}{1 + e^{\beta \mathbf{x}_i}}, \quad \text{mit } \beta \mathbf{x}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_M x_{iM}$$

Sofern die Unabhängigkeit (z.B. durch eine geeignete Fallrekrutierung) sichergestellt ist, kann bei einer logistischen Regressionsanalyse also nur die zweite Annahme verletzt sein (z.B. durch einen Spezifikationsfehler im linearen Teil des Modells).

Um präzise Schätzungen und zuverlässige Signifikanztests zu erhalten, ist auf eine ausreichende **Stichprobengröße** zu achten. In der Literatur finden sich u.a. folgende Empfehlungen:

- Urban (1993, S. 13) nennt 50 Fälle als minimale Stichprobengröße, rechnet aber erst ab 100 Fällen mit einer zufrieden stellenden Präzision.
- Nach Backhaus et al. (2008, S. 288) sollte jede Kategorie der abhängigen Variablen mindestens 25 Fällen enthalten, bei einer „größeren“ Anzahl von unabhängigen Variablen jedoch mehr.
- Bei Hosmer & Lemeshow (2000, S. 346) und Norušis (2005, S. 319) wird das Zehnfache der Anzahl zu schätzender Parameter als minimale Häufigkeit der schwächer besetzten Kriteriumskategorie genannt. Bei einem dichotomen Kriterium und 4 metrischen Prädiktoren in einem Modell mit Ordinateabschnitt (also insgesamt 5 Parametern) sollten also beide Kriteriumskategorien minimal 50 Fälle enthalten.

Für Modellgültigkeitstests auf der Basis von Pearson- oder Devianz-Residuen müssen alle K Prädiktorwertekombinationen *mehrfach* besetzt sein. Häufig wird gefordert, dass die erwarteten Häufigkeiten unter dem zu prüfenden Modell bei allen ($2 \cdot K$) Zellen größer als 1 und bei mindestens 80% aller Zellen größer als 5 sein sollen. Der bei überwiegend einfach besetzten Prädiktorwertekombinationen anwendbare Hosmer-Lemeshow-Modellgültigkeitstest benötigt über die obigen Empfehlungen hinaus keine Voraussetzungen bei den Zellhäufigkeiten aus.

Insbesondere bei der Analyse von *seltenen* Attributen ist eine spezielle Liberalität der binären logistischen Regressionsanalyse hinsichtlich der Stichprobenziehung von Vorteil: Man kann aus den Teilpopulationen *mit* bzw. *ohne* das zu untersuchende Attribut (z.B. eine seltene Krankheit) Stichproben mit unterschiedlichen Quoten ziehen (z.B. 10 % aus der Patientenpopulation und 1 % aus der Kontrollpopulation), um stabilere Ergebnisse (z.B. kleinere Standardfehler) im Vergleich zu einer einfachen Zufallsstichprobe aus der Gesamtpopulation zu erhalten. Die essentiellen Ergebnisse (z.B. die Regressionskoeffizienten mit Ausnahme des konstanten Terms) sind gegenüber einer solchen kriteriumsgesteuerten Quotierung invariant (Allison 1999, S. 78ff; Norušis 2008, S. 63). Bei der linearen Regressionsanalyse sind analoge Strategien der Stichprobenziehung mit Problemen verbunden.

2.1.3 Äquivalente Modellformulierungen

Aus dem Populationsmodell (1) ergibt sich durch äquivalente Umformung folgende Darstellung für das logarithmierte Verhältnis aus den beiden beteiligten komplementären Wahrscheinlichkeiten:

$$\ln\left(\frac{P(Y=1)}{P(Y=0)}\right) = \beta X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_M X_M \quad (2)$$

Für den Wahrscheinlichkeitsquotienten $\left(\frac{P(Y=1)}{P(Y=0)}\right)$ schlägt Urban (1993, S. 25) die Bezeichnung *Gewinnchance* vor; in der angelsächsischen Literatur spricht man von den *odds*. Der logarithmierte Wahrscheinlichkeitsquotient wird generell als **Logit** bezeichnet.

In der folgenden Tabelle sind zur Illustration der Beziehungen zwischen den beteiligten Begriffen für einige Einserwahrscheinlichkeiten die zugehörigen Odds- und Logit-Werte angegeben:

$P(Y=1)$	$\frac{P(Y=1)}{P(Y=0)}$	$\ln\left[\frac{P(Y=1)}{P(Y=0)}\right]$
0,90	9	2,20
0,75	3	1,10
0,5	1	0
0,25	0,3	-1,10
0,10	0,1	-2,20

Obwohl auf der rechten Seite von Gleichung (2) ein *lineares* Modell steht, ist das in der Einleitung angesprochene Wertebereichsproblem bei der linearen Modellierung von Wahrscheinlichkeiten durch die Logit-Definition überwunden:

- Der Wahrscheinlichkeitsquotient nimmt Werte von 0 bis $+\infty$ (positiv unendlich) an.
- Durch das Logarithmieren ergibt sich ein Wertebereich von $-\infty$ bis $+\infty$.

Über die Logit-Formulierung des Modells lassen sich seine Koeffizienten analog zur gewöhnlichen linearen Regressionsgleichung interpretieren: Bei der binären logistischen Regression gibt der Koeffizient β_m an, wie sich das Logit verändert, wenn der Prädiktor X_m um eine Einheit erhöht wird, und alle anderen Prädiktoren unverändert bleiben. Diese Betrachtungsweise setzt (wie bei der linearen Regression) voraus, dass X_m bei keiner Interaktion beteiligt ist.

Durch Anwendung der Exponentialfunktion auf Gleichung (2) erhalten wir:

$$\frac{P(Y=1)}{P(Y=0)} = e^{\beta X} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_M X_M} = e^{\beta_0} e^{\beta_1 X_1} \dots e^{\beta_M X_M} \quad (3)$$

Die Gleichung zur Erklärung der Wahrscheinlichkeitsquotienten (engl. *odds*) bietet eine alternative Interpretationsmöglichkeit für die Regressionskoeffizienten. Der Ausdruck e^{β_m} gibt den *Faktor* an, um den sich das Wahrscheinlichkeitsverhältnis ändert, wenn der Prädiktor X_m um eine Einheit erhöht wird, und alle anderen unverändert bleiben:

$$e^{\beta_0} e^{\beta_1 X_1} \dots e^{\beta_m (X_m+1)} \dots e^{\beta_M X_M} = e^{\beta_0} e^{\beta_1 X_1} \dots e^{\beta_m X_m} \dots e^{\beta_M X_M} \cdot e^{\beta_m}$$

Weil e^{β_m} gerade dem Quotienten aus den Odds für $(X_1, X_2, \dots, X_m+1, \dots, X_M)$ und $(X_1, X_2, \dots, X_m, \dots, X_M)$

$$e^{\beta_m} = \frac{e^{\beta_0} e^{\beta_1 X_1} \dots e^{\beta_m (X_m+1)} \dots e^{\beta_M X_M}}{e^{\beta_0} e^{\beta_1 X_1} \dots e^{\beta_m X_m} \dots e^{\beta_M X_M}}$$

entspricht, wird der Ausdruck (nicht nur) in der angelsächsischen Literatur meist als *odds ratio* bezeichnet. Urban (1993, S. 40f) spricht von der *Effektgröße*.

Aus dem Verhalten der Exponentialfunktion folgt unmittelbar:

- Ist β_m positiv (also $e^{\beta_m} > 1$), dann steigt das Wahrscheinlichkeitsverhältnis ($P(Y = 1)$) wird größer).
- Ist β_m negativ (also $e^{\beta_m} < 1$), dann sinkt das Wahrscheinlichkeitsverhältnis ($P(Y = 1)$) wird kleiner).

Wird ein metrischer Prädiktor vor der binären logistischen Regressionsanalyse (BLR) standardisiert, liefert e^{β_m} den Effekt bei einer Erhöhung des Wertes um eine Standardabweichung.

Oft wird das BLR-Modell über die Logit-Gleichung (2) eingeführt bzw. definiert. Man überzeugt sich leicht davon, dass aus dieser Modellformulierung sofort die Gleichung (1) folgt.

2.1.4 Ein möglicher Entstehungshintergrund

Bei vielen binären Kriteriumsvariablen kann man sich vorstellen, dass ihre Werte durch das Dichotomisieren einer latenten metrischen Variablen entstanden sind. Für die manifeste Kriteriumsvariable Y und die zugehörige latente Variable η nimmt man also folgendes „Messmodell“ an:

$$Y = \begin{cases} 0, & \text{falls } \eta \leq \tau \\ 1, & \text{falls } \eta > \tau \end{cases}$$

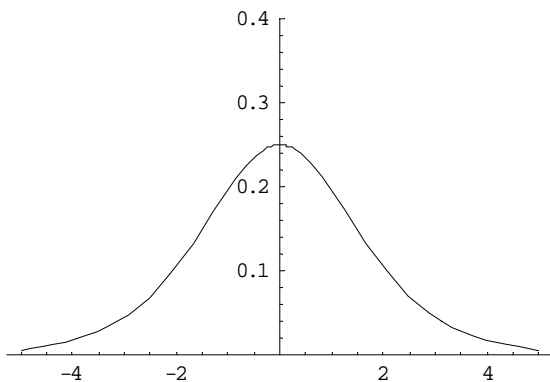
Für die Regression der latenten Variablen η auf die (manifesten) Prädiktorvariablen X_1 bis X_M setzt man ein lineares Modell mit der Residualvariablen ε an:

$$\eta = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_M X_M + \sigma \varepsilon \quad (4)$$

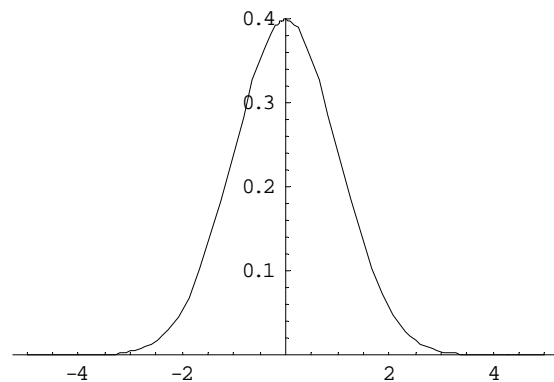
Schließlich wird angenommen, dass ε für jede Prädiktorwertekombination eine **logistische Verteilung** mit der folgenden Dichtefunktion besitzt:

$$f(w) = \frac{e^w}{(1 + e^w)^2}$$

Diese Verteilung hat den Erwartungswert 0 sowie die Varianz $\frac{\pi^2}{3}$ (mit der Kreiszahl $\pi = 3,1415 \dots$) und ähnelt der Normalverteilung, lädt aber im Vergleich zu dieser mehr Masse an den Rändern ab:



Dichte der logistischen Verteilung (Varianz $\frac{\pi^2}{3}$)



Dichte der Standardnormalverteilung (Varianz 1)

Damit die Varianz der Residualvariablen einen beliebigen Wert annehmen kann, enthält die Gleichung (4) noch den freien Parameter σ als Vorfaktor zu ε .

Vermutlich kommt Ihnen das Integral der logistischen Dichte, also die zugehörige Verteilungsfunktion, vertraut vor:

$$F(w) = \frac{e^w}{1 + e^w}$$

Das zugehörige Diagramm kennen Sie schon aus Abschnitt 2.1.1.

Bei festen Prädiktorausprägungen gilt für die Wahrscheinlichkeit $P(Y = 1) = P(\eta > \tau)$:

$$P(\eta > \tau) = P(\gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_M X_M + \sigma \varepsilon > \tau)$$

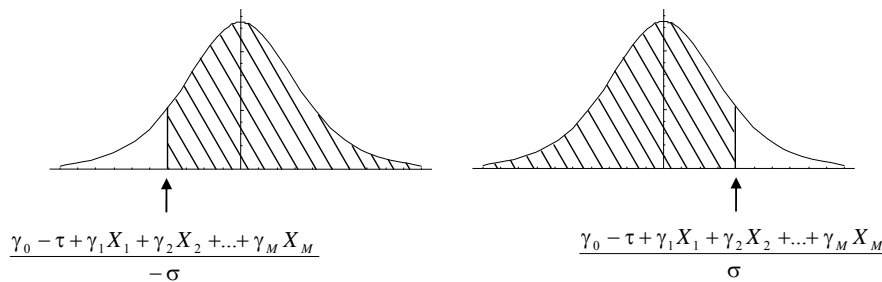
Einfaches Umstellen ergibt.

$$P(\eta > \tau) = P(-\sigma \varepsilon < \gamma_0 - \tau + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_M X_M)$$

Dividiert man die Ungleichung in der rechten Ereignisdefinition durch $(-\sigma)$, wechselt der Vergleichsoperator ($\sigma > 0$):

$$P(\eta > \tau) = P\left(\varepsilon > \frac{\gamma_0 - \tau + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_M X_M}{-\sigma}\right)$$

Aus der Symmetrie und Stetigkeit der logistischen Verteilung¹



folgt:

$$P(\eta > \tau) = P\left(\varepsilon \leq \frac{\gamma_0 - \tau + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_M X_M}{\sigma}\right)$$

Mit den Definitionen:

$$\begin{aligned} \beta_0 &:= \frac{\gamma_0 - \tau}{\sigma} \\ \beta_m &:= \frac{\gamma_m}{\sigma}, m = 1, \dots, M \end{aligned} \quad (5)$$

erhalten wir:

$$P(\eta > \tau) = P(\varepsilon \leq \beta \mathbf{X})$$

Die Wahrscheinlichkeit dafür, dass die Zufallsvariable ε einen Wert kleiner oder gleich $\beta \mathbf{X}$ annimmt, ist identisch mit dem Wert ihrer Verteilungsfunktion an dieser Stelle:

$$P(\varepsilon \leq \beta \mathbf{X}) = \frac{e^{\beta \mathbf{X}}}{1 + e^{\beta \mathbf{X}}}$$

Insgesamt erhalten wir die Modellgleichung (1):

$$P(Y = 1) = \frac{e^{\beta \mathbf{X}}}{1 + e^{\beta \mathbf{X}}}$$

¹ Aus dem für viele Leser wohl wenig vertrauten Begriff der Stetigkeit folgt, dass von den überabzählbar unendlich vielen Ausprägungen einer Variablen mit stetiger Verteilung jede einzelne Ausprägung die Wahrscheinlichkeit Null besitzt:

$$P(\varepsilon = \omega) = 0 \Rightarrow P(\varepsilon < \omega) = P(\varepsilon \leq \omega)$$

Die Koeffizienten des logistischen Modells für die manifeste (dichotome) Kriteriumsvariable sind also im Vergleich zu den Koeffizienten des Modells für die korrespondierende latente Variable um den Faktor $\frac{1}{\sigma}$ gemindert. Mit Ausnahme des konstanten Terms hängen sie **nicht** davon ab, bei welchem Schwellenwert τ die latente Variable η dichotomisiert wurde.

Weil σ in realen Studien unbekannt ist, lassen sich die Koeffizienten γ_m ($m = 1, \dots, M$) in Gleichung (4) nicht aus den Koeffizienten β_m ($m = 1, \dots, M$) in Gleichung (1) berechnen, doch wegen

$$\gamma_m = \sigma \beta_m, m = 1, \dots, M$$

ist ein Signifikanztest zur Hypothese

$$H_0 : \beta_m = 0$$

äquivalent zu einem Test der Hypothese

$$H_0 : \gamma_m = 0$$

Um Missverständnisse zu vermeiden, soll noch einmal betont werden, dass die in diesem Abschnitt präsentierte Herleitung der Modellgleichung (1) keinesfalls eine *Voraussetzung* für die Verwendung der binären logistischen Regression ist. Sie hilft jedoch beim Verständnis und bei der Einordnung des Verfahrens.

2.1.5 Vergleich mit der Probit-Analyse

Aus dem „logistischen“ Modell für eine latente und metrische Kriteriumsvariable (siehe Gleichung 4) ergibt sich sofort das analoge Modell der so genannten **Probit-Analyse**, wenn für die Residualvariable ε an Stelle der logistischen Verteilung eine Standardnormalverteilung angenommen wird.

Dementsprechend ergibt sich die Probit-Modellgleichung aus der BLR-Variante, indem die logistische Verteilungsfunktion durch das Normalverteilungs-Analogon ersetzt wird:

$$P(Y = 1) = \Phi(\beta X), \text{ mit } \Phi(\beta X) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta X} e^{-\frac{w^2}{2}} dw$$

Daraus erhält man sofort die meist verwendete Darstellungsform des Probit-Modells:

$$\Phi^{-1}(P(Y = 1)) = \beta X$$

Wegen der starken Verwandtschaft ihrer Modellgleichungen werden die BLR- und die Probit-Analyse in der Regel weitgehend äquivalente Ergebnisse produzieren (siehe z.B. Menard 1995, S. 59).

SPSS unterstützt die Probit-Analyse in den Prozeduren **PROBIT** (verfügbar über **Analysieren > Regression > Probit**) und **PLUM** (verfügbar über **Analysieren > Regression > Ordinal**, siehe Abschnitt 4).

2.2 Anwendungsbeispiel

Zur Erläuterung der logistischen Regression für dichotome abhängige Variablen wird ein künstlicher Beispieldatensatz verwendet. Wir stellen uns vor, dass bei einer medizinischen Untersuchung zu den Ursachen von Durchblutungsstörungen an einer Stichprobe der Größe $N = 200$ folgende Variablen erhoben worden sind:

- Kriteriumsvariable:

DBS	Vorliegen einer Durchblutungsstörung (1 = ja, 0 = nein)
-----	---
- Regressoren:
 - ABWIG Abweichung vom Idealgewicht (gemessen in kg)
 - BEWEG Körperliche Betätigung (Skala von 1 bis 6)
 - DRUCK Diastolischer Blutdruck (gemessen in mm/Hg)
 - STRESS Stress (Skala von 1 bis 6)
 - ERBE Erbliche Vorbelastung (1 = ja, 0 = nein)
 - RAUCHER (1 = aktiver Raucher, 2 = ehemaliger Raucher, 3 = Nichtraucher)

In der zu untersuchenden künstlichen Population gilt für eine latente Variable η im Sinn von Abschnitt 2.1.4:

$$\eta = -4 + 0,06 \cdot \text{ABWIG} - 0,75 \cdot \text{BEWEG} + 0,033 \cdot \text{DRUCK} + 1,1 \cdot \text{STRESS} + 1,55 \cdot \text{ERBE} + 4 \cdot \text{RAUCHER1} + 1 \cdot \text{RAUCHER2} + 1,5 \cdot \varepsilon$$

$$\text{RAUCHER1} = \begin{cases} 1, & \text{falls RAUCHEN} = 1 \\ 0, & \text{sonst} \end{cases}$$

$$\text{RAUCHER2} = \begin{cases} 1, & \text{falls RAUCHEN} = 2 \\ 0, & \text{sonst} \end{cases}$$

ε hat für jede beliebige Prädiktorwertekombination eine logistische Verteilung mit der Verteilungsfunktion:

$$F(w) = \frac{e^w}{1 + e^w}$$

Als Schwellenwert für den Übergang von der latenten Variablen η zur manifesten abhängigen Variablen DBS wird 1 verwendet:

$$\text{DBS} = \begin{cases} 0, & \text{falls } \eta \leq 1 \\ 1, & \text{falls } \eta > 1 \end{cases}$$

Nach den Überlegungen von Abschnitt 2.1.4 gilt in der künstlichen Population also ein BLR-Modell gemäß Gleichung (1).

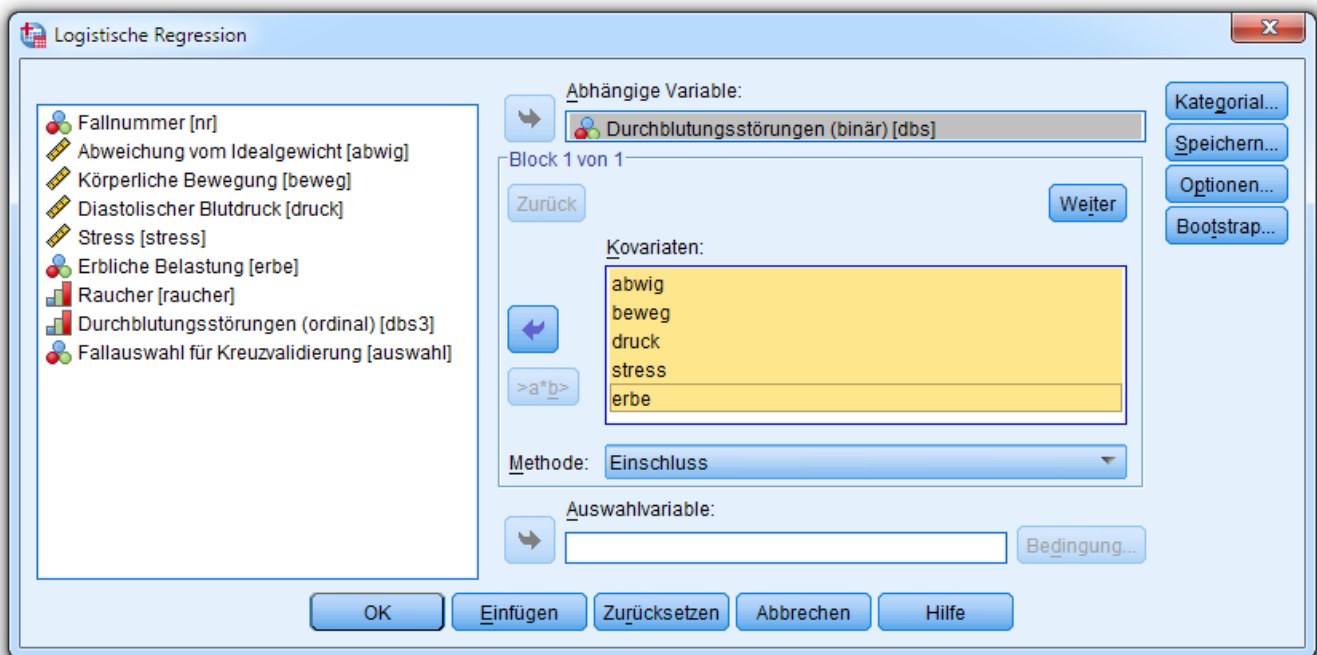
Sie finden die simulierten Daten in der Datei **DBS.SAV** an der im Vorwort vereinbarten Stelle.¹

Mit Hilfe der Datei **DBS.SAV** können wir in SPSS nach dem Menübefehl

Analyse > Regression > Binär logistisch

in folgender Dialogbox eine binäre logistische Regression anfordern:

¹ Medizinisch gebildete Leser mögen eventuelle erfahrungswidrige oder gar medizinisch ausgeschlossene Wertekonstellationen in der künstlichen Stichprobe nachsehen.

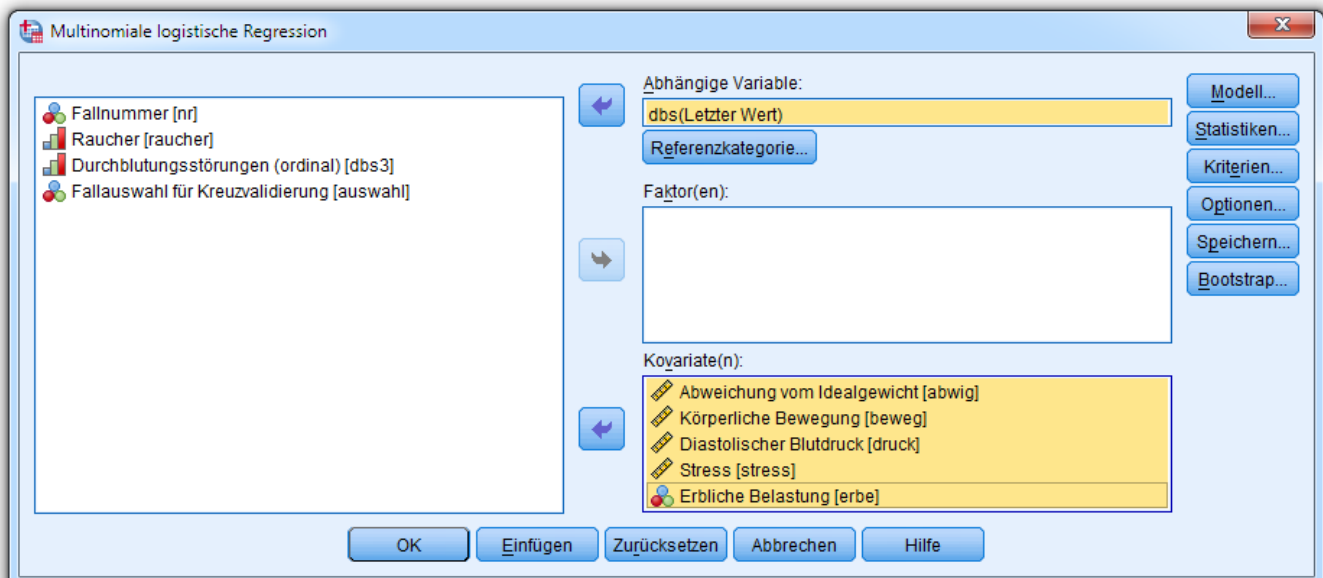


Aus didaktischen Gründen wird der Prädiktor RAUCHER bis zum Abschnitt 2.7 ignoriert.

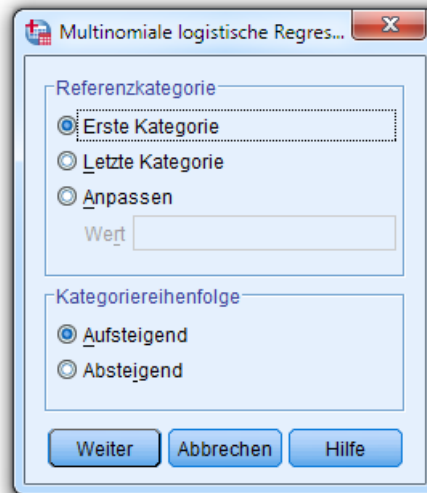
In unserer Demostudie treten kaum mehrfach besetzte Wertekombinationen auf, weil auch präzise erfasste metrische Regressoren mit zahlreichen Ausprägungen (z.B. ABWIG, DRUCK) zum Einsatz kommen. Daher werden wir primär mit der für Individualdaten konzipierten Prozedur LOGISTIC REGRESSION arbeiten. Für einige im Beispiel sinnvolle und zulässige Ergebnisse benötigen wir jedoch die für aggregierte Daten konzipierte und per Menübefehl

Analyse > Regression > Multinomial logistisch

erreichbare Prozedur NOMREG:

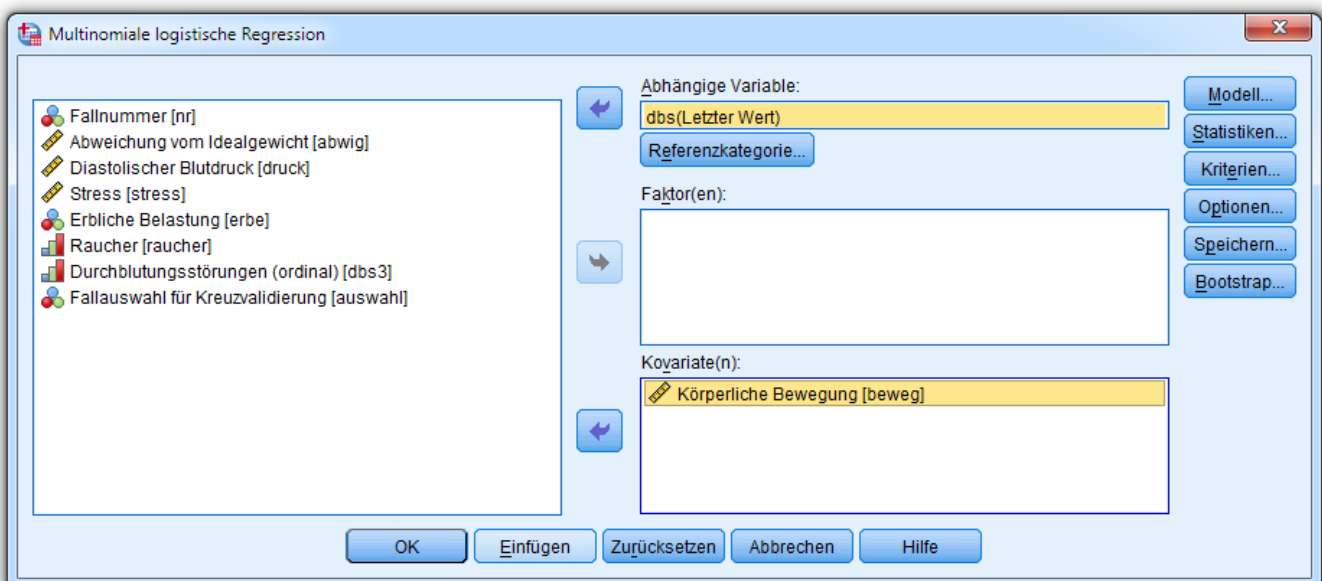


Um eine inhaltlich unbedeutende, bei der Interpretation jedoch leicht störende Invertierung der Vorzeichen bei den geschätzten Regressionskoeffizienten (im Vergleich zu den Ergebnissen von LOGISTIC REGRESSION) zu verhindern, wird im NOMREG-Aufruf bei der abhängigen Variablen die **Referenzkategorie** über den gleichnamigen Schalter geändert:



Warum auch der kategoriale Prädiktor ERBE bedenkenlos als **Kovariate** angemeldet werden darf, wird in Abschnitt 3.3 erläutert.

Zur Demonstration einiger methodischer Details (z.B. Modellgültigkeitstest über Pearsons *Goodness of Fit* – Statistik) betrachten wir ein reduziertes Modell mit mehrfach besetzten Prädiktorwertekombinationen, wobei dieselbe Variable DBS als Kriterium verwendet wird, im Design aber lediglich die Variable BEWEG verbleibt:



Weil der metrische Regressor BEWEG relativ grob gemessen ist (in 6 Stufen), resultiert ein Modell mit aggregierbaren Daten.

2.3 Schätzung der Parameter

2.3.1 Die Maximum-Likelihood-Methode

Während in der linearen Regressionsanalyse Kleinst-Quadrat-Schätzer Verwendung finden, welche die Summe der quadrierten Abweichungen zwischen den beobachteten und den vom Modell vorhergesagten Werten minimieren, kommt in der logistischen Regressionsanalyse die Maximum-Likelihood-Methode zum

Einsatz.¹ Hier werden Parameterschätzungen bestimmt, welche die Wahrscheinlichkeit der beobachteten Daten unter dem parametrisch spezifizierten Modell maximieren.

Die gemäß Modellgleichung (1) von den als fest gegeben anzunehmenden Regressorwerten des i -ten Falles (gesammelt im Vektor \mathbf{x}_i) und vom Parametervektor $\boldsymbol{\beta}$ abhängige wahre Wahrscheinlichkeit $P(Y_i = 1)$ werde mit π_i bezeichnet:

$$\pi_i := P(Y_i = 1) = \frac{e^{\boldsymbol{\beta}\mathbf{x}_i}}{1 + e^{\boldsymbol{\beta}\mathbf{x}_i}}, \text{ mit } \boldsymbol{\beta}\mathbf{x}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_M x_{iM}$$

Für eine konkrete Stichprobenrealisation (y_1, y_2, \dots, y_N mit $y_i \in \{0, 1\}, i = 1, \dots, N$) der Zufallsvariablen Y_i zu den N (unabhängigen!) Beobachtungen ergibt sich dann folgende Wahrscheinlichkeit:

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_N = y_N) = \prod_{i=1}^N \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

Beobachtungen mit realisierter 1 gehen mit π_i in das Produkt ein, die Beobachtungen mit realisierter 0 hingegen mit $(1 - \pi_i)$.

Wir ersetzen die unbekannten Parameter in $\boldsymbol{\beta}$ durch die frei schätzbaren Werte b_m ($m = 0, 1, \dots, M$) im Vektor \mathbf{b} und bezeichnen mit $L_i(\mathbf{b})$ die Likelihood für $(Y_i = 1)$ unter der Annahme $\boldsymbol{\beta} = \mathbf{b}$:

$$L_i(\mathbf{b}) := \frac{e^{\mathbf{b}\mathbf{x}_i}}{1 + e^{\mathbf{b}\mathbf{x}_i}}, \text{ mit } \mathbf{b}\mathbf{x}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_M x_{iM}$$

Für die gesamte Stichprobe ergibt sich dann die folgende **Likelihood-Funktion**:

$$L(\mathbf{b}) := \prod_{i=1}^N (L_i(\mathbf{b}))^{y_i} (1 - L_i(\mathbf{b}))^{1-y_i} \quad (6)$$

Das unmittelbar plausible Prinzip der Maximum-Likelihood-Schätzung besteht darin, denjenigen Vektor \mathbf{b} zu bestimmen, welcher die Likelihood-Funktion maximiert.

Um die Suche nach dem Maximum zu erleichtern, geht man zum *Logarithmus* über, was aufgrund der Monotonie dieser Funktion zulässig ist:

$$LL(\mathbf{b}) := \ln(L(\mathbf{b})) = \sum_{i=1}^N y_i \ln(L_i(\mathbf{b})) + (1 - y_i) \ln(1 - L_i(\mathbf{b})) \quad (7)$$

Aus dem Produkt in Gleichung (6) ist eine *Summe* geworden, was die Extremwertbestimmung erleichtert, die mit einem iterativen numerischen Verfahren (z.B. Newton-Raphson) vorgenommen wird.

Als Ergebnis erhält man den Vektor $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_M)$ mit dem ML-Schätzungen der Parameter. Daraus ergeben sich sofort die ML-Schätzungen $\hat{\pi}_i$ der Wahrscheinlichkeiten $P(Y_i = 1)$:

$$\hat{\pi}_i := \frac{e^{\hat{\boldsymbol{\beta}}\mathbf{x}_i}}{1 + e^{\hat{\boldsymbol{\beta}}\mathbf{x}_i}}$$

Mit dieser Vereinbarung können wir die Likelihood an der Stelle $\hat{\boldsymbol{\beta}}$ so schreiben:

$$L(\hat{\boldsymbol{\beta}}) = \prod_{i=1}^N \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1-y_i}$$

¹ Es zeigt sich übrigens, dass die Kleinst-Quadrat-Schätzer in der linearen Regressionsanalyse unter der üblichen Annahme normalverteilter Residuen auch Maximum-Likelihood-Schätzer sind.

Für das in Abschnitt 2.2 vorgestellte Beispiel (Durchblutungsstörungen) liefert LOGISTIC REGRESSION nach einer entsprechenden Aufforderung in der **Optionen**-Subdialogbox das folgende **Iteration-protokoll**:

		-2 Log-Likelihood	Koeffizienten					
			Konstante	abwig	beweg	druck	stress	erbe
Schritt 1	1	179,465	-1,452	,028	-,500	,022	,352	,886
	2	163,976	-2,257	,049	-,773	,034	,536	1,420
	3	161,847	-2,713	,061	-,915	,041	,630	1,701
	4	161,785	-2,809	,063	-,945	,042	,649	1,760
	5	161,785	-2,813	,063	-,946	,042	,650	1,762
	6	161,785	-2,813	,063	-,946	,042	,650	1,762

Nach lediglich 6 Iterationen stellt sich eine Konvergenz aller Parameterschätzungen auf stabile Werte ein, und das Verfahren stoppt.

Erfreulicherweise bietet die allgemeine Maximum-Likelihood-Theorie (siehe z.B. Rao 1973) einige für uns außerordentlich nützliche Ergebnisse:

- Über die Matrix der zweiten partiellen Ableitungen der Log-Likelihood-Funktion an der Stelle $\hat{\beta}$ gewinnt man eine Schätzung der Varianz-Kovarianzmatrix der Parameterschätzer, so dass sich Konfidenzintervalle und Tests konstruieren lassen (siehe z.B. Agresti 1990, S. 112ff; Hosmer & Lemeshow 2000, S. 34ff).
- Die mit -2 vormultiplizierte Log-Likelihood an der Stelle $\hat{\beta}$, im folgenden mit $-2LL(\hat{\beta})$ bezeichnet, spielt bei vielen Signifikanztests und Goodness-of-Fit - Indizes eine wichtige Rolle (siehe unten).

Allerdings gelten die Verteilungsaussagen der Maximum-Likelihood-Theorie generell nur *approximativ*, d.h. für $N \rightarrow \infty$. In Abschnitt 2.1.2 finden sich Empfehlungen zur minimal erforderlichen Stichprobengröße.

In Abschnitt 4.8 beschäftigen wir uns mit numerischen Schätzproblemen, die durch spezielle Muster in den Daten verursacht werden (z.B. Multikollinearität, leere Zellen bei kategorialen Prädiktoren).

2.3.2 Alternative Verfahren

Als Methode zur Parameterschätzung hat sich in der logistischen Regressionsanalyse die Maximum-Likelihood-Methode (ML-Methode) weitgehend durchgesetzt. Weil dies vor allem auch für aktuelle Statistik-Programmpakete wie SPSS gilt, werden alternative Verfahren in diesem Manuskript nicht behandelt (siehe z.B. Hosmer & Lemeshow 2000, S. 43ff).

Bei *kleinen* Stichproben (siehe Abschnitt 2.1.2) ist die Maximum-Likelihood-Technologie unbefriedigend, weil sie nur *approximative* Konfidenzintervalle zu Parameterschätzungen und Überschreitungswahrscheinlichkeiten zu Hypothesentests liefert. Während sich die meisten Statistikprogrammpakete (wie auch SPSS) bei der logistischen Regression auf ML-Methoden beschränken, liefert das Programm **LogXact** (siehe <http://www.cytel.com/>) exakte Ergebnisse über eine Generalisierung von Fischers exaktem Test für 4-Felder-Kontingenztabellen (Allison 1999, S. 47f). Außerdem liefert LogXact sinnvolle Ergebnisse für problematische Daten mit einer quasi-vollständigen Trennung (siehe Abschnitt 5.3), während die ML-Methode hier versagt. Wegen des enormen Rechenaufwands eignet sich LogXact nur für kleine Stichproben. In großen Stichproben sind die approximativen ML-Ergebnisse allerdings akzeptabel, und mit einer quasi-vollständigen Trennung ist kaum zu rechnen.

2.4 Beurteilung der Modellgültigkeit

Bevor die Schätz- und Testergebnisse zum gesamten Modell bzw. zu einzelnen Regressoren interpretiert werden, sollte zunächst die Modellgültigkeit anhand von diversen diagnostischen Informationen überprüft werden. Als potentielle Schwachstellen sind u.a. zu beachten:

- **Spezifikationsfehler**

Die Logit-Modellformulierung in Gleichung (2) kann sich als fehlerhaft erweisen. Eventuell sind hier nichtlineare oder multiplikative Beziehungen angemessener. Wie bei der linearen Regression sind auch bei der logistischen Regression die Parameterschätzungen verzerrt, wenn relevante Regressoren im Modell fehlen, die mit vorhandenen Regressoren korreliert sind (*omitted-variable-error*, siehe z.B. Baltes-Götz 1994, S. 1-3f). Eine Analyse dieses Problems setzt natürlich voraus, dass die potentiell relevanten Regressoren erfasst worden sind. Überflüssige Regressoren können die Vertrauensintervalle zu Regressionskoeffizienten und die β -Fehler von Signifikanztests vergrößern, speziell bei Korrelationen mit anderen Regressoren. Im Zusammenhang mit der Beurteilung einzelner Regressoren werden wir und mit fehlenden und mit irrelevanten Prädiktoren beschäftigen (siehe Abschnitt 2.6.3).

Aus der Entscheidung für die logistische Verteilungsfunktion zur Anpassung der Modellprognose an den Wertebereich von Wahrscheinlichkeiten resultieren in der Regel keine Einschränkungen für die Modellgültigkeit (vgl. Abschnitt 2.1.5). Wer Alternativen ausprobieren möchte, findet sie in den SPSS-Prozeduren PLUM und PROBIT.

- **Schwächen des Modells bei bestimmten Teilstichproben**

Bei einer Analyse der Residuen lassen sich Teilstichproben identifizieren, deren Verhalten vom Modell schlecht erklärt werden kann.

- **Einzelfälle mit starkem Einfluss auf die Schätzergebnisse**

Wenn einzelne Fälle die Schätzergebnisse stark beeinflussen, ist die Generalisierbarkeit in Frage gestellt.

2.4.1 Globale Modellgültigkeitstests

In der Literatur zur logistischen Regression sind einige **Goodness-of-Fit - Statistiken** vorgeschlagen worden, die global beurteilen sollen, wie gut ein geschätztes Modell zu den Daten passt. Dabei gibt man sich nicht mit deskriptiven Indizes zufrieden, sondern versucht zu Modellgültigkeitstests zu kommen. Bei der Auswahl einer Goodness-of-Fit – Statistik ist unbedingt zu berücksichtigen, ob Individualdaten oder aggregierte Daten vorliegen.

2.4.1.1 Pearson- χ^2 -Statistik

Über die folgende **Pearson- χ^2 -Statistik** kann bei *aggregierten* Daten mit K mehrfach besetzten Prädiktorwertekombinationen ein Modellgültigkeitstest konstruiert werden (siehe z.B. Hosmer & Lemeshow 2000, S. 145):

$$\chi_p^2 := \sum_{k=1}^K \frac{(\tilde{y}_k - h_k \hat{\pi}_k)^2}{h_k \hat{\pi}_k (1 - \hat{\pi}_k)} \quad (8)$$

Für eine mit Häufigkeit h_k realisierte Prädiktorwertekombination (engl.: *covariate pattern*) wird die erwartete Häufigkeit $h_k \hat{\pi}_k$ ermittelt, wobei $\hat{\pi}_k$ die vom geschätzten Modell für die k -te Wertekombination prognostizierte Wahrscheinlichkeit zum Einserereignis ist. Mit \tilde{y}_k soll bei aggregierten Daten die beobachtete Anzahl von Einsen in der k -ten Wertekombination ausgedrückt werden. Im Zähler eines χ_p^2 - Summanden steht also die quadrierte Abweichung der beobachteten Häufigkeit \tilde{y}_k von ihrer Erwartung

$h_k \hat{\pi}_k$ unter dem Modell. Im Nenner steht die geschätzte Varianz der $B(\hat{\pi}_k, h_k)$ - binomialverteilten Variablen \tilde{y}_k . Bei der Wurzel aus einem χ_p^2 - Summanden handelt es sich offenbar um ein standardisiertes Residuum nach klassischer Bauart, das anschließend als **Pearson-Residuum** bezeichnet und mit rp_k notiert werden soll. Insgesamt enthält χ_p^2 die Summe der quadrierten Pearson-Residuen zu den K Prädiktorwertkombinationen:

$$\chi_p^2 = \sum_{k=1}^K rp_k^2 \quad \text{mit} \quad rp_k := \frac{\tilde{y}_k - h_k \hat{\pi}_k}{\sqrt{h_k \hat{\pi}_k (1 - \hat{\pi}_k)}}$$

Alternativ kann man die χ_p^2 - Statistik auch als Diskrepanzmaß nach dem Muster:

$$\sum \frac{(\text{beobachtete Häufigkeit} - \text{erwartete Häufigkeit})^2}{\text{erwartete Häufigkeit}}$$

interpretieren, wobei über $(2 \cdot K)$ Zellen (für *beide* Kategorien des Kriteriums) zu summieren ist:

$$\begin{aligned} \sum_{k=1}^K \frac{(\tilde{y}_k - h_k \hat{\pi}_k)^2}{h_k \hat{\pi}_k} + \sum_{k=1}^K \frac{((h_k - \tilde{y}_k) - h_k (1 - \hat{\pi}_k))^2}{h_k (1 - \hat{\pi}_k)} &= \sum_{k=1}^K \frac{(\tilde{y}_k - h_k \hat{\pi}_k)^2 (1 - \hat{\pi}_k) + (h_k \hat{\pi}_k - \tilde{y}_k)^2 \hat{\pi}_k}{h_k \hat{\pi}_k (1 - \hat{\pi}_k)} \\ &= \sum_{k=1}^K \frac{(\tilde{y}_k - h_k \hat{\pi}_k)^2}{h_k \hat{\pi}_k (1 - \hat{\pi}_k)} \\ &= \chi_p^2 \end{aligned}$$

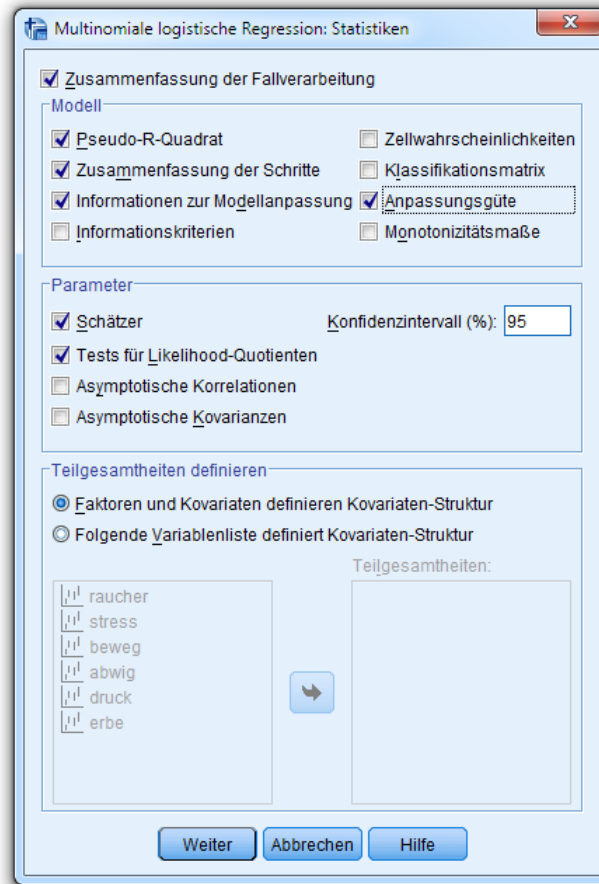
Pearsons- χ^2 -Statistik ist bei Gültigkeit des Modells χ^2 -verteilt, wenn alle erwarteten Häufigkeiten „groß genug“ sind (vgl. Abschnitt 2.1.2). Die Anzahl der Freiheitsgrade beträgt $K - (M + 1)$, wobei K für die Anzahl der Prädiktorwertkombinationen steht und M für die Anzahl der Designvariablen.¹

Ist die Verteilungsasymptotik akzeptabel, kann ein Modellgültigkeitstest auf Basis der Pearson- χ^2 -Statistik vorgenommen werden, der zwischen folgenden Hypothesen entscheidet:

- H_0 : Das logistische Modell gemäß Gleichung (1) ist gültig. Die Wahrscheinlichkeiten $P(Y_i = 1)$ genügen seinen Restriktionen.
- H_1 : Das logistische Modell ist ungültig.

Sinnvollerweise steht der Pearson- χ^2 -Statistik in der SPSS-Prozedur zur binären logischen Regression, die mit Individualdaten arbeitet, *nicht* zur Verfügung. Bei der multinomialen Variante kann er in der **Statistiken**-Subdialogbox über das Kontrollkästchen **Anpassungsgüte** angefordert werden:

¹ Ein nominalskaliertem Regressor mit s Ausprägungen wird im Design durch $(s - 1)$ Kodiervariablen repräsentiert (vgl. Abschnitt 2.7). Ein Modellgültigkeitstest ist nur dann möglich, wenn weniger als $K - 1$ Designvariablen vorhanden sind.



Mit dem in Abschnitt 2.2 beschriebenen Modell zur Erklärung von Durchblutungsstörungen kann *kein* Modellgültigkeitstest auf Basis der Pearson-Statistik durchgeführt werden, weil praktisch alle Regressor-Wertekombinationen nur einfach besetzt sind. Daher betrachten wir das reduzierte Modell mit BEWEG als einzigem (metrischem!) Regressor (vgl. Abschnitt 2.2). Trotzdem erhalten wir von NOMREG eine Warnung im Hinblick auf den angeforderten Modellgültigkeitstest:

Warnungen

Es gibt 1 (8,3%) Zellen (d.h. Niveaus der abhängigen Variablen für Teilgesamtheiten) mit der Häufigkeit Null.

Bei einer (BEWEG \times DBS) - Kombination liegt die *beobachtete* Häufigkeit 0 vor. Für die Anwendbarkeit des Modellgültigkeitstest via Pearsons- χ^2 -Statistik sind allerdings die *erwarteten* Häufigkeiten der (BEWEG \times DBS) - Kombinationen relevant (siehe unten). Folglich dürfen wir die Warnung ignorieren.

Wir erhalten eine Pearson-Statistik von 2,507, die bei 4 Freiheitsgraden eine Überschreitungswahrscheinlichkeit von 0,643 besitzt, und können das lineare Modell daher akzeptieren:

Güte der Anpassung

	Chi-Quadrat	Freiheitsgrade	Signifikanz
Pearson	2,507	4	,643
Abweichung	2,531	4	,639

Es liegen 4 (= 6 – 2) Freiheitsgrade vor, denn ...

- Es sind 6 Prädiktorwertekombinationen (die BEWEG-Ausprägungen) vorhanden ($K = 6$).
- Das Design enthält einen Prädiktor ($M = 1$).

Wer die Berechnung der Pearson- χ^2 -Statistik nachvollziehen und die erwarteten Häufigkeiten kontrollieren möchte, kann in der NOMREG-Subdialogbox **Statistik** über das Kontrollkästchen **Zellwahrscheinlichkeiten** die folgende Tabelle mit den beobachteten und prognostizierten Häufigkeit sowie den Pearson-Residuen anfordern:

Beobachtete und vorhergesagte Häufigkeiten

Körperliche Bewegung	Durchblutungsstörungen (binär)	Häufigkeit			Prozent	
		Beobachtet	Vorhergesagt	Pearson-Residuum	Beobachtet	Vorhergesagt
1	Nein	1	,314	1,266	20,0%	6,3%
	Ja	4	4,686	-1,266	80,0%	93,7%
2	Nein	4	4,239	-,126	14,8%	15,7%
	Ja	23	22,761	,126	85,2%	84,3%
3	Nein	22	21,850	,040	34,4%	34,1%
	Ja	42	42,150	-,040	65,6%	65,9%
4	Nein	39	40,754	-,430	56,5%	59,1%
	Ja	30	28,246	,430	43,5%	40,9%
5	Nein	23	22,418	,275	82,1%	80,1%
	Ja	5	5,582	-,275	17,9%	19,9%
6	Nein	7	6,425	,791	100,0%	91,8%
	Ja	0	,575	-,791	,0%	8,2%

Prozentwerte basieren auf der beobachteten Gesamthäufigkeit in jeder Teilgrundgesamtheit.

Die Summe der quadrierten Pearson-Residuen zu den 6 Prädiktorwertkombinationen ergibt den Indexwert:

$$1,266^2 + 0,126^2 + 0,040^2 + 0,430^2 + 0,275^2 + 0,791^2 \approx 2,51$$

Nach Formel (8) haben wir die K quadrierten Pearson-Residuen zu den Einserkategorien zu addieren. Dass die Pearson-Residuen zu den Null-Kategorien gerade den negativen Wert $-rp_k$ annehmen, ergibt sich sofort:

$$\frac{(h_k - \tilde{y}_k) - h_k(1 - \hat{\pi}_k)}{\sqrt{h_k \hat{\pi}_k (1 - \hat{\pi}_k)}} = \frac{-\tilde{y}_k + h_k \hat{\pi}_k}{\sqrt{h_k \hat{\pi}_k (1 - \hat{\pi}_k)}} = -rp_k$$

In zwei Zellen (BEWEG = 1, DBS = 0 bzw. BEWEG = 6, DBS = 1) ist die erwartete Häufigkeit sehr klein (0,314 bzw. 0,575), so dass die Interpretierbarkeit des Modellgültigkeitstests nach der Empfehlung aus Abschnitt 2.1.2 etwas fraglich ist.

2.4.1.2 Devianz-Statistik

Für aggregierte Daten kann auch die **Devianz-Statistik** als GFI-Index (*Goodness of Fit*) und Basis für einen Modellgültigkeitstest verwendet werden. Hier vergleicht man die maximale Likelihood der Daten unter dem betrachteten Modell mit der Likelihood L_S des *saturierten Modells*, das für jede Prädiktorwertkombination eine eigene Einserwahrscheinlichkeit erlaubt, welche über die relative Häufigkeit geschätzt wird. Mit den Bezeichnungen wie in den Abschnitten 2.3.1 und 2.4.1.1 gilt für die Likelihood des parametrisch spezifizierten Modells:

$$L(\hat{\beta}) = \prod_{i=1}^N \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1-y_i} = \prod_{k=1}^K \hat{\pi}_k^{\tilde{y}_k} (1 - \hat{\pi}_k)^{h_k - \tilde{y}_k}$$

Die Likelihood L_S des saturierten Modells ist definiert durch:

$$L_S := \prod_{k=1}^K \left(\frac{\tilde{y}_k}{h_k} \right)^{\tilde{y}_k} \left(\frac{h_k - \tilde{y}_k}{h_k} \right)^{h_k - \tilde{y}_k}$$

Die folgendermaßen definierte **Devianz D**:

$$D := -2 \ln \left(\frac{L(\hat{\beta})}{L_S} \right)$$

ist bei Gültigkeit des Modells approximativ χ^2 -verteilt mit $K - (M + 1)$ Freiheitsgraden. Dies folgt aus generellen Theoremen über Likelihood-Quotienten, die wir noch mehrfach ausnutzen werden (vgl. Abschnitt 2.5.1).

Wie die in Abschnitt 2.4.1.1 enthaltene SPSS-Tabelle **Güte der Anpassung** zeigt, erreicht das DBS-Partialmodell mit dem einzigen Regressor BEWEG eine Devianz (deutsch: *Abweichung*) von 2,531, so dass bei 4 Freiheitsgraden eine Überschreitungswahrscheinlichkeit von 0,639 resultiert. Folglich kann die Nullhypothese (also das Modell) beibehalten werden.

Auch der Modellgültigkeitstest per Devianz-Statistik ist nur bei aggregierten Daten mit mehrfach besetzten Regressor-Wertekombinationen anwendbar, hat also dieselben Voraussetzungen wie der im letzten Abschnitt beschriebene Pearson-Anpassungstest. Nach den Ergebnissen von Stelzl (2000) ist die Pearson-Variante in der Regel aufgrund der besseren Verteilungsapproximation zu bevorzugen.

Beide GFI-Statistiken bzw. Modellgültigkeitstests können in der **Statistiken**-Subdialogbox der SPSS-Prozedur NOMREG über das Kontrollkästchen **Anpassungsgüte** angefordert werden (siehe Abschnitt 2.4.1.1).

2.4.1.3 Hosmer-Lemeshow - Statistik

Für Individualdaten mit überwiegend nur einfach besetzten Regressor-Wertekombinationen schlagen Hosmer & Lemeshow (2000, S. 147f) vor, aufgrund der prognostizierten Wahrscheinlichkeiten L (z.B. 10) annähernd gleich stark besetzte Gruppen zu bilden und dann analog zur **Pearson- χ^2 -Statistik** erwartete und beobachtete Häufigkeiten zu vergleichen. Wenn mit $\bar{\pi}_l$ das arithmetische Mittel der prognostizierten Wahrscheinlichkeiten zu allen Fällen in der l -ten Gruppe bezeichnet wird, lässt sich die **Hosmer-Lemeshow – Goodness-of-Fit – Statistik** so notieren:

$$\chi_{HL}^2 := \sum_{l=1}^L \frac{(\tilde{y}_l - h_l \bar{\pi}_l)^2}{h_l \bar{\pi}_l (1 - \bar{\pi}_l)}$$

Analog zu Formel (8) steht h_l für die Anzahl der Fälle in Gruppe l und \tilde{y}_l für die hier beobachtete Anzahl von Einsen, wobei jedoch die Gruppen diesmal nicht über Prädiktorwertekombinationen festgelegt sind, sondern über Intervalle bzgl. der prognostizierten Wahrscheinlichkeiten.

Analog zu der in Abschnitt 2.4.1.1 vorgeführten Umstellung kann man auch χ_{HL}^2 so notieren, dass Diskrepanzbeiträge aus $2 \cdot L$ Zellen zu addieren sind:

$$\sum_{l=1}^L \frac{(\tilde{y}_l - h_l \bar{\pi}_l)^2}{h_l \bar{\pi}_l (1 - \bar{\pi}_l)} = \sum_{l=1}^L \frac{(\tilde{y}_l - h_l \bar{\pi}_l)^2}{h_l \bar{\pi}_l} + \sum_{l=1}^L \frac{((h_l - \tilde{y}_l) - h_l (1 - \bar{\pi}_l))^2}{h_l (1 - \bar{\pi}_l)}$$

Bei unserem Durchblutungsbeispiel mit dem in Abschnitt 2.2 beschriebenen Prädiktorensatz (jedoch noch ohne RAUCHER) werden die 200 Fälle nach der prognostizierten Einserswahrscheinlichkeit in 10 exakt gleich stark besetzte Gruppen aufgeteilt, alle h_l sind also gleich 20.

Kontingenztafel für Hosmer-Lemeshow-Test

		Durchblutungsstörungen (binär) = Nein		Durchblutungsstörungen (binär) = Ja		Gesamt
		Beobachtet	Erwartet	Beobachtet	Erwartet	
Schritt 1	1	19	19,657	1	,343	20
	2	18	18,218	2	1,782	20
	3	16	16,678	4	3,322	20
	4	18	13,839	2	6,161	20
	5	9	10,395	11	9,605	20
	6	8	7,829	12	12,171	20
	7	4	5,166	16	14,834	20
	8	3	2,765	17	17,235	20
	9	1	1,194	19	18,806	20
	10	0	,258	20	19,742	20

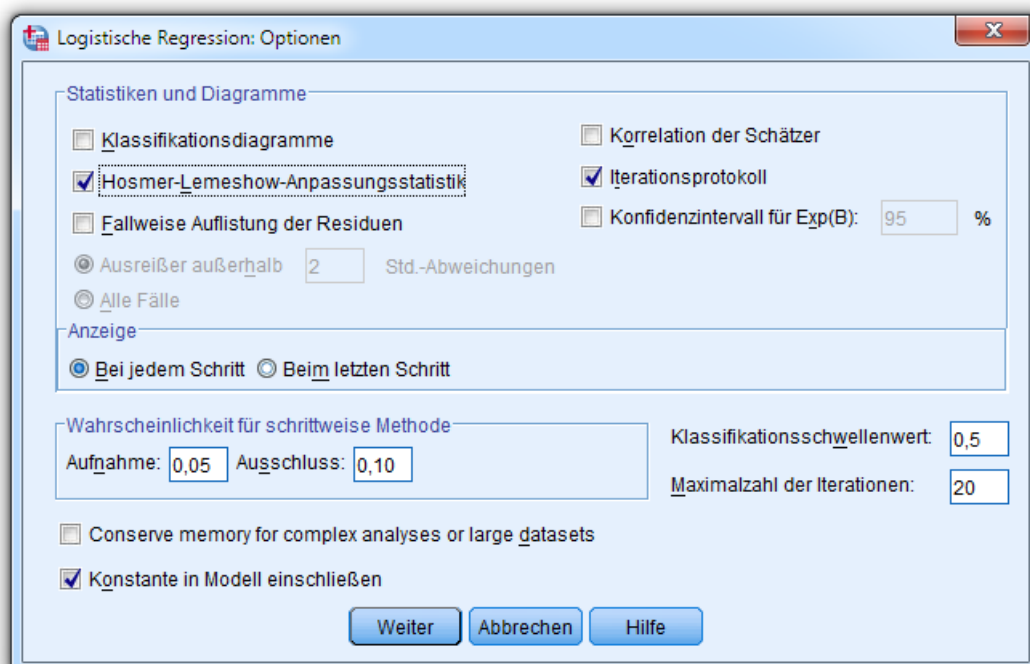
Durch Kreuzen dieser Aufteilungsvariablen mit dem dichotomen Kriterium entstehen 10×2 Zellen mit folgenden erwarteten Häufigkeiten:

- $h_i \pi_i$ für die Ja-Varianten
- $h_i(1 - \pi_i)$ für die Nein-Varianten

In jeder Zelle wird die Differenz *Beobachtung* – *Erwartung* quadriert und durch die Erwartung dividiert. Über alle 20 Zellen summiert ergibt sich die Hosmer-Lemeshow – Goodness-of-Fit – Statistik.

Mit Hilfe von Simulationen haben Hosmer & Lemeshow ermittelt, dass χ^2_{HL} bei korrektem Modell einer χ^2 -Verteilung mit $L - 2$ Freiheitsgraden folgt, sofern alle Prädiktorwertkombinationen nur einfach besetzt sind. Die Autoren vermuten, dass die Approximation auch dann noch verlässlich ist, wenn nur wenige Kombinationen mehrfach auftreten.

SPSS bietet den Hosmer-Lemeshow – Modelgültigkeitstest in der Prozedur zur binären logistischen Regression, anzufordern in der **Optionen**-Subdialogbox:



Im DBS-Beispiel liefert der Test zur χ^2_{HL} -Statistik keinen Anlass, die Modellgültigkeit zu bezweifeln:

Hosmer-Lemeshow-Test

Schritt	Chi-Quadrat	df	Sig.
1	6,609	8	,579

2.4.2 Untersuchung von Residuen und Einflussindikatoren

2.4.2.1 Residuendiagnostik

Anhand von Residuen kann man lokale Anpassungsschwächen eines Modells aufspüren. Leider ist die Residuendiagnostik bei der logistischen Regression mit einigen Schwierigkeiten belastet (z.B. im Vergleich zur linearen Regressionsanalyse). Insbesondere gelten die Individualresiduen als problematisch, so dass meist empfohlen wird, die Residuen für Prädiktorwertekombinationen (covariate patterns) zu betrachten (siehe z.B. Hosmer & Lemeshow 2000, S. 170; Ryan 1997, S. 284). Zudem sollten die einzelnen Wertekombinationen mit adäquater Häufigkeit in der Stichprobe vertreten sein, was in Abhängigkeit von der Stichprobengröße, Anzahl der Regressoren und Messgenauigkeit eventuell kaum zu realisieren ist. In unserer Stichprobe ($N = 200$) treten durch die relativ genau gemessenen Regressoren ABWIG (in kg) und DRUCK (in mm/Hg) kaum mehrfach besetzte Wertekombinationen auf.

Ein technisches Problem besteht zudem darin, dass SPSS nur Individualresiduen produziert, so dass man die Residuen zu Prädiktorwertekombinationen etwas aufwändig (z.B. per Aggregation, siehe unten) ermitteln muss. Fallbezogen kann SPSS etliche Residualvarianten erzeugen, wobei die BLR-Prozedur in der **Speichern**-Subdialogbox entsprechende Wünsche entgegen nimmt:



In diesem Manuskript werden nur die Pearson- und die Devianz-Residuen berücksichtigt, die in enger Beziehung zu bereits diskutierten Goodness-of-Fit – Statistiken stehen (siehe Abschnitt 2.4.1) und in obiger SPSS-Dialogbox unter den Bezeichnungen **Standardisiert** bzw. **Abweichung** zu finden sind.

2.4.2.1.1 Devianz-Residuen

Die in Abschnitt 2.4.1.2 vorgestellte Devianz-Statistik lässt sich als Summe von Beiträgen der K Prädiktorwertekombinationen schreiben, sofern alle Einserhäufigkeiten \tilde{y}_k größer 0 und kleiner als h_k sind:

$$\begin{aligned}
 D &= -2 \ln \left(\frac{L(\hat{\beta})}{L_s} \right) = 2 \ln \left(\frac{L_s}{L(\hat{\beta})} \right) \\
 &= 2 \ln \left(\prod_{k=1}^K \left(\frac{\tilde{y}_k}{\hat{\pi}_k h_k} \right)^{\tilde{y}_k} \left(\frac{h_k - \tilde{y}_k}{h_k (1 - \hat{\pi}_k)} \right)^{h_k - \tilde{y}_k} \right) \\
 &= \sum_{k=1}^K \left[2 \tilde{y}_k \ln \left(\frac{\tilde{y}_k}{\hat{\pi}_k h_k} \right) + 2 (h_k - \tilde{y}_k) \ln \left(\frac{h_k - \tilde{y}_k}{h_k (1 - \hat{\pi}_k)} \right) \right] \quad \text{falls } 0 < \tilde{y}_k < h_k
 \end{aligned}$$

Wie in Abschnitt 2.4.1.1 bedeuten (jeweils für die k -te Prädiktorwertekombination):

- h_k Gruppenstärke
- $\hat{\pi}_k$ vom Modell prognostizierte Einserwahrscheinlichkeit
- \tilde{y}_k beobachtete Anzahl von Einsen

Die Wurzel des k -ten Summanden, versehen mit dem Vorzeichen von $(\tilde{y}_k - h_k \hat{\pi}_k)$, bezeichnet man als **Devianzresiduum** rd_k zur k -ten Prädiktorwertekombination (siehe z.B. Hosmer & Lemeshow 2000, S. 146). Es gilt trivialerweise:

$$D = \sum_{k=1}^K rd_k^2$$

Bei der Definition des Devianzresiduums rd_i für einen *Einzelfall* muss man die Gruppenformel modifizieren, um das ungültige Logarithmus-Argument 0 zu vermeiden:

$$rd_i := \begin{cases} -\sqrt{2|\ln(1 - \hat{\pi}_i)|} & \text{für } y_i = 0 \\ \sqrt{2|\ln(\hat{\pi}_i)|} & \text{für } y_i = 1 \end{cases}$$

Mit dieser Definition arbeitet die BLR-Prozedur in SPSS, wenn **Abweichungs**-Residuen angefordert werden.

Die Summe der quadrierten Devianz-Individualresiduen ergibt übrigens gerade $-2LL(\hat{\beta})$ (vgl. Abschnitt 2.3.1):

$$\sum_{i=1}^N rd_i^2 = -2LL(\hat{\beta})$$

Aus diesem Grund wird die Größe $-2LL(\hat{\beta})$ aus der logistischen Regression oft mit der Fehlerquadratsumme aus der linearen Regression verglichen (siehe Abschnitt 2.5.1).

Bei Individualdaten korrelieren die Devianzresiduen hoch mit den Pearson-Residuen (z.B. 0,91 im kompletten DBS-Modell) und zeigen daher im Wesentlichen dasselbe Verhalten. Daher können wir uns anschließend auf eine Diskussion der Pearson-Residuen beschränken.

2.4.2.1.2 Pearson-Residuen

Das **Pearson-Residuum** für die k -te Prädiktorwertekombination ist definiert durch:

$$rp_k := \frac{\tilde{y}_k - h_k \hat{\pi}_k}{\sqrt{h_k \hat{\pi}_k (1 - \hat{\pi}_k)}} \quad (9)$$

Folglich ist die in Abschnitt 2.4.1.1 vorgestellte Pearson- χ^2 -Statistik gerade die Summe der quadrierten Pearson-Residuen zu allen Prädiktorwertekombinationen.

Das Pearson-Residuum für einen *Einzelfall* ergibt sich aus Formel 9 als Spezialfall mit Häufigkeit 1:

$$rp_i := \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i (1 - \hat{\pi}_i)}}$$

Es wird in SPSS als **standardisiertes Residuum** bezeichnet und enthält ...

- im Zähler die $B(\hat{\pi}_i, 1)$ – binomialverteilte Zufallsvariable y_i , die durch Subtrahieren ihrer modelgemäßen Erwartung $\hat{\pi}_i$ zentriert wird
- im Nenner die geschätzte Standardabweichung von y_i

Eine analoge Standardisierung ist auch in der Gruppenformel enthalten, so dass alle Pearson-Residuen bei gültigem Modell approximativ den Erwartungswert 0 und die Varianz 1 haben.

2.4.2.1.3 Untersuchung von Individualresiduen

Als Beleg für die oben erwähnte Überlegenheit der Gruppenresiduen wollen wir betrachten, wie sich die beiden Varianten des Pearson-Residuums verhalten, wenn alle Fälle in einer Prädiktorwertekombination den selben Kriteriumswert 0 erreichen (vgl. Hosmer & Lemeshow 2000, S. 170). Für die Individualresiduen gilt dann:

$$rp_i = \frac{-\hat{\pi}_i}{\sqrt{\hat{\pi}_i (1 - \hat{\pi}_i)}} = -\sqrt{\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}}$$

Das Residuum zur Gruppe ist betragsmäßig um den Faktor $\sqrt{h_k}$ größer:¹

$$rp_k = \frac{-h_k \hat{\pi}_k}{\sqrt{h_k \hat{\pi}_k (1 - \hat{\pi}_k)}} = -\sqrt{h_k} \sqrt{\frac{\hat{\pi}_k}{1 - \hat{\pi}_k}}$$

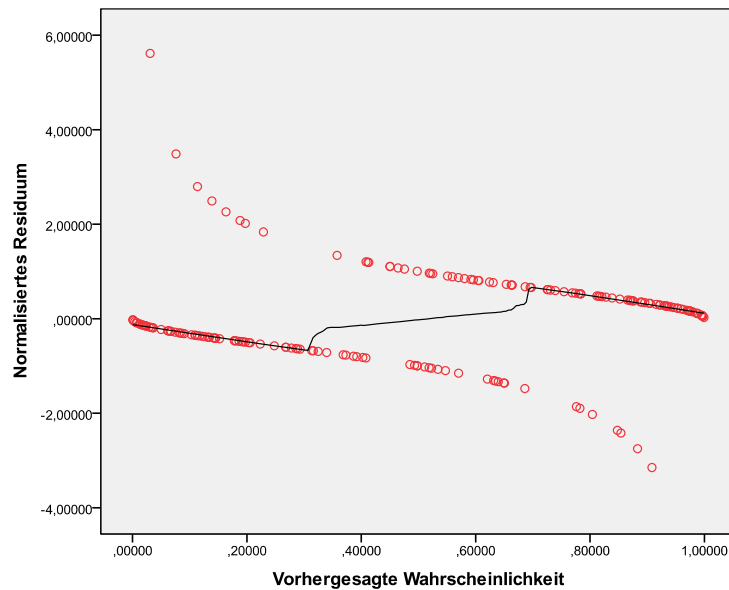
Wenn z.B. bei einer prognostizierten Wahrscheinlichkeit von 0,5 alle 9 Fälle einer Prädiktorwertekombination zur Nullkategorie des Kriteriums gehören, kann man von einem lokalen Anpassungsproblem des Modells ausgehen. Während dieses Problem im Gruppenresiduum von -3 deutlich zum Ausdruck kommt (3 Standardabweichungen vom Mittelwert entfernt), zeigen alle 9 Individual-Residuen den unauffälligen Wert -1.

Für die von LOGISTIC REGRESSION aufgrund der obigen **Speichern**-Subdialogbox (siehe Abschnitt 2.4.2.1) in der Arbeitsdatei erzeugten Variablen PRE_1 und ZRE_1 mit den Modellprognosen bzw. Pearson-Individualresiduen zum DBS-Modell (ohne den Prädiktor RAUCHER) kann man z.B. über den folgenden Menübefehl

Diagramme > Veraltete Dialogfelder > Streu-/Punkt-Diagramm > Einfach

¹ Das Pearson-Gruppenresiduum ist generell identisch mit dem $\sqrt{h_k}$ - fachen des mittleren Individualresiduums.

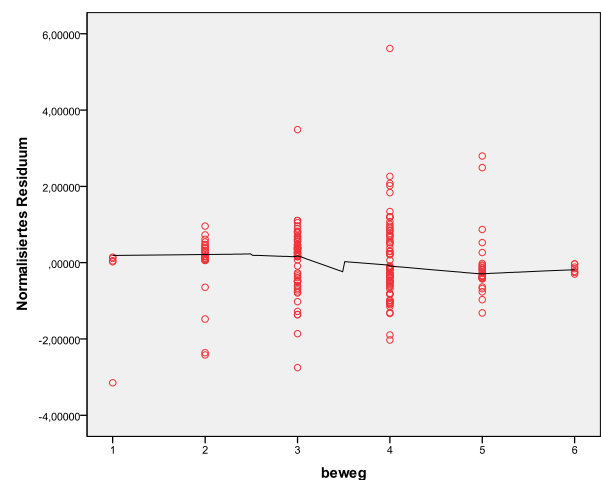
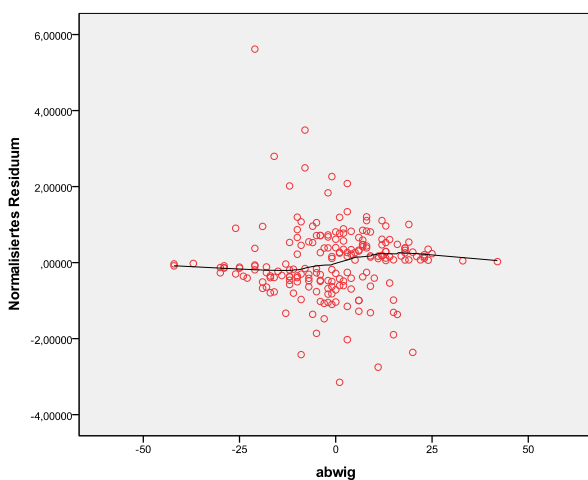
ein Streudiagramm anfordern. Es unterscheidet sich stark von analogen Diagrammen aus einer linearen Regressionsanalyse:

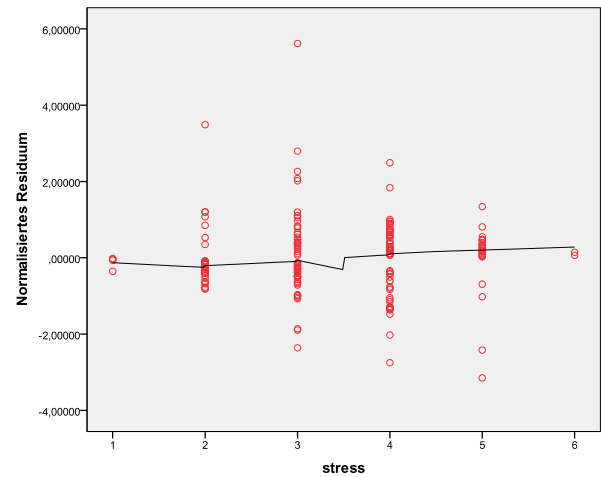
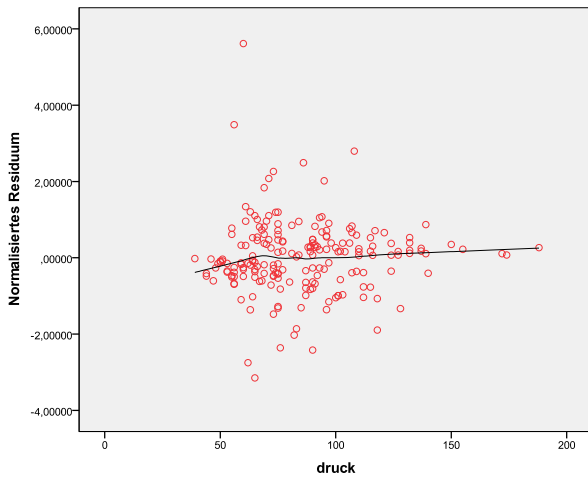


Das prägnante Muster entsteht, weil für alle Fälle mit einer bestimmten prognostizierten Wahrscheinlichkeit nur zwei Beobachtungswerte (0 oder 1) und infolgedessen auch nur zwei Residualwerte auftreten können.

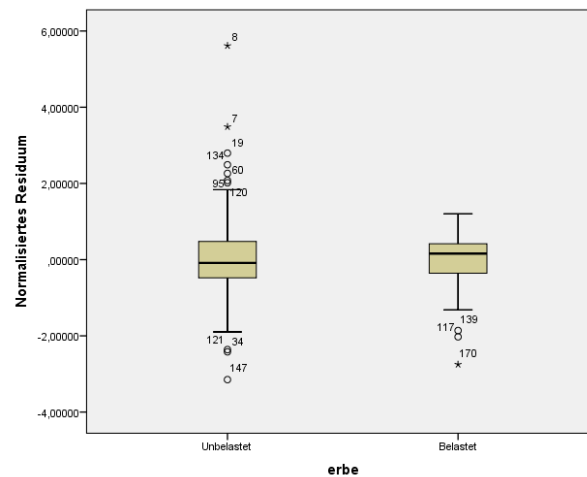
Trotz dieser Besonderheiten müssen auch die individuellen Pearson-Residuen aus der binären logistischen Regression bei gültigem Modell für jede Modellprognose einen Mittelwert von Null besitzen, d.h. die Funktion der bedingten Erwartungen muss annähernd konstant in der Höhe Null verlaufen (Fox & Weisberg 2011, S. 320). Um diese Forderung graphisch beurteilen zu können, wurde eine **Anpassungslinie** eingezeichnet und durch Zuweisung der **Anpassungsmethode Loess** (*locally estimated scatterplot smoothing*) stückweise optimal an die Daten angepasst. Diese Anpassungslinie zeigt keine nennenswerten Abweichungen von der erwarteten Geraden.

Auch in den Streudiagrammen der Pearson-Individualresiduen mit jeweils einem einzelnen metrischen Prädiktor sollte die **Loess**-Anpassungslinie nicht wesentlich vom waagerechten Verlauf in der Höhe Null abweichen. Wir erhalten die folgenden Diagramme:





Beim kategorialen Prädiktor ERBE ist ein Boxplot geeignet, um das Verhalten der Pearson-Residuen zu beschreiben:



Wer bei den metrischen Prädiktoren einen Signifikanztest zur Beurteilung auf Anpassungsschwächen des BLR-Modells bevorzugt, kann über den Menübefehl

Analysieren > Regression > Kurvenanpassung

für die Regression der Pearson-Residuen auf den fraglichen Prädiktor ein lineares, quadratisches und kubisches Modell erproben, z.B.:

Modellzusammenfassung und Parameterschätzer

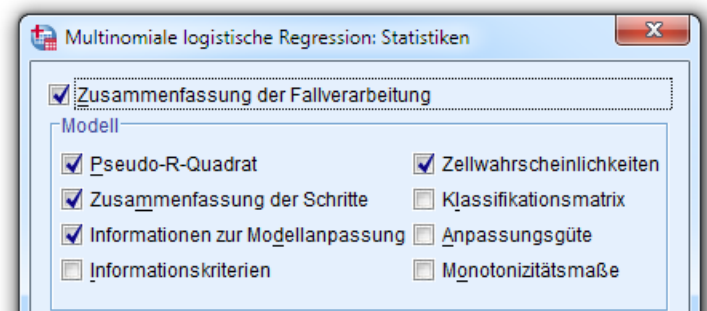
Abhängige Variable: Normalisiertes Residuum

Gleichung	Modellzusammenfassung					Parameterschätzer			
	R-Quadrat	F	Freiheitsgrad e 1	Freiheitsgrad e 2	Sig.	Konstante	b1	b2	b3
Linear	,000	,001	1	198	,977	,022	,000		
Quadratisch	,000	,010	2	197	,990	,016	1,772E-5	3,298E-5	
Kubisch	,000	,020	3	196	,996	,015	-,001	4,596E-5	1,868E-6

Die unabhängige Variable ist abwig.

2.4.2.1.4 Untersuchung von Gruppenresiduen

Zum reduzierten DBS-Modell (mit BEWEG als einzigem Prädiktor) kann die Prozedur NOMREG die Pearson-Residuen der Gruppen in einer Tabelle ausgeben (wählbar über das Kontrollkästchen **Zellwahrscheinlichkeiten** in der **Statistiken**-Subdialogbox):



Bei hinreichend stark besetzten Prädiktorwertekombinationen folgen die standardisierten Gruppenresiduen eines gültigen Modells approximativ einer Standardnormalverteilung, so dass man durch Vergleich mit den Perzentilen dieser Verteilung lokale Anpassungsdefizite aufdecken kann. Im Beispiel sind die Beträge aller Residuen deutlich kleiner als das 97,5 - Perzentil der Standardnormalverteilung ($\approx 1,96$):

Beobachtete und vorhergesagte Häufigkeiten

Körperliche Bewegung	Durchblutungsstörungen (binär)	Häufigkeit			Prozent	
		Beobachtet	Vorhergesagt	Pearson-Residuum	Beobachtet	Vorhergesagt
1	Nein	1	,314	1,266	20,0%	6,3%
	Ja	4	4,686	-1,266	80,0%	93,7%
2	Nein	4	4,239	-,126	14,8%	15,7%
	Ja	23	22,761	,126	85,2%	84,3%
3	Nein	22	21,850	,040	34,4%	34,1%
	Ja	42	42,150	-,040	65,6%	65,9%
4	Nein	39	40,754	-,430	56,5%	59,1%
	Ja	30	28,246	,430	43,5%	40,9%
5	Nein	23	22,418	,275	82,1%	80,1%
	Ja	5	5,582	-,275	17,9%	19,9%
6	Nein	7	6,425	,791	100,0%	91,8%
	Ja	0	,575	-,791	,0%	8,2%

Prozentwerte basieren auf der beobachteten Gesamthäufigkeit in jeder Teilgrundgesamtheit.

Mit etwas Handarbeit lässt sich im Beispiel auch ein Streudiagramm mit den Pearson-Residuen und den mittleren Prognosewerten der BEWEG-Gruppen erstellen:¹

- Man lässt die geschätzten Zugehörigkeitswahrscheinlichkeiten zu den Kriteriumskategorien in neue Variablen der Arbeitsdatei sichern, z.B. über das Kontrollkästchen **Geschätzte Wahrscheinlichkeiten für abhängige Variable** in der NOMREG-Subdialogbox **Speichern**.

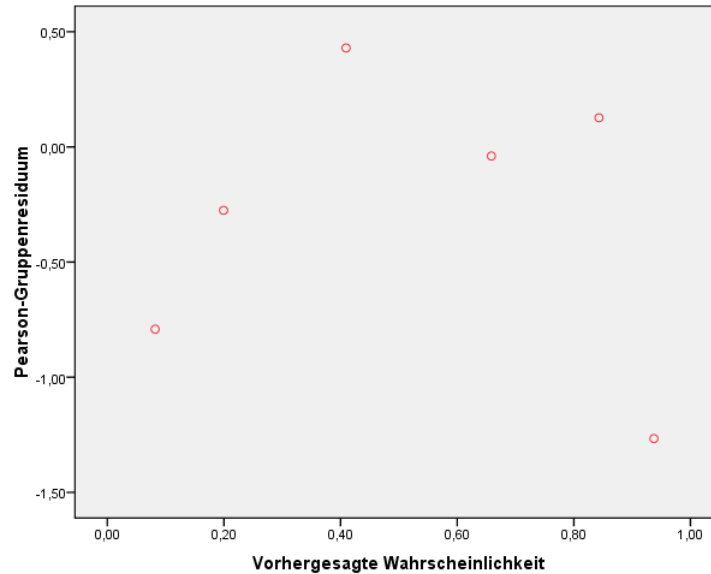
¹ Die vorgeschlagene Schrittfolge lässt sich mit der folgenden SPSS-Syntax automatisieren:

```
NOMREG
  dbs WITH beweg /SAVE ESTPROB.
AGGREGATE
  /OUTFILE=* /BREAK=beweg
  /pred = MEAN(est2_1) /y = SUM(dbs) /h=N.
COMPUTE resid = (y - pred * h)/sqrt(h*pred*(1-pred)).
GRAPH
  /SCATTERPLOT(BIVAR)=pred WITH resid.
```

Beachten Sie bitte, dass die von NOMREG erstellten Variablen mit den Zellwahrscheinlichkeiten (im Programm benötigt: EST2_1) bei jedem Aufruf innerhalb einer Sitzung neue Namen erhalten.

- Man aggregiert die Daten mit der **Break**-Variablen BEWEG.
- Man ermittelt die Pearson-Residuen per COMPUTE-Kommando gemäß Formel (9).
- Man erstellt ein Streudiagramm.

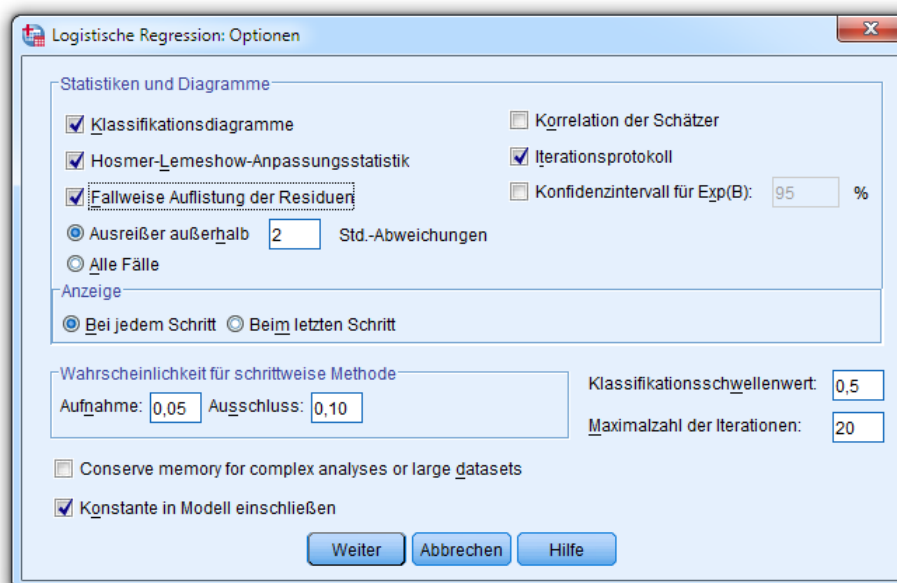
Erwartungsgemäß zeigen sich im Diagramm keine prägnanten Abweichungen vom erwünschten Muster einer diffusen Streuung der Pearson-Residuen um die Parallele zur X-Achse mit Ordinatenabschnitt 0:



2.4.2.1.5 Ausreißer-Diagnose über Individualresiduen

Trotz der oben diskutierten Schwächen von Individualresiduen kann es sich lohnen, die extremen Exemplare dieser Gattung zu inspizieren. So lernt man z.B. Fälle kennen, bei denen trotz sehr hoher prognostizierter Wahrscheinlichkeit das Einserereignis *nicht* aufgetreten ist. Es ist dann zu prüfen, ob eine Modellschwäche oder eine Besonderheit bei Einzelfällen vorliegt.

Die BLR-Prozedur von SPSS bietet in der **Optionen**-Subdialogbox die **Fallweise Auflistung der Residuen** an, wobei man sich auf **Ausreißer** (betragsmäßig extreme Werte) beschränken kann:



In unserem DBS-Beispiel (mit allen Prädiktoren außer RAUCHER) resultiert die folgende Tabelle:

Fallweise Liste^b

Fall	Ausgewählter Status ^a	Beobachtet	Vorhergesagt	Vorhergesagte Gruppe	Temporäre Variable	
		dbs			Resid	ZResid
7	S	J**	,076	N	,924	3,487
8	S	J**	,031	N	,969	5,614
19	S	J**	,113	N	,887	2,795
34	S	N**	,854	J	-,854	-2,420
134	S	J**	,139	N	,861	2,491
147	S	N**	,908	J	-,908	-3,149
170	S	N**	,883	J	-,883	-2,750

a. S = Ausgewählte, U = Nicht ausgewählte Fälle und ** = Falsch klassifizierte Fälle.

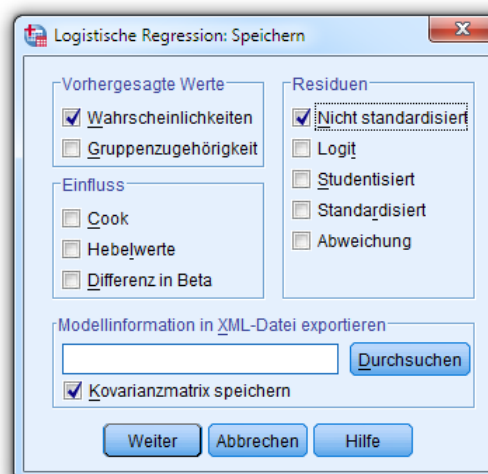
b. Fälle mit studentisierten Residuen größer als 2,000 werden aufgelistet.

Die Anzahl der extremen Werte bewegt sich im „normalen“ Rahmen: Von 200 Fällen erreichen nur 7 ein standardisiertes Residuum im zweiseitigen 5%-Ablehnungsbereich der Standardnormalverteilung (Betrag > 1,96). Folglich besteht kein Anlass, einzelne Fälle als Ausreißer auszuschließen. Hier handelt es sich zugegebenermaßen um eine unsaubere Argumentation, weil nach obigen Überlegungen die Normalverteilungstheorie zur Beurteilung der standardisierten *Individualresiduen* kaum genutzt werden kann.

Beim Fall Nr. 8 mit dem extremen Pearson-Residuum von 5,614 zeigt sich die „Ungerechtigkeit“ des Pseudozufallszahlengenerators. Der Fall ist von Durchblutungsstörungen betroffen, obwohl seine Prädiktorwerte für ein sehr geringes Risiko von 0,03 sprechen: 21 kg Untergewicht, überdurchschnittliche Bewegung (Wert 4 von 6), sehr niedriger Blutdruck von 60 mm/Hg, durchschnittlicher Stress (3 von 6), keine erbliche Belastung. Es zeigt sich, dass auch bei einem gültigen Modell durch puren Zufall ein extremes Residuum auftreten kann.

2.4.2.1.6 Trennwert-Optimierung per Residuen-Diagramm

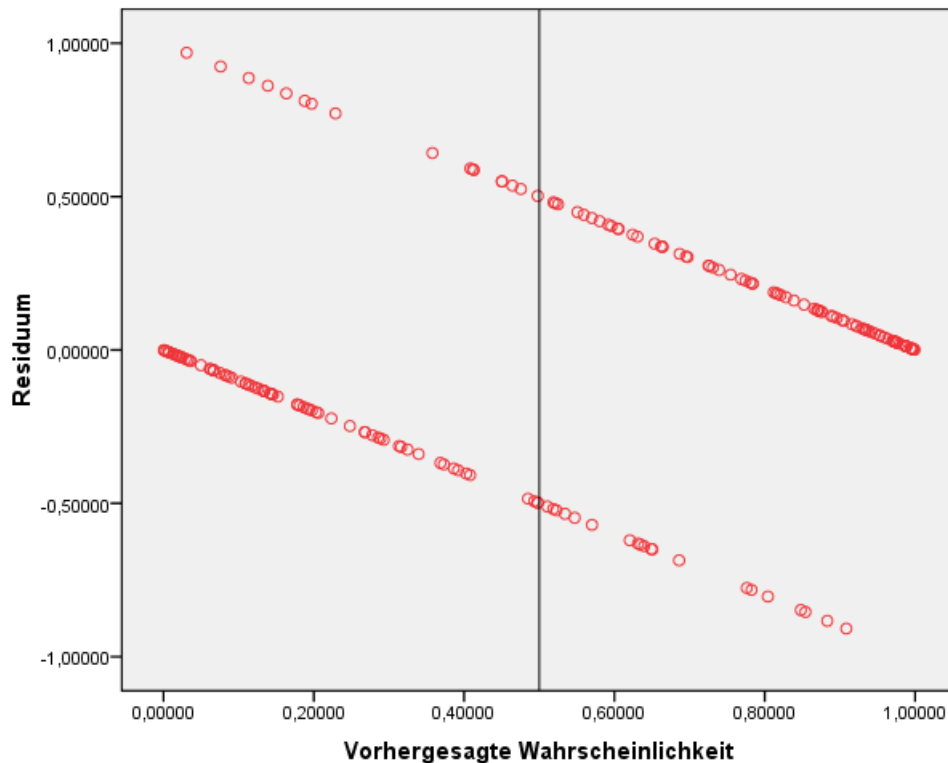
Am Ende der Ausführungen zur Residualdiagnostik soll noch darauf hingewiesen werden, dass man den Plot der einfachen (nicht standardisierten) Residuen gegen die prognostizierten Wahrscheinlichkeiten auch zur Bestimmung eines optimalen **Trennwertes** für die Prognose der Kriteriumsgruppenzugehörigkeit verwenden kann (vgl. Abschnitt 2.5.3). In der Prozedur zur binären logistischen Regression fordert man die benötigten Variablen per **Speichern**-Subdialogbox an:



Mit den (z.B. bei der ersten Modellierung innerhalb einer Sitzung als PRE_1 bzw. RES_1 bezeichneten) Variablen erstellt man ein Streudiagramm, z.B. über:

Diagramme > Veraltete Dialogfelder > Streu-/Punkt-Diagramm > Einfach

Beim folgenden Exemplar wurde über das **Optionen**-Menü des Diagramm-Editors eine **Bezugslinie für die x-Achse** ergänzt:



Links oben (Residuum > 0,5; Prognose falsch negativ) sowie rechts unten (Residuum < -0,5; Prognose falsch positiv) befinden sich die vom Modell über den Trennwert 0,5 falsch klassifizierten Fälle.

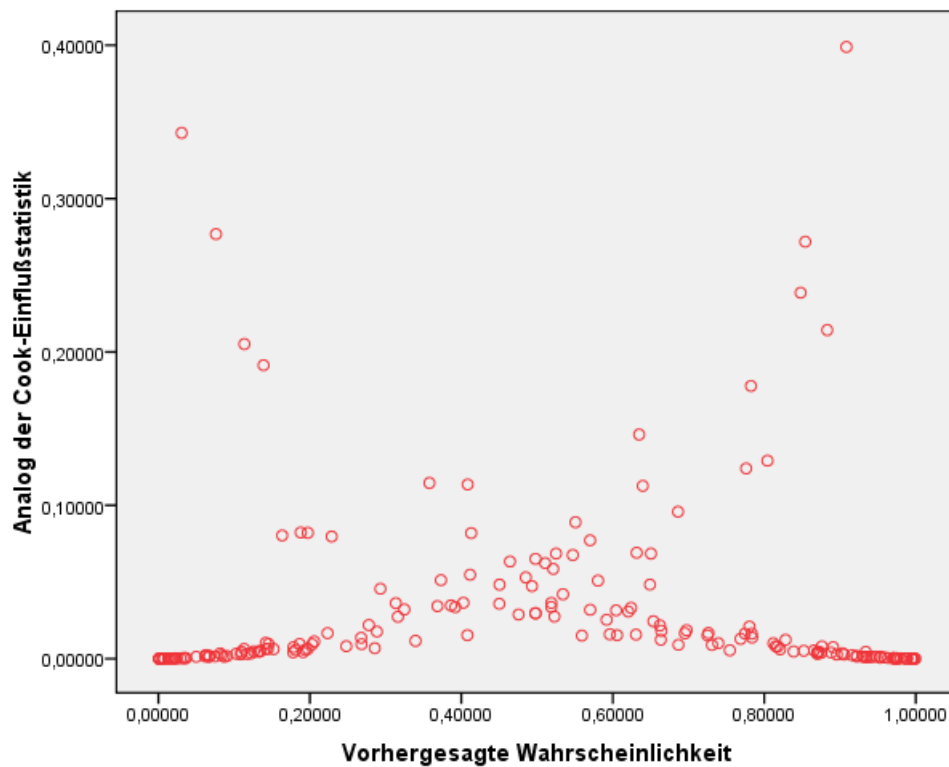
2.4.2.2 Einflussreiche Fälle

Oft haben Ausreißer (Fälle mit betragsmäßig großen Individualresiduen) einen überdurchschnittlichen Einfluss auf die Parameterschätzungen. Allerdings kann ein großer Einfluss auch durch extreme Prädiktorwerte zustande kommen, und nicht jeder einflussreiche Fall muss auch als Ausreißer in Erscheinung treten. Er kann die Schätzungen so „in seinem Sinne“ beeinflussen, dass sein Residuum unauffällig bleibt. Daher muss bei jeder Analyse neben der Ausreißer-Diagnostik auch eine Kontrolle auf Fälle mit ungebührlichem Einfluss auf die Schätzergebnisse vorgenommen werden. Mit zunehmender Stichprobengröße wird es allerdings unwahrscheinlicher, dass einzelne Fälle das Ergebnis dominieren.

SPSS bietet in der BLR-Prozedur einige Maße an, die den Einfluss eines Falles auf die Parameterschätzungen quantifizieren sollen. Sie sind wie die Residuen in der **Speichern**-Subdialogbox anzufordern:



In folgendem Diagramm sind **Cook's Distanzen** aus dem DBS-Beispiel (mit allen Prädiktoren außer RAUCHER) in Abhängigkeit von der Modellprognose zu sehen:



SPSS speichert diese Informationen aufgrund der abgebildeten Dialogbox in der neuen Variablen COO_1.

Alle Fälle halten einen großen Abstand zum kritischen Wert 1, den Hosmer & Lemeshow (2000, S. 180) nennen.

2.5 Beurteilung der Modellrelevanz

2.5.1 Der Likelihood-Quotiententest zur globalen Nullhypothese

Bei der logistischen Regression werden die Parameter β_m in der Regel nach dem Maximum-Likelihood-Prinzip geschätzt, d.h. die Schätzer $\hat{\beta}_m$ werden so bestimmt, dass die Likelihood (Wahrscheinlichkeit) der Stichprobendaten unter dem parametrisch spezifizierten Modell maximal wird (vgl. Abschnitt 2.3.1). Für viele Aussagen ist die Likelihood allerdings weniger geeignet als die mit -2 multiplizierte, logarithmierte Variante (Bezeichnung: -2LL). Je besser ein Modell zu den Daten passt, desto größer wird seine Likelihood, und desto kleiner folglich die Größe -2LL, die somit als *Fehlermaß* aufgefasst werden kann.

Die -2LL - Statistiken taugen vor allem zum *Vergleich* verschiedener Modelle, die in einer bestimmten Spezialisierungs- bzw. Generalisierungsbeziehung zueinander stehen. Zur Konkurrenz stehen dabei:

- Ein *uneingeschränktes*, als gültig akzeptiertes Modell
- Ein *eingeschränktes* Modell, das sich aus dem uneingeschränkten Modell durch Streichen von Parametern, also durch zusätzliche Restriktionen, ergibt

Wenn in dieser Situation auch das eingeschränkte Modell gilt, dann folgt die Differenz zwischen dem -2LL - Wert des eingeschränkten Modells (-2LL(E)) und dem -2LL - Wert des uneingeschränkten Modells (-2LL(U)) bei hinreichend großer Stichprobe (siehe Abschnitt 2.1.2) einer χ^2_{df} -Verteilung, wobei df die Anzahl der gestrichenen Parameter ist. Damit taugt die -2LL - Differenz als Prüfstatistik für einen Signifikanztest zum Hypothesenpaar:

- H_0 : Das eingeschränkte Modell ist gültig, d.h. die gestrichenen Parameter sind alle gleich 0.
- H_1 : Das eingeschränkte Modell ist falsch, d.h. mindestens ein gestrichener Parameter ist von 0 verschieden.

Derart konstruierte Tests arbeiten nach dem so genannten **Likelihood-Quotienten** – Prinzip (engl.: *likelihood ratio*). Dass es sich bei der Differenz

$$-2LL(E) - (-2LL(U))$$

tatsächlich um einen logarithmierten und anschließend mit -2 multiplizierten Likelihood-Quotienten handelt, lässt sich mit Hilfe der Rechenregeln für den Logarithmus zeigen:

$$-2 \ln \left(\frac{L(E)}{L(U)} \right) = -2 [\ln(L(E)) - \ln(L(U))] = -2LL(E) - (-2LL(U))$$

Bei der Konstruktion eines **globalen Modelltests** nach dem Likelihood-Quotienten-Prinzip sind folgende Modelle bzw. -2LL – Ausdrücke beteiligt:

- Beim uneingeschränkten Modell handelt es sich hier um das vollständige Modell, dessen logarithmierte Likelihood wir mit $LL(\hat{\beta})$ bezeichnen.
- Beim eingeschränkten Modell handelt es sich um das so genannte *Basismodell* mit dem konstanten Term als einzigem Parameter. Die logarithmierte Likelihood des Basismodells soll mit $LL(0)$ bezeichnet werden.

Nach obigen Überlegungen erlaubt die Differenz $-2LL(0) - (-2LL(\hat{\beta}))$ einen Test der folgenden globalen Nullhypothese:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_M = 0 \quad (10)$$

Offenbar entspricht dieser Test dem globalen F-Test in der linearen Regressionsanalyse.

Für unser DBS-Beispiel (mit allen Prädiktoren außer RAUCHER) ergibt sich eine Prüfgröße von 115,154, die mit einer sehr kleinen Überschreitungswahrscheinlichkeit verbunden ist, so dass die globale

Nullhypothese klar scheitert. Das Ergebnis erscheint in der folgenden Tabelle aus der BLR-Ausgabe gleich dreimal, weil alle Regressoren simultan in das Modell aufgenommen wurden:

Omnibus-Tests der Modellkoeffizienten

		Chi-Quadrat	df	Sig.
Schritt 1	Schritt	115,154	5	,000
	Block	115,154	5	,000
	Modell	115,154	5	,000

Den $-2LL$ – Wert des vollständigen Modells erfahren wir in folgender Tabelle:

Modellzusammenfassung

Schritt	-2 Log-Likelihood	Cox & Snell R-Quadrat	Nagelkerkes R-Quadrat
1	161,785 ^a	,438	,584

a. Schätzung beendet bei Iteration Nummer 6, weil die Parameterschätzer sich um weniger als ,001 änderten.

Den $-2LL$ – Wert des Basismodells (= 276,939) liefert die BLR-Prozedur über das Iterationsprotokoll (Anforderung via **Optionen**-Subdialogbox) zum Anfangsblock:

Iterationsprotokoll^{a, b, c}

		-2 Log-Likelihood	Koeffizienten
Iteration			Konstante
Schritt 0	1	276,939	,080
	2	276,939	,080

a. Konstante in das Modell einbezogen.
b. Anfängliche -2 Log-Likelihood: 276,939
c. Schätzung beendet bei Iteration Nummer 2, weil die Parameterschätzer sich um weniger als ,001 änderten.

Manche Autoren sehen folgende Entsprechungen zwischen den $-2LL$ – Werten der logistischen Regression und den Quadratsummen der linearen Regression:

Logistische Regression	Lineare Regression
-2LL(Basismodell)	Totale Quadratsumme (SST)
-2LL(Vollständiges Modell)	Fehlerquadratsumme (SSE) (siehe Abschnitt 2.4.2.1.1)
-2LL(Basismodell) – (-2LL(Vollständiges Modell))	Aufgeklärte Quadratsumme (SSR)

2.5.2 Pseudo- R^2 -Statistiken

In der oben wiedergegebenen Tabelle **Modellzusammenfassung** präsentiert die SPSS-Prozedur zur binären logistischen Regression auch zwei Maßzahlen, die in Analogie zum Determinationskoeffizienten (R^2) der linearen Regression den Anteil der vom BLR-Modell erklärten Kriteriumsvarianz schätzen sollen:

Generalisiertes R^2 nach Cox & Snell

Die von Cox & Snell (1989) vorgeschlagene R^2 -Variante ist folgendermaßen definiert:

$$R_{cs}^2 := 1 - \left(\frac{L(0)}{L(\hat{\beta})} \right)^{\frac{2}{N}}$$

Dabei ist $L(0)$ die Likelihood des Basismodells und $L(\hat{\beta})$ die Likelihood des Modells mit den geschätzten Koeffizienten. R_{cs}^2 basiert unmittelbar auf der χ^2 -Prüfgröße $-2LL(0) - (-2LL(\hat{\beta}))$ für den globalen Modelltest (vgl. Abschnitt 2.5.1):

$$1 - \exp\left(-\frac{-2LL(0) - (-2LL(\hat{\beta}))}{N}\right) = 1 - \left(\frac{\exp(LL(0))}{\exp(LL(\hat{\beta}))} \right)^{\frac{2}{N}} = 1 - \left(\frac{L(0)}{L(\hat{\beta})} \right)^{\frac{2}{N}} = R_{cs}^2$$

Auch das klassische R^2 der linearen Regressionsanalyse besitzt eine Darstellung wie im linken Term. Zudem kann R_{cs}^2 für *jedes* per Maximum-Likelihood geschätzte Regressionsmodell berechnet werden und verdient daher die Bezeichnung *generalisiertes R^2* .

Da $L(\hat{\beta})$ als Produkt von Wahrscheinlichkeiten nach oben durch den Wert 1 beschränkt ist, kann R_{cs}^2 maximal den Wert

$$R_{cs\max}^2 := 1 - L(0)^{\frac{2}{N}}$$

erreichen. Im Nullhypothesenmodell ist für jeden Fall die Modellprognose gerade mit dem Stichprobenanteil der Einsen identisch, der mit $\hat{\pi}$ bezeichnet werden soll. Damit gilt:

$$\begin{aligned} R_{cs\max}^2 &= 1 - \left[\hat{\pi}^{\hat{\pi}N} (1 - \hat{\pi})^{N - \hat{\pi}N} \right]^{\frac{2}{N}} \\ &= 1 - \hat{\pi}^{2\hat{\pi}} (1 - \hat{\pi})^{2 - 2\hat{\pi}} \end{aligned}$$

Bei $\hat{\pi} = 0,5$ resultiert z.B. ein maximales R_{cs}^2 von 0,75.

Pseudo- R^2 nach Nagelkerke

Um den gewohnten R^2 -Wertebereich von 0 bis 1 zu erhalten, hat **Nagelkerke** (1991) die folgende Definition vorgeschlagen:

$$R_N^2 := \frac{R_{cs}^2}{R_{cs\max}^2}$$

Pseudo- R^2 nach McFadden

In der Prozedur zur multinomialen logistischen Regression liefert SPSS auch die R^2 -Variante nach **McFadden**:

$$R_M^2 := 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

Verwendet man die am Ende von Abschnitt 2.5.1 angeführten Analogien zwischen den Quadratsummen der linearen Regression und bestimmten $-2LL$ – Statistiken der logistischen Regression, dann hat McFaddens Maß eine Darstellung, die auch für den Determinationskoeffizienten der linearen Regression gilt:

$$1 - \frac{SSE}{SST}$$

Für unser DBS-Modell (mit allen Prädiktoren bis auf RAUCHER) liefert die NOMREG-Prozedur nach einem Aufruf mit der in Abschnitt 2.2 beschriebenen Dialogbox:

Pseudo-R-Quadrat

Cox und Snell	,438
Nagelkerke	,584
McFadden	,416

Die Pseudo-R²-Maße können sowohl aus Individualdaten als auch (mit identischem Ergebnis) aus aggregierten Daten berechnet werden. In der Ausgabe der Prozedur LOGIST REGRESSION, die schon in Abschnitt 2.5.1 zu sehen war, fehlt der Index von McFadden.

Bei vergleichbaren Modellen fallen die Pseudo-R²-Maße meist geringer aus als der Determinationskoeffizient der linearen Regressionsanalyse (Norušis 2005, S. 328), so dass z.B. beim R_M² von **McFadden** Werte zwischen 0,2 und 0,4 als sehr erfolgreich gelten (Tabachnik & Fidell 2007, S. 460).

Bei den Pseudo-R²-Maßen der logistischen Regression fehlt eine Korrektur analog zum adjustierten R² der linearen Regression. Sie wachsen also grundsätzlich mit der Anzahl der Prädiktoren.

2.5.3 Prädiktive Effizienz

Oft sind bei einem Forschungsprojekt nicht nur theoretische Überlegungen von Bedeutung, sondern auch diagnostische Aufgaben. Dann sollte ein entwickeltes Modell auch nach seiner Klassifikationsleistung beurteilt werden. Die Rate korrekter Klassifikationen (*Correct Classification Rate*) ist oft wichtiger als die Pseudo-R²-Maße, und es kann in beiden Disziplinen zu recht unterschiedlichen Bewertungen eines Modells kommen.

2.5.3.1 Die Klassifikationstabelle

Die logistische Regression liefert gemäß Modellgleichung (1) für jeden Fall eine Schätzung $\hat{\pi}_i$ für die Wahrscheinlichkeit, zur Einsergruppe zu gehören. Es liegt nahe, die Fälle aufgrund ihrer $\hat{\pi}_i$ -Werte wie folgt zu klassifizieren:

$$\hat{\pi}_i \begin{cases} \leq 0,5 \rightarrow \text{Gruppe 0} \\ > 0,5 \rightarrow \text{Gruppe 1} \end{cases} \quad (11)$$

Nun kann durch Vergleich mit den tatsächlichen Gruppenzugehörigkeiten die Klassifikationsleistung eines Modells beurteilt werden. Unser gegenwärtiges DBS-Modell schneidet folgendermaßen ab:

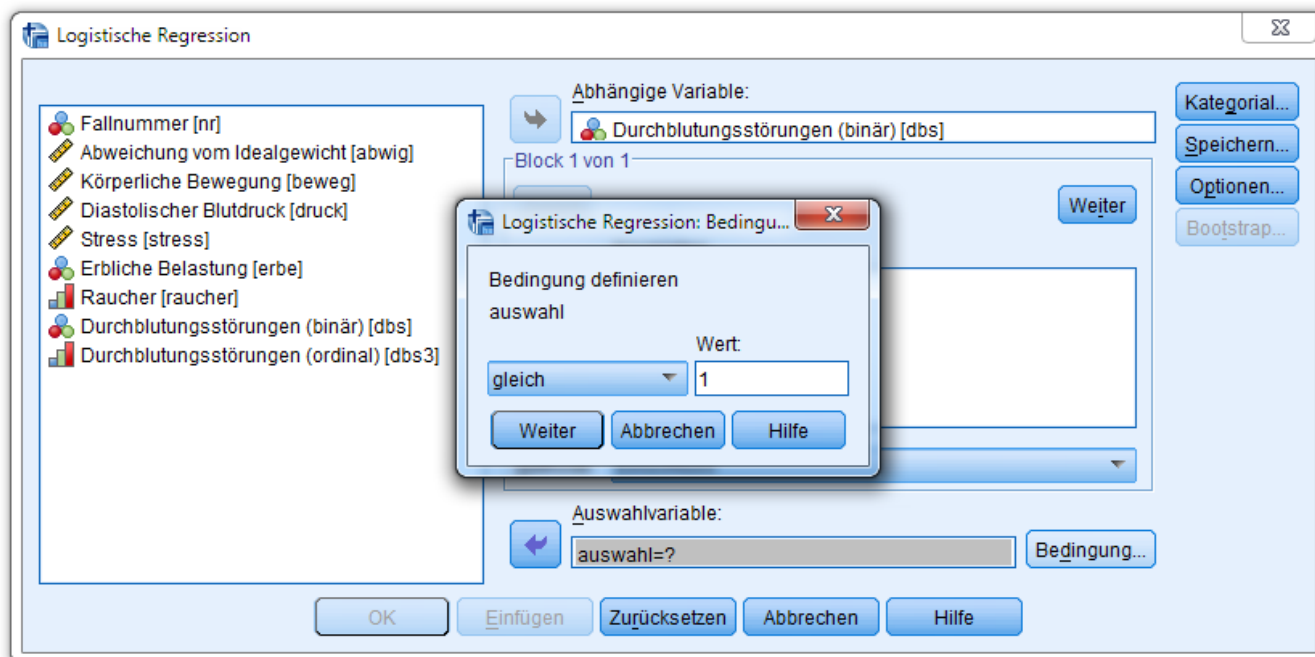
Klassifizierungstabelle^a

Beobachtet			Vorhergesagt		
			Durchblutungsstörungen (binär)		Prozentsatz der Richtigen
			Nein	Ja	
Schritt 1	Durchblutungsstörungen	Nein	76	20	79,2
		Ja	17	87	83,7
Gesamtprozentsatz					81,5

a. Der Trennwert lautet ,500

Die Fälle aus Gruppe 0 (keine Durchblutungsstörungen diagnostiziert) werden zu 79,2% richtig klassifiziert, die Fälle aus Gruppe 1 zu 83,7%. Insgesamt werden 81,5% aller Fälle von unserem Modell richtig eingeordnet. Diese Trefferrate ist zu vergleichen mit der Leistung des Basismodells (ohne Prädiktoren), das jeden Fall unbeschieden in die stärker besetzte Gruppe 0 einordnet, was bei unseren Daten zu einer Trefferrate von 52,0 % führen würde.

Bei der Arbeit mit solchen Klassifikationstabellen, die auch bei Diskriminanzanalysen gern verwendet werden, ist generell zu bedenken, dass die Trefferraten in der Stichprobe, aus der das Modell stammt, *überschätzt* werden. Einen realistischen Eindruck vermittelt nur die **Kreuzvalidierung** des Modells anhand einer unabhängigen Stichprobe. SPSS ermöglicht die Auswahl einer Teilstichprobe zur Modellgewinnung, so dass die ausgeschlossenen Fälle für eine echte Kreuzvalidierung zur Verfügung stehen:



Bei den folgenden Kreuzvalidierungsergebnissen wurde die Zugehörigkeit zur Schätzstichprobe (Anteil ca. 72%) bzw. zur Kreuzvalidierungsstichprobe (Anteil ca. 28%) über die Variable AUSWAHL gesteuert:

Klassifizierungstabelle^c

Beobachtet			Vorhergesagt				
			Ausgewählte Fälle ^a			Nicht ausgewählte Fälle ^b	
			Durchblutungsstörungen (binär)		Prozentsatz der Richtigen	Durchblutungsstörungen (binär)	
			Nein	Ja		Nein	Ja
Schritt 1	Durchblutungsstörungen (binär)	Nein	55	14	79,7	20	7
		Ja	12	53	81,5	7	32
Gesamtprozentsatz					80,6		

a. Ausgewählte Fälle Fallauswahl für Kreuzvalidierung EQ 1

b. Nicht ausgewählte Fälle Fallauswahl für Kreuzvalidierung NE 1

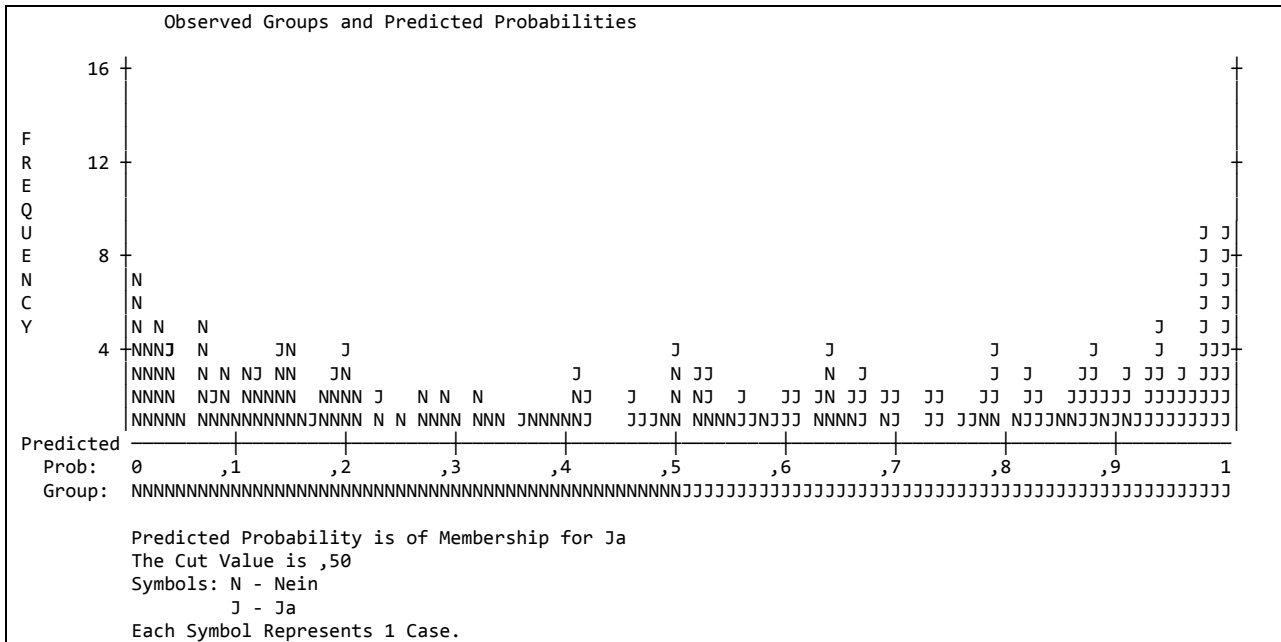
c. Der Trennwert lautet ,500

Die in der Kreuzvalidierungsstichprobe nur geringfügig reduzierte Trefferrate spricht für die diagnostische Tauglichkeit des Modells.

2.5.3.2 Klassifikationsdiagramm

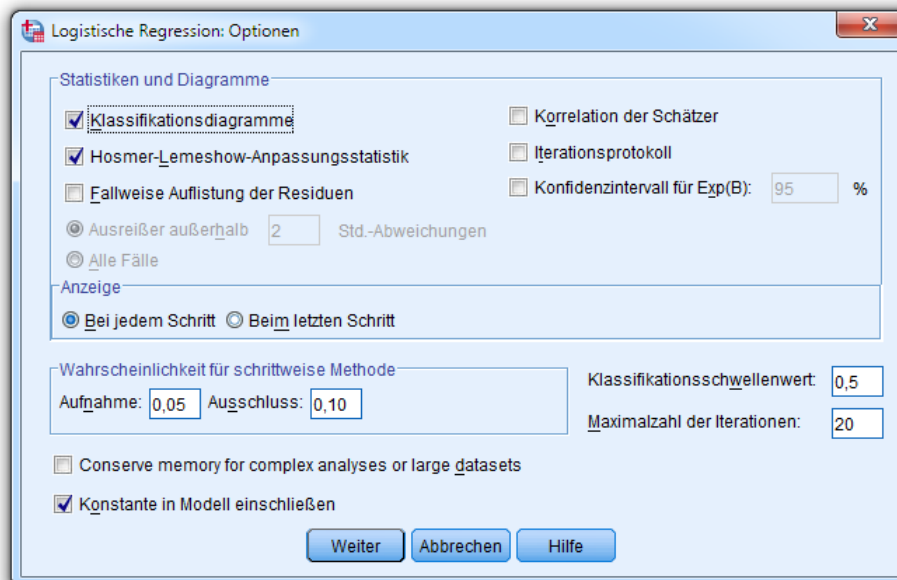
Im folgenden **Klassifikationsdiagramm** werden die beobachteten Gruppenzugehörigkeiten y_i in Abhängigkeit von der geschätzten Wahrscheinlichkeit $\hat{\pi}_i$ dargestellt. Daraus lässt sich z.B. ablesen, ob die falsch negativen Klassifikationen (eine vorhandene Durchblutungsstörung wird nicht prognostiziert)

überwiegend bei knappen Entscheidungen ($\hat{\pi}_i$ nahe 0,5) aufgetreten sind. Wie wir bereits aus Abschnitt 2.4.2.1.5 wissen, hat es in unserer Stichprobe allerdings einen Fall (Nr. 8) erwischt, dem das Modell ein sehr kleines Risiko (0,031) zugeordnet hatte.



Aufgrund eines solchen Plots könnte man die Standardklassifikationsregel in Formel (11) so abändern, dass vor allem die mit höheren Kosten verbundene Fehldiagnose vermieden wird. Für denselben Zweck kann man auch den Plot der einfachen (nicht standardisierten) Residuen gegen die prognostizierten Wahrscheinlichkeiten verwenden (siehe Abschnitt 2.4.2.1).

In der SPSS-Prozedur zur binären logistischen Regression wird das **Klassifikationsdiagramm** in der **Optionen**-Subdialogbox angefordert:



Hier kann man auch den **Klassifikationsschwellenwert** ändern.

2.6 Beurteilung der einzelnen Regressoren

2.6.1 Regressionskoeffizienten und Effektgrößen

Für unser gegenwärtiges DBS-Modell liefert SPSS folgende Schätzergebnisse:

Variablen in der Gleichung							95% Konfidenzintervall für EXP(B)	
		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	
Schritt 1 ^a	abwig	,063	,020	10,265	1	,001	1,066	1,025 1,108
	beweg	-,946	,252	14,091	1	,000	,388	,237 ,636
	druck	,042	,009	19,947	1	,000	1,043	1,024 1,063
	stress	,650	,236	7,609	1	,006	1,915	1,207 3,040
	erbe	1,762	,476	13,709	1	,000	5,826	2,292 14,809
	Konstante	-2,813	1,429	3,874	1	,049	,060	

a. In Schritt 1 eingegebene Variablen: abwig, beweg, druck, stress, erbe.

Wir erfahren z.B. über den Regressor ABWIG, dass bei einer Erhöhung um eine Einheit (1 kg) das logarithmierte Wahrscheinlichkeitsverhältnis (siehe Gleichung (2)) um 0,063 ansteigt.

Rechts neben den geschätzten Koeffizienten stehen die zugehörigen Standardfehler.

Bei der Interpretation eines Regressionskoeffizienten ist seine Skalierung bzw. Maßeinheit zu berücksichtigen. In der obigen Tabelle vermisst man eine zum *standardisierten* Regressionskoeffizienten (Beta-Gewicht) der linearen Regression analoge Statistik. Menard (1995, S. 46) berichtet über entsprechende Definitionsvorschläge und beschreibt auch, wie man diese standardisierten Koeffizienten in SPSS (mit einigem Aufwand) zu berechnen sind. Wenn es nur darum geht, die Beiträge der metrischen Regressoren vergleichbar zu machen, kann man diese Regressoren *vor* der Analyse standardisieren, was z.B. Tabachnik und Fidell (2007, S. 469) vorschlagen. Weil beim Kriterium keine „Standardisierung“ stattfindet, erhält man allerdings keine Koeffizienten im Variationsbereich von Beta-Gewichten. Außerdem wird z.B. beim Standardisieren der Variablen DRUCK die Maßeinheit *mm/Hg* ersetzt durch *DRUCK-Standardabweichung in der betrachteten Stichprobe*. Folglich hängt der Regressionskoeffizient nicht nur vom Einfluss des Blutdrucks auf das Kriterium ab, sondern auch von der Varianz des Blutdrucks in der untersuchten Stichprobe. Spätestens beim Vergleich von verschiedenen Stichproben bzw. Populationen sind die Ergebnisse zu standardisierten Variablen weniger geeignet.

Urban (1993, S. 38f) empfiehlt, die Beiträge der Regressoren über die Wald-Statistik (siehe Abschnitt 2.6.2) zu vergleichen, und begründet diese Empfehlung mit dem Hinweis, dass die Wald-Statistik bei metrischen Regressoren invariant gegenüber Änderungen der Skala (Varianz) ist.

In der letzten Spalte der obigen Tabelle befindet sich für jeden Regressor der von Urban (1993, S. 40f) als *Effektgröße*, von anderen Autoren als *odds ratio* bezeichnete Ausdruck e^{β_m} . Er gibt an, um welchen Faktor sich das Wahrscheinlichkeitsverhältnis

$$\frac{P(Y = 1)}{P(Y = 0)}$$

ändert, wenn der Prädiktor X_m um eine Einheit steigt und alle anderen konstant bleiben (vgl. Abschnitt 2.1.3). In der Demostudie führt z.B. die Steigerung des STRESS-Indikators um eine Einheit zu einem nahezu verdoppelten Wahrscheinlichkeitsverhältnis.

Die Konfidenzintervalle zu den Effektgrößen müssen bei der SPSS- Prozedur zur binären logistischen Regression in der **Optionen**-Subdialogbox angefordert werden.

Über die Reduktion der Pseudo- R^2 -Statistiken beim Entfernen eines Regressors aus dem Modell kann man seinen eigenständigen Erklärungsbeitrag beurteilen.

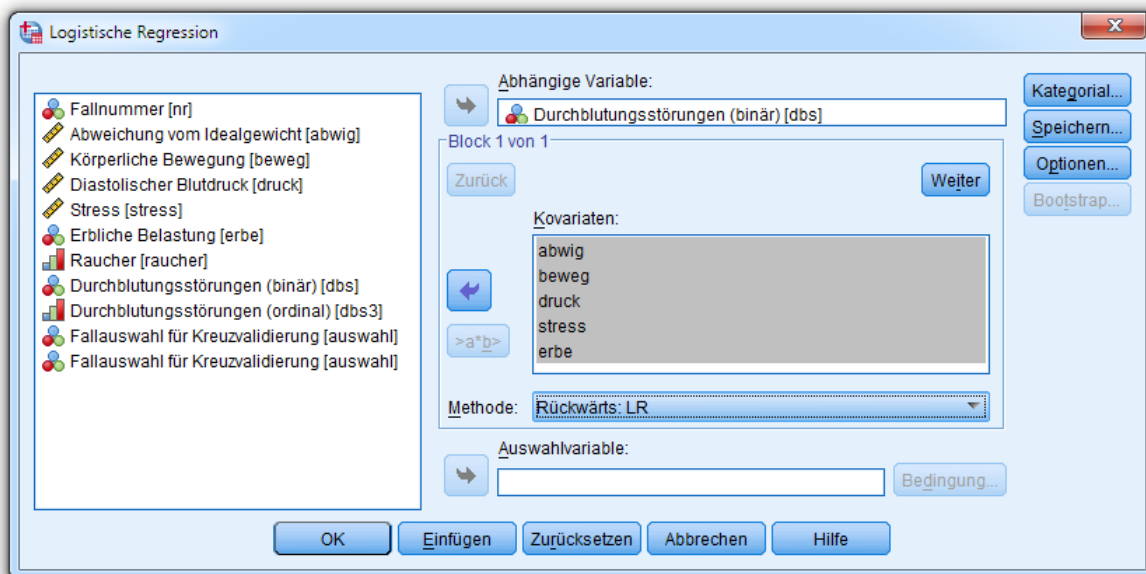
2.6.2 Signifikanz

Zum Testen der Nullhypothese, dass ein Regressor irrelevant ist, kann die **Wald-Statistik** verwendet werden, die bei hinreichend großer Stichprobe (siehe Abschnitt 2.1.2) annähernd χ^2 -verteilt ist. Bei metrischen und dichotomen Regressoren, die jeweils *einen* Freiheitsgrad besitzen, ist die Wald-Statistik gerade identisch mit dem quadrierten Quotienten aus dem geschätzten Regressionskoeffizienten und seinem Standardfehler. Bei einem *Faktor* (nominalskalierten Regressor) mit ($k > 2$) Kategorien folgt die Wald-Statistik unter der Nullhypothese einer χ^2 -Verteilung mit ($k - 1$) Freiheitsgraden (siehe Abschnitt 2.7). In der **Sig**-Spalte sind die empirischen Überschreitungswahrscheinlichkeiten der Wald-Tests zu den Regressoren angegeben.

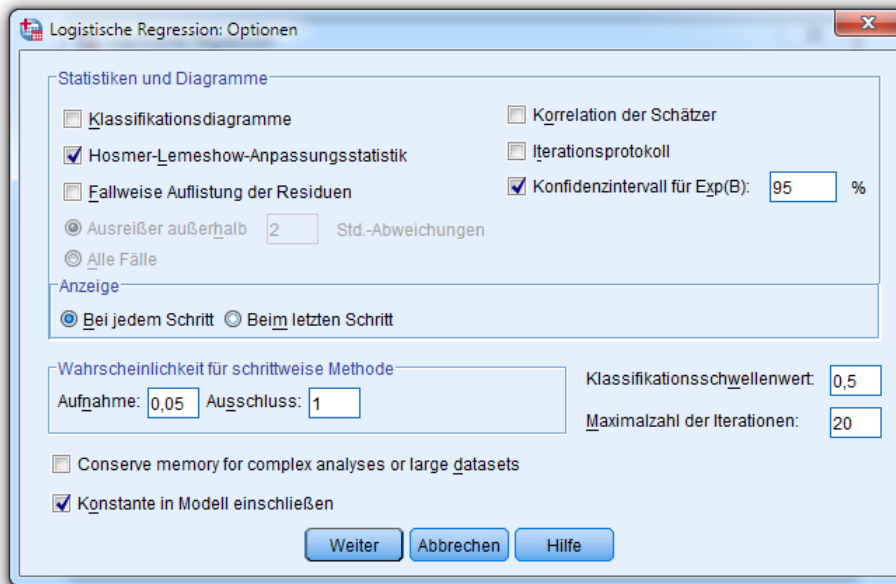
Bei Regressoren mit betragsmäßig großen Werten ist die Wald-Statistik nach unten verfälscht, so dass die Signifikanztests mit einem erhöhten Fehlerrisiko zweiter Art belastet sind (Norušis 2005, S. 329).

Generell sollten auch zur Beurteilung der einzelnen Regressoren Likelihood-Quotienten-Tests durchgeführt werden, die genauer arbeiten als alle konkurrierenden Methoden (Menard 1995, S. 38). Bei ihrer Konstruktion nach dem in Abschnitt 2.5.1 beschriebenen Prinzip sind im eingeschränkten Modell genau die Parameter der zu testenden Regressoren gestrichen.

Um die Prozedur LOGISTIC REGRESSION mit einem bei Norušis (2005, S. 329) beschriebenen Trick zur Berechnung der Likelihood-Quotienten-Tests zu überreden, wählt man die zur automatischen Modellsuche (siehe Abschnitt 2.9.2) gedachte Methode **Rückwärts: LR**



und setzt im **Optionen**-Dialog das **Ausschluss**-Kriterium auf 1:



Im Ergebnis startet SPSS bei der „Modellsuche“ mit allen Prädiktoren, verwendet Likelihood-Quotienten-Tests (ab jetzt kurz bezeichnet als: *LQ-Tests*) zur Beurteilung der Prädiktoren und stoppt sofort, weil kein Prädiktor mit seinem P-Level im LQ-Test das Ausschlusskriterium überbietet.

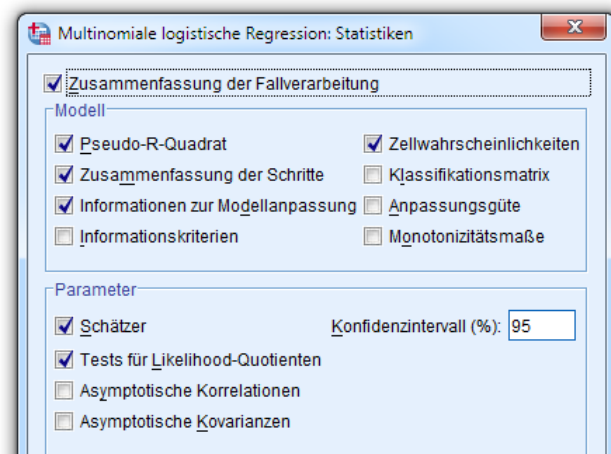
Wir erhalten die Tabelle

Modellieren, wenn Term entfernt

Variable		Log-Likelihood des Modells	Änderung der -2 Log-Likelihood	df	Signifikanz der Änderung
Schritt 1	abwig	-86,549	11,313	1	,001
	beweg	-89,253	16,720	1	,000
	druck	-94,818	27,851	1	,000
	stress	-84,962	8,138	1	,004
	erbe	-88,831	15,876	1	,000

und finden dort alle Ergebnisse der Wald-Tests bestätigt.

Die Prozedur NOMREG zur multinomialen logistischen Regression bietet in der **Statistiken**-Subdialogbox über ein (per Voreinstellung markiertes) Kontrollkästchen Likelihood-Quotienten-Tests zu den Prädiktoren an:



In unserem Beispiel resultiert die folgende Tabelle mit längst bekannten Testentscheidungen:

Likelihood-Quotienten-Tests				
Effekt	Kriterien für die Modellanpassung	Likelihood-Quotienten-Tests		
	-2 Log-Likelihood für reduziertes Modell	Chi-Quadrat	Freiheitsgrade	Signifikanz
Konstanter Term	165,864	4,079	1	,043
abwig	173,098	11,313	1	,001
beweg	178,506	16,720	1	,000
druck	189,637	27,851	1	,000
stress	169,923	8,138	1	,004
erbe	177,661	15,876	1	,000

Die Chi-Quadrat-Statistik stellt die Differenz der -2 Log-Likelihoods zwischen dem endgültigen Modell und einem reduziertem Modell dar. Das reduzierte Modell wird berechnet, indem ein Effekt aus dem endgültigen Modell weggelassen wird. Hierbei liegt die Nullhypothese zugrunde, nach der alle Parameter dieses Effekts 0 betragen.

2.6.3 Fehlende bzw. irrelevante Prädiktoren

Wir betrachten ein Modell, das lediglich den Prädiktor STRESS enthält, der mit den nun fehlenden Prädiktoren ABWIG, BEWEG und DRUCK korreliert. Im Vergleich zu den realistischen Schätzergebnissen zum korrekten Modell (vgl. Abschnitt 2.6.1) erhalten wir für den STRESS-Regressionskoeffizienten einen deutlich erhöhten Wert (1,04 statt 0,65):

Variablen in der Gleichung							95% Konfidenzintervall für EXP(B)	
		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	
Schritt 1 ^a	stress	1,041	,180	33,525	1	,000	2,831	1,990
	Konstante	-3,456	,629	30,176	1	,000	,032	4,026

a. In Schritt 1 eingegebene Variablen: stress.

Ein analoges Problem ist auch für das oben gelegentlich betrachtete Modell mit BEWEG als einzigem Prädiktor anzunehmen.

Um die Beeinträchtigung der Schätz- und Testergebnisse durch **irrelevante Prädiktoren** beobachten zu können, nehmen wir vorübergehend drei solche Prädiktoren in das Modell auf, von denen zwei (IRRE1 und IRRE2) recht hoch mit relevanten Prädiktoren korrelieren:¹

Korrelationen						
		abwig	beweg	druck	stress	erbe
irre1	Korrelation nach Pearson	-,592**	,464**	-,159*	-,538**	,015
	Signifikanz (2-seitig)	,000	,000	,025	,000	,829
	N	200	200	200	200	200
irre2	Korrelation nach Pearson	-,598**	,458**	-,126	-,526**	,052
	Signifikanz (2-seitig)	,000	,000	,076	,000	,464
	N	200	200	200	200	200
irre3	Korrelation nach Pearson	-,121	,160*	-,077	-,085	,148*
	Signifikanz (2-seitig)	,088	,024	,277	,229	,037
	N	200	200	200	200	200

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

¹ Sie finden die simulierten Daten in der Datei **DBS MIT IRRELEVANTEN PRÄDIKTOREN.SAV** an der im Vorwort vereinbarten Stelle.

Außerdem reduzieren wir die Stichprobengröße auf 150, was sich auf die Ergebnisse zum korrekten Modell kaum auswirkt:

Variablen in der Gleichung									
		Regressions- koeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	95% Konfidenzintervall für EXP(B)	
								Unterer Wert	Oberer Wert
Schritt 1	abwig	,051	,022	5,444	1	,020	1,053	1,008	1,099
	beweg	-,709	,276	6,617	1	,010	,492	,287	,845
	druck	,036	,010	14,029	1	,000	1,037	1,017	1,056
	stress	,700	,267	6,880	1	,009	2,013	1,193	3,396
	erbe	1,629	,522	9,730	1	,002	5,099	1,832	14,192
	Konstante	-3,201	1,585	4,077	1	,043	,041		

Durch die Aufnahme der irrelevanten Variablen wird die Interpretation der Ergebnisse erschwert:

Variablen in der Gleichung									
		Regressions-koeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	95% Konfidenzintervall für EXP(B)	
								Unterer Wert	Oberer Wert
Schritt 1	abwig	,051	,026	3,802	1	,051	1,052	1,000	1,107
	beweg	-,738	,307	5,763	1	,016	,478	,262	,873
	druck	,040	,011	14,116	1	,000	1,041	1,019	1,063
	stress	,759	,293	6,715	1	,010	2,136	1,203	3,793
	erbe	1,634	,559	8,545	1	,003	5,125	1,713	15,331
	irre1	,420	,364	1,325	1	,250	1,521	,745	3,108
	irre2	-,830	,331	6,303	1	,012	,436	,228	,834
	irre3	,532	,291	3,334	1	,068	1,701	,962	3,010
	Konstante	-3,712	1,722	4,643	1	,031	,024		

Während der relevante Prädiktor ABWIG die Signifikanzgrenze verpasst, wird ein irrelevanter Prädiktor als signifikant beurteilt. Außerdem werden die Standardfehler zu den relevanten Prädiktoren größer.

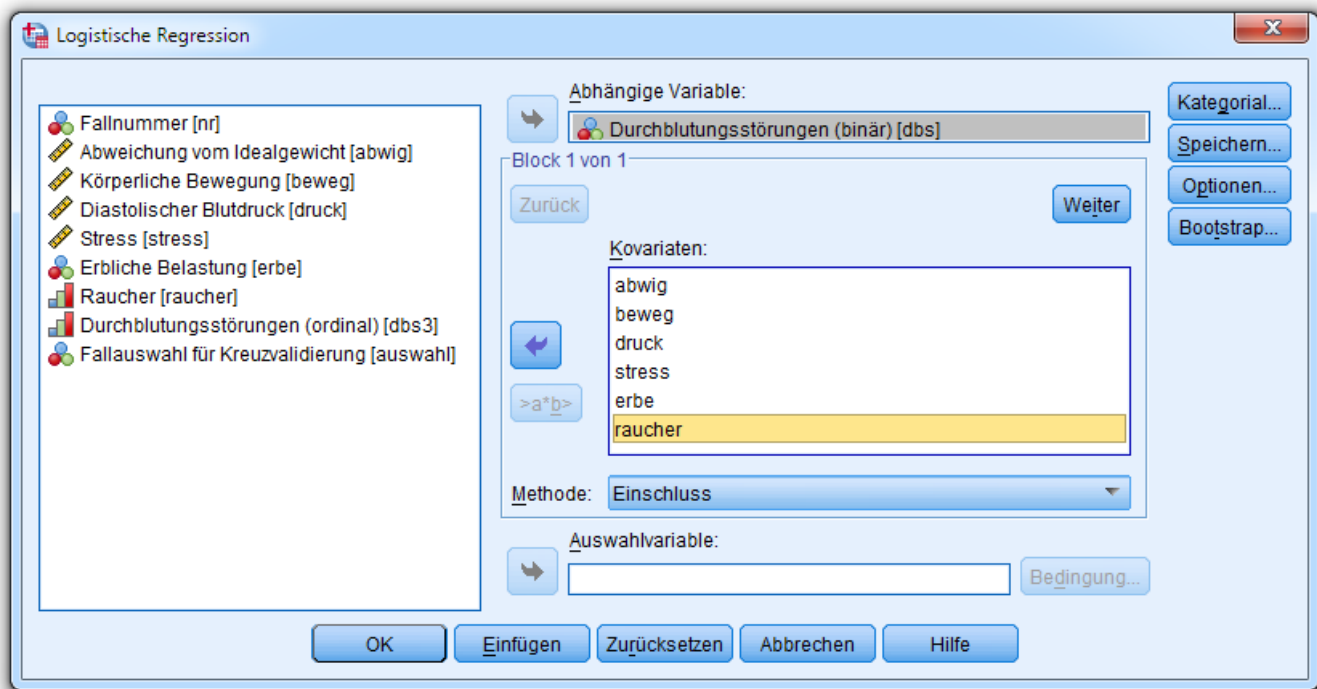
2.7 Nominalskalierte Regressoren mit mehr als zwei Kategorien

Analog zum Vorgehen bei der linearen Regression können auch bei der logistischen Regression nominalskalierte Regressoren einbezogen werden. Die *dichotom*-kategoriale Variable ERBE haben wir bisher schon als Prädiktor verwendet. Per (0,1)-Kodierung wurde für einen leicht interpretierbaren Regressionskoeffizienten gesorgt: β_5 steht für den Logit-Zuwachs bei Vorliegen des Risikofaktors (siehe Gleichung 2). Dichotome kategoriale Regressoren lassen sich bei der logistischen Regression also genauso verwenden wie metrische Regressoren (Kovariaten).

Nun wollen wir unser Modell um eine nominalskalierte Variable mit drei Kategorien erweitern, so dass einige Überlegungen zur Bildung geeigneter Kodiervariablen nötig werden. Es handelt sich um die vom Leser sicher längst erwartete Variable RAUCHER, die mit folgenden Kategorien „erhoben“ worden ist:

Kategorie	kodierter Wert in DBS.SAV
Raucher	1
Ehemaliger Raucher	2
Nichtraucher	3

Wir fordern mit der folgenden Dialogbox eine neue BLR-Analyse an:



Zwar sind in unserem konkreten Beispiel die Kategorien geordnet, aber es ist trotzdem nicht sinnvoll, RAUCHER durch einen *einzelnen* Regressor (z.B. mit den Werten 1, 2, 3) zu repräsentieren, da über die *Abstände* zwischen den 3 Kategorien nichts bekannt ist. Wir müssen also wie im Fall einer nominalskalierten Variablen mit S (> 2) Stufen vorgehen. Um die Information einer solchen Variablen vollständig in das Design einzubringen, müssen $S - 1$ neue Variablen, so genannte *Kodiervariablen*, geeignet definiert und als Regressoren verwendet werden. In unserem konkreten Beispiel RAUCHER ist also das Modell um 2 Kodiervariablen mit zugehörigen Regressionsgewichten β_6 und β_7 zu erweitern. In der Logit-Formulierung lautet das erweiterte Modell:

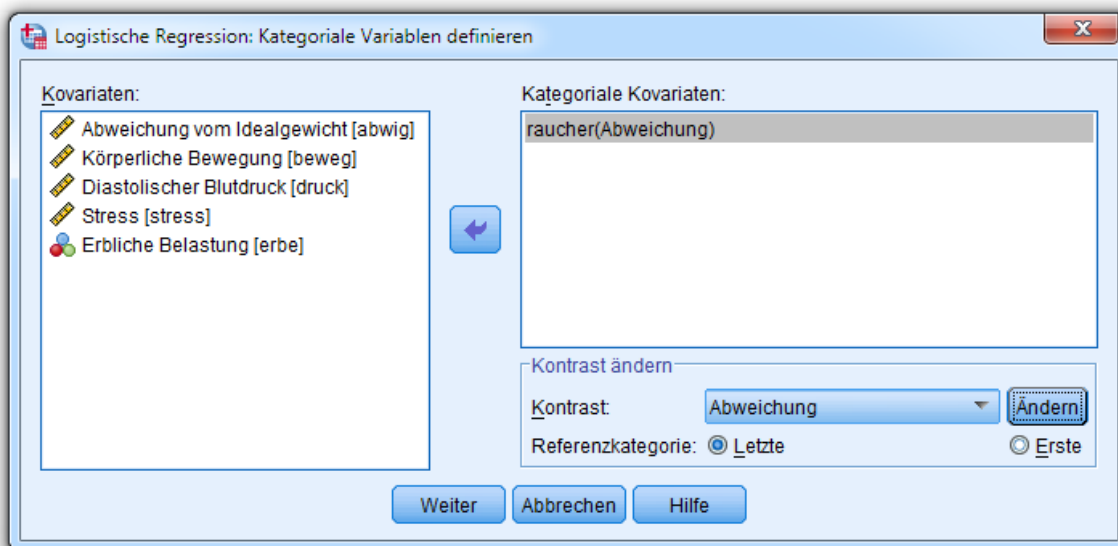
$$\ln\left(\frac{P(Y = 1)}{P(Y = 0)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7$$

Nach der Erweiterung kann unser Modell für jede Raucher-kategorie einen eigenen Logit-Effekt schätzen, der bei allen Wertekombination der anderen Prädiktoren zu den sonstigen Effekten addiert wird.

In Signifikanztests zum Prädiktor RAUCHER wird die folgende Nullhypothese geprüft:

$$H_0: \beta_6 = \beta_7 = 0$$

Damit SPSS die Variable RAUCHER bei der binären logistischen Regression als kategorial behandelt und durch 2 automatisch gebildete Kodiervariablen in das Design einbindet, muss man in der zugehörigen Dialogbox mit dem Schalter **Kategorial** die folgende Subdialogbox anfordern und RAUCHER in die Liste der kategorialen Variablen befördern:



Bzgl. der genauen Definition der Kodiervariablen zu einem nominalskalierten Prädiktor mit mehr als zwei Kategorien bietet SPSS mehrere Alternativen, von denen die Bedeutung der Regressionskoeffizienten zu den Kodiervariablen abhängt. In der Dialogbox **Kategoriale Variable definieren** wählt man einen **Kontrast** aus der versteckten Liste sowie eine **Referenzkategorie** und quittiert über den Schalter **Ändern**. Ist nur *ein* kategorialer Prädiktor vorhanden, übersieht man leicht den Schalter **Ändern**, so dass die gesamte Dialogbox ohne Effekt bleibt.

Im Beispiel sollen für den Prädiktor RAUCHER Abweichungskontraste mit der letzten Kategorie als Referenz verwendet werden. Welche Kodiervariablen SPSS daraufhin bildet, wird in der Ausgabe protokolliert:

Codierungen kategorialer Variablen

		Häufigkeit	Parameterkodierung	
			(1)	(2)
Raucher	Raucher	48	1,000	,000
	Ehem. Raucher	46	,000	1,000
	Nichtraucher	106	-1,000	-1,000

Anschließend wird die Bedeutung der Regressionskoeffizienten β_6 und β_7 zu den so definierten RAUCHER-Kodierungsvariablen X_6 und X_7 erläutert. Wir betrachten eine beliebige, aber fest gewählte, Wertkombination x_1, \dots, x_5 für die Regressoren X_1 bis X_5 und erhalten mit der Abkürzung

$$\kappa := \beta_0 + \beta_1 x_1 + \dots + \beta_5 x_5$$

für die Raucherkategorien folgende Modellaussagen:

Raucher-kategorie	X_6	X_7	Logit
Raucher	1	0	$L_1 := \kappa + \beta_6$
Ehem. Raucher	0	1	$L_2 := \kappa + \beta_7$
Nichtraucher	-1	-1	$L_3 := \kappa - \beta_6 - \beta_7$

Als *ungewichtetes* Mittel der Modellprognosen für die drei Raucher-kategorien erhalten wir:

$$\frac{(\kappa + \beta_6) + (\kappa + \beta_7) + (\kappa - \beta_6 - \beta_7)}{3} = \kappa$$

Nun wird klar:

- β_6 ist die Abweichung des Logits zur Kategorie *Raucher* vom (ungewichteten) Mittel der drei Logits
- β_7 ist die Abweichung des Logits zur Kategorie *Ehem. Raucher* vom (ungew.) Mittel der drei Logits
- die Abweichung des Logits zur Kategorie *Nichtraucher* vom (ungewichteten) Mittel beträgt $(-\beta_6 - \beta_7)$

In diesem Zusammenhang soll noch vor der Verwechslung zwischen den eben behandelten Kodiervariablen und den zugehörigen **Kontrastvektoren** gewarnt werden. Die Nullhypothese

$$H_0 : \beta_6 = 0$$

kann nach obigen Überlegungen äquivalent auch folgendermaßen formuliert werden:

$$L_1 - \frac{1}{3}(L_1 + L_2 + L_3) = 0$$

$$\Leftrightarrow \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} = 0$$

Genau der Kontrastvektor $(\frac{2}{3} \quad -\frac{1}{3} \quad -\frac{1}{3})$ ist z.B. im SPSS-Regressions-Handbuch (SPSS 2011, S. 41) zur Abweichungskodierung angegeben. Er gibt Aufschluss darüber, welche Linearkombination der Kategorien-Logits durch die Nullhypothese zu einem Kodiervariablen-Regressionsgewicht gleich 0 setzt wird. In der SPSS-Ausgabe zur binären logistischen Regression erscheinen jedoch nicht die Kontrastvektoren, sondern die Kodiervariablen, also die korrespondierenden Spalten der Designmatrix (vgl. Nichols 1993).

In unserem Beispiel erhalten wir für das erweiterte Modell folgende Schätzergebnisse:

Variablen in der Gleichung

	Regressions- koeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1 ^a						
abwig	,043	,021	3,998	1	,046	1,044
beweg	-,828	,265	9,790	1	,002	,437
druck	,042	,010	18,055	1	,000	1,043
stress	,509	,254	4,001	1	,045	1,663
erbe	1,856	,502	13,685	1	,000	6,400
raucher			13,069	2	,001	
raucher(1)	1,787	,539	11,011	1	,001	5,974
raucher(2)	-,581	,369	2,483	1	,115	,559
Konstante	-2,175	1,540	1,995	1	,158	,114

a. In Schritt 1 eingegebene Variablen: abwig, beweg, druck, stress, erbe, raucher.

Wie die Ergebniszeile zu RAUCHER zeigt, lehnt der Wald-Test die Nullhypothese ($\beta_6 = \beta_7 = 0$) deutlich ab. RAUCHER ist also insgesamt ein signifikanter Risikofaktor. Von den beiden Kodiervariablen hat nur RAUCHER(1) ein signifikantes Gewicht ($b_6 = 1,787$, $p = 0,001$). Das logarithmierte Wahrscheinlichkeitsverhältnis $\ln\left(\frac{P(Y=1)}{P(Y=0)}\right)$ liegt also bei Rauchern signifikant über dem (ungewichteten) Mittel.

Bei diesen Personen ist der Wahrscheinlichkeitsquotient $\frac{P(Y=1)}{P(Y=0)}$ gegenüber dem *geometrischen* Mittel der 3 Wahrscheinlichkeitsquotienten um den Faktor e^{b_6} ($= 5,974$, siehe Spalte **Exp(B)**) erhöht, denn:

$$L_1 = \frac{1}{3}(L_1 + L_2 + L_3) + \beta_6$$

$$\Leftrightarrow e^{L_1} = e^{\frac{1}{3}(L_1 + L_2 + L_3) + \beta_6} = e^{\beta_6} e^{\frac{1}{3}(L_1 + L_2 + L_3)} = e^{\beta_6} \sqrt[3]{e^{L_1} e^{L_2} e^{L_3}}$$

erhöht. Ferner rechnet man leicht nach, dass bei Rauchern der Wahrscheinlichkeitsquotient $\frac{P(Y=1)}{P(Y=0)}$ im Vergleich zu Nichtrauchern um den Faktor

$$\frac{e^{\kappa+b_6}}{e^{\kappa-b_6-b_7}} = e^{2b_6+b_7} = e^{b_6} e^{b_6} e^{b_7} = 19,96$$

erhöht ist.

Demgegenüber scheinen Ex-Raucher mit einem mittleren Risiko belastet zu sein (kein signifikanter Logit-Kontrast zum ungewichteten Mittel). Im Vergleich zu Nichtrauchern ist ihr Wahrscheinlichkeitsquotient immerhin noch um den Faktor

$$\frac{e^{\kappa+b_7}}{e^{\kappa-b_6-b_7}} = e^{b_6+2b_7} = e^{b_6} e^{b_7} e^{b_7} = 1,87$$

erhöht.

Für die Nichtraucher kann man ausrechnen, dass Ihr Logit im Vergleich zum (ungewichteten) Mittel um

$$b_6 + b_7 = 1,787 - 0,581 = 1,206$$

erniedrigt ist. Ihr Wahrscheinlichkeitsquotient $\frac{P(Y=1)}{P(Y=0)}$ ist im Vergleich zum geometrischen Mittel der 3 Wahrscheinlichkeitsquotienten um den Faktor $e^{-b_6-b_7}$ ($= 0,3$) reduziert.

Für das erweiterte Modell ermittelt SPSS einen akzeptablen Gültigkeitstest

Hosmer-Lemeshow-Test

Schritt	Chi-Quadrat	df	Sig.
1	10,588	8	,226

und hohe Pseudo- R^2 - Statistiken:

Modellzusammenfassung

Schritt	-2 Log-Likelihood	Cox & Snell R-Quadrat	Nagelkerkes R-Quadrat
1	141,995	,491	,655

Die globale Nullhypothese wird erwartungsgemäß deutlich abgelehnt,

Omnibus-Tests der Modellkoeffizienten

	Chi-Quadrat	df	Sig.
Schritt 1	134,944	7	,000
Block	134,944	7	,000
Modell	134,944	7	,000

und die Klassifikationsleistung steigt an:

Klassifizierungstabelle^a

Beobachtet			Vorhergesagt		
			Durchblutungsstörungen (binär)		Prozentsatz der Richtigen
			Nein	Ja	
Schritt 1	Durchblutungsstörungen (binär)	Nein	83	13	86,5
		Ja	18	86	82,7
Gesamtprozentsatz					84,5

a. Der Trennwert lautet ,500

Als Alternative zur oben beschriebenen Abweichungskodierung kommt vor allem die so genannte **Indikator-** oder **Dummy-Kodierung** in Frage, die SPSS im Ausgabefenster folgendermaßen protokolliert:

Codierungen kategorialer Variablen

		Häufigkeit	Parameterkodierung	
			(1)	(2)
Raucher	Raucher	48	1,000	,000
	Ehem. Raucher	46	,000	1,000
	Nichtraucher	106	,000	,000

Hier steht β_6 bzw. β_7 für die Logit-Differenz zwischen der ersten bzw. zweiten RAUCHER-Kategorie und der Referenzkategorie, wobei die Prozedur LOGISTIC REGRESSION per Voreinstellung die *letzte* Kategorie (mit dem höchsten Wert, im Beispiel: Nichtraucher) als Referenz verwendet. Für diese Rolle ist z.B. die in vielen Studien vorhandene Kontrollgruppe prädestiniert.

In der folgender Tabelle können die oben mühsam berechneten odds-Faktoren für Raucher und Ex-Raucher im Vergleich zu den Nichtrauchern direkt abgelesen werden:

Variablen in der Gleichung

	Regressions- koeffizient B	Standard- fehler	Wald	df	Sig.	Exp(B)
Schritt 1						
abwig	,043	,021	3,998	1	,046	1,044
beweg	-,828	,265	9,790	1	,002	,437
druck	,042	,010	18,055	1	,000	1,043
stress	,509	,254	4,001	1	,045	1,663
erbe	1,856	,502	13,685	1	,000	6,400
raucher			13,069	2	,001	
raucher(1)	2,994	,834	12,898	1	,000	19,956
raucher(2)	,625	,482	1,677	1	,195	1,868
Konstante	-3,381	1,543	4,800	1	,028	,034

2.8 Interaktionen

Die aus der Varianz- oder Regressionsanalyse wohlbekannte Interaktion, wobei der Effekt eines Regressors auf das Kriterium durch einen anderen Regressor (oder auch durch mehrere) moderiert wird, lässt sich auch in ein logistisches Modell integrieren. Man kann Interaktionen zwischen beliebigen metrischen oder kategorialen Regressoren spezifizieren, wobei SPSS *Produktvariablen* als zusätzliche Regressoren in das Design aufnimmt. Wir beschränken uns anschließend auf die Zweifachwechselwirkung zwischen zwei kategorialen Regressoren. Jaccard (2001) widmet den Interaktionseffekten in der logistischen Regression ein komplettes

Buch und beschreibt weitere Spezialfälle, z.B. Dreifachwechselwirkung, Beteiligung von metrischen Regressoren.

2.8.1 Interaktionen zwischen nominalskalierten Regressoren

Wenn wir in unserem DBS-Beispiel etwa vermuten, dass der Effekt des Rauchens durch erbliche Veranlagung moderiert wird, und einen entsprechenden Interaktionseffekt spezifizieren, dann bildet SPSS zu der Veranlagungs-Kodiervariablen X_5 und den beiden Raucher-Kodiervariablen X_6 und X_7 die beiden Wechselwirkungs-Kodiervariablen X_8 und X_9 als Produkte $X_5 \cdot X_6$ und $X_5 \cdot X_7$.

2.8.1.1 Bedeutung der Regressionsgewichte bei Indikatorkodierung

Bei Verwendung der Indikatorkodierung für die beteiligten Variablen ERBE und RAUCHER (vgl. Abschnitt 2.7) erhält man insgesamt folgende Kodiervariablen:

Merkmalskombination		Kodiervariablen				
Erbliche Veranlag.	Raucher	X_5	X_6	X_7	$X_8 = X_5 \cdot X_6$	$X_9 = X_5 \cdot X_7$
Nein	Ja	0	1	0	0	0
Nein	Ex	0	0	1	0	0
Nein	Nein	0	0	0	0	0
Ja	Ja	1	1	0	1	0
Ja	Ex	1	0	1	0	1
Ja	Nein	1	0	0	0	0

Beim Signifikanztest für den Interaktionseffekt wird die Nullhypothese geprüft, dass die zu X_8 bzw. X_9 gehörigen Regressionsgewichte β_8 bzw. β_9 *beide* gleich Null sind.

Die Bedeutung der Regressionskoeffizienten in Modellen mit Interaktionseffekten hängt von den verwendeten Kodierungsmethoden ab. Um die Bedeutung für die Situation in obiger Tabelle zu erläutern, wollen wir für die Variablen X_1 bis X_4 die beliebige, aber fest gewählte Wertkombination x_1, \dots, x_4 betrachten. Ferner soll mit L_{ij} das unter diesen Randbedingungen vom Modell für die Veranlagungs-Raucher-Kombination (i, j) prognostizierte Logit bezeichnet werden. Mit der Abkürzung

$$\kappa := \beta_0 + \beta_1 x_1 + \dots + \beta_4 x_4$$

schätzt unser Modell für die sechs (ERBE \times RAUCHER) – Merkmalskombinationen (siehe obige Tabelle) folgende Logits:

$$L_{01} = \kappa + \beta_6$$

$$L_{02} = \kappa + \beta_7$$

$$L_{03} = \kappa$$

$$L_{11} = \kappa + \beta_5 + \beta_6 + \beta_8$$

$$L_{12} = \kappa + \beta_5 + \beta_7 + \beta_9$$

$$L_{13} = \kappa + \beta_5$$

Nach diesen Vorbereitungen kann die Bedeutung der Regressionskoeffizienten β_5 bis β_9 präzise angegeben werden:

- β_5 steht für den Unterschied zwischen den ERBE-Logits bei Personen aus der RAUCHER - Referenzkategorie 3 (Nein):

$$\beta_5 = L_{13} - L_{03}$$

β_5 steht also *nicht* für den „Haupteffekt“ von ERBE, sondern für den bedingten Effekt von ERBE gegeben eine bestimmte RAUCHER-Kategorie.

- β_6 steht für den Logit-Unterschied zwischen den RAUCHER-Gruppen 1 und 3 gegeben die ERBE-Kategorie 0 (Nein):

$$\beta_6 = L_{01} - L_{03}$$

β_6 steht also *nicht* für den Kontrast zwischen den RAUCHER-Kategorien 1 und 3, sondern für den bedingten Kontrast gegeben eine bestimmte ERBE-Kategorie.

- β_7 steht für den Logit-Unterschied zwischen den RAUCHER-Gruppen 2 und 3 gegeben die ERBE-Kategorie 0 (Nein):

$$\beta_7 = L_{02} - L_{03}$$

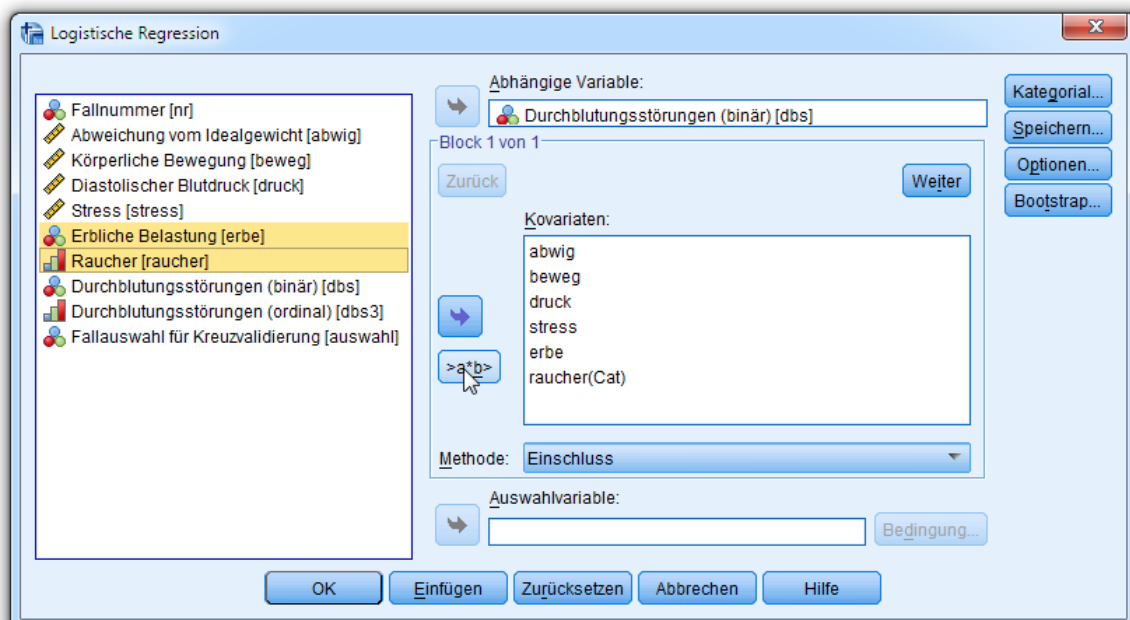
- β_8 beschreibt, wie sich der Logit-Unterschied zwischen den erblich belasteten Rauchern und den erblich belasteten Nichtrauchern unterscheidet vom Logit-Unterschied zwischen den erblich *unbelasteten* Rauchern und den erblich *unbelasteten* Nichtrauchern:

$$\beta_8 = (L_{11} - L_{13}) - (L_{01} - L_{03}) = (\beta_6 + \beta_8) - \beta_6$$

- β_9 beschreibt, wie sich der Logit-Unterschied zwischen den erblich belasteten Ex-Rauchern und den erblich belasteten Nichtrauchern unterscheidet vom Logit-Unterschied zwischen den erblich *unbelasteten* Ex-Rauchern und den erblich *unbelasteten* Nichtrauchern:

$$\beta_9 = (L_{12} - L_{13}) - (L_{02} - L_{03}) = (\beta_7 + \beta_9) - \beta_7$$

Um bei der SPSS-Prozedur zur binären logistischen Regression eine Wechselwirkung zu spezifizieren, markiert man in der Variablenliste beide Interaktionspartner, was bei festgehaltener **Strg**-Taste gelingt. Dann nimmt man über den Schalter **>a*b>** den Wechselwirkungsterm ins Design auf:



Im konkreten Beispiel muss der Prädiktor RAUCHER als kategorial definiert werden. Wenn die Regressionsgewichte die oben diskutierte Bedeutungen haben sollen, müssen die RAUCHER-Kodiervariablen nach der Indikatormethode gebildet werden. Beide Einstellungen nimmt man nach einem Mausklick

auf den Schalter **Kategorial** in der Subdialogbox **Kategoriale Variablen definieren** vor (vgl. Abschnitt 2.7).

Der Wald-Test zur ERBE \times RAUCHER – Wechselwirkung, d.h. zur kombinierten Nullhypothese

$$\beta_8 = \beta_9 = 0$$

verfehlt mit einer Überschreitungswahrscheinlichkeit von 0,227 deutlich die Signifikanzgrenze:

Variablen in der Gleichung

	Regressions- koeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1 ^a						
abwig	,040	,021	3,578	1	,059	1,041
beweg	-,857	,271	9,968	1	,002	,424
druck	,042	,010	17,752	1	,000	1,043
stress	,543	,259	4,408	1	,036	1,722
erbe	2,004	,604	11,003	1	,001	7,421
raucher			11,116	2	,004	
raucher(1)	3,720	1,119	11,043	1	,001	41,248
raucher(2)	,586	,577	1,034	1	,309	1,798
erbe * raucher			2,962	2	,227	
erbe by raucher(1)	-2,680	1,633	2,692	1	,101	,069
erbe by raucher(2)	,229	1,096	,044	1	,834	1,258
Konstante	-3,506	1,604	4,776	1	,029	,030

a. In Schritt 1 eingegebene Variablen: abwig, beweg, druck, stress, erbe, raucher, erbe * raucher .

Damit ist wohl das Haupteffektsmodell aus Abschnitt 2.7 die bessere Lösung.

2.8.1.2 Bedeutung der Regressionsgewichte bei Abweichungskodierung

Nachdem wir in Abschnitt 2.7 die Abweichungs-Kontraste für RAUCHER erläutert haben, soll nun vorgeführt werden, wie sich diese Kodierung auf die Bedeutung der Regressionskoeffizienten in einem Modell mit Wechselwirkung auswirkt. Wer sich auf die (in SPSS voreingestellte) Indikatorkodierung beschränken möchte, kann den Rest des aktuellen Abschnitts 2.8.1.2 überspringen.

Bei Verwendung der Indikatorkodierung für die Variable ERBE und der Abweichungskodierung für die Variable RAUCHER resultieren in einem Modell mit Wechselwirkung zwischen den beiden Variablen die folgenden Kodiervariablen:

Merkmalskombination		Kodiervariablen				
Erbliche Veranlag.	Raucher	X_5	X_6	X_7	$X_8 = X_5 \cdot X_6$	$X_9 = X_5 \cdot X_7$
Nein	Ja	0	1	0	0	0
Nein	Ex	0	0	1	0	0
Nein	Nein	0	-1	-1	0	0
Ja	Ja	1	1	0	1	0
Ja	Ex	1	0	1	0	1
Ja	Nein	1	-1	-1	-1	-1

Beim Signifikanztest für den Interaktionseffekt wird die Nullhypothese geprüft, dass die zu X_8 bzw. X_9 gehörigen Regressionsgewichte β_8 bzw. β_9 *beide* gleich Null sind.

Um die Bedeutung der Regressionskoeffizienten zu erläutern, wollen wir für die Variablen X_1 bis X_4 die beliebige, aber fest gewählte Wertkombination x_1, \dots, x_4 betrachten. Ferner soll mit L_{ij} das unter diesen Randbedingungen vom Modell für die Veranlagungs-Raucher-Kombination (i, j) prognostizierte Logit bezeichnet werden. Mit der Abkürzung

$$\kappa := \beta_0 + \beta_1 x_1 + \dots + \beta_4 x_4$$

schätzt unser Modell für die sechs (ERBE \times RAUCHER) – Merkmalskombinationen (siehe obige Tabelle) folgende Logits:

$$L_{01} = \kappa + \beta_6$$

$$L_{02} = \kappa + \beta_7$$

$$L_{03} = \kappa - \beta_6 - \beta_7$$

$$L_{11} = \kappa + \beta_5 + \beta_6 + \beta_8$$

$$L_{12} = \kappa + \beta_5 + \beta_7 + \beta_9$$

$$L_{13} = \kappa + \beta_5 - \beta_6 - \beta_7 - \beta_8 - \beta_9$$

Mit $\bar{L}_{0.}$ bzw. $\bar{L}_{1.}$ soll das *ungewichtete* Mittel der Logits aus den drei erblich unbelasteten bzw. belasteten Zellen bezeichnet werden, also:

$$\bar{L}_{0.} := \frac{1}{3}(L_{01} + L_{02} + L_{03})$$

$$\bar{L}_{1.} := \frac{1}{3}(L_{11} + L_{12} + L_{13})$$

Aus obigen Gleichungen folgt:

$$\bar{L}_{0.} = \kappa$$

$$\bar{L}_{1.} = \kappa + \beta_5$$

Nach diesen etwas mühsamen Vorbereitungen kann die Bedeutung der Regressionskoeffizienten β_5 bis β_9 präzise angegeben werden:

- β_5 ist die Differenz zwischen dem mittleren Logit aus den drei erblich belasteten und dem mittleren Logit aus den drei unbelasteten Zellen:

$$\beta_5 = \bar{L}_1 - \bar{L}_0$$

- β_6 beschreibt, wie sich die erblich unbelasteten Raucher vom Mittel aller erblich unbelasteten Personen unterscheiden:

$$\beta_6 = L_{01} - \bar{L}_0$$

- β_7 beschreibt, wie sich die erblich unbelasteten Ex-Raucher vom Mittel aller erblich unbelasteten Personen unterscheiden:

$$\beta_7 = L_{02} - \bar{L}_0$$

- β_8 beschreibt, wie sich die Abweichung der erblich belasteten Raucher vom Gesamtmittel aller belasteten Personen unterscheidet von der Abweichung der erblich unbelasteten Raucher vom Mittel aller erblich unbelasteten Personen:

$$\beta_8 = (L_{11} - \bar{L}_1) - (L_{01} - \bar{L}_0)$$

- β_9 beschreibt, wie sich die Abweichung der erblich belasteten Ex-Raucher vom Gesamtmittel aller erblich belasteten Personen unterscheidet von der Abweichung der erblich unbelasteten Ex-Raucher vom Mittel aller erblich unbelasteten Personen:

$$\beta_9 = (L_{12} - \bar{L}_1) - (L_{02} - \bar{L}_0)$$

Auf den Signifikanztest zum Wechselwirkungseffekt wirkt sich die Änderung der Kodierungstechnik *nicht* aus:

Variablen in der Gleichung		Regressions- koeffizientB	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1	abwig	,040	,021	3,578	1	,059	1,041
	beweg	-,857	,271	9,968	1	,002	,424
	druck	,042	,010	17,752	1	,000	1,043
	stress	,543	,259	4,408	1	,036	1,722
	erbe	1,187	,643	3,413	1	,065	3,279
	raucher			11,116	2	,004	
	raucher(1)	2,284	,727	9,869	1	,002	9,818
	raucher(2)	-,849	,471	3,246	1	,072	,428
	erbe * raucher			2,962	2	,227	
	erbe by raucher(1)	-1,863	1,083	2,959	1	,085	,155
	erbe by raucher(2)	1,046	,827	1,600	1	,206	2,847
	Konstante	-2,071	1,588	1,700	1	,192	,126

2.8.2 „Haupteffekte“ in Modellen mit Wechselwirkung

In der Tabelle **Variablen in der Gleichung** präsentiert die Prozedur LOGISTIC REGRESSION u.a. eine Signifikanzbeurteilung zur Variablen ERBE. Konkret geht es dabei um den Parameter β_5 , dessen Bedeutung aber stark vom Kodierungsschema für den ERBE-Interaktionspartner RAUCHER abhängt (vgl. Abschnitte 2.8.1.1 und 2.8.1.2). Es darf also nicht verwundern, dass in unserem Beispiel ganz verschiedene Überschreitungswahrscheinlichkeiten resultieren:

Kodierungsschema RAUCHER	Überschreitungswahrscheinlichkeit β_5
Indikator	0,001
Abweichung	0,065

Für den „Haupteffekt“ RAUCHER erhalten wir im Beispiel unabhängig von der RAUCHER-Kodierung die identische Überschreitungswahrscheinlichkeit 0,004. Ändert man jedoch die ERBE-Kodierung durch eine schlichte Vertauschung der Werte 0 und 1, dann resultiert für RAUCHER der sehr verschiedene Wert 0,528:

Variablen in der Gleichung							
		Regressions- koeffizient-B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1	abwig	,040	,021	3,578	1	,059	1,041
	beweg	-,857	,271	9,968	1	,002	,424
	druck	,042	,010	17,752	1	,000	1,043
	stress	,543	,259	4,408	1	,036	1,722
	erbe	-2,004	,604	11,003	1	,001	,135
	raucher			1,277	2	,528	
	raucher(1)	1,040	1,203	,747	1	,387	2,829
	raucher(2)	,816	,932	,767	1	,381	2,261
	erbe * raucher			2,962	2	,227	
	erbe by raucher(1)	2,680	1,633	2,692	1	,101	14,583
	erbe by raucher(2)	-,229	1,096	,044	1	,834	,795
	Konstante	-1,502	1,558	,930	1	,335	,223

Wir haben aufwendig die Bedeutung der Regressionskoeffizienten in Abhängigkeit von den gewählten Kodierungsmethoden geklärt. Mit dieser Vorarbeit lassen sich auch die dubiosen „Haupteffektbeurteilungen“ zu RAUCHER verstehen. Die mitgeteilte Überschreitungswahrscheinlichkeit gehört zum Wald-Test zur kombinierten Nullhypothese, dass die zu den RAUCHER-Kodiervariablen X_6 und X_7 gehörigen Regressionsgewichte β_6 bzw. β_7 beide gleich Null seien. Diese Hypothese bezieht sich auf zwei bedingte Kontraste (vgl. Abschnitte 2.8.1.1 und 2.8.1.2) und ist in ihrem Gehalt stark von der ERBE-Kodierung abhängig (welche Kategorie hat den Wert 0).

Über das Subkommando TEST der Prozedur NOMREG (vgl. Abschnitt 3.7) lässt sich diese kombinierte Nullhypothese explizit formulieren:

```
NOMREG dbs (BASE=LAST ORDER=ASCENDING) BY erbe raucher WITH abwig beweg druck stress
  /FULLFACTORIAL
  /INTERCEPT=INCLUDE
  /TEST (0, 0) = ALL 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0;
                  ALL 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0.
```

Es resultiert das erwartete Testergebnis:

Testergebnisse				
Durchblutungsstörungen (binär)	Kontraste	Wald	Freiheitsgrade	Signifikanz
Nein	Alle	1,277	2	,528

Damit sollte klar geworden sein, dass man bei einem logistischen Regressionsmodell mit Wechselwirkung nicht über die „Haupteffekte“ der Interaktionspartner reden sollte. Stattdessen sind bedingte Effekte von Interesse, deren genaue Bedeutung bei kategorialen Regressoren von den verwendeten Kodierungstechniken abhängt. Unter bestimmten Voraussetzungen (Zentrierung bei metrischen Regressoren, Verwendung der gewichteten Effektkodierung bei nominalskalierten Regressoren) lässt sich für einen Interaktionspartner sein mittlerer bedingter Effekt schätzen, den man als Ersatz für den nicht definierten „Haupteffekt“ verwenden kann (Baltes-Götz 2009).

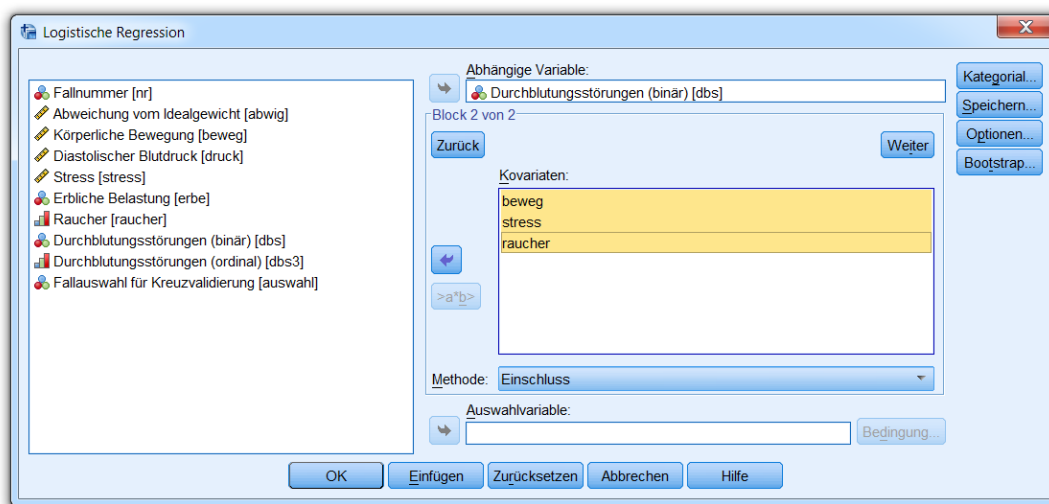
2.9 Strategien zur Modellbildung

2.9.1 Signifikanztests zu Prädiktorblöcken

Die SPSS-Prozedur zur *linearen* Regression erlaubt eine sukzessive Modellerweiterung um Blöcke von Regressoren, wobei mit einem F-Test beurteilt werden kann, ob der zuletzt einbezogene Block von Regressoren eine signifikante R^2 -Verbesserung bewirkt hat (*hierarchische Regressionsanalyse*). Ein analoger Test ist auch in der logistischen Regression verfügbar. Die zugehörige Prüfgröße ist durch die -2LL - Differenz zwischen einem erweiterten Modell und einem Ausgangsmodell definiert und folgt unter der Nullhypothese einer χ^2_{df} -Verteilung, wobei df der Anzahl zusätzlicher Parameter im erweiterten Modell entspricht (siehe Abschnitt 2.5.1).

Die folgende Dialogbox gehört zu einer logistischen Regressionsanalyse der DBS-Daten mit zwei Blöcken von Prädiktoren:

- In einem ersten Block werden die Regressoren ABWIG, DRUCK und ERBE aufgenommen.
- Nach einem Mausklick auf **Weiter** folgen in einem zweiten Block die Regressoren BEWEG, STRESS und RAUCHER:



In der Ausgabe zieht SPSS nach jedem Block Zwischenbilanz und liefert in der Tabelle mit **Omnibus-Tests der Modellkoeffizienten** neben der Bezeichnung **Block** den Likelihood-Quotiententest zum Paket mit den neu aufgenommenen Regressoren. Für den *zweiten* Block erfahren wir:

Omnibus-Tests der Modellkoeffizienten

		Chi-Quadrat	df	Sig.
Schritt 1	Schritt	52,875	4	,000
	Block	52,875	4	,000
	Modell	134,944	7	,000

Da wir in den früheren Analysen vom Basismodell (mit dem Ordinatenabschnitt als einzigem Parameter) direkt zum vollständigen Modell übergegangen sind, stimmten die **Block**- und die **Modell**-Testergebnisse stets überein.

2.9.2 Automatische Modellsuche

Innerhalb eines Blocks, der auch *alle* Regressoren umfassen darf, kann man LOGISTIC REGRESSION automatisch nach einem guten Modell suchen lassen. Es wird schrittweise anhand von Signifikanztests entschieden, ob Regressoren aufgenommen oder entfernt werden sollen, wobei zwei Strategien zur Verfügung stehen (Beschreibung nach Norušis 2005, S. 338ff):

- **Vorwärts**

Ausgehend vom Modell *ohne* den fraglichen Block wird in jedem Schritt darüber entschieden, ob ein (weiterer) Regressor aus dem Block aufgenommen werden sollte. Dazu wird für jeden noch verfügbaren Regressor die Nullhypothese geprüft, dass er nach Aufnahme in das aktuelle Modell den Regressionskoeffizienten 0 besitzt. Unterschreitet das kleinste dabei ermittelte p -Level das Aufnahmekriterium (Voreinstellung: 0,05), wird der zugehörige Prädiktor ins Modell integriert. Nach jeder Aufnahme wird durch analoge Nullhypothesenprüfungen darüber entschieden, ob vorhandene Prädiktoren überflüssig geworden sind. Überschreitet das größte ermittelte p -Level das Ausschlusskriterium (Voreinstellung: 0,10), wird der zugehörige Prädiktor entfernt. Anschließend wird das Modell neu geschätzt und nach weiteren Entlassungskandidaten gesucht. Nach einer Entlassungswelle beginnt die nächste Suche nach einem Aufnahmekandidaten. Das Verfahren endet,

...

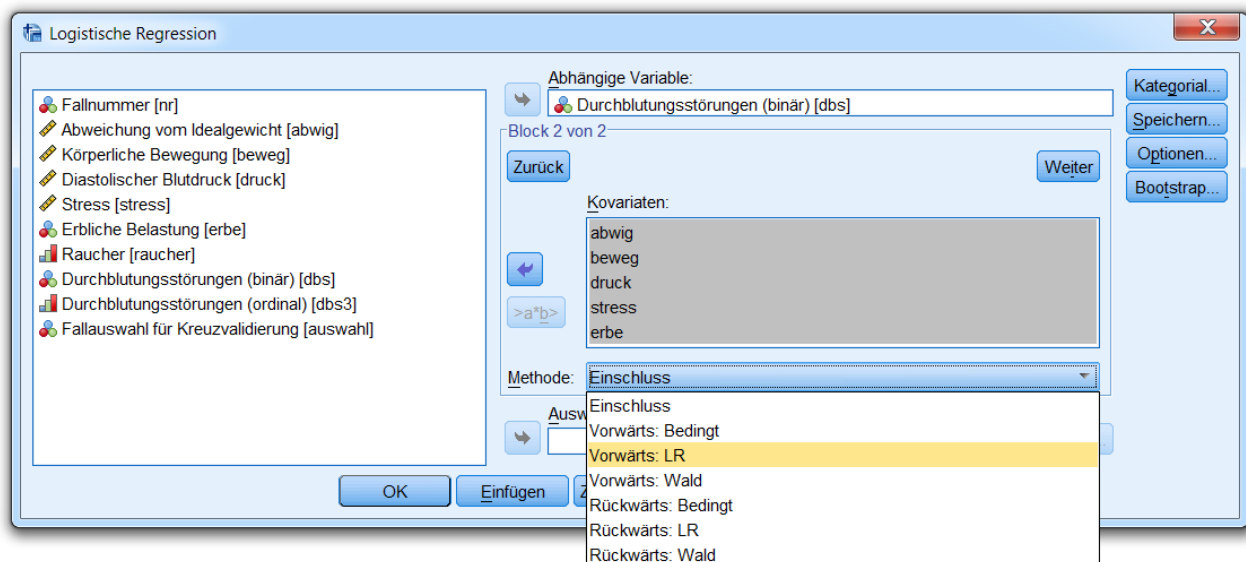
- wenn keine weitere Variable das Aufnahme- oder Entlassungskriterium erfüllt,
- oder wenn das Verfahren auf ein bereits zuvor betrachtetes Modell stößt, also zu zirkulieren droht.

- **Rückwärts**

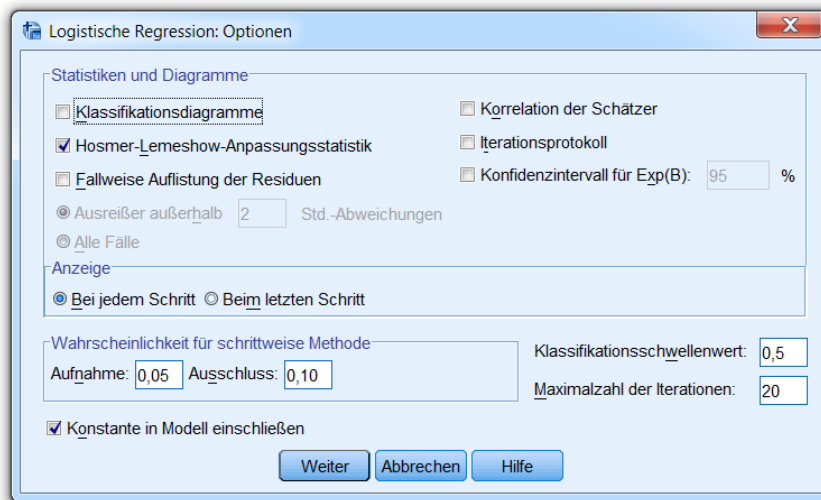
Zunächst wird der gesamte Block aufgenommen. Dann wird Schritt für Schritt geprüft, ob ein Regressor entfernt werden sollte. Nach jeder Elimination wird geprüft, ob ein früher ausgeschlossener Prädiktor wieder aufgenommen werden sollte. Bei einer relativ kleinen Stichprobe ist ein Kompletmodell mit vielen Prädiktoren wegen seiner problematischen Präzision eventuell ein schlechter Ausgangspunkt für die Modellsuche (Norusis 2005, S. 342).

Bei den Signifikanztests zur *Aufnahme* eines Prädiktors verwendet LOGISTIC REGRESSION grundsätzlich die mit geringem Rechenaufwand zu ermittelnde **Score-Statistik**. Bei den Signifikanztests zur *Entfernung* eines Prädiktors besteht eine Wahlmöglichkeit: Neben dem klar zu bevorzugenden Likelihood-Quotienten-Test bietet SPSS auch den Wald-Test und den so genannten *bedingten Test* an, wobei es sich um eine Likelihood-Quotienten - Variante mit reduziertem Rechenaufwand handelt.

Die Auswahlstrategie und das Testverfahren für die Entfernung von Prädiktoren wählt man über die **Methoden**-Liste:



Die kritischen Überschreitungswahrscheinlichkeiten für die Aufnahme bzw. für den Ausschluss von Prädiktoren werden in der **Optionen**-Subdialogbox eingestellt:



Um der automatischen Modellsuche eine knifflige Aufgabe zu stellen, erweitern wir die obige Liste der Prädiktoren (alle außer RAUCHER) um drei irrelevante Variablen, von denen zwei (IRRE1 und IRRE2) recht hoch mit relevanten Prädiktoren korrelieren, und reduzieren die Stichprobe auf die ersten 150 Fälle (siehe Abschnitt 2.6.3). Wie die folgende Tabelle mit dem Verlauf der **VORWÄRTS**-Modellsuche zeigt, nimmt LOGISTIC REGRESSION eine irrelevante Variable ins Modell auf (IRRE2) und lässt einen relevanten Prädiktor weg (ABWIG):

Variablen nicht in der Gleichung				Wert	df	Sig.
Schritt 1	Variablen	abwig		10,096	1	,001
		beweg		8,626	1	,003
		druck		14,834	1	,000
		erbe		7,470	1	,006
		irre1		5,814	1	,016
		irre2		11,217	1	,001
		irre3		3,967	1	,046
	Gesamtstatistik			44,224	7	,000
Schritt 2	Variablen	abwig		9,596	1	,002
		beweg		10,238	1	,001
		erbe		8,191	1	,004
		irre1		3,571	1	,059
		irre2		10,310	1	,001
		irre3		5,560	1	,018
	Gesamtstatistik			32,718	6	,000
Schritt 3	Variablen	abwig		4,997	1	,025
		beweg		7,354	1	,007
		erbe		10,506	1	,001
		irre1		,521	1	,470
		irre3		5,802	1	,016
	Gesamtstatistik			24,423	5	,000
Schritt 4	Variablen	abwig		5,463	1	,019
		beweg		8,359	1	,004
		irre1		,281	1	,596
		irre3		3,545	1	,060
	Gesamtstatistik			15,165	4	,004
	Variablen		abwig	2,623	1	,105
Schritt 5			irre1	,281	1	,596
			irre3	3,185	1	,074
	Gesamtstatistik			7,218	3	,065

In der Spalte **Wert** sind die Prüfgrößen des Score-Tests angegeben, den LOGISTIC REGRESSION als Aufnahmeprüfung verwendet.

2.9.3 Empfehlungen zur Modellbildung

Eine automatische Modellsuche kann zwar als explorative Technik in bestimmten Situationen sinnvoll sein, doch in der Regel sollte die Modellkonstruktion unter Verwendung von inhaltlichen und statistischen Informationen durch einen kreativen und rationalen Forscher vorgenommen werden. Dazu einige Empfehlungen (vgl. Kleinbaum 1994):

- Man startet mit *allen* Regressoren, die nach inhaltlichen Überlegungen relevant sind. Es gilt, den *omitted-variable-error* zu vermeiden (vgl. Abschnitt 2.4). Dann werden sukzessive irrelevante Variablen entfernt.
- Sind im Modell Wechselwirkungen vorgesehen, werden diese zuerst geprüft.
- Bei einer signifikanten Wechselwirkung müssen alle enthaltenen Regressoren im Modell verbleiben. Wer komplexe Überlegungen vermeiden will, sollte generell mit *hierarchisch wohlgeformten Modellen* arbeiten, die zu jedem Interaktionsterm auch alle Komponenten von niedrigerer Ordnung enthalten (Cohen et al. 2003, S. 284; Jaccard 2001).

3 Die multinomiale logistische Regression

In diesem Abschnitt betrachten wir eine Verallgemeinerung der binären logistischen Regression, welche die Analyse von nominalskalierten Kriterien mit mehr als zwei Kategorien erlaubt. Der Bequemlichkeit halber wird im Manuskript gelegentlich die Abkürzung *MLR* für die multinomiale logistische Regression verwendet.

3.1 Populationsmodell

Im Fall eines nominalskalierten Kriteriums mit J Kategorien ist für $J - 1$ Gruppen die Wahrscheinlichkeit in Abhängigkeit von den Prädiktoren X_1 bis X_M (intervallskaliert oder durch Kodierung von kategorialen Variablen entstanden) zu modellieren. Damit liegt auch die Wahrscheinlichkeit für die letzte Gruppe fest

Statt (wie im binären Fall) direkt Modellgleichungen für die Wahrscheinlichkeiten $P(Y = j)$ anzusetzen, starten wir mit einer äquivalenten Modellformulierung über die folgenden $J - 1$ Logit-Gleichungen, wobei im Nenner der Wahrscheinlichkeitsquotienten jeweils die Referenzkategorie J auftritt (vgl. Abschnitt 2.1.3):

$$\ln\left(\frac{P(Y = j)}{P(Y = J)}\right) = \beta_j \mathbf{X}, \text{ mit } \beta_j := [\beta_{j0}, \beta_{j1}, \beta_{j2}, \dots, \beta_{jM}], \quad j = 1, \dots, J - 1, \text{ und } \mathbf{X} := \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} \quad (12)$$

Dies ist eine nahe liegende Generalisierung des binären logistischen Modells, das als Spezialfall (mit $J = 2$) enthalten ist.

Durch Anwendung der Exponentialfunktion erhalten wir äquivalente Gleichungen über $J - 1$ Wahrscheinlichkeitsquotienten (odds):

$$\frac{P(Y = j)}{P(Y = J)} = e^{\beta_j \mathbf{X}} = e^{\beta_{j0} + \beta_{j1}X_1 + \beta_{j2}X_2 + \dots + \beta_{jM}X_M} = e^{\beta_{j0}} e^{\beta_{j1}X_1} \dots e^{\beta_{jM}X_M}, \quad j = 1, \dots, J - 1$$

Zu einem Prädiktor X_m sagt der Exponentialfunktionswert $e^{\beta_{jm}}$ des Parameters β_{jm} aus, um welchen Faktor sich der j -te Wahrscheinlichkeitsquotient ändert, wenn sich X_m um eine Einheit erhöht, und alle anderen Prädiktoren unverändert bleiben.

$$e^{\beta_{jm}(X_m+1)} = e^{\beta_{jm}X_m + \beta_{jm}} = e^{\beta_{jm}X_m} e^{\beta_{jm}}$$

Im binären Fall haben wir von der Effektgröße e^{β_m} eines Prädiktors gesprochen; im multinomialen Fall sind $J - 1$ Effektgrößen $e^{\beta_{jm}}$ vorhanden (vgl. Abschnitt 2.6).

Für die Wahrscheinlichkeit der Referenzkategorie J in Abhängigkeit von den Prädiktoren im Vektor \mathbf{X} behauptet das Modell:

$$\frac{1}{P(Y = J)} = \frac{\sum_{j=1}^J P(Y = j)}{P(Y = J)} = 1 + \sum_{j=1}^{J-1} \frac{P(Y = j)}{P(Y = J)} = 1 + \sum_{j=1}^{J-1} e^{\beta_j \mathbf{X}} \Rightarrow P(Y = J) = \frac{1}{1 + \sum_{j=1}^{J-1} e^{\beta_j \mathbf{X}}}$$

Damit gilt für die anderen $J - 1$ Wahrscheinlichkeiten:

$$P(Y = j) = e^{\beta_j \mathbf{X}} P(Y = J) = \frac{e^{\beta_j \mathbf{X}}}{1 + \sum_{j=1}^{J-1} e^{\beta_j \mathbf{X}}}, \quad j = 1, \dots, J-1 \quad (13)$$

Diese Darstellungen der Wahrscheinlichkeiten liefern sofort wiederum das Logit-Modell (12):

$$\ln\left(\frac{P(Y = j)}{P(Y = J)}\right) = \ln\left(\frac{\frac{e^{\beta_j \mathbf{X}}}{1 + \sum_{j=1}^{J-1} e^{\beta_j \mathbf{X}}}}{\frac{1}{1 + \sum_{j=1}^{J-1} e^{\beta_j \mathbf{X}}}}\right) = \ln(e^{\beta_j \mathbf{X}}) = \beta_j \mathbf{X}, \quad j = 1, \dots, J-1$$

3.2 Stichprobenmodell

Beobachtet man bei N Fällen mit den Prädiktorwertekombinationen \mathbf{x}_i jeweils die Kriteriumskategorie, kommt folgendes MLR-Stichprobenmodell zum Einsatz:

- Es sind N unabhängige Zufallsvariablen Y_i mit Werten aus der Menge $\{1, 2, \dots, J\}$ vorhanden.
- Für die Wahrscheinlichkeiten der J Kriteriumskategorien gilt bei Y_i :

$$P(Y_i = j) = \begin{cases} \frac{e^{\beta_j \mathbf{x}_i}}{1 + \sum_{j=1}^{J-1} e^{\beta_j \mathbf{x}_i}} & \text{für } j = 1, \dots, J-1 \\ \frac{1}{1 + \sum_{j=1}^{J-1} e^{\beta_j \mathbf{x}_i}} & \text{für } j = J \end{cases} \quad \text{mit } \beta_j \mathbf{x}_i = \beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \dots + \beta_{jM}x_{iM}$$

Bei K mehrfach besetzten Prädiktorwertekombination kann man die Verhältnisse in der Stichprobe auch so beschreiben: Wenn für die Prädiktorwertekombination \mathbf{x}_k mit h_k die Gruppenstärke, mit \tilde{y}_{kj} die beobachtete Häufigkeit der j -ten Kriteriumskategorie und mit \tilde{Y}_{kj} die zugehörige Zufallsvariable bezeichnet wird, dann sind bei unabhängigen Beobachtungen die J -dimensionalen Zufallsvariablen

$$\tilde{Y}_k := (\tilde{Y}_{k1}, \tilde{Y}_{k2}, \dots, \tilde{Y}_{kJ}), \quad k = 1, \dots, K$$

$M(h_k; \pi_{k1}, \pi_{k2}, \dots, \pi_{kJ})$ - multinomialverteilt, und für die Wahrscheinlichkeiten π_{kj} behauptet das Modell der multinomialen logistischen Regression:

$$\pi_{kj} = \frac{e^{\beta_j \mathbf{x}_k}}{1 + \sum_{j=1}^{J-1} e^{\beta_j \mathbf{x}_k}}, \quad j = 1, \dots, J-1, \quad \text{und} \quad \pi_{kJ} = \frac{1}{1 + \sum_{j=1}^{J-1} e^{\beta_j \mathbf{x}_k}}$$

Bei zentralen Ergebnissen einer logistischen Regressionsanalyse (z.B. bei der Parameterschätzung) bestehen zwischen Algorithmen für individuelle oder gruppierte Daten jedoch nur rechentechnische Unterschiede. Im Manuskript wird speziell darauf hingewiesen, wenn *mehrfach* besetzte Prädiktorwertekombinationen erforderlich oder schädlich sind.

Wie schon in Abschnitt 2.1.2 erläutert, ist die logistische Regressionsanalyse auch dann anwendbar, wenn keine Zufallsstichprobe aus der Gesamtpopulation vorliegt, sondern aus den J Teilpopulationen zu den

Kriteriumsausprägungen jeweils eine Stichprobe mit individueller Quote gezogen wurde. Dieses Vorgehen kommt z.B. dann in Frage, wenn sich die Teilpopulationen in der Größe erheblich unterscheiden.

Zur erforderlichen Stichprobengröße können die Angaben aus Abschnitt 2.1.2 übernommen werden.

Ist die Unabhängigkeit sichergestellt und die Stichprobengröße ausreichend, kann bei einem MLR-Modell also nur die Linearität in den Logit-Gleichungen (12) verletzt sein.

Wir werden im weiteren Verlauf von Abschnitt 3 ohne Beschränkung der Allgemeinheit aus Gründen der terminologischen Vereinfachung meist den Fall $J = 3$ betrachten, so dass zwei Vektoren β_1 und β_2 mit zusammen $2(M + 1)$ Regressionskoeffizienten beteiligt sind.

3.3 Anwendungsbeispiel

Wie schon im binären Fall wird auch zur Demonstration der multinomialen logistischen Regression ein künstlicher Datensatz verwendet. Zwar kommt hier weniger Spannung auf als bei echten Daten, doch beim Erzeugen der Daten beschäftigt man sich gründlich mit den vom Modell behaupteten empirischen Gesetzmäßigkeiten und gewinnt einen guten Eindruck von den potentiellen Einsatzmöglichkeiten des Verfahrens.

Wir stellen uns vor, dass in einer Studie Bewohner einer Kleinstadt nach ihrer Vorliebe für eine von drei Kneipen befragt werden. Die so entstehende nominalskalierte Kriteriumsvariable mit drei Kategorien soll den Namen KNEIPE erhalten. Außerdem verwenden wir ...

- als kategorialen Prädiktor eine fingierte Geschlechtsvariable (GESCHL) mit folgender Kodierung:
 - 1: Frau
 - 0: Mann
- als metrischen Prädiktor eine fingierte Altersvariable (ALTER)

Im Vergleich zur Referenzkneipe Nummer 3 spricht die erste Kneipe eher Publikum reiferen Alters an, so dass der logarithmierte Wahrscheinlichkeitsquotient

$$\ln\left(\frac{P(\text{KNEIPE} = 1)}{P(\text{KNEIPE} = 3)}\right)$$

mit dem Alter wächst. In der zweiten Kneipe fühlen sich vor allem junge Leute wohl, so dass deren Präferenz gegenüber der dritten Kneipe

$$\ln\left(\frac{P(\text{KNEIPE} = 2)}{P(\text{KNEIPE} = 3)}\right)$$

mit dem Alter sinkt. Außerdem soll die zweite Kneipe noch eine spezielle Attraktion für Frauen besitzen, z.B. durch einen Preisnachlass auf alle Getränke.

Wir legen in unserer künstlichen Population über zwei Logit-Gleichungen das folgende wahre Modell fest:

$$\begin{aligned}\ln\left(\frac{P(\text{KNEIPE} = 1)}{P(\text{KNEIPE} = 3)}\right) &= -3,5 + 0,07 \text{ ALTER} \\ \ln\left(\frac{P(\text{KNEIPE} = 2)}{P(\text{KNEIPE} = 3)}\right) &= 1,5 - 0,05 \text{ ALTER} + 1,3 \text{ GESCHL}\end{aligned}$$

Gemäß Abschnitt 3.1 gilt z.B. für den Effekt (das *odds ratio*) von Geschlecht in der zweiten Gleichung:

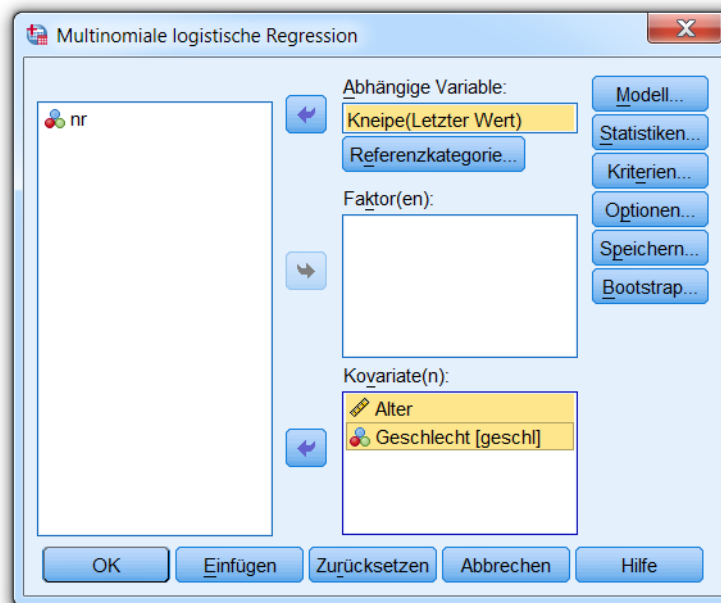
$$e^{\beta_{2G}} = \frac{P(\text{KNEIPE} = 2 | \text{ALTER}, \text{GESCHL} = 1)}{P(\text{KNEIPE} = 2 | \text{ALTER}, \text{GESCHL} = 0)} \bigg/ \frac{P(\text{KNEIPE} = 3 | \text{ALTER}, \text{GESCHL} = 1)}{P(\text{KNEIPE} = 3 | \text{ALTER}, \text{GESCHL} = 0)} \approx 3,67$$

Eine Zufallsstichprobe mit 125 Fällen aus der künstlichen Population finden Sie in der Datei **KNEIPE.SAV** an der im Vorwort vereinbarten Stelle.

Die multinomiale logistische Regressionsanalyse der Beispieldaten (durch die SPSS-Prozedur NOMREG) wird nach

Analyse > Regression > Multinomial logistisch

folgendermaßen angefordert:



Über den Schalter **Statistiken** aktivieren wir noch die folgenden optionalen Ausgaben:

- **Klassifikationsmatrix**
- **Anpassungsgüte**

Statt die (dichotome!) GESCHL-Variable (wie schon den Prädiktor ERBE im DBS-Beispiel) als **Kovariate** zu behandeln, kann sie auch als **Faktor** einbezogen werden. Dabei kommt es zu sachlich irrelevanten Änderungen der Schätzergebnisse: Die GESCHL-Regressionskoeffizienten in den beiden Logit-Gleichungen wechseln das Vorzeichen, und bei den Konstanten kommt es zu kompensierenden Änderungen. Mit der folgenden GESCHL-Kodierung

- 1: Frau
- 2: Mann

lassen sich die oberflächlich Abweichungen der Schätzergebnisse verhindern. NOMREG verwendet nämlich bei nominalskalierten Prädiktoren die Indikator- (alias Dummy-) Kodierung (vgl. Abschnitt 2.7), wobei die Kategorie mit dem höchsten Wert als Referenz fungiert, also den Wert 0 bei allen Kodiervariablen erhält. Bei der eben vorgeschlagenen GESCHL-Rekodierung resultiert letztlich eine Indikatorvariable mit den erwünschten Werten:

- 1: Frau
- 0: Mann

Weil die Behandlung als Faktor bei dichotomen Regressoren keinen Vorteil bringt, kann man sich den Aufwand sparen.

3.4 Parameterschätzung

Im Unterschied zu Abschnitt 2.3.1 betrachten wir anschließend *keine* Individualdaten, sondern Prädiktorwertekombinationen, weil die SPSS-Prozedur NOMREG grundsätzlich mit aggregierten Daten rechnet. Daher gelangen wir zu einer Likelihood-Funktion, die sich leicht (um eine Konstante) vom Gegenstück für Individualdaten unterscheidet, sofern tatsächlich mehrfach besetzte Prädiktorwertekombinationen vorhanden sind.

Für die Prädiktorwertekombination \mathbf{x}_k werde mit h_k die Gruppenstärke und mit \tilde{y}_{kj} die beobachtete Häufigkeit der j -ten Kriteriumskategorie bezeichnet, die als Realisation einer Zufallsvariable \tilde{Y}_{kj} aufgefasst werden kann. Für eine konkrete Stichprobenrealisation $(\tilde{y}_{11}, \tilde{y}_{12}, \tilde{y}_{13}, \tilde{y}_{21}, \dots, \tilde{y}_{K1}, \tilde{y}_{K2}, \tilde{y}_{K3})$ der Zufallsvariablen \tilde{Y}_{kj} zu K Prädiktorwertekombinationen und 3 Kriteriumskategorien ergibt sich aus dem Modell der multinomialen logistischen Regression folgende Wahrscheinlichkeit:

$$\prod_{k=1}^K \frac{h_k!}{\tilde{y}_{k1}! \cdot \tilde{y}_{k2}! \cdot \tilde{y}_{k3}!} \left[\frac{e^{\beta_1 \mathbf{x}_k}}{1 + e^{\beta_1 \mathbf{x}_k} + e^{\beta_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k1}} \left[\frac{e^{\beta_2 \mathbf{x}_k}}{1 + e^{\beta_1 \mathbf{x}_k} + e^{\beta_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k2}} \left[\frac{1}{1 + e^{\beta_1 \mathbf{x}_k} + e^{\beta_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k3}}$$

Auch in Abschnitt 2.3.1 zum binären Modell hätte man analog die Wahrscheinlichkeit

$$P(\tilde{Y}_1 = \tilde{y}_1, \dots, \tilde{Y}_K = \tilde{y}_K)$$

an Stelle von

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_N = y_N), y_i \in \{0,1\}$$

betrachten können. Dort wurden nicht zuletzt mit Rücksicht auf die SPSS-Prozedur LOGISTIC REGRESSION Individualdaten bevorzugt. Beide Ansätze unterscheiden sich nur um eine triviale Konstante, doch sorgen die unterschiedlichen Likelihood-Berechnungen für einige Verwirrung, so dass wir das Thema in Abschnitt 3.8 noch einmal aufgreifen.

Mit den frei schätzbaren Vektoren \mathbf{b}_1 und \mathbf{b}_2 an Stelle von β_1 und β_2 erhalten wir für das multinomiale Modell die folgende **Likelihood-Funktion**:

$$L(\mathbf{b}_1, \mathbf{b}_2) := \prod_{k=1}^K \frac{h_k!}{\tilde{y}_{k1}! \cdot \tilde{y}_{k2}! \cdot \tilde{y}_{k3}!} \left[\frac{e^{\mathbf{b}_1 \mathbf{x}_k}}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k1}} \left[\frac{e^{\mathbf{b}_2 \mathbf{x}_k}}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k2}} \left[\frac{1}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k3}}$$

Gesucht sind Parametervektoren $\hat{\beta}_1 = (\hat{\beta}_{10}, \hat{\beta}_{11}, \dots, \hat{\beta}_{1M})$ und $\hat{\beta}_2 = (\hat{\beta}_{20}, \hat{\beta}_{21}, \dots, \hat{\beta}_{2M})$, welche die Likelihood-Funktion maximieren. Um die Extremwertbestimmung zu erleichtern, die mit einem iterativen Verfahren (z.B. Newton-Raphson) erfolgt, geht man zum Logarithmus über:

$$\begin{aligned} LL(\mathbf{b}_1, \mathbf{b}_2) &:= \ln[L(\mathbf{b}_1, \mathbf{b}_2)] \\ &= \ln \left(\prod_{k=1}^K \frac{h_k!}{\tilde{y}_{k1}! \cdot \tilde{y}_{k2}! \cdot \tilde{y}_{k3}!} \right) + \sum_{k=1}^K \left(\tilde{y}_{k1} \ln \left[\frac{e^{\mathbf{b}_1 \mathbf{x}_k}}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right] + \tilde{y}_{k2} \ln \left[\frac{e^{\mathbf{b}_2 \mathbf{x}_k}}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right] + \tilde{y}_{k3} \ln \left[\frac{1}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right] \right) \\ &= \ln \left(\prod_{k=1}^K \frac{h_k!}{\tilde{y}_{k1}! \cdot \tilde{y}_{k2}! \cdot \tilde{y}_{k3}!} \right) + \sum_{k=1}^K \left(\tilde{y}_{k1} \mathbf{b}_1 \mathbf{x}_k + \tilde{y}_{k2} \mathbf{b}_2 \mathbf{x}_k - h_k \ln(1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}) \right) \end{aligned}$$

Als Ergebnis erhält man die ML-Schätzungen der Parameter, aus denen sich sofort die ML-Schätzungen der Wahrscheinlichkeiten $P(Y = j)$ zu den Prädiktorwertekombination \mathbf{x}_k ergeben.

Für unser Kneipenbeispiel mit den bekannten Populationsverhältnissen liefert die Zufallsstichprobe folgende Schätzer:

Parameterschätzer								
Kneipe ^a	B	Standardfehler	Wald	Freiheitsgrade	Signifikanz	Exp(B)	95% Konfidenzintervall für Exp(B)	
							Untergrenze	Obergrenze
1 Konstanter Term	-3,083	1,034	8,886	1	,003			
Alter	,067	,021	10,559	1	,001	1,069	1,027	1,113
geschl	-,271	,484	,314	1	,575	,762	,295	1,969
2 Konstanter Term	1,485	,878	2,861	1	,091			
Alter	-,048	,020	5,624	1	,018	,953	,916	,992
geschl	1,085	,470	5,329	1	,021	2,961	1,178	7,441

a. Die Referenzkategorie lautet: 3.

Trotz der nicht allzu großen Stichprobe liegen alle Schätzer relativ nahe bei den wahren Populationsparametern.

Weil unser Modell garantiert stimmt, dürfen wir uns schon vor einer Modellgültigkeitsprüfung mit der Parameterinterpretation beschäftigen. Über die erste Logit-Gleichung erfahren wir aus den Schätzergebnissen, dass bei einem Alterszuwachs von 10 Jahren das Wahrscheinlichkeitsverhältnis

$$\frac{P(\text{KNEIPE} = 1)}{P(\text{KNEIPE} = 3)}$$

um den Faktor $e^{10 \cdot 0,067} \approx 1,95$ wächst.

Zum Einfluss des Prädiktors GESCHL können wir aus den Schätzergebnissen für die zweite Logit-Gleichung direkt ablesen, dass bei Frauen das Wahrscheinlichkeitsverhältnis

$$\frac{P(\text{KNEIPE} = 2)}{P(\text{KNEIPE} = 3)}$$

ungefähr um den Faktor 2,96 höher liegt als bei Männern.

3.5 Modellgültigkeit

Wie Sie bereits wissen, sind Modellgültigkeitstests über die Pearson- bzw. Devianz-Statistik nur bei Daten mit mehrfach besetzten Prädiktorwertkombinationen möglich. Damit ist bei unseren Daten nicht zu rechnen, weil mit ALTER ein relativ präzise gemessener metrischer Regressor beteiligt ist. SPSS kritisiert, dass bei zahlreichen Prädiktorwertkombinationen leere Kriteriumszellen auftreten:

Warnungen

Es gibt 105 (52,2%) Zellen (d.h. Niveaus der abhängigen Variablen für Teilgesamtheiten) mit der Häufigkeit Null.

Kritisch für die Aussagekraft der Modellgültigkeitstests über die Pearson- bzw. Devianz-Statistik sind zwar nicht geringe *beobachtete* Häufigkeiten, sondern geringe *erwartete* Häufigkeiten, doch sprechen leere Kriteriumszellen bei zahlreichen Prädiktorwertkombinationen gegen ausreichende erwartete Häufigkeiten unter dem zu prüfenden Modell. Die eigentlich unzulässigen Tests kommen aber trotzdem zum korrekten Ergebnis:

Güte der Anpassung

	Chi-Quadrat	Freiheitsgrade	Signifikanz
Pearson	133,111	128	,361
Abweichung	135,183	128	,315

Durch dieses Beispiel sollen Sie aber keinesfalls dazu ermuntert werden, die beiden Tests unverdrossen auf Individualdaten (mit überwiegend nur einfach besetzten Prädiktorwertekombinationen) anzuwenden. Die bessere Alternative besteht darin, für jede Logit-Gleichung ein binäres Partialmodell mit einer geeignet eingeschränkten Stichprobe schätzen zu lassen, um den hier verfügbaren Hosmer-Lemeshow-Test nutzen zu können (vgl. Hosmer & Lemeshow 2000, S. 281). Dieses Verfahren ist kaum bedenklich, weil aus binären Partialmodellen konsistente¹ Schätzer resultieren, die in vielen Situationen sogar relativ effizient² sind (Hosmer & Lemeshow 2000, S. 278). In unserem Fall resultiert für die erste Logit-Gleichung:

Hosmer-Lemeshow-Test			
Schritt	Chi-Quadrat	df	Sig.
1	11,474	8	,176

und für die zweite Gleichung:

Hosmer-Lemeshow-Test			
Schritt	Chi-Quadrat	df	Sig.
1	7,255	8	,509

Dass alle Tests *für* unser Modell votieren, durften wir erwarten.

Über das Kontrollkästchen **Zellwahrscheinlichkeiten** in der **Statistiken**-Subdialogbox kann man NOMREG dazu bewegen, die Pearson-Residuen des Modells zu tabellieren. Diese sind jedoch nur bei *mehrfach* besetzten Prädiktorwertekombinationen sinnvoll zu interpretieren (siehe Abschnitt 2.4.2.1), so dass wir bei unserem Beispiel auf eine Inspektion verzichten.

3.6 Beurteilung der Modellrelevanz

Die globale Nullhypothese

$$H_0 : \beta_{j1} = \beta_{j2} = \dots = \beta_{jM} = 0, j = 1, \dots, J - 1$$

wird per Likelihood-Quotienten-Test deutlich zurückgewiesen:

Informationen zur Modellanpassung				
Modell	Kriterien für die Modellanpassung	Likelihood-Quotienten-Tests		
	-2 Log-Likelihood	Chi-Quadrat	Freiheitsgrade	Signifikanz
Nur konstanter Term	221,291			
Endgültig	178,819	42,472	4	,000

Weil die globale Nullhypothese in zwei Logit-Gleichungen jeweils zwei Parameter (zu ALTER und GESCHL) negiert, resultiert eine approximativ χ^2 - verteilte Prüfgröße mit 4 Freiheitsgraden.

Die Pseudo-R²-Statistiken basieren auf den Likelihood-Werten verschiedener Modelle und sind im multinomialen Fall analog zum binären einsetzbar (vgl. Abschnitt 2.5.2). Wir erhalten:

Pseudo-R-Quadrat	
Cox und Snell	,288
Nagelkerke	,324
McFadden	,155

Ebenso uneingeschränkt vom binären Fall zu übernehmen ist die Klassifikationstabelle:

¹ Konsistente Schätzer kommen den wahren Werten mit wachsender Stichprobengröße beliebig nahe.

² Effiziente Schätzer besitzen im Vergleich zu potentiellen Alternativen eine relativ kleine Varianz.

Klassifikation				
Beobachtet	Vorhergesagt			
	1	2	3	Prozent richtig
1	28	7	6	68,3%
2	6	31	8	68,9%
3	11	15	13	33,3%
Prozent insgesamt	36,0%	42,4%	21,6%	57,6%

Das Publikum der Kneipe 3 kann offenbar mit den verfügbaren Prädiktoren schlecht klassifiziert werden.

3.7 Beurteilung der einzelnen Regressoren

Zur Signifikanzbeurteilung der einzelnen Parameter stehen für jede Logit-Gleichung Wald-Tests zur Verfügung. In unserem Beispiel werden alle in der künstlichen Population von 0 verschiedenen Parameter als signifikant beurteilt (siehe Tabelle in Abschnitt 3.4).

Zusätzlich liefert NOMEQ auch Likelihood-Quotienten-Tests, die *beide* Logit-Gleichungen simultan berücksichtigen:

Likelihood-Quotienten-Tests				
Effekt	Kriterien für die Modellanpassung	Likelihood-Quotienten-Tests		
	-2 Log-Likelihood für reduziertes Modell	Chi-Quadrat	Freiheitsgrade	Signifikanz
Konstanter Term	201,185	22,365	2	,000
Alter	212,144	33,325	2	,000
geschl	187,186	8,366	2	,015

Über das NOMREG-Subkommando TEST, dessen Leistungen nur per Syntax verfügbar sind, lassen sich Hypothesen über beliebige Linearkombinationen von Parametern prüfen. Z.B. kann man:

- zwei (oder mehr) Parameter auf Identität testen,
- einen Parameter gegen einen speziellen Wert testen (statt gegen 0),
- mehrere Hypothesen simultan testen.

Mit folgender Syntax

```
NOMREG
  kneipe WITH alter geschl
  /MODEL /INTERCEPT = INCLUDE /PRINT = FIT PARAMETER SUMMARY LRT
  /TEST (2) = ALL 0 0 1.
```

wird für das Kneipenbeispiel die Hypothese

$$H_0: \beta_{2G} = 2$$

getestet, die einen bestimmten Geschlechtseffekt in Logit-Gleichung 2 behauptet. NOMREG protokolliert zunächst seine Interpretation der Syntax:

Kontrastkoeffizienten	
	C1
Konstanter Term	0
Alter	0
geschl	1

In der folgenden Tabelle interessieren nur die Ergebnisse für die *zweite* Logit-Gleichung:

Kontrastergebnisse

Kneipe	Kontraste	Schätzung	Standardfehler	Testwert	Wald	Freiheitsgrade	Signifikanz	Exp (Schätzung)	95% Konfidenzintervall für Exp (Schätzung)	
									Untergrenze	Obergrenze
1	C1	-,271	,484	2,000	22,015	1	,000	,762	,295	1,969
2	C1	1,085	,470	2,000	3,783	1	,052	2,961	1,178	7,441

Der Wald-Test scheitert knapp an der Signifikanzgrenze, so dass die (tatsächlich falsche) Nullhypothese nicht abgelehnt werden kann.

Neben den separaten Wald-Tests für die einzelnen Logit-Gleichungen liefert NOMREG auch einen simultanen Wald-Test über *alle* Gleichungen, der aber im Beispiel nicht interessiert:

Testergebnisse

Kneipe	Kontraste	Wald	Freiheitsgrade	Signifikanz
Alle	C1	22,027	2	,000

Nähere Informationen zum NOMREG-Subkommando TEST findet man in der SPSS-Hilfe über

Hilfe > Befehlssyntax-Referenz > NOMREG

3.8 Log-Likelihood - Varianten

Bei Vergleichen mit der Literatur oder mit den Ergebnisse anderer Programme (z.B. SAS, Stata) stößt man bei den mitgeteilten Log-Likelihood – Werten zu verschiedenen Modellen auf drastische Unterschiede, die aber nicht auf Fehler sondern auf unterschiedliche Definitionen zurückgehen. Sogar die beiden SPSS-Prozeduren LOGISTIC REGRESSION und NOMREG unterscheiden sich in dieser Hinsicht. Es scheint also erforderlich, etwas Mühe in die Aufklärung der Unterschiede zu investieren.

Aus Abschnitt 3.4 kennen wir schon folgende Darstellung für die Likelihood im Modell der multinomialen logistischen Regression:

$$\begin{aligned}
 L(\mathbf{b}_1, \mathbf{b}_2) &= \prod_{k=1}^K \frac{h_k!}{\tilde{y}_{k1}! \cdot \tilde{y}_{k2}! \cdot \tilde{y}_{k3}!} \left[\frac{e^{\mathbf{b}_1 \mathbf{x}_k}}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k1}} \left[\frac{e^{\mathbf{b}_2 \mathbf{x}_k}}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k2}} \left[\frac{1}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k3}} \\
 &= \underbrace{\prod_{k=1}^K \frac{h_k!}{\tilde{y}_{k1}! \cdot \tilde{y}_{k2}! \cdot \tilde{y}_{k3}!}}_{\text{Konstante}} \underbrace{\prod_{k=1}^K \left[\frac{e^{\mathbf{b}_1 \mathbf{x}_k}}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k1}} \left[\frac{e^{\mathbf{b}_2 \mathbf{x}_k}}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k2}} \left[\frac{1}{1 + e^{\mathbf{b}_1 \mathbf{x}_k} + e^{\mathbf{b}_2 \mathbf{x}_k}} \right]^{\tilde{y}_{k3}}}_{\text{Kern}}
 \end{aligned}$$

Weil der linke Faktor offenbar nicht von den Parametern abhängt und daher sowohl bei der ML-Schätzung wie auch bei Modellvergleichen irrelevant ist, lassen manche Programme bzw. Autoren diese multinomiale Konstante unter den Tisch fallen und beschränken sich auf den (oft als *Kern* bezeichneten) rechten Faktor. Ist keine einzige Prädiktorwertekombination mehrfach besetzt ($K = N$), dann ist die multinomiale Konstante gleich 1 und ihre Berücksichtigung folglich irrelevant.

Auch die von LOGISTIC REGRESSION praktizierte Verarbeitung auf *Individualdaten* und Betrachtung der Wahrscheinlichkeit $P(Y_1 = y_1, Y_2 = y_2, \dots, Y_N = y_N)$ an Stelle von $P(\tilde{Y}_1 = \tilde{y}_1, \tilde{Y}_2 = \tilde{y}_2, \dots, \tilde{Y}_K = \tilde{y}_K)$ führt zu einer Likelihood-Funktion *ohne* das binomiale Analogon zur obigen Konstanten.

Im folgenden Beispiel aus Hosmer & Lemeshow (2000, S. 266) zum Effekt von familiärer Erfahrung mit Brustkrebs (Variable HIST) auf die Teilnahme an der Früherkennung per Mammographie (Variable ME)

HIST * ME Kreuztabelle

			ME			Gesamt
			Never	Within 1 Year	Over 1 Year	
HIST	No	Anzahl	220	85	63	368
		% innerhalb von HIST	59,8%	23,1%	17,1%	100,0%
	Yes	Anzahl	14	19	11	44
		% innerhalb von HIST	31,8%	43,2%	25,0%	100,0%
Gesamt	Anzahl	234	104	74	412	
	% innerhalb von HIST	56,8%	25,2%	18,0%	100,0%	

liefert die SPSS-Prozedur NOMREG u.a. folgende Tabelle mit -2LL - Werten:

Informationen zur Modellanpassung

Modell	Kriterien für die Modellanpassung	Likelihood-Quotienten-Tests		
	-2 Log-Likelihood	Chi-Quadrat	Freiheitsgrade	Signifikanz
Nur konstanter Term	32,515			
Endgültig	19,657	12,858	2	,002

Bei Hosmer & Lemeshow (2000, S. 267) findet sich demgegenüber eine *Log-Likelihood* von -396,17.

Es zeigt sich, dass die SPSS-Prozedur NOMREG die mit (-2) vormultiplizierte „komplette“ Likelihood ausgibt, während sich Hosmer & Lemeshow auf den Kern beschränken:

$$LL(Parameter) = \ln(multinomialeKonstante) + \ln(Kern)$$

$$= \ln\left(\frac{368!}{220! 85! 63!} \frac{44!}{14! 19! 11!}\right) + \ln(Kern)$$

$$= 386,341 - 396,17$$

$$= -9,829$$

4 Die ordinale logistische Regression

Zur Aufklärung von ordinalen Kriterien durch ein Modell mit nominalen und/oder metrische Regressoren steht in SPSS die Prozedur **PLUM** (*PoLytomous Universal Model*) bereit. Neben der ordinalen *logistischen* Regression, auf die wir uns beschränken, werden auch ordinale Varianten anderer Modelle unterstützt (z.B. das ordinale Probit-Modell).

Der Bequemlichkeit halber wird im Manuskript gelegentlich die Abkürzung *OLR* für die ordinale logistische Regression verwendet.

Prinzipiell kann ein ordinale Kriterium auch mit der multinomialen logistischen Regression (MLR) untersucht werden, doch bringt die OLR erhebliche Vorteile, weil sie die ordinale Struktur im Kriterium berücksichtigt und daher deutlich weniger Parameter benötigt:

- Die Ergebnisse sind leichter zu interpretieren.
- Die Hypothesentests haben eine größere Teststärke.

Allerdings können die im ordinalen Modell enthaltenen Restriktionen auch falsch sein, z.B. wegen einer falschen Annahme über die Anordnung der Kriteriumskategorien.

Für die Erweiterung der BLR auf ordinale Kriterien sind mehrere Vorschläge erarbeitet worden (siehe Allison 1999, S. 133ff; Hosmer & Lemeshow 2000, S. 288ff), wobei wir uns auf das meist verwendete und in der SPSS-Prozedur PLUM unterstützte *kumulative Logit-Modell* (McCullagh 1980) beschränken.

4.1 Das kumulative Logit-Modell

Bei einem ordinalen Kriterium mit J Stufen erklärt das kumulative Logit-Modell in $J - 1$ Gleichungen jeweils die *kumulative* Wahrscheinlichkeit $P(Y \leq j)$ für $j = 1, 2, \dots, J - 1$:

$$P(Y \leq j) = \frac{e^{\alpha_j - \beta \mathbf{X}}}{1 + e^{\alpha_j - \beta \mathbf{X}}}, \text{ mit } \boldsymbol{\beta} := [\beta_1, \beta_2, \dots, \beta_M] \text{ und } \mathbf{X} := \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix}, j = 1, \dots, J - 1 \quad (14)$$

Jede einzelne Gleichung ist im Wesentlichen von derselben Form, die schon im binären Fall zum Einsatz kam (siehe Gleichung (1)). Offenbar ist das BLR-Modell als Spezialfall mit $J = 2$ im OLR-Modell enthalten.

Allerdings ist eine terminologische Abweichung im Vergleich zum BLR-Modell zu beachten: Im Modell (14) wird die Wahrscheinlichkeit der Zugehörigkeit zu einer Klasse mit Index *kleiner* oder gleich j modelliert. Andererseits erwartet man von einem ordinalen Regressionsmodell intuitiv, dass ein Prädiktor mit „kriteriums-steigernder Wirkung“ einen *positiven* Regressionsparameter besitzt. Als einfache Lösung hat sich eingebürgert, die Regressionsparameter im Modell (bis auf die konstanten Terme) mit einem *negativen* Vorzeichen zu versehen. Damit behauptet die j -te Modellgleichung bei Wertsteigerungen eines positiv wirkenden Prädiktors eine sinkende Wahrscheinlichkeit der Kriteriumskategorien mit Index *kleiner* oder gleich j . Man kann sich aber nicht generell darauf verlassen, dass alle Statistik-Programmpakete mit dieser Modellvariante arbeiten. Die SAS-Prozedur LOGISTIC tut es z.B. *nicht* und liefert daher im Vergleich zur SPSS-Prozedur PLUM betragsgleiche Schätzer mit umgekehrten Vorzeichen.

Dass der (im Vergleich zu früheren analogen Definitionen um β_0 gekürzte) Parametervektor $\boldsymbol{\beta}$ *ohne* Index für die Gleichung auskommt, stellt eine zentrale Eigenschaft des kumulativen Modells dar: Während es für jede Gleichung einen frei schätzbaren konstanten Term α_j enthält, sind die restlichen Regressionskoeffizienten in allen Gleichungen *identisch*. Diese Annahme *paralleler Regressionen* stellt die entscheidende Restriktion der kumulativen OLR gegenüber der MLR dar und muss bei jeder Anwendung des

Modells getestet werden. Begründet man das kumulative Logit-Modell durch ein lineares „Hintergrundmodell“ für eine metrische Variable η , aus der die ordinale Kriteriumsvariable Y durch vergrößerndes Messen hervorgegangen ist, dann erscheint die Annahme paralleler Regressionen als sehr plausibel (siehe unten).

Analog zum binären Fall kann man die Modellformulierung in Formel (14) äquivalent auch durch die folgenden Gleichungssysteme über Odds (Wahrscheinlichkeitsquotienten) oder Logits (logarithmierte Wahrscheinlichkeitsquotienten) ersetzen, die u.a. bei der Interpretation der Regressionskoeffizienten von Nutzen sind:

$$\frac{P(Y \leq j)}{P(Y > j)} = e^{\alpha_j - \beta \mathbf{X}} = e^{\alpha_j - \beta_1 X_1 - \beta_2 X_2 - \dots - \beta_M X_M} = e^{\alpha_j} e^{-\beta_1 X_1} e^{-\beta_2 X_2} \dots e^{-\beta_M X_M}, j = 1, \dots, J-1$$

$$\ln \left(\frac{P(Y \leq j)}{P(Y > j)} \right) = \alpha_j - \beta \mathbf{X}, j = 1, \dots, J-1$$

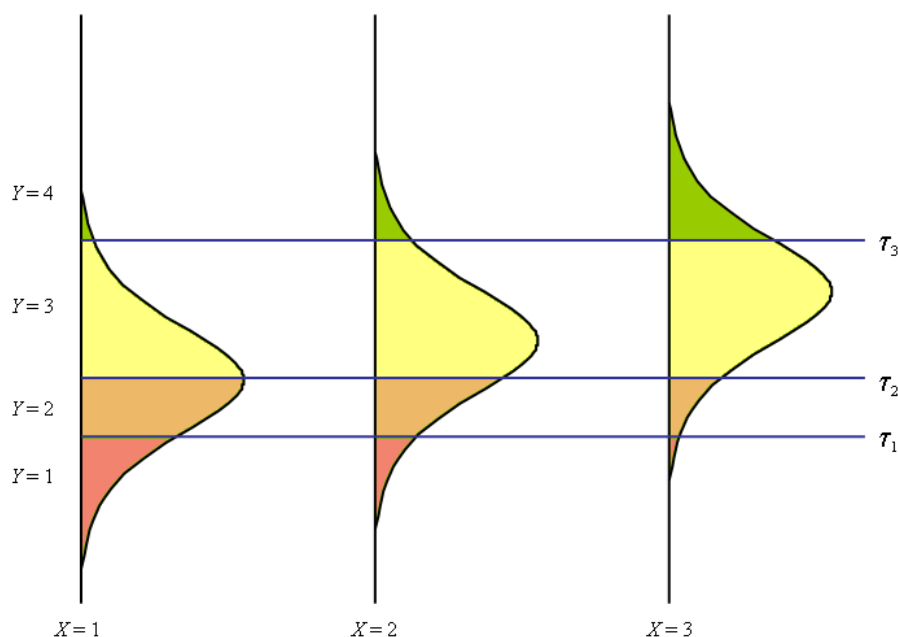
Auch das ordinale logistische Modell kann aus einem linearen Modell für eine latente kontinuierliche Variable η mit logistisch verteiltem Fehleranteil hergeleitet werden (vgl. Abschnitt 2.1.4), wobei die manifeste Kriteriumsvariable Y aus der latenten Variablen durch ein Treppenfunktions-Messmodell mit Schwellenwerten hervorgeht:

$$\eta = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_M X_M + \sigma \varepsilon \quad (15)$$

$$Y = \begin{cases} 1, & \text{falls } \tau_0 = -\infty < \eta \leq \tau_1 \\ 2, & \text{falls } \tau_1 < \eta \leq \tau_2 \\ \vdots & \\ J, & \text{falls } \tau_{J-1} < \eta < \tau_J = \infty \end{cases}$$

Wie im binären Fall gilt ausdrücklich, dass die Unterstellung einer solchen Herkunft *keine* Voraussetzung für die Verwendung des ordinalen logistischen Modells ist.

Im folgenden Beispiel hängt die latente Kriteriumsvariable η von einem Prädiktor X ab und geht durch „vergrößernde Messung“ in ein manifestes Kriterium mit vier Stufen über:



Die farblich gekennzeichneten Flächenanteile stehen für die Wahrscheinlichkeiten der vier Y -Kategorien bei einem bestimmten X -Wert. Offenbar sinkt mit wachsendem X -Wert z.B. die Wahrscheinlichkeit der ersten Y -Kategorie, während für die vierte Y -Kategorie das Gegenteil gilt.

Analog zur BLR-Situation besteht auch im OLR-Modell mit latentem Hintergrund eine einfache Beziehung zwischen den γ -Koeffizienten in der Regressionsgleichung für η und den Koeffizienten im Modell (14). Für die Wahrscheinlichkeit

$$P(Y > j) = P(\eta > \tau_j), \quad j = 1, 2, \dots, J - 1$$

gilt:

$$P(\eta > \tau_j) = P(\gamma_0 - \tau_j + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_M X_M > -\sigma \varepsilon)$$

Wie in Abschnitt 2.1.4 erhalten wir:

$$P(\eta > \tau_j) = P\left(\varepsilon \leq \frac{\gamma_0 - \tau_j + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_M X_M}{\sigma}\right)$$

Mit den Definitionen:

$$\begin{aligned} \alpha_j &:= \frac{\tau_j - \gamma_0}{\sigma} \\ \beta_m &:= \frac{\gamma_m}{\sigma}, \quad m = 1, \dots, M \end{aligned} \tag{16}$$

gilt:

$$P(\eta > \tau_j) = P(\varepsilon \leq -\alpha_j + \beta \mathbf{X})$$

Die Wahrscheinlichkeit dafür, dass die Zufallsvariable ε einen Wert kleiner oder gleich $(-\alpha_j + \beta \mathbf{X})$ annimmt, ist identisch mit dem Wert ihrer Verteilungsfunktion an dieser Stelle:

$$P(\varepsilon \leq -\alpha_j + \beta \mathbf{X}) = \frac{e^{-\alpha_j + \beta \mathbf{X}}}{1 + e^{-\alpha_j + \beta \mathbf{X}}}$$

Bei Betrachtung der Wahrscheinlichkeit $P(Y \leq j)$ ergibt sich Modell (14):

$$P(Y \leq j) = 1 - P(Y > j) = 1 - \frac{e^{-\alpha_j + \beta \mathbf{X}}}{1 + e^{-\alpha_j + \beta \mathbf{X}}} = \frac{1}{1 + e^{-\alpha_j + \beta \mathbf{X}}} = \frac{e^{\alpha_j - \beta \mathbf{X}}}{1 + e^{\alpha_j - \beta \mathbf{X}}}$$

Wir haben gerade gesehen, dass die aus einem latenten Modell abgeleiteten Koeffizienten der j -ten Gleichung des kumulativen Logit-Modells mit Ausnahme des konstanten Terms *nicht* von der Wahl des Schwellenwertes τ_j abhängen. Weil sich die $J - 1$ Gleichungen des kumulativen OLR-Modells nur durch den zur Dichotomisierung der latenten Variablen verwendeten Schwellenwert τ_j unterscheiden, ergibt sich unter dieser Perspektive zwingend das kumulative Logit-Modell (14) mit einem einzigen β -Vektor.

Aus der Annahme paralleler Regressionen folgt für zwei feste Prädiktorwertkombinationen \mathbf{x}_1 und \mathbf{x}_2 , dass der Quotient aus ihren Wahrscheinlichkeitsverhältnissen (ihr odds ratio) *nicht* von der Grenzkategorie j abhängt:

$$\frac{P(Y \leq j | \mathbf{X} = \mathbf{x}_1) / P(Y > j | \mathbf{X} = \mathbf{x}_1)}{P(Y \leq j | \mathbf{X} = \mathbf{x}_2) / P(Y > j | \mathbf{X} = \mathbf{x}_2)} = e^{\beta(\mathbf{x}_2 - \mathbf{x}_1)}$$

Daher wird das kumulative Logit-Modell in der angelsächsischen Literatur auch als *proportional-odds model* - Modell bezeichnet.

4.2 Anwendungsbeispiel

Als Beispiel greifen wir die *Risikofaktoren für Durchblutungsstörungen* wieder auf (Datei **DBS.SAV**, siehe Abschnitt 2.2), wobei aber ein *dreistufiges* Kriterium (Variable DBS3) verwendet wird:¹

- 0 Keine diagnostizierte Durchblutungsstörung
- 1 Periphere Durchblutungsstörungen
- 2 Koronare Herzkrankheit, Schlaganfall oder Vorstufe

Abweichend von den in Abschnitt 2.2 beschriebenen Verhältnissen steht in der künstlichen Population das manifeste und ordinale Kriterium DBS3 in folgender Beziehung zur latenten und stetigen Variablen η :

$$\text{DBS3} = \begin{cases} 0, & \text{falls } -\infty < \eta \leq 1 \\ 1, & \text{falls } 1 < \eta \leq 5 \\ 2, & \text{falls } 5 < \eta < \infty \end{cases}$$

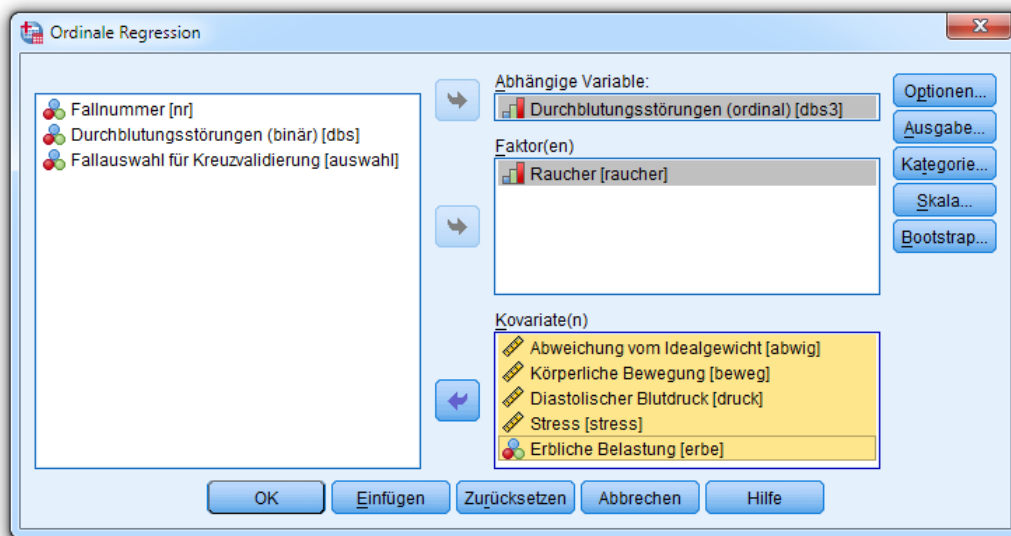
Das Modell für die latente Variable η lautet unverändert:

$$\eta = -4 + 0,06 \cdot \text{ABWIG} - 0,75 \cdot \text{BEWEG} + 0,033 \cdot \text{DRUCK} + 1,1 \cdot \text{STRESS} + 1,55 \cdot \text{ERBE} + 4 \cdot \text{RAUCHER1} + 1 \cdot \text{RAUCHER2} + 1,5 \cdot \varepsilon$$

Wir beziehen alle Prädiktoren (ABWIG, BEWEG, DRUCK, STRESS, ERBE, RAUCHER) ein und fordern eine OLR mit der SPSS-Prozedur PLUM nach

Analysieren > Regression > Ordinal

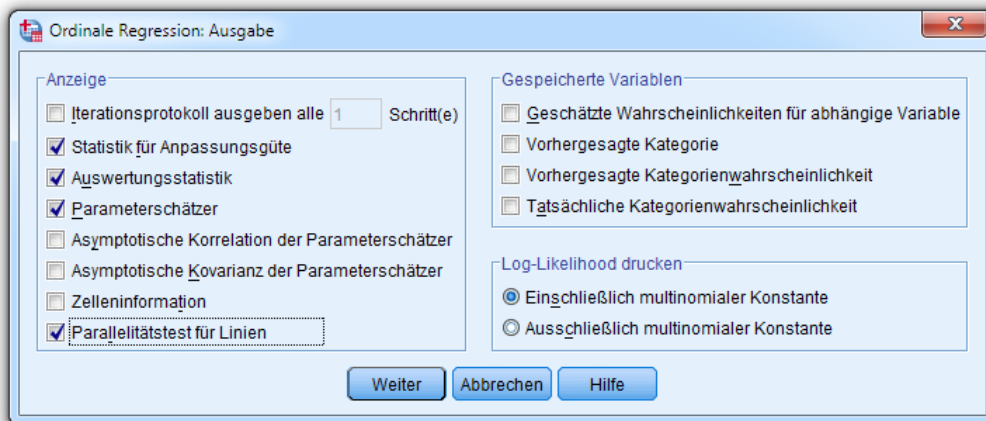
in folgender Dialogbox an:



Hier wird der dichotome Prädiktor ERBE als **Kovariate** behandelt. Deklariert man ihn als **Faktor**, kommt es zu sachlich irrelevanten Änderungen der Schätzergebnisse: Der Regressionskoeffizient wechselt das Vorzeichen, und bei den Schwellen kommt es zu kompensatorischen Änderungen.

¹ Aufgrund der BLR-Historie haben die Kategorien unseres Kriteriums numerische Werte ab 0, was eine leichte Diskrepanz zur obigen Formulierung des allgemeinen OLR-Modells schafft (mit $j = 1, 2, \dots, J$).

Über den Schalter **Ausgabe** fordern wir zusätzlich den **Parallelitätstest für Linien** an:



4.3 Parameterschätzung

Die Parameterschätzungen werden wie üblich nach dem ML-Prinzip ermittelt. Weil sich keine wesentlichen Unterschiede zur Situation bei der BLR und der MLR ergeben, ersparen wir uns diesmal die Formeln. Für unser Beispiel mit bekannten Populationsparametern liefert die Zufallsstichprobe folgende Schätzer:

Parameterschätzer

		Schätzer	Standardfehler	Wald	Freiheitsgrade	Sig.	Konfidenzintervall 95%	
							Untergrenze	Obergrenze
Schwelle	[dbs3 = 0]	2,745	1,272	4,655	1	,031	,251	5,238
	[dbs3 = 1]	6,000	1,365	19,318	1	,000	3,325	8,676
Lage	abwig	,050	,018	8,206	1	,004	,016	,084
	beweg	-,802	,215	13,901	1	,000	-1,224	-,380
	druck	,035	,007	23,666	1	,000	,021	,049
	stress	,490	,209	5,481	1	,019	,080	,900
	erbe	1,563	,390	16,100	1	,000	,800	2,327
	[raucher=1]	2,590	,504	26,416	1	,000	1,602	3,577
	[raucher=2]	,761	,423	3,229	1	,072	-,069	1,590
	[raucher=3]	0 ^a	.	.	0	.	.	.

Verknüpfungsfunktion: Logit.

a. Dieser Parameter wird auf Null gesetzt, weil er redundant ist.

Für $P(Y \leq 0)$ liefert das Modell die folgende Schätzung:

$$\frac{e^{2,745 - \hat{\beta}X}}{1 + e^{2,745 - \hat{\beta}X}}$$

$$\text{mit } \hat{\beta}X = 0,05 \cdot \text{ABWIG} - 0,802 \cdot \text{BEWEG} + 0,035 \cdot \text{DRUCK} + 0,49 \cdot \text{STRESS} + 1,563 \cdot \text{ERBE} + 2,59 \cdot \text{RAUCHER1} + 0,761 \cdot \text{RAUCHER2}$$

Trotz der nicht allzu großen Stichprobe liegen alle Schätzer akzeptabel nahe bei den wahren Populationsparametern, wobei wegen Beziehung (16) und $\sigma = 1,5$ jeweils das 1,5-fache der Schätzer zum Vergleich herangezogen werden muss.

SPSS bezeichnet die geschätzten Ordinatenabschnitte (im Beispiel $\hat{\alpha}_1$ und $\hat{\alpha}_2$) als *Schwellen*, sieht also darin offenbar Schätzer für die Trennwerte τ_1 und τ_2 , die den Übergang von η zu Y definieren. Allerdings sind die Trennwerte nicht ohne weiteres aus den Ordinatenabschnitten α_j des OLR-Modells zu gewinnen. In der folgenden Bestimmungsgleichung (siehe oben)

$$\alpha_j = \frac{\tau_j - \gamma_0}{\sigma} \Leftrightarrow \tau_j = \alpha_j \sigma + \gamma_0$$

stecken zwei Identifikationsprobleme:

- Das Fehlergewicht σ ist unbekannt.
- Auch bei bekanntem Fehlergewicht σ könnte man aus α_j nicht gleichzeitig den Ordinatenabschnitt γ_0 und den Schwellenwert τ_j berechnen.

Diese Identifikationsprobleme sind nur durch Annahmen zu lösen. Weil die Statistik-Programmpakete dabei verschiedene Wege gehen, ermitteln sie bei gleichen Daten unterschiedliche Schätzwerte für die Schwellen (vgl. Long 1997, S. 122). SPSS kommt über die Annahmen

$$\begin{aligned}\sigma &= 1 \\ \gamma_0 &= 0\end{aligned}$$

zum Ergebnis:

$$\tau_j = \alpha_j$$

Mit dem (in echten Studien natürlich nicht vorhandenen) Wissen über γ_0 ($= -4$) und σ ($= 1,5$) liefert das geschätzte Modell andere Schwellen, die recht nahe bei den wahren Werten (1 und 5) liegen:

$$\hat{\tau}_1 = \hat{\alpha}_1 \sigma + \gamma_0 \approx 2,745 \cdot 1,5 - 4 \approx 0,118$$

$$\hat{\tau}_2 = \hat{\alpha}_2 \sigma + \gamma_0 \approx 6,0 \cdot 1,5 - 4 \approx 5,0$$

Weil unser Modell garantiert stimmt, dürfen wir uns schon vor einer Diskussion der Modellgültigkeitsprüfung mit der Parameterinterpretation beschäftigen. Wir erfahren z.B., dass die Wahrscheinlichkeitsquotienten

$$\frac{P(\text{DBS} \leq 0)}{P(\text{DBS} > 0)} \quad \text{bzw.} \quad \frac{P(\text{DBS} \leq 1)}{P(\text{DBS} > 1)}$$

bei Rauchern um den Faktor $e^{-2,59} \approx 0,075$ kleiner eingeschätzt werden als bei Nichtrauchern. Bei jeder beliebigen DBS-Dichotomisierung ist also die geschätzte Wahrscheinlichkeit der gesünderen Seite bei Rauchern deutlich kleiner. Bei dieser Interpretation ist zu beachten:

- PLUM arbeitet bei nominalskalierten Regressoren mit der Indikator- bzw. Dummy-Kodierung und verwendet die letzte Kategorie als Referenz. So resultieren im Beispiel für den Faktor RAUCHER zwei Kodiervariablen mit zugehörigen Parametern, welche die erste (Raucher) bzw. zweite (Ex-Raucher) Kategorie mit der dritten Kategorie (Nichtraucher) vergleichen. Folglich kann der z.B. der Logit-Unterschied zwischen Rauchern und Nichtrauchern direkt als $\hat{\beta}_6$ aus den Schätzergebnissen abgelesen werden. In der folgenden (mit LOGISTIC REGRESSION erzeugten) Tabelle werden die Indikator-Kodiervariablen beschrieben:

Codierungen kategorialer Variablen

		Häufigkeit	Parameterkodierung	
			(1)	(2)
raucher	Raucher	48	1,000	,000
	Ehem. Raucher	46	,000	1,000
	Nichtraucher	106	,000	,000

- Beim Berechnen der Effektgrößen ist zu beachten, dass die β -Koeffizienten mit einem negativen Vorzeichen in das OLR-Modell eingehen.

Für ein um 3 Skalenpunkte höheres Maß an Körperbewegung wird für die Wahrscheinlichkeitsverhältnisse eine Änderung um den Faktor $e^{3 \cdot 0,802} \approx 11,09$ in der gesunden Richtung geschätzt.

4.4 Modellgültigkeit

4.4.1 Parallelität

Zunächst wird im Parallelitätstest wird das OLR-Modell in Gleichung (14) mit einer liberalisierten Variante verglichen, die für jede Logitgleichung einen *eigenen* Vektor β_j mit Regressionsgewichten schätzt:

$$P(Y \leq j) = \frac{e^{\alpha_j - \beta_j X}}{1 + e^{\alpha_j - \beta_j X}}, \text{ mit } \beta_j := [\beta_{j1}, \beta_{j2}, \dots, \beta_{jM}], j = 1, \dots, J-1$$

Bei einem dreistufigen Kriterium (mit zwei Logit-Gleichungen) prüft der Parallelitätstest die folgende Nullhypothese:

$$H_0 : \beta_{11} = \beta_{21}, \beta_{12} = \beta_{22}, \dots, \beta_{1M} = \beta_{2M}$$

Weil durch die Restriktionen im parallelen Modell gerade M Parameter eingespart werden, erhalten wir in unserem Beispiel eine Prüfgröße mit 7 Freiheitsgraden:

Parallelitätstest für Linien^a

Modell	-2 Log-Likelihood	Chi-Quadrat	Freiheitsgrade	Sig.
Nullhypothese	242,131			
Allgemein	238,409	3,722	7	,811

Die Nullhypothese gibt an, daß die Lageparameter (Steigungskoeffizienten) über die Antwortkategorien übereinstimmen.

a. Verknüpfungsfunktion: Logit.

Der Test spricht deutlich für die Annahme paralleler Regressionen, so dass die Globalbeurteilung des OLR-Modells für unser Beispiel insgesamt positiv ausfällt.

Im Fall einer unhaltbaren Parallelitätsannahme kann eventuell eine Modifikation des Modells Abhilfe schaffen (z.B. die nichtlineare Transformation eines metrischen Prädiktors). Sollte die in Abschnitt 4.8 diskutierte Homogenitätsannahme problematisch sein, kann mit SPSS-PLUM ein Lokations-Skalen - Modell geschätzt werden.

Bei O'Connell (2006, S. 47ff) werden weitere Optionen diskutiert, z.B.:

- Separate Betrachtung der binären Bestandteile von Modell (14)
So erhält man allerdings weder ein sparsames Modell, noch eindeutige Schätzungen zu den Wahrscheinlichkeiten der Kriteriumsausprägungen.
- Verwendung der multinomialen logistischen Regression
Dabei wird allerdings die ordinale Struktur des Kriteriums ignoriert.

4.4.2 Globale Modellgültigkeit

Auch bei der OLR ermöglichen die Pearson- und die Devianz-Statistik jeweils die Konstruktion eines globalen Modellgültigkeitstests. Aus den in Abschnitt 2.4.1 diskutierten Gründen sind die Tests aber nur bei mehrfach besetzten Prädiktorwertekombinationen anwendbar, in unserem Beispiel also kaum. Sie kommen aber trotzdem zum korrekten Ergebnis:

Anpassungsgüte

	Chi-Quadrat	Freiheitsgrade	Sig.
Pearson	369,068	389	,759
Abweichung	242,131	389	1,000

Verknüpfungsfunktion: Logit.

Durch dieses Beispiel sollen Sie aber keinesfalls dazu ermuntert werden, die beiden Tests auf Individualdaten (mit überwiegend einfach besetzten Prädiktorwertkombinationen) anzuwenden. Ersatzweise wird vorgeschlagen, die Gültigkeit des OLR-Modells (14) in zwei Schritten zu testen:

- Man bildet für jede einzelne Logit-Gleichung durch geeignete Dichotomisierung des Kriteriums ein binäres Modell und nutzt den hier verfügbaren Hosmer-Lemeshow - Test (siehe Abschnitt 2.4.1.3).
- Sind die dichotomen Modelle akzeptiert, ist noch die Annahme paralleler Regressionen zu prüfen. Dazu bietet die PLUM-Prozedur einen Test, der auch bei Individualdaten anwendbar ist (siehe unten).

Für das erste binäre Partialmodell ist in unserem Beispiel folgendes Kriterium zu verwenden:

$$DBS2a = \begin{cases} 1, & \text{falls } DBS3 = 0 \\ 0, & \text{sonst} \end{cases}$$

Abgesehen von der (letztlich beliebigen) Kategoriennummerierung handelt es sich natürlich um dasselbe Kriterium, das wir im Zusammenhang mit der BLR untersucht haben. Wie wir bereits wissen, akzeptiert der Hosmer-Lemeshow-Test dieses binäre Modell:

Hosmer-Lemeshow-Test

Schritt	Chi-Quadrat	df	Sig.
1	10,588	8	,226

Für das zweite binäre Partialmodell wird folgendes Kriterium verwendet:

$$DBS2b = \begin{cases} 1, & \text{falls } DBS3 \leq 1 \\ 0, & \text{sonst} \end{cases}$$

Auch dieses Modell besteht den Gültigkeitstest:

Hosmer-Lemeshow-Test

Schritt	Chi-Quadrat	df	Sig.
1	3,753	8	,879

4.4.3 Lokale Modellanalyse

Wer auch eine *lokale* Modellanalyse vornehmen möchte, kann über das Kontrollkästchen **Zellinformationen** in der PLUM-Subdialogbox **Ausgabe** Pearson-Residuen sowie beobachtete und erwartete Zellenhäufigkeiten anfordern, wobei eine Interpretation der Residuen nur bei mehrfach besetzten Prädiktorwertkombinationen sinnvoll ist, in unserem Beispiel also sicher nicht.

4.5 Beurteilung der Modellrelevanz

Die globale Nullhypothese

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_M = 0$$

wird für unser Beispiel per Likelihood-Quotienten-Test deutlich zurückgewiesen:

Information zur Modellanpassung

Modell	-2 Log-Likelihood	Chi-Quadrat	Freiheitsgrade	Sig.
Nur konstanter Term	414,544			
Final	242,131	172,413	7	,000

Verknüpfungsfunktion: Logit.

Weil von der globalen Nullhypothese 7 Regressionsgewichte auf 0 fixiert werden, resultiert eine approximativ χ^2 - verteilte Prüfgröße mit 7 Freiheitsgraden.

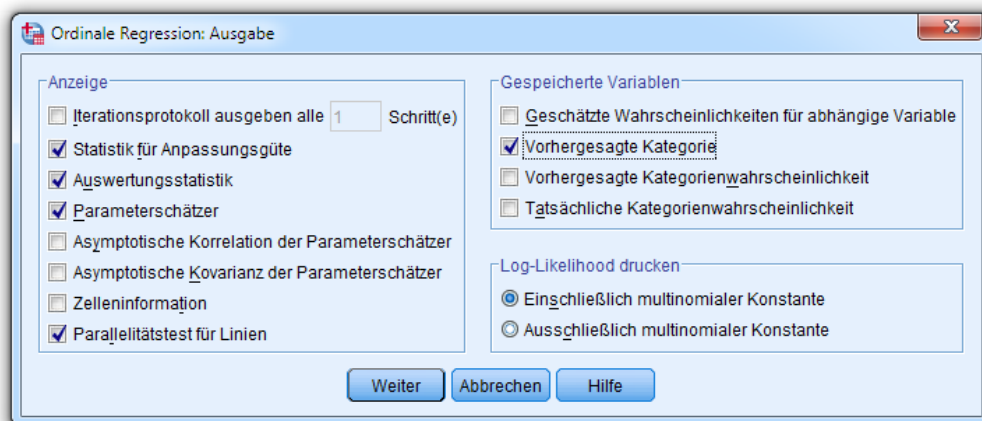
Auch bei der OLR können die in Abschnitt 2.5.2 diskutierten Pseudo-R²-Statistiken berechnet werden, die auf den Likelihood-Werten verschiedener Modelle basieren. Wir erhalten:

Pseudo R-Quadrat

Cox und Snell	,578
Nagelkerke	,661
McFadden	,416

Verknüpfungsfunktion:
n: Logit.

Leider gibt die Prozedur PLUM keine Klassifikationstabelle aus, kann aber immerhin die vorhergesagte Kategorie in eine neue Variable schreiben (Subdialogbox **Ausgabe**):



Nach dieser Vorarbeit erhält man eine brauchbare Klassifikationstabelle über den folgenden Menübefehl

Analysieren > Deskriptive Statistiken > Kreuztabellen

Über die Subdialogbox **Zellen** kann man auch noch die Prozentanteile richtiger Klassifikationen berechnen lassen:

Durchblutungsstörungen (ordinal) * Vorhergesagte Antwortkategorie Kreuztabelle

			Vorhergesagte Antwortkategorie			Gesamt
			Nein	Peripher	Koronar	
DBS3	Nein	Anzahl	83	12	1	96
		% innerhalb von DBS3	86,5%	12,5%	1,0%	100,0%
	Peripher	Anzahl	18	40	7	65
		% innerhalb von DBS3	27,7%	61,5%	10,8%	100,0%
	Koronar	Anzahl	1	13	25	39
		% innerhalb von DBS3	2,6%	33,3%	64,1%	100,0%
Gesamt	Anzahl	102	65	33	200	
	% innerhalb von DBS3	51,0%	32,5%	16,5%	100,0%	

Im Beispiel werden 148 (83 + 40 + 25) von 200 Fällen (= 74 %) richtig eingeordnet. Einen realistischen Eindruck von der Diagnoseleistung eines Modells vermittelt allerdings nur eine Kreuzvalidierung, die bei einer OLR mit SPSS leider etwas Handarbeit erfordert: Man berechnet für die Fälle der Kreuzvalidierungsstichprobe mit Hilfe der geschätzten Parameter die prognostizierten Wahrscheinlichkeiten der Kriteriumskategorien und wählt die Kategorie mit dem größten Wert als Modellprognose.

4.6 Beurteilung der einzelnen Regressoren

Zur Signifikanzbeurteilung der einzelnen Parameter stehen Wald-Tests zur Verfügung (siehe Abschnitt 2.6). In unserem Beispiel werden alle Regressoren mit Ausnahme der zweiten RAUCHER-Kodiervariablen als signifikant beurteilt (siehe Tabelle in Abschnitt 4.3).

Wie die SPSS-Prozedur NOMREG besitzt auch PLUM ein Subkommando TEST, das per Syntax die Prüfung von Hypothesen über beliebige Linearkombinationen von Parametern erlaubt. Man kann z.B.:

- zwei (oder mehr) Parameter auf Identität testen,
- einen Parameter gegen einen beliebigen Wert testen (statt gegen 0),
- mehrere Hypothesen simultan testen.

Um in unserer Beispielstudie zu einem Gesamturteil über den kategorialen Regressor RAUCHER zu kommen, sollte man die Regressionskoeffizienten β_6 und β_7 zu den RAUCHER-Kodiervariablen einem simultanen Test unterziehen, dessen Nullhypothese lautet:

$$H_0 : \beta_6 = \beta_7 = 0$$

Ein solcher Test kann mit der folgenden Syntax angefordert werden:

```
PLUM dbs3 BY raucher WITH abwig beweg druck stress erbe
  /TEST (0, 0) = ALL 0 0 0 0 0 0 0 1 0 0;
                ALL 0 0 0 0 0 0 0 0 1 0.
```

PLUM protokolliert zunächst seine Interpretation der Syntax:

Kontrastkoeffizienten

		C1	C2
Schwelle	[dbs3 = 0]	0	0
	[dbs3 = 1]	0	0
Lage	abwig	0	0
	beweg	0	0
	druck	0	0
	stress	0	0
	erbe	0	0
	[raucher=1]	1	0
	[raucher=2]	0	1
	[raucher=3]	0	0

Anschließend erscheinen separate Wald-Tests für die *einzelnen* Kontraste bzw. Parameter, die im konkreten Beispiel aber keine Neuigkeiten enthalten:

Kontrastergebnisse

Kontraste	Schätzer	Standardfehler	Testwert	Wald	Freiheitsgrade	Sig.	Konfidenzintervall 95%	
							Untergrenze	Obergrenze
C1	2,590	,504	0	26,416	1	,000	1,602	3,577
C2	,761	,423	0	3,229	1	,072	-,069	1,590

Verknüpfungsfunktion: Logit.

Im simultanen Wald-Test über *beide* Parameter wird die RAUCHER-Nullhypothese deutlich verworfen:

Testergebnisse

Wald	Freiheitsgrade	Sig.
26,775	2	,000

Verknüpfungsfunktion: Logit.

Nähere Informationen zum PLUM-Subkommando TEST findet man in der SPSS-Hilfe über

Hilfe > Befehlssyntax-Referenz > NOMREG

Über das Kontrollkästchen **Zellinformation** in der **Ausgabe**-Subdialogbox kann man PLUM dazu bewegen, die Pearson-Residuen des Modells zu tabellieren. Diese sind jedoch nur bei *mehrfach* besetzten Prädiktorwertekombinationen sinnvoll zu interpretieren (siehe Abschnitt 2.4.2.1), so dass wir bei unserem Beispiel auf eine Inspektion verzichten.

4.7 Vergleiche mit alternativen Auswertungsverfahren

Zum Urteil über die Leistungsfähigkeit der OLR-Analyse können die Ergebnisse alternativer Auswertungsverfahren einen Beitrag leisten.

4.7.1 Multinomiale logistische Regression

Ignoriert man in unserem Beispiel die ordinale Struktur des Kriteriums und führt eine MLR-Analyse durch, dann erhält man grundsätzlich vergleichbare, aber in Details weniger prägnante Ergebnisse, die zudem schwerer zu interpretieren sind.

Parameterschätzer

			Standard- fehler	Wald	Freiheits- grade	Signifikanz	Exp(B)	95% Konfidenzintervall für Exp(B)		
								Untergrenze	Obergrenze	
dbs3 ^a	Nein	Konstanter Term	6,941	2,483	7,814	1	,005			
		abwig	-,097	,034	8,189	1	,004	,908	,850	,970
		beweg	1,469	,404	13,246	1	,000	4,344	1,970	9,581
		druck	-,062	,014	20,310	1	,000	,940	,915	,966
		stress	-,700	,396	3,120	1	,077	,497	,228	1,080
		erbe	-2,740	,733	13,956	1	,000	,065	,015	,272
		[raucher=1]	-4,842	1,092	19,651	1	,000	,008	,001	,067
		[raucher=2]	-1,658	,884	3,521	1	,061	,190	,034	1,077
		[raucher=3]	0 ^b	.	.	0
Peripher		Konstanter Term	3,530	2,132	2,741	1	,098			
		abwig	-,060	,028	4,409	1	,036	,942	,891	,996
		beweg	,715	,331	4,670	1	,031	2,043	1,069	3,907
		druck	-,022	,010	4,713	1	,030	,978	,958	,998
		stress	-,219	,332	,435	1	,509	,803	,419	1,539
		erbe	-,981	,578	2,882	1	,090	,375	,121	1,164
		[raucher=1]	-2,188	,766	8,157	1	,004	,112	,025	,503
		[raucher=2]	-1,147	,809	2,011	1	,156	,318	,065	1,550
		[raucher=3]	0 ^b	.	.	0

a. Die Referenzkategorie lautet: Koronar.

b. Dieser Parameter wird auf Null gesetzt, weil er redundant ist.

Die Nullhypothese zum Prädiktor STRESS kann im Likelihood-Quotienten-Test *nicht* verworfen werden:

Likelihood-Quotienten-Tests

Effekt	Kriterien für die Modellanpassung	Likelihood-Quotienten-Tests		
	-2 Log-Likelihood für reduziertes Modell	Chi- Quadrat	Freiheits- grade	Signifikanz
Konstanter Term	237,160	,000	0	.
abwig	246,202	9,041	2	,011
beweg	253,180	16,020	2	,000
druck	266,364	29,204	2	,000
stress	241,609	4,448	2	,108
erbe	255,945	18,785	2	,000
raucher	267,539	30,379	4	,000

Die höheren Pseudo-R² – Statistiken sprechen *nicht* für eine Überlegenheit des multinomialen Modells, weil bei deren Berechnung die 7 zusätzlichen Parameter (im Vergleich zum OLR-Modell) *nicht* in Rechnung gestellt werden:

Pseudo-R-Quadrat

Cox und Snell	,588
Nagelkerke	,673
McFadden	,428

4.7.2 Lineare Regressions- bzw. Varianzanalyse

Bei einem dreistufigen Kriterium sind die Annahmen der linearen Regressions- bzw. Varianzanalyse über die Normalverteilung der Residuen zweifellos grob verletzt. Trotzdem wurde zu Vergleichszwecken über

Analysieren > Allgemeines lineares Modell > Univariat

eine Kovarianzanalyse mit unseren Daten gerechnet. Aufgrund der bei DBS3 im Vergleich zur latenten Variablen η erheblich eingeschränkten Standardabweichung (0,77 versus 4,74) sind die geschätzten Regressionskoeffizienten betragsmäßig deutlich kleiner als die wahren Parameter im Modell für die latente Variable, sie stehen jedoch untereinander in plausiblen Relationen:

Abhängige Variable: db3

Parameter	B	Standardfehler	T	Signifikanz	95% Konfidenzintervall	
					Untergrenze	Obergrenze
Konstanter Term	,089	,253	,352	,725	-,409	,587
abwig	,008	,003	2,369	,019	,001	,014
beweg	-,161	,041	-3,980	,000	-,241	-,081
druck	,007	,001	5,132	,000	,004	,010
stress	,084	,043	1,966	,051	,000	,169
erbe	,317	,077	4,109	,000	,165	,470
[raucher=1]	,727	,105	6,941	,000	,520	,933
[raucher=2]	,207	,094	2,200	,029	,021	,393
[raucher=3]	0 ^a

a. Dieser Parameter wird auf Null gesetzt, weil er redundant ist.

Zwar sind die Ergebnisse trotz der grob verletzten Normalverteilungsannahme insgesamt nur mäßig verzerrt, doch wird wie bei der MLR-Analyse der Effekt des Prädiktors STRESS „übersehen“: Die bekanntermaßen falsche Nullhypothese kann nicht verworfen werden. Der (unkorrigierte) Determinationskoeffizient liegt mit 0,597 ungefähr im selben Bereich wie die Pseudo- R^2 - Maße der OLR.

4.8 Lokations-Skalen - Modell

In Abschnitt 4.1 haben wir das lineare „Hintergrundmodell“ für eine metrische latente Variable η auf die ordinale logistische Regression angewendet:

$$\eta = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_M X_M + \sigma \varepsilon$$

Hier wird angenommen, dass die Residualvariable ε für *jede* Prädiktorwertkombination eine logistische Verteilung mit der identischen Varianz $\frac{\pi^2}{3}$ besitzt (Homogenitätsannahme), wobei der Parameter σ für die Anpassung an die Residualvarianz in der aktuell untersuchten Population zuständig ist. Wie die weitere Argumentation in den Abschnitten 2.1.4 und 4.1 zeigt, geht neben γ_m auch σ in den Parameter β_m des Modells für das manifeste ordinale Kriterium ein:

$$\beta_m := \frac{\gamma_m}{\sigma}, m = 1, \dots, M$$

Im bisherigen (homoskedastischen) Modell kann der Parameter σ nicht geschätzt werden.

Wird eine Studie in zwei Populationen (mit jeweils gültiger Homogenitätsannahme) durchgeführt, liegen eventuell zwei verschiedene σ - Werte vor. Findet man für einen Regressionskoeffizienten in beiden Populationen verschiedene Werte, kann nicht entschieden werden, ob verschiedene Wirkungen vorliegen, oder unterschiedliche Residualvarianzen (mit-)verantwortlich sind (Williams, R. 2009).

Für jede Anwendung der logistischen Regressionsanalyse (mit einem nominalen oder ordinalen Kriterium) hat eine Verletzung der Varianzhomogenitätsannahme zwei unerfreuliche Konsequenzen (Keele & Park 2006):

- unter- oder überschätzte Standardfehler mit entsprechenden Folgen für die Inferenzstatistik (Konfidenzintervalle, Hypothesentests)
- verzerrte, inkonsistente Parameterschätzungen

Während bei einer linearen Regressionsanalyse (mit einem metrischem Kriterium) „nur“ die Inferenzstatistik unter verletzter Homoskedastizität leidet, sind bei der logistischen Regression auch die Parameterschätzungen betroffen, was aufgrund des in Gleichung (16) beschriebenen Zusammenhangs zwischen den Parametern β_m , γ_m und σ verständlich wird.

Die SPSS-Prozedur PLUM unterstützt ein erweitertes Modell, das heterogene Varianzen erlaubt. Für einen Fall i mit den Prädiktorwerten im Vektor \mathbf{x}_i

$$\mathbf{x}_i := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix}$$

und einer Auswahl seiner Prädiktorwerte im Vektor \mathbf{z}_i

$$\mathbf{z}_i := \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_Q \end{bmatrix}$$

wird der bisherige Modellkern

$$\alpha_j - \beta \mathbf{x}_i$$

nun als **Lokationsmodell** bezeichnet und durch das so genannte **Skalenmodell** mit Parametervektor λ

$$e^{\lambda \mathbf{z}_i}, \lambda := [\lambda_1, \lambda_2, \dots, \lambda_Q]$$

dividiert, um einen individuellen Skalierungsfaktor (eine individuelle Varianz) zu ermöglichen:

$$\frac{\alpha_j - \beta \mathbf{x}_i}{e^{\lambda \mathbf{z}_i}}$$

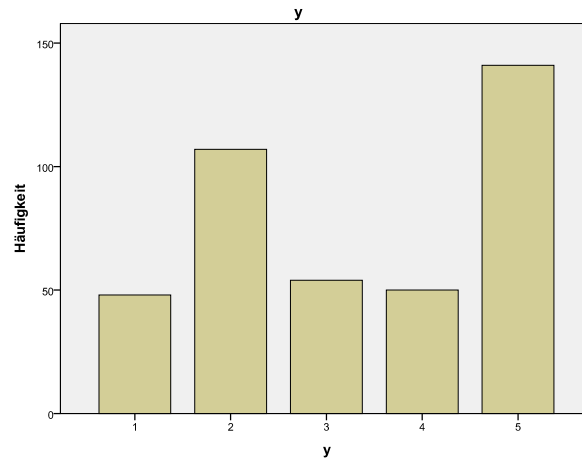
In der Regel wird man sich im Varianz- bzw. Skalenmodell auf wenige Prädiktoren beschränken.

Mit der Abkürzung F_L für die logistische Verteilungsfunktion, ergibt sich das folgende Lokations-Skalen – Modell:

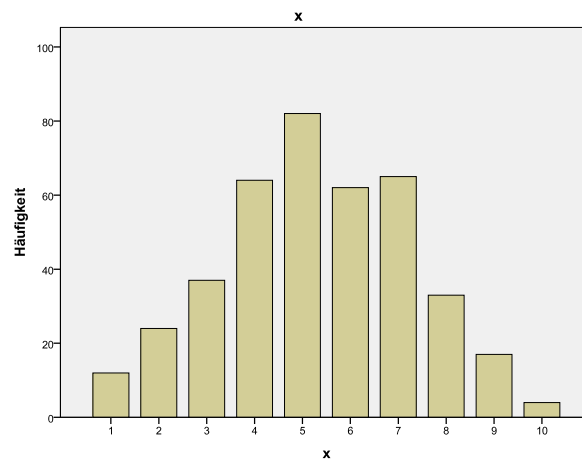
$$P(y_i \leq j) = F_L \left(\frac{\alpha_j - \beta \mathbf{x}_i}{e^{\lambda \mathbf{z}_i}} \right), \quad i = 1, \dots, N, \quad j = 1, \dots, J - 1 \quad (17)$$

Es wird von einigen englischsprachigen Autoren auch als *Heterogenous Choice Model* bezeichnet (z.B. Williams 2009).

In einer Simulationsstudie betrachten wir ein ordinales Kriterium Y



sowie einen metrischen Regressor X



und einen gleichverteilten dichotomen Regressor G .

Für die latente Variable η gilt das wahre Modell

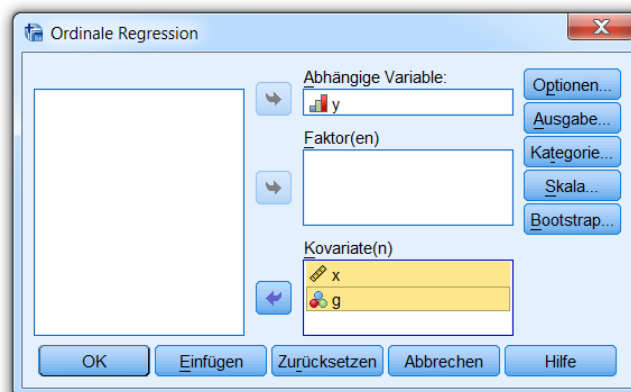
$$\eta = X + G + e^G \varepsilon$$

mit einer logistisch verteilten Residualvariablen, deren Varianz von G abhängt.

Nach dem Menübefehl

Analysieren > Regression > Ordinal

werden die Variablen gewählt:



Weil das voreingestellte Lokationsmodell mit zwei Haupteffekten korrekt ist, muss es nicht geändert werden.

Bei ignorierter Heteroskedastizität wird das Modell verworfen,

Anpassungsgüte			
	Chi-Quadrat	Freiheitsgrade	Sig.
Pearson	136,591	74	,000
Abweichung	137,585	74	,000

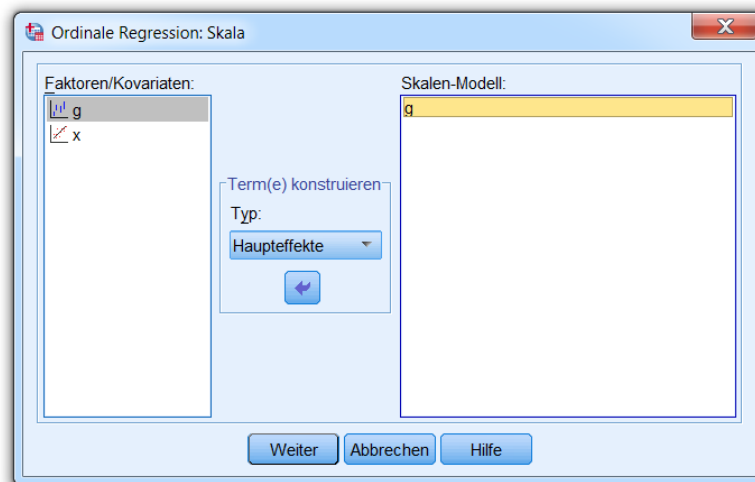
Verknüpfungsfunktion: Logit.

und die Parameterschätzer sind stark verzerrt:

Parameterschätzer								
		Schätzer	Standardfehler	Wald	Freiheitsgrade	Sig.	Konfidenzintervall 95%	
							Untergrenze	Obergrenze
Schwelle	[y = 1]	,906	,246	13,547	1	,000	,424	1,388
	[y = 2]	2,914	,273	113,921	1	,000	2,379	3,449
	[y = 3]	3,499	,286	149,415	1	,000	2,938	4,060
	[y = 4]	3,933	,297	175,883	1	,000	3,351	4,514
Lage	x	,586	,048	148,994	1	,000	,492	,680
	g	,319	,174	3,363	1	,067	-,022	,659

Verknüpfungsfunktion: Logit.

Nun definieren wir im Dialog der ordinalen Regression nach einem Klick auf den Schalter **Skala** das Skalenmodell:



und erhalten daraufhin eine günstige Modellbeurteilung

Anpassungsgüte			
	Chi-Quadrat	Freiheitsgrade	Sig.
Pearson	69,416	73	,597
Abweichung	67,124	73	,672

Verknüpfungsfunktion: Logit.

sowie sinnvolle Schätzergebnisse:

Parameterschätzer

		Schätzer	Standardfehler	Wald	Freiheitsgrade	Sig.	Konfidenzintervall 95%	
							Untergrenze	Obergrenze
Schwelle	[y = 1]	1,845	,337	29,913	1	,000	1,184	2,506
	[y = 2]	5,074	,464	119,411	1	,000	4,164	5,984
	[y = 3]	6,023	,510	139,417	1	,000	5,023	7,023
	[y = 4]	6,757	,547	152,449	1	,000	5,685	7,830
Lage	x	1,011	,086	136,957	1	,000	,842	1,181
	g	,923	,365	6,403	1	,011	,208	1,638
Skala	g	,936	,117	64,393	1	,000	,708	1,165

Verknüpfungsfunktion: Logit.

Williams (2008) warnt vor folgenden Problemen mit Lokations-Skalen - Modellen:

- Instabile Schätzungen
- Der Ansatz erlaubt empirisch nicht unterscheidbare Modelle (mit identischem Fit), die gravierend unterschiedliche theoretische Implikationen besitzen.

Erfolgreiche Anwendungen werden u.a. aus dem Bereich der Signalentdeckungsmodelle berichtet (siehe z.B. DeCarlo 2003).

Gelegentlich wird eine Heteroskedastizität vorgetäuscht durch Spezifikationsfehler (z.B. eine vergessene Wechselwirkung).

5 Numerische Schätzprobleme

In diesem Abschnitt geht es um numerische Probleme der Maximum-Likelihood-Schätzung bei einer logistischen Regressionsanalyse, die durch spezielle Muster in den Daten verursacht werden.

5.1 Multikollinearität

Allzu hohe Korrelationen zwischen den Regressoren führen zu großen Standardfehlern bei den geschätzten Regressionskoeffizienten, worunter auch die Power der Tests zu den Regressionskoeffizienten leidet. Jeder Regressor sollte möglichst viel „Eigenständigkeit“ in das Design einbringen, die bei metrischen Regressoren über ihre quadrierte multiple Korrelation mit den restlichen Regressoren beurteilt werden kann. Nach Menard (1995, S. 66) sollte dieser Wert nicht über 0,80 liegen.

5.2 Quasi-vollständige Trennung

Bei *kategorialen* Prädiktoren können je nach Modell leere Zellen dafür sorgen, dass die iterative Suche nach dem Maximum der Likelihood-Funktion misslingt. Dies passiert z.B. dann, wenn nur *ein* kategorialer Prädiktor vorhanden ist, und alle Fälle mit einer bestimmten Prädiktorausprägung zur selben Kriteriumskategorie gehören. Weil hier eine Teilstichprobe aufgrund der Informationen im Design perfekt zugeordnet werden kann, wird die Konstellation in SPSS als *quasi-vollständige Trennung* bezeichnet.

Im folgenden Beispiel gehören alle 21 Fälle mit der Ausprägung A eines kategorialen Prädiktors zur Kriteriumsstufe 1, so dass die Zelle (A, 0) leer bleibt:

KatReg * Krit Kreuztabelle

Anzahl		Krit		Gesamt
		0	1	
KatReg	A	0	21	21
	B	19	10	29
	C	15	2	17
	Gesamt	34	33	67

Hier liefert das ML-Schätzverfahren unbrauchbare Ergebnisse, was an überhöhten Schätzwerten und vor allem an *extrem großen Standardfehlern* zu erkennen ist, z.B. in der Ausgabe von LOGISTIC REGRESSION:

Variablen in der Gleichung

		Regressions- koeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1 ^a	KatReg			2,621	2	,270	
	KatReg(1)	23,218	8770,825	,000	1	,998	12116061321,384
	KatReg(2)	1,373	,848	2,621	1	,105	3,947
	Konstante	-2,015	,753	7,164	1	,007	,133

a. In Schritt 1 eingegebene Variablen: KatReg.

Bei der berichteten Analyse wurde für den kategorialen Regressor die voreingestellte Indikator-Kodierung mit der dritten Stufe als Referenz gewählt:

Codierungen kategorialer Variablen

		Häufigkeit	Parameterkodierung	
			(1)	(2)
KatReg	A	21	1,000	,000
	B	29	,000	1,000
	C	17	,000	,000

Folglich gilt für den Regressionskoeffizienten β_1 zur ersten Kontrastvariablen:

$$\beta_1 = \ln\left(\frac{P(Y = 1 | KATREG = A)}{P(Y = 0 | KATREG = A)}\right) - \ln\left(\frac{P(Y = 1 | KATREG = C)}{P(Y = 0 | KATREG = C)}\right)$$

Als Schätzwert für

$$\frac{P(Y = 1 | KATREG = A)}{P(Y = 0 | KATREG = A)}$$

tritt im Beispiel aber der undefinierte Wert

$$\frac{1}{0}$$

auf, so dass die Probleme des iterativen Schätzverfahrens verständlich sind.

Im Gegensatz zu LOGISTIC REGRESSION bemerkt die Prozedur NOMREG das Problem und warnt:

Warnungen

Bei den Daten liegt möglicherweise eine quasi-vollständige Trennung vor. Entweder existieren die Maximum-Likelihood-Schätzungen nicht, oder einige Parameterschätzungen sind unendlich.

Die Prozedur NOMREG wird trotz obiger Warnungen fortgesetzt. Die nachfolgend angezeigten Ergebnisse basieren auf der letzte Iteration. Die Gültigkeit der Modellanpassung ist ungewiss.

Allerdings sind alle in Abschnitt 4.8 behandelten Schätzprobleme auch ohne Warnungen der Software an überhöhten Standardfehlern zuverlässig zu erkennen.

Eine zulässige (wenngleich nicht immer sympathische) Maßnahme zum Entfernen leerer Zellen besteht darin, die betroffene Prädiktorkategorie mit einer anderen zu kombinieren. Wenn im Beispiel die kritische Kategorie A mit der Nachbarkategorie B zusammengefasst wird, hat der ML-Algorithmus kein Problem mehr:

Variablen in der Gleichung

		Regressions- koeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1 ^a	KatRegB(1)	2,504	,807	9,627	1	,002	12,237
	Konstante	-2,015	,753	7,164	1	,007	,133

a. In Schritt 1 eingegebene Variablen: KatRegB.

Dass in einer (nicht allzu kleinen Stichprobe) alle Fälle einer Regressorenkategorie zur selben Kriteriumskategorie gehören, dürfte eher selten passieren. Schon eher ist bei einer durch Kreuzen mehrerer Regressoren resultierenden Zelle mit der Kriteriums-„Monotonie“ zu rechnen. In folgender Situation

				Krit	
				0	1
				Anzahl	Anzahl
KatReg1	1	KatReg2	0	8	2
			1	5	5
	2	KatReg2	0	6	2
			1	2	10
	3	KatReg2	0	4	1
			1	0	5

macht das Haupteffektmodell keine Probleme,

Variablen in der Gleichung

		Regressions- koeffizient B	Standard- fehler	Wald	df	Sig.	Exp(B)
Schritt 1	KatReg1			2,993	2	,224	
	KatReg1(1)	-1,439	,970	2,201	1	,138	,237
	KatReg1(2)	-,311	,943	,109	1	,741	,732
	KatReg2(1)	-2,463	,719	11,731	1	,001	,085
	Konstante	1,805	,892	4,094	1	,043	6,082

weil in den folgenden Kreuztabellen keine leeren Zellen auftreten:

KATREG1 × KRIT

Anzahl		Krit		
		0	1	Gesamt
KatReg1	1	13	7	20
	2	8	12	20
	3	4	6	10
	Gesamt	25	25	50

KATREG2 × KRIT

Anzahl		Krit		
		0	1	Gesamt
KatReg2	0	18	5	23
	1	7	20	27
	Gesamt	25	25	50

Nach Erweiterung des Modells um die Wechselwirkung der beiden kategorialen Regressoren sorgt jedoch die leere Zelle in der Tabelle

KATREG1 × KATREG2 × KRIT

für massive Schätzprobleme:

Variablen in der Gleichung

		Regressions- koeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1	KatReg1			2,590	2	,274	
	KatReg1(1)	-21,203	17974,849	,000	1	,999	,000
	KatReg1(2)	-19,593	17974,849	,000	1	,999	,000
	KatReg2(1)	-22,589	17974,850	,000	1	,999	,000
	KatReg1 * KatReg2			,762	2	,683	
	KatReg1(1) by KatReg2(1)	21,203	17974,850	,000	1	,999	1615474059,928
	KatReg1(2) by KatReg2(1)	19,881	17974,850	,000	1	,999	430793082,647
	Konstante	21,203	17974,849	,000	1	,999	1615474059,928

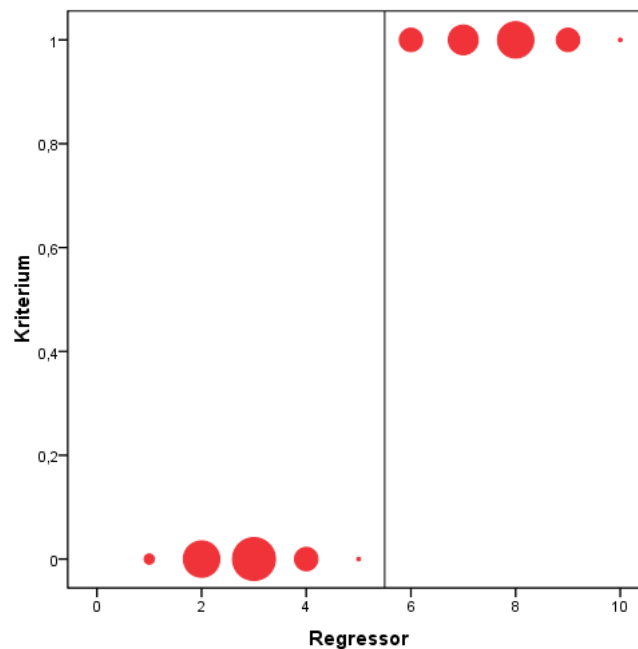
Um zu einem Urteil über die Wechselwirkung zu gelangen, kann man die Kategorien 2 und 3 von KATREG1 zusammenfassen. Im Beispiel ergibt sich dabei, dass der Interaktionsterm vermutlich überflüssig ist:

Variablen in der Gleichung

		Regressions- koeffizient B	Standard- fehler	Wald	df	Sig.	Exp(B)
Schritt 1	KatReg1b(1)	-2,015	,983	4,200	1	,040	,133
	KatReg2(1)	-3,219	1,000	10,361	1	,001	,040
	KatReg1b(1) by KatReg2(1)	1,833	1,423	1,658	1	,198	6,250
	Konstante	2,015	,753	7,164	1	,007	7,500

5.3 Vollständige Trennung

Maximum-Likelihood-Schätzprobleme treten auch dann auf, wenn das logistische Modell eine vollständige Separation der Kriteriumskategorien erlaubt. Im folgenden (künstlichen) Beispiel kann das dichotome Kriterium perfekt aufgrund eines (metrischen) Regressors vorgesagt werden:



Man hätte eigentlich Anlass zur Freude, wenn nicht numerische Probleme im Schätzverfahren für kuriose Ergebnisse sorgen würden:

Variablen in der Gleichung

		Regressions- koeffizient B	Standard- fehler	Wald	df	Sig.	Exp(B)
Schritt 1	regr	33,289	3273,024	,000	1	,992	286431455405860,400
	Konstante	-182,588	18009,595	,000	1	,992	,000

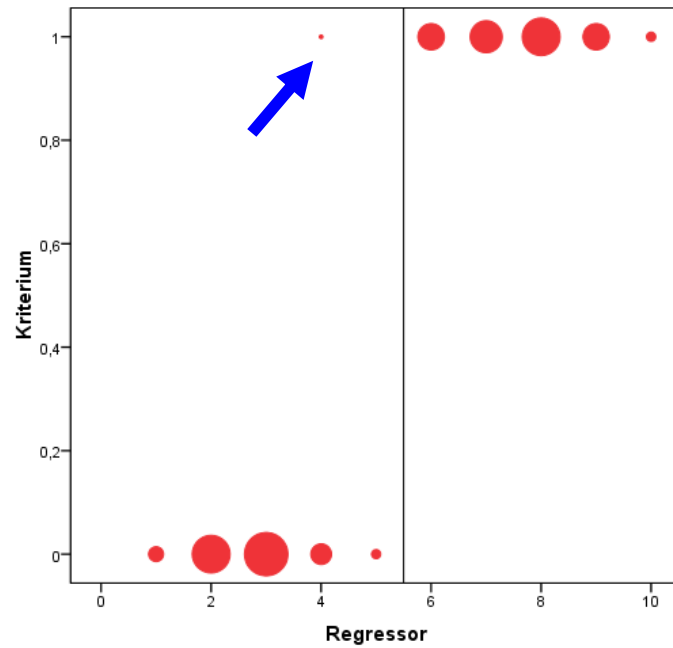
Im Gegensatz zu LOGISTIC REGRESSION macht die Prozedur NOMREG wiederum explizit auf das Problem aufmerksam:

Warnungen

Die Log-Likelihood-Werte sind nahe Null. Bei den Daten könnte eine vollständige Trennung vorliegen. Die Maximum-Likelihood-Schätzungen existieren nicht.

Die Prozedur NOMREG wird trotz obiger Warnungen fortgesetzt. Die nachfolgend angezeigten Ergebnisse basieren auf der letzte Iteration. Die Gültigkeit der Modellanpassung ist ungewiss.

Sobald man die (eigentlich wünschenswerte) perfekte Trennung durch Verändern *eines* Falles aufhebt,



arbeitet der ML-Algorithmus korrekt:

Variablen in der Gleichung

		Regressions- koeffizient B	Standard- fehler	Wald	df	Sig.	Exp(B)
Schritt 1	regr	2,475	,917	7,279	1	,007	11,878
	Konstante	-12,384	4,655	7,077	1	,008	,000

Allzu große Sorgen muss man sich wegen der vollständigen Trennung, die mit einem perfekten Determinationskoeffizienten von 1,0 in linearen Regressionsmodellen vergleichbar ist, bei realen Anwendungen allerdings nicht machen.

6 Anhang

6.1 Symbolverzeichnis

Symbol	Bedeutung
M	Anzahl der Prädiktorvariablen im Design (intervallskaliert oder durch Kodierung von nominalskalierten Variablen entstanden)
m	Index für die Prädiktorvariablen im Design
J	Anzahl der Kategorien des Kriteriums
j	Index für die Kategorien des Kriteriums
N	Anzahl der Beobachtungen
i	Index für die Beobachtungen
Y_i	Zufallsvariable zur Beobachtung i
y_i	Stichprobenergebnis für Beobachtung i bei der binären logistischen Regression, $y_i \in \{0, 1\}$
\mathbf{X}	Spaltenvektor mit den Prädiktorvariablen X_1 bis X_M im Populationsmodell (mit einer führenden 1 für die Konstante) $\mathbf{X} = \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \cdot \\ \cdot \\ X_M \end{bmatrix}$
\mathbf{x}_i	Spaltenvektor mit den Stichprobenwerten der Prädiktorvariablen X_1 bis X_M bei Beobachtung i (mit einer führenden 1 für die Konstante) $\mathbf{x}_i = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \cdot \\ \cdot \\ x_{iM} \end{bmatrix}$
β	Zeilenvektor mit den Modellparametern der binären oder ordinalen logistischen Regression
β_j	Zeilenvektor mit den Modellparametern der j -ten Gleichung eines multinomialen logistischen Regressionsmodells
K	Anzahl der Prädiktorwertkombinationen
k	Index für die Prädiktorwertkombinationen
\tilde{y}_k, \tilde{Y}_k	Stichprobenhäufigkeit der Einserkategorie für die Prädiktorwertkombination k bei der binären logistischen Regression bzw. zugehörige Zufallsvariable
$\tilde{y}_{kj}, \tilde{Y}_{kj}$	Stichprobenhäufigkeit der Kriteriumskategorie j für die Prädiktorwertkombination k bei der multinomialen oder ordinalen logistischen Regression bzw. zugehörige Zufallsvariable

6.2 SPSS-Programme zu den Beispielen

Die im Anhang enthaltenen SPSS-Programme stehen an der im Vorwort angegebenen Stelle zur Verfügung.

6.2.1 SPSS-Syntaxdatei zum DBS-Beispiel

Mit dem folgenden SPSS-Programm wurden die simulierten Daten zum Beispiel *Durchblutungsstörungen* erzeugt:

```
set seed = 555.
input program.
+ loop #i = 1 to 200.
- compute hilf = normal(1).
- compute abwig = rnd(-1*hilf * 9 + normal(9)).
- compute beweg = hilf * 0.6 + normal(0.8).
- recode beweg (lo thru -2 = 1) (-2 thru -1 = 2) (-1 thru 0 = 3)
  (0 thru 1 = 4) (1 thru 2 = 5) (2 thru hi = 6).
- compute druck = rnd(80 - 6*hilf + (normal(1)**2+normal(1)**2+normal(1)**2+
  normal(1)**2+normal(1)**2+normal(1)**2 - 5)*7).
- compute stress = normal(0.7) - hilf * 0.7.
- recode stress (lo thru -2 = 1) (-2 thru -1 = 2) (-1 thru 0 = 3)
  (0 thru 1 = 4) (1 thru 2 = 5) (2 thru hi = 6).
- compute erbe = uniform(1).
- recode erbe (0 thru 0.3 = 1) (0.3 thru 1 = 0).
- compute raucher = normal(0.7) - hilf * 0.7.
- recode raucher (lo thru 0.1 = 3) (0.1 thru 0.7 = 2) (0.7 thru hi = 1).
- do if (raucher = 3).
-   compute raucheff = 0.
- else if (raucher = 2).
-   compute raucheff = 1.
- else if (raucher = 1).
-   compute raucheff = 4.
- end if.
* resid besitzt eine logistische Verteilung.
- compute resid = idf.logistic(uniform(1), 0, 1).
- compute bed_erw = -4 + 0.06 * abwig - 0.75 * beweg + 0.033 * druck + 1.1 * stress +
  1.55 * erbe + raucheff .
- compute latent = bed_erw + 1.5* resid.
- compute dbs = (latent > 1).
- do if (latent <= 1).
-   compute dbs3 = 0.
- else if (latent <= 5).
-   compute dbs3 = 1.
- else.
-   compute dbs3 = 2.
- end if.
- end case.
+ end loop.
+ end file.
end input program.

compute nr = $casenum.
compute auswahl = (uniform(1) <= 0.7).

variable labels nr 'Fallnummer'.
variable labels auswahl 'Fallauswahl für Kreuzvalidierung'.
variable labels abwig 'Abweichung vom Idealgewicht'.
variable labels beweg 'Körperliche Bewegung'.
variable labels druck 'Diastolischer Blutdruck'.
variable labels stress 'Stress'.
variable labels erbe 'Erbliche Belastung'.
variable labels raucher 'Raucher'.

variable labels dbs 'Durchblutungsstörungen (binär)'.
variable labels dbs3 'Durchblutungsstörungen (ordinal)'.

value labels erbe 0 'Unbelastet' 1 'Belastet'.
value labels raucher 1 'Raucher' 2 'Ehem. Raucher' 3 'Nichtraucher'.
value labels dbs auswahl 0 'Nein' 1 'Ja'.
value labels dbs3 0 'Nein' 1 'Peripher' 2 'Koronar'.
```

```

formats  beweg, stress to raucheff dbs dbs3 auswahl (f1.0) abwig (f2.0) nr druck (f3.0).
variable width  nr abwig to raucher dbs dbs3 auswahl (5).
variable level  nr erbe dbs auswahl (nominal) /raucher dbs3 (ordinal) /abwig beweg druck stress
(scale).
variable role
  /input  abwig beweg druck stress erbe raucher
  /target  dbs dbs3
  /none nr.

save  outfile='dbs.sav'
/keep = nr abwig to raucher dbs dbs3 auswahl.

```

6.2.2 SPSS-Syntaxdatei zum Beispiel für die multinomiale Regression

Mit dem folgenden SPSS-Programm wurden die simulierten Daten zum Kneipen-Beispiel erzeugt:

```

set seed = 111318054.
input program.
+ loop #i = 1 to 125.
- compute Alter = rnd(20 + 50 * uniform(1)).
- compute geschl = (uniform(1) > 0.5).
- compute linfunk1 = -3.5 + 0.07 * alter.
- compute linfunk2 = 1.5 - 0.05 * alter + 1.3 * geschl.
- compute ew1 = exp(linfunk1)/(1+exp(linfunk1)+exp(linfunk2)).
- compute ew2 = exp(linfunk2)/(1+exp(linfunk1)+exp(linfunk2)).
* - compute ew3 = 1-ew1-ew2.
- compute sel = uniform(1).
- do if (sel <= ew1).
-   compute Kneipe = 1.
- else if (sel <= ew1+ew2).
-   compute kneipe = 2.
- else.
-   compute kneipe = 3.
- end if.
- end case.
+ end loop.
+ end file.
end input program.

compute nr = $casenum.
formats alter, geschl, kneipe (f1.0), nr (f2.0).

variable Labels geschl 'Geschlecht'.
value labels geschl 1 'Frau' 0 'Mann'.

variable width  nr alter geschl kneipe (5).
variable level  nr geschl kneipe (nominal) /alter (scale).
variable role
  /input  geschl alter
  /target  kneipe
  /none nr.

save  outfile='kneipe.sav'
/keep = nr alter, geschl, kneipe.

```


Literatur

- Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.
- Allison, P. D. (1999). *Logistic Regression Using the SAS System*. Cary, NC: SAS Institute.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2008). *Multivariate Analysemethoden* (12. Aufl.). Berlin: Springer.
- Baltes-Götz, B. (1994). *Einführung in die Analyse von Strukturgleichungsmodellen mit LISREL 7 und PRELIS unter SPSS*. Online-Dokumentation: <http://www.uni-trier.de/index.php?id=22734>
- Baltes-Götz, B. (2009). *Moderatoranalyse per multipler Regression mit SPSS*. Online-Dokument: <http://www.uni-trier.de/index.php?id=22528>
- Baltes-Götz, B. (2012). *Statistisches Praktikum mit SPSS Statistics 20*. Online-Dokument: <http://www.uni-trier.de/index.php?id=22552>
- Bühl, A. (2012). *SPSS 20. Einführung in die moderne Datenanalyse*. München: Pearson Studium.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analyses for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Cox, D. R. & Snell, E. J. (1989). *The Analysis of Binary Data*. London: Chapman and Hall.
- DeCarlo, L.T. (2003). Using the PLUM procedure of SPSS to fit unequal variance and generalized signal detection models, *Behavior Research Methods, Instruments, & Computers*, 35(1), 49-56. Online-Dokument: <http://www.columbia.edu/~ld208/brmic03.pdf> (Abgerufen: 11.06.2012)
- Fox, J. & Weisberg, S. (2011). *An R Companion to Applied Regression*. Thousand Oaks, CA: Sage.
- Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). New York: Wiley & Sons.
- Jaccard, J. (2001). *Interaction Effects in Logistic Regression*. Thousand Oaks, CA: Sage.
- Keele, L. & Park, D.K. (2006). *Difficult Choices: An Evaluation of Heterogenous Choice Models*. Online-Dokument: <http://www.nd.edu/~rwilliam/oglm/ljk-021706.pdf> (Abgerufen: 11.06.2012)
- Kleinbaum, D. G. (1994). *Logistic Regression. A Self-Learning Text*. New York: Springer.
- Liang, K.-L. & Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- Long, J. S. (1997). *Regression Models for Categorical and limited Dependent Variables*. Thousand Oaks, CA: Sage.
- McCullagh, P. (1980). Regression Models for Ordinal Data, *Journal of the Royal Statistical Society, Series B*, 42, 109-142.
- Menard, S. (1995). *Applied Logistic Regression Analysis*. Thousand Oaks: Sage.
- Nagelkerke, N.J. D. (1991). A note on the general definition of the coefficient of determination. *Biometrika*, 78, 691-692.
- Norušis, M.J. (2005). *SPSS 14.0. Statistical Procedures Companion*. Upper Saddle River, NJ: Prentice Hall.
- Norušis, M.J. (2008). *SPSS 16.0. Advanced Statistical Procedures Companion*. Upper Saddle River, NJ: Prentice Hall.
- Nichols, D. (1993). Interpreting MANOVA parameter estimates. *Keywords*, 50, 8-14.
- O'Connell, A. (2006). *Logistic Regression Models for Ordinal Response Variables*. Thousand Oaks: Sage.
- Ryan, T. P. (1997). *Modern Regression Methods*. New York: Wiley & Sons.

- IBM Corp. (2011). *IBM SPSS Regression 20*. Online-Dokument:
<http://publib.boulder.ibm.com/infocenter/spsstat/v20r0m0/index.jsp> (Abgerufen: 11.06.2012)
- Stelzl, I. (2000). What sample sizes are needed to get correct significance levels for log-linear models?
Methods of Psychological Research Online, 2000, Vol. 5, No. 2, 95-116.
- Tabachnik, B.G. & Fidell, L.S. (2007, 5th ed.). *Using multivariate statistics*. Boston: Pearson Education.
- Urban, D. (1993). *Logit-Analyse: Statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen*. Stuttgart: Fischer.
- Williams, R. (2008). *Ordinal regression models: Problems, solutions, and problems with the solutions*.
German Stata User Group Meetings, June 27, 2008, Online-Dokument:
<http://www.stata.com/meeting/germany08/GSUG2008.pdf> (Abgerufen: 11.06.2012)
- Williams, R. (2009). *Using Heterogenous Choice Models to Compare Logit and Probit Coefficient Across Groups*. Online-Dokument:
http://nd.edu/~rwilliam/oglm/RW_Hetero_Choice.pdf (Abgerufen: 11.06.2012)

Stichwortverzeichnis

Abweichungskodierung	48	McFadden	39
Basismodell	36	Modellgleichung	
Binäre logistische Regression	9	binäre logist. Regr.	9
Block		multinomiale logist. Regr.	62
von Regressoren	58	ordinale logist. Regr.	72
Cook's Distanz	35	Modellgüte	36
Correct Classification Rate	39	Multikollinearität	89
covariate pattern	21, 26	Multinomiale	
Cox	38	logistische Regression	62
Devianz	24	Multinomialverteilung	63
Devianzresiduen	27	Nagelkerke	38
Devianz-Residuen	27	Newton-Raphson	19
Diskriminanzanalyse	6	NOMREG	8, 69
Dummy-Kodierung	51, 77	odds	11
Effektgröße	12, 43	Odds	73
Effiziente Schätzer	68	odds ratio	12, 43, 74
Einflussreiche Fälle	35	odds-Gleichung	11
Fehlerquadratsumme	28	Parallele Regressionen	72
F-Test der lin. Regr.	37	Parameterschätzung	20
Generalisiertes R^2	38	Pearson- χ^2 -Statistik	21
GENLOG	6	Pearson-Residuen	28
Gewinnchance	11	PLUM	8, 72, 81
Goodness-of-fit-Statistiken	21	TEST	81
Gruppenresiduen	31	PoLytomous Universal Model	8
Hierarchische Regressionsanalyse	58	Probit-Analyse	14
Hosmer-Lemeshow-Goodness-of-Fit-Test	25	proportional-odds model	74
Indikatorkodierung	77	Pseudo- R^2 -Statistiken	38
Indikator-Kodierung	51	R^2 -Statistiken	38
Individualresiduen	28	Referenzkategorie	51
Interaktion	52	Regressionskoeffizienten	42
Kategoriale Regressoren	46	Residuen	
Kern	70	Devianz	27
Klassifikationsdiagramm	41	Pearson	28
Klassifikationstabelle	39	standardisierte	28
Kodiervariablen	47	SAS	70, 72
Konsistente Schätzer	68	Saturiertes Modell	24
Kontrastvektoren	49	Schwellen	77
Kreuzvalidierung	40	Score-Statistik	59
Kumulatives Logit-Modell	72	Separation	
Leere Zellen	89	quasi-vollständige	89
Likelihood	36	vollständige	92
Likelihood-Funktion	19	Skalenmodell	85
Likelihood-Ratio – Test	36	Snell	38
lineare Regressionsanalyse	6	Standardfehler	42
Loess	29	Standardisierte Residuen	28
LOGISTIC REGRESSION	7	Stata	70
Logistische Verteilung	12	Stichprobengröße	10
Logistische Verteilungsfunktion	9	Stichprobenmodell	10
Logistischen Verteilung	84	TEST-Subkommando	69
Logit	11	Trennung	
Logit-Gleichung	11	quasi-vollständige	89
Log-lineare Modelle	6	vollständige	92
LogXact	20	Trennwert	34, 41
Lokations - Skalen - Modelle	84	Wald-Statistik	43
Lokationsmodell	85		
Maximum-Likelihood-Schätzmethode	18		