Universität der Bundeswehr München

Fakultät für Humanwissenschaften
Sozialwissenschaftliche Methodenlehre
Prof. Dr. Daniel Lois

Logistische Regression (in SPSS)

Stand: April 2015 (V2.0)

Inhaltsverzeichnis

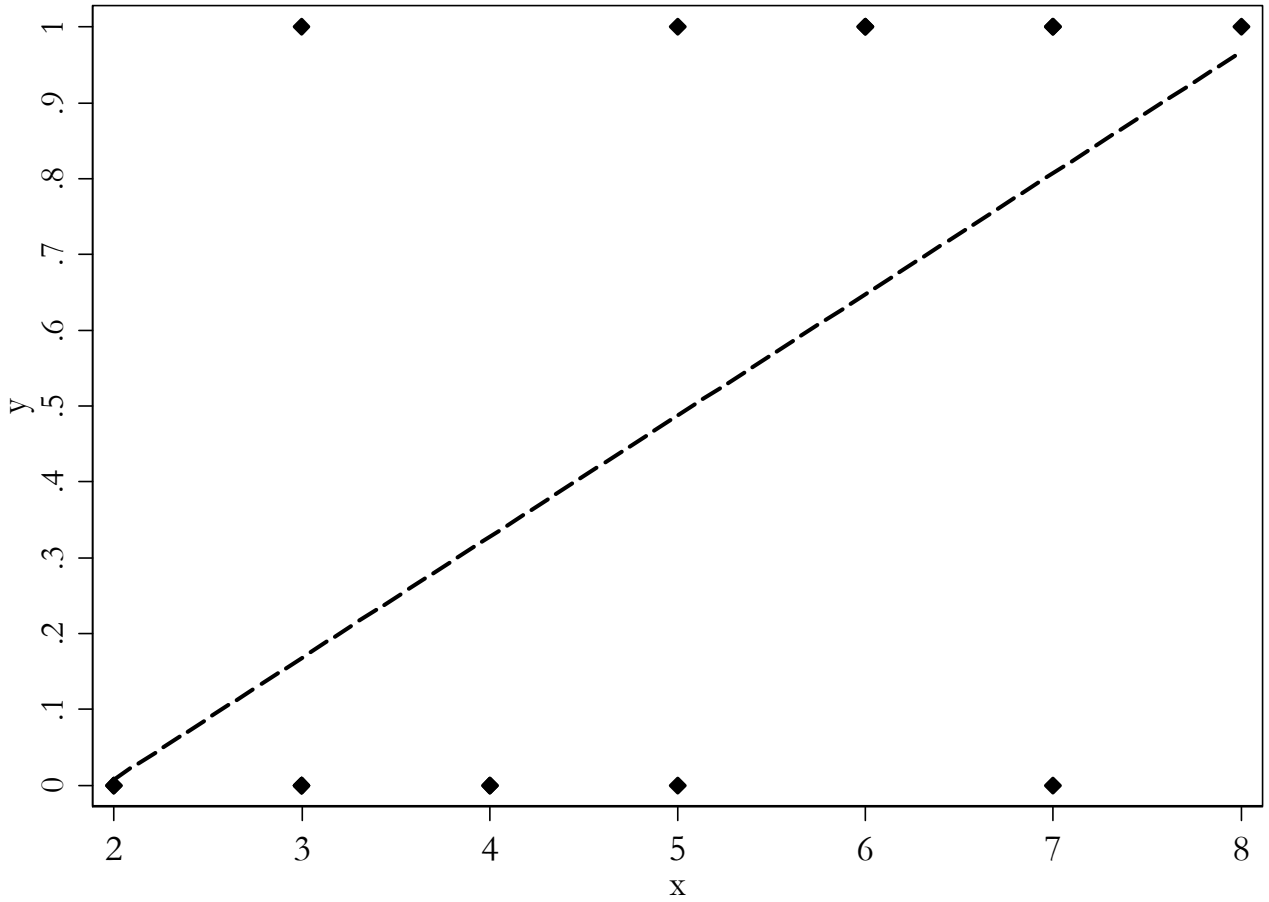
Grundlagen

- Mit logistischen Regressionsmodellen wird die Abhängigkeit **nominaler** abhängiger Variablen (z.B. Teilnahme an Weiterbildung (ja/nein) von anderen unabhängigen Variablen, die ein beliebiges Messniveau aufweisen können, untersucht
- Ist die abhängige Variable dichotom (zwei Ausprägungen), wird die binäre logistische Regression angewandt
- Bei mehrstufig kategorialen abhängigen Variablen (z.B. Familienstand), kommt die multinomiale logistische Regression zum Einsatz (wird hier nicht behandelt), die eine Erweiterung des binären Modells darstellt

Grundlagen

- Die logistische Regression wird anstelle der linearen Regression eingesetzt, da einige Voraussetzungen für die Anwendung einer linearen Regression (BLUE-Annahmen) bei nominalen AV nicht gegeben sind
- Wird eine lineare Regression auf eine dichotome AV gerechnet (lineares Wahrscheinlichkeitsmodell), kommt es zu einer Heteroskedastizität der Residuen, da die Abweichungen zwischen Vorhersage- und Beobachtungswerten im mittleren Wertebereich einer metrischen UV zwangsläufig am größten sind (siehe nächste Folie)

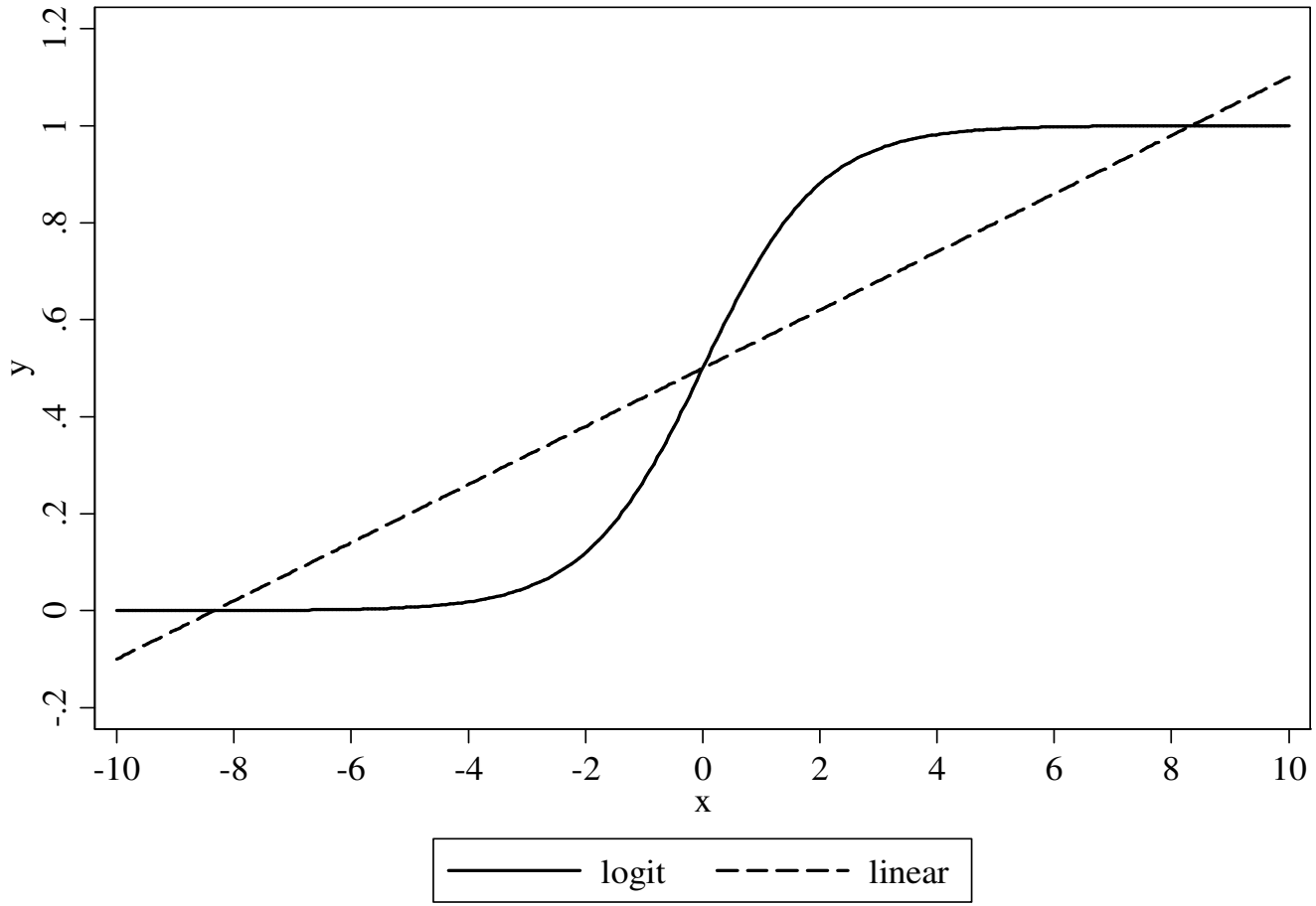
Grundlagen



Grundlagen

- Außerdem können die mit dem linearen Modell geschätzten Wahrscheinlichkeiten außerhalb des für Wahrscheinlichkeiten zulässigen Bereichs [0,1] liegen
- Die folgende Grafik zeigt einen positiven Zusammenhang zwischen einer dichotomen abhängigen Variablen mit den Ausprägungen 0 und 1 und einer metrischen UV
- Die Vorhersagewerte der linearen Regression übersteigen 1, wenn die Werte der unabhängigen Variablen über 8 liegen; außerdem werden sie bei geringen x-Werten kleiner als 0
- Um dieses Problem zu vermeiden, verläuft die Regressionskurve bei der logistischen Regression nicht linear, sondern langgestreckt S-förmig und nähert sich asymptotisch an die Extremwerte 1 und 0 an

Grundlagen



Grundlagen

- Fiktive Beispieldaten (n = 20):
  - AV: Teilnahme an Weiterbildung („wb“, 0 = nein, 1 = ja)
  - Arbeiter (1 = ja, 0 = Angestellter)

arbeiter * wb Kreuztabelle

			wb		
			,00	1,00	Gesamt
arbeiter	,00	Anzahl	3	7	10
		% von arbeiter	30,0%	70,0%	100,0%
	1,00	Anzahl	7	3	10
		% von arbeiter	70,0%	30,0%	100,0%
	Gesamt	Anzahl	10	10	20
		% von arbeiter	50,0%	50,0%	100,0%

Grundlagen

- Die **Wahrscheinlichkeit**, an Weiterbildung teilgenommen zu haben, entspricht der Anzahl der Teilnehmern dividiert durch n: 10/20 = 0,5
- Die Teilnahmewahrscheinlichkeit insgesamt beträgt demnach 50%
- Die Teilnahmewahrscheinlichkeit für Arbeiter (30%) ist dabei wesentlich kleiner als die der Angestellten (70%):
  - Arbeiter: 3/10 = 0,3
  - Angestellte: 7/10 = 0,7

Grundlagen

- Die **Chance**, an Weiterbildung teilgenommen zu haben, entspricht der Anzahl der Teilnehmer dividiert durch die Anzahl der Nichtteilnehmer: 10/10 = 1,0
- Die Chance, an Weiterbildung teilgenommen zu haben, ist somit genauso groß wie die Chance, nicht teilgenommen zu haben
- Die Teilnahmechance für Arbeiter (0,43 : 1) ist wiederum wesentlich kleiner als die der Angestellten (2,33 : 1)
  - Arbeiter: 3/7 = 0,43
  - Angestellte: 7/3 = 2,33

Grundlagen

- Formal ist eine Chance (odds) definiert als Verhältnis von zwei Wahrscheinlichkeiten (im Beispiel: Wahrscheinlichkeit einer Teilnahme durch Gegenwahrscheinlichkeit):

$$Chance_{\text{Teilnahme}} = \frac{Wahrscheinlichkeit_{\text{Teilnahme}}}{Wahrscheinlichkeit_{\text{Nichtteilnahme}}}$$

- Allgemein: Wahrscheinlichkeit, dass die abhängige Variable den Wert 1 annimmt dividiert durch Gegenwahrscheinlichkeit:

$$odds = \frac{P(y_i = 1)}{1 - P(y_i = 1)}$$

Grundlagen

- Das **Odds Ratio** ist ein Verhältnis zweier Chancen
- Die Teilnahmechance der Arbeiter betrug 0,43 : 1 und die der Angestellten 2,33 : 1
- Das Odds Ratio für den Vergleich von Arbeitern mit Angestellten beträgt daher: 0,43 / 2,33 = 0,18
- Die Teilnahmechance der Arbeiter beträgt nur das 0,18-fache der Teilnahmechance der Angestellten
- Das Odds Ratio für den Vergleich von Angestellten mit Arbeitern beträgt entsprechend: 2,33 / 0,43 = 5,42
- Die Teilnahmechance der Angestellten ist 5,4 mal höher als die der Arbeiter

Grundlagen

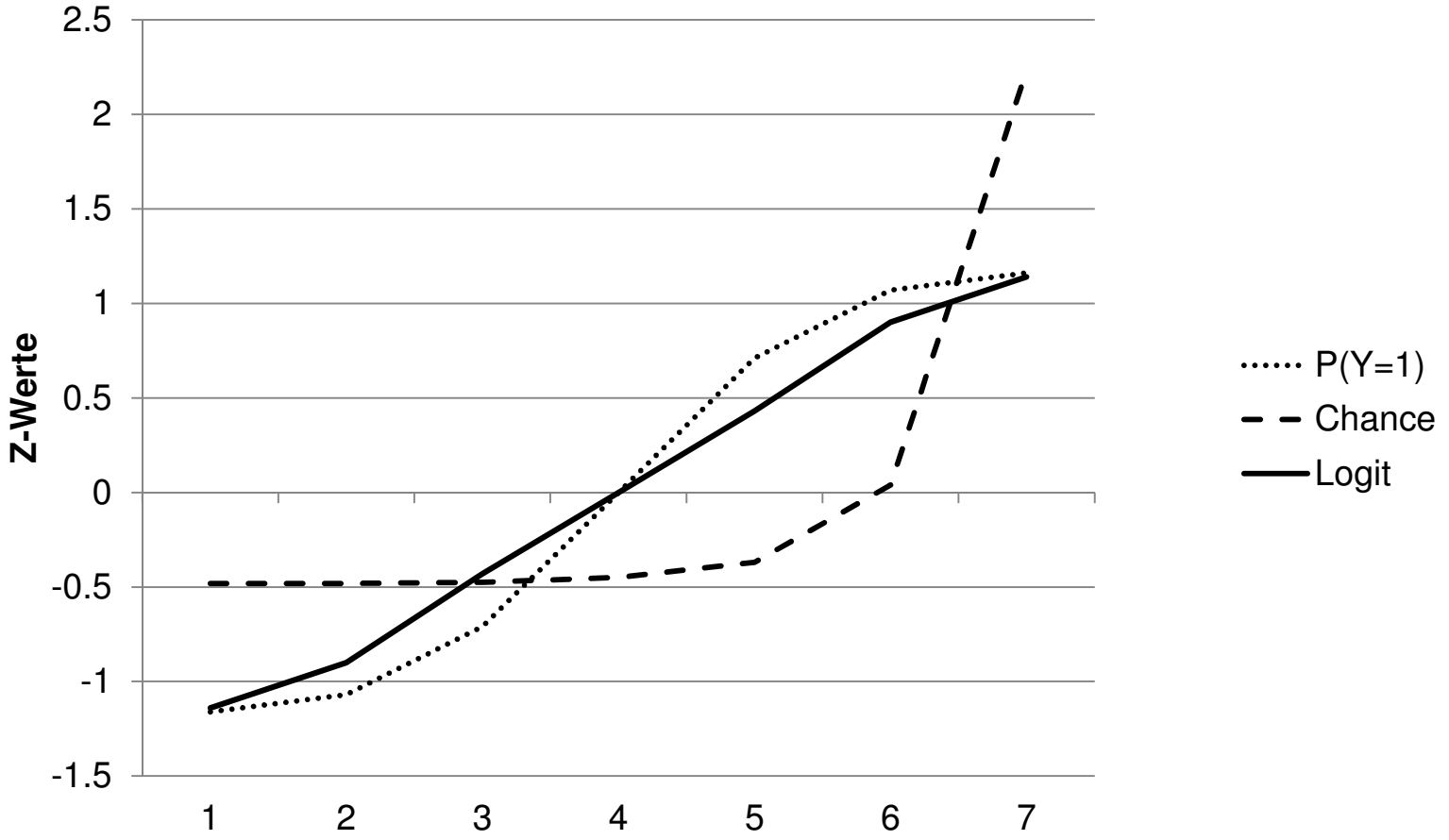
- Die logarithmierte Chance eignet sich sehr gut als abhängige Variable für ein Regressionsmodell, da sie nach oben und unten nicht begrenzt und zudem symmetrisch (um den Ursprung) ist
- Die Chancen (odds) sind dagegen nicht symmetrisch, da sie sich im Bereich < 1 aufgrund der unteren Grenze sehr viel langsamer verändern als im nach oben hin offenen Bereich größer als eins (siehe nächste Folien)

Grundlagen

Tabelle: Wahrscheinlichkeiten, Odds und Logits im Vergleich

P(Y=1)	Chance (odd) = p(y=1)/1-p(y=1)	Logit / ln(odd)
.01	1/99 = .01	-4.60
.05	5/95 = .05	-2.94
.20	20/80 = .25	-1.39
.50	50/50 = 1.00	0.00
.80	80/20 = 4.00	1.39
.95	95/5 = 19.00	2.94
.99	99/1 = 99.00	4.60

Grundlagen



Grundlagen

- Formal lässt sich das logistische Regressionsmodell wie folgt ausdrücken:

$$\ln\left(\frac{P(y_i=1)}{1-P(y_i=1)}\right) = L_i = b_0 + b_1 x_{i1} + \ldots + b_j x_{ij}$$

- Der Ausdruck links vom Gleichheitszeichen steht für die logarithmierte Chance, dass die 0/1-codierte abhängige Variable für Untersuchungseinheiten i = 1,2,…, n den Wert 1 annimmt
- Der Ausdruck rechts vom Gleichheitszeichen kommt uns aus der linearen Regression bekannt vor: Der Einfluss von Kovariaten (x_i) auf das Logit wird auch hier durch eine Linearkombination berechnet, d.h. durch eine Regressionskonstante (b_0) und Regressionsgewichte (b_j)

Grundlagen

- Erst in einem zweiten Schritt kann in der logistischen Regression die Wahrscheinlichkeit berechnet werden, dass die 0/1-codierte AV den Wert 1 annimmt
- Hierzu wird das zuvor berechnete Logit (L, logarithmierte Chance) in folgende Gleichung eingesetzt:

$$P = \frac{1}{1 + e^{-L}}$$

- e steht für die Eulersche Zahl, die Basis des natürlichen Logarithmus (ungefähr = 2,718)
- Die Wahrscheinlichkeiten bilden nun die Vorhersagewerte der logistischen Regression und bestimmen grafisch den Verlauf der Regressionskurve

Grundlagen

- Werden die Beispieldaten in SPSS eingegeben und eine logistische Regression berechnet („wb“ codiert mit 1 für Teilnahme und 0 für Nichtteilnahme, „arbeiter“ codiert mit 1 für Arbeiter und 0 für Angestellte) erhält man (u.a.) folgenden Output:

Variablen in der Gleichung

		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1	arbeiter	-1,695	,976	3,015	1	,082	,184
	Konstante	,847	,690	1,508	1	,220	2,333

Grundlagen

- Der Logit-Koeffizient für Arbeiter (b = -1,695) besagt, dass die logarithmierte Chance, an Weiterbildung teilgenommen zu haben, um 1,695 Einheiten niedriger liegt als bei Angestellten
- Allgemein gibt der b-Koeffizient damit an, wie sich die logarithmierte Chance für y = 1 (hier: Teilnahme) verändert, wenn sich die unabhängige Variable um eine Einheit erhöht
- Die Konstante entspricht der logarithmierten Teilnahmechance für x = 0, d.h. für Angestellte
- Wird diese logarithmierte Teilnahmechance entlogarithmiert, erhalten wir die weiter oben bereits berechnete Chance: $e^{0,847} = 2,33$

Grundlagen

- Die Teilnahmechance für Arbeiter erhalten wir, indem für zunächst die logarithmierte Teilnahmechance für Arbeiter, d.h. das Logit ausrechnen: 0,847-1,695 = -0,848
- Wird dieses Logit entlogarithmiert ($e^{-0,848} = 0,43$) erhalten wir die bereits bekannte Teilnahmechance von 0,43 : 1
- Dass Odds Ratio für den Vergleich von Arbeitern mit Angestellten wird in der Spalte „Exp(B)“ ausgegeben und beträgt wiederum 0,18 : 1
- Es handelt sich hier um den entlogarithmierten Logit-Koeffizienten für den Effekt von Arbeiter: $e^{-1,695} = 0,18$

Grundlagen

- Schließlich können auch Wahrscheinlichkeiten anhand des Modelloutputs berechnet werden
- Zum Beispiel beträgt die Wahrscheinlichkeit, dass Arbeiter an Weiterbildung teilgenommen haben, 30%
- Berechnung:

$$P = \frac{1}{1 + e^{0,848}} = 0,3$$

Grundlagen

- Die Wald-Statistik (äquivalent zur t-Statistik in der linearen Regression) testet die Nullhypothese, dass der jeweilige Regressionskoeffizient b in der Grundgesamtheit 0 ist
- Berechnung der Wald-Statistik im Beispiel:

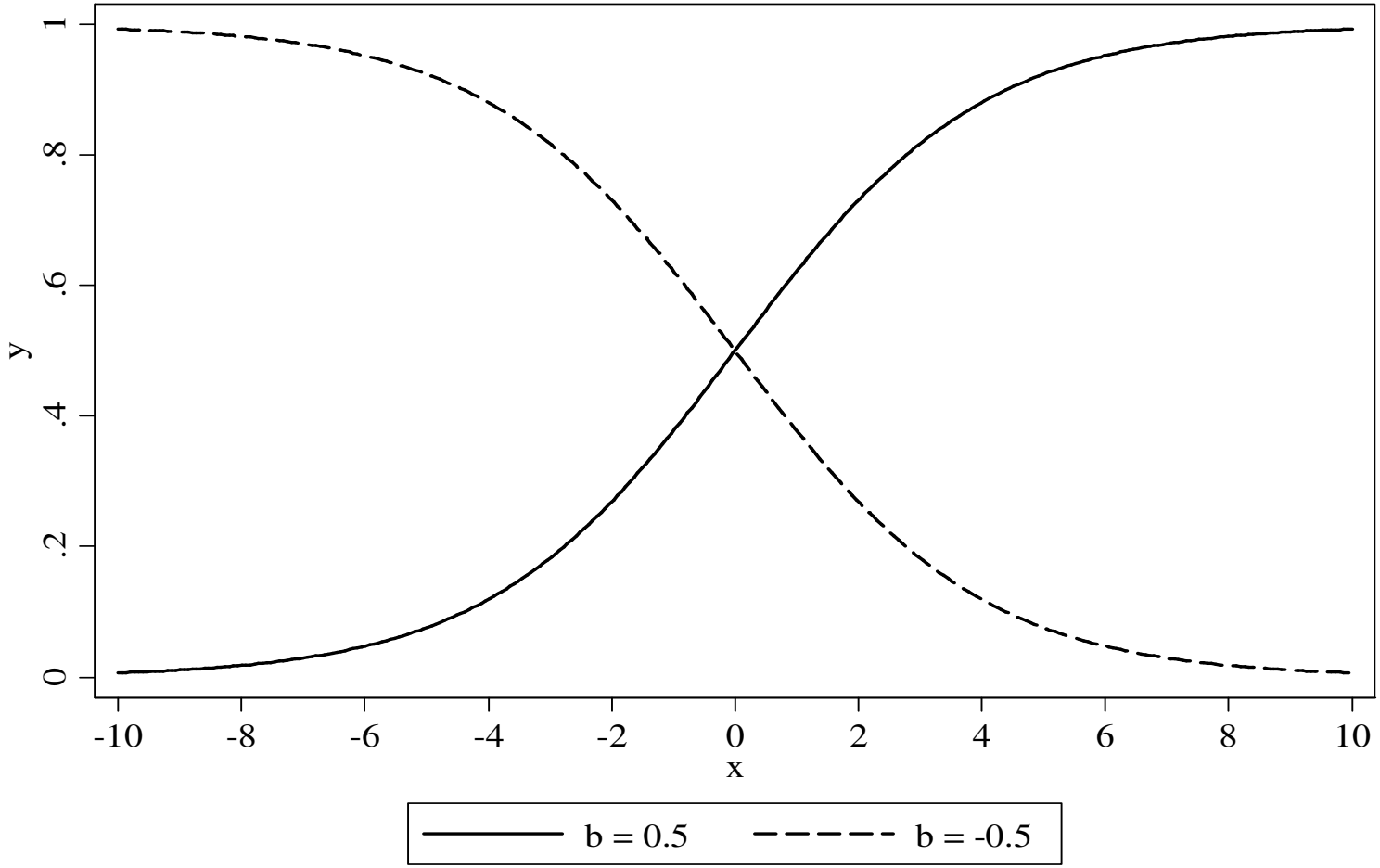
$$Wald = \left(\frac{b}{S.E._b} \right)^2 = \left(\frac{-1{,}695}{0{,}976} \right)^2 = 3{,}02$$

- Für jeden Wald-Wert wird in der Spalte „p“ – wie gewöhnlich – eine Wahrscheinlichkeit für die Ablehnung einer Nullhypothese angegeben, die tatsächlich wahr ist (der Wert im Beispiel, 8,2%, liegt über der konventionellen Signifikanzgrenze von 5%)

Grundlagen

- Auf der folgenden Grafik ist zu sehen, wie sich unterschiedliche b- bzw. Logitkoeffizienten auf den Verlauf der Regressionskurve auswirken
- Bei einem $b > 0$ steigt die Regressionskuve von links unten nach rechts oben, bei einem $b < 0$ fällt sie von links oben nach rechts unten
- Bei $b = 0$ (nicht dargestellt) verläuft die Regressionskurve der logistischen Regression genau wie bei der linearen Regression parallel zur x-Achse

Grundlagen



Grundlagen

- Aus dieser Darstellung ergeben sich bestimmte Richtlinien für die Ergebnisinterpretation in logistischen Regressionen:
  - Bei b > 0 steigen die logarithmierten Chancen für y = 1 um b Einheiten, wenn die unabhängige Variable um eine Einheit ansteigt
  - Das heißt: Bei b > 0 lässt sich zu den Wahrscheinlichkeiten nur sagen, dass sie bei Anstieg der Kovariate steigen bzw. bei b < 0 sinken
  - Sie steigen/sinken – bei metrischen Kovariaten – jedoch nicht um b, da die Regressionskurve nicht linear, sondern s-förmig verläuft (!)
  - Bei einem Odds Ration > 1 steigen die Chancen für y = 1 um $|1-e^b|$ wenn die unabhängige Variable um eine Einheit ansteigt
  - Bei einem Odds Ratio < 1 sinken die Chancen für y = 1 entsprechend um $|1-e^b|$ und bei einem Odds Ratio = 1 besteht kein Zusammenhang zwischen der Kovariate und der Chance für y = 1

Grundlagen

Übersicht: Auswirkungen positiver und negativer Regressionskoeffizienten auf die Eintrittswahrscheinlichkeit des Ereignisses y = 1

	Logit (**ln** der Chance)	Exp(b) (Odds Ratio)	Odds Ratio…	P(y = 1)
b > 0	steigt um b	$e^b > 1$	...steigt pro Einheit von x **um** den Faktor l1- e^b l (**auf** e^b)	steigt
b < 0	sinkt um b	$e^b < 1$	...sinkt pro Einheit von x **um** den Faktor l1- e^b l (**auf** e^b)	sinkt

Logit-Funktion und Modellfit

- Bei der linearen Regression werden die Regressionsparameter nach der Methode der kleinsten Quadrate bestimmt
- Bei der logistischen Regression erfolgt die Schätzung nach der Maximum-Likelihood-Methode
- Diese folgt in etwa folgender Logik: Gegeben sind Daten einer Stichprobe (d.h. die Information, bei welchen Personen die abhängige Variable den Wert 1 oder 0 annimmt sowie die Informationen zu unabhängigen Variablen)
- Es sind nun diejenigen Regressionskoeffizienten gesucht, bei denen das Auftreten der Stichprobendaten am wahrscheinlichsten ist

Logit-Funktion und Modellfit

- Gesucht wird iterativ diejenige Kombination von Regressionskoeffizienten, welche eine sog. Likelihood-Funktion maximiert und damit die beste Trennung zwischen den Ausprägungen der abhängigen Variable (0 und 1) bewirkt
- Hier wird die logarithmierte Likelihood-Funktion (LL) verwendet (für i = 1,…n Untersuchungsobjekte):

$$LL = \ln(L) = \sum_{i=1}^{n} \ln(P(y_i = 1))^{y_i} + \sum_{i=1}^{n} \ln(1 - P(y_i = 1))^{1-y_i}$$

Logit-Funktion und Modellfit

- Die beiden zu addierenden Summen entsprechen den verschiedenen Vorhersagewerten für die beiden Ausprägungen der abhängigen Variablen ($y_i = 1$ und $y_i = 0$)
- Die Funktion kann Werte zwischen 0 (bestmögliches Modell) und minus unendlich annehmen
- Der Wert 0 wird erreicht, wenn für alle beobachteten Fälle $y_i = 1$ auch eine Wahrscheinlichkeit P = 1 durch das Regressionsmodell vorhergesagt wird und für alle beobachteten Fälle $y_i = 0$ eine Wahrscheinlichkeit von P = 0

Logit-Funktion und Modellfit

- Je größer die Differenzen zwischen vorhergesagten Wahrscheinlichkeiten und den Beobachtungswerten sind, desto kleiner ist auch die Funktion LL
- In SPSS wird die Funktion LL allerdings mit -2 multipliziert, was den Wert „-2LL“ ergibt
- Diese Funktion kann Werte zwischen 0 und plus unendlich annehmen. Bei -2LL = 0 handelt es sich wiederum um ein perfekt angepasstes Modell
- Je größer -2LL wird, desto schlechter ist das Modell an die Daten angepasst
- Die nächste Folie zeigt die Berechnung der -2LL-Funktion

wb	arbeiter	P(y=1)	LL-Wert für y=0: (ln(1-p(y=1))	LL-Wert für y=1: ln(p(y=1))
0	1	0,3	-0,357	
0	1	0,3	-0,357	
0	1	0,3	-0,357	
0	1	0,3	-0,357	
0	1	0,3	-0,357	
0	1	0,3	-0,357	
0	1	0,3	-0,357	
0	0	0,7	-1,204	
0	0	0,7	-1,204	
0	0	0,7	-1,204	
1	1	0,3		-1,204
1	1	0,3		-1,204
1	1	0,3		-1,204
1	0	0,7		-0,357
1	0	0,7		-0,357
1	0	0,7		-0,357
1	0	0,7		-0,357
1	0	0,7		-0,357
1	0	0,7		-0,357
1	0	0,7		-0,357
			-6,111	-6,111

$$LL = \sum_{i=1}^{n} \ln(P(y_i = 1))^{y_i} + \sum_{i=1}^{n} \ln(1 - P(y_i = 1))^{1-y_i} = -6{,}111 - 6{,}111 = -12{,}222$$

Logit-Funktion und Modellfit

- Wird der Wert der LL-Funktion mit -2 multipliziert (-12,222*-2), erhält man (bis auf einen kleinen Rundungsfehler) den entsprechenden -2LL-Wert aus dem SPSS-Output:

Modellzusammenfassung

Schritt	-2 Log-Likelihood	Cox & Snell R-Quadrat	Nagelkerkes R-Quadrat
1	24,435[a]	,152	,202

a. Schätzung beendet bei Iteration Nummer 3, weil die Parameterschätzer sich um weniger als ,001 änderten.

Logit-Funktion und Modellfit

- Auf der -2LL-Funktion basieren zwei wesentliche Kennziffern zur Güte des Gesamtmodells:
  - Der „Omnibus-Test der Modellkoeffizienten“ (äquivalent zum F-Test der linearen Regression)
  - Pseudo-R^2-Maßzahlen zur Erklärungskraft des Modells – z.B. die hier vorgestellte Version nach McFadden

Logit-Funktion und Modellfit

Omnibus-Tests der Modellkoeffizienten

		Chi-Quadrat	df	Sig.
Schritt 1	Schritt	3,291	1	,070
	Block	3,291	1	,070
	Modell	3,291	1	,070

Logit-Funktion und Modellfit

- Der „Omnibus-Test der Modellkoeffizienten“ testet die Nullhypothese, dass alle b_1-Koeffizienten des Modells in der Grundgesamtheit = 0 sind
- Der Wert berechnet sich aus der X^2-verteilten Differenz zwischen dem -2LL-Wert des Modells ohne unabhängige Variablen („Nur konstanter Term“) und des -2LL-Wertes für das Modell mit unabhängigen Variablen
- Im Beispiel kann die Nullhypothese, dass alle b_1-Koeffizienten (im Beispiel nur einer) gleich 0 sind, nicht zurückgewiesen werden, da ein Chi-Quadrat von 3,29 bei einem Freiheitsgrad (= Anzahl der b_1-Koeffizienten) nur tendenziell signifikant ist

Logit-Funktion und Modellfit

- Wer die Berechnung nachvollziehen will, kann durch Aktivierung der Option „Iterationsprotokoll“ im Menü „Optionen“ den folgenden Output erzeugen
- Der -2LL-Wert des Nullmodells (27,726) steht in einer Fußnote unterhalb der Tabelle

Iterationsprotokoll[a,b,c,d]

Iteration		-2 Log-Likelihood	Koeffizienten	
			Constant	arbeiter
Schritt 1	1	24,444	,800	-1,600
	2	24,435	,847	-1,694
	3	24,435	,847	-1,695

a. Methode: Einschluß

b. Konstante in das Modell einbezogen.

c. Anfängliche -2 Log-Likelihood: 27,726

d. Schätzung beendet bei Iteration Nummer 3, weil die Parameterschätzer sich

27,726-24,435 = 3,291

Logit-Funktion und Modellfit

- Ein aus der -2LL-Funktion abgeleitetes Maß für die Modellgüte ist das Pseudo-R^2 nach McFadden

$$PseudoR^2_{McFadden} = 1 - \frac{-2LL(\text{Endmodell})}{-2LL(\text{Konstantenmodell})} = 1 - \frac{24{,}435}{27{,}726} = 0{,}12$$

- Es handelt sich bei diesem Pseudo-R^2 um ein PRE (proportional reduction of error)-Maß
- Der Minimalwert von 0 bedeutet entsprechend, dass die unabhängigen Variablen die Erklärungskraft des Modells nicht verbessern können
- Der Maximalwert 1 wird erreicht, wenn das Modell perfekt an die beobachteten Daten angepasst ist

Logit-Funktion und Modellfit

- Im Beispiel wird die Erklärungskraft des Modells bei der Hinzuziehung des einzigen Prädiktors „arbeiter“ um 12% gegenüber dem Nullmodell erhöht
- Die Werte für McFaddens Pseudo-R^2 sind in der Regel kleiner als diejenigen des R^2-Wertes in der linearen Regression und alternativer Pseudo-R^2-Werte wie die in SPSS ausgegebenen Varianten nach Cox & Snell (0,15) sowie Nagelkerke (0,20)
- In Analogie zum korrigierten R^2 der linearen Regression, das die Zahl der Regressionsparameter berücksichtigt, kann auch das Pseudo-R^2 nach McFadden (bei mehr als einem Prädiktor) adjustiert werden, indem man vom -2LL-Wert des Endmodells die Anzahl der Modellparameter (Freiheitsgrade im Omnibustest) abzieht

Anwendungsbeispiel

- Der Output zeigt ein auf Mikrozensusdaten basierendes logistisches Regressionsmodell mit der abhängigen Variablen „Teilnahme an beruflicher Weiterbildung“ im Jahr 2002 (1 = ja) und den unabhängigen Dummy-Variablen „Beschäftigung im öffentlichen Dienst“ (1 = ja) und „Berufswechsel im letzten Jahr“ (1 = ja) sowie dem metrischen Alter

Variablen in der Gleichung

		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1[a]	alter	-,040	,003	215,960	1	,000	,961
	öffentlich(1)	,675	,063	115,284	1	,000	1,965
	brwechsel(1)	,313	,102	9,504	1	,002	1,368
	Konstante	-1,426	,107	179,299	1	,000	,240

a. In Schritt 1 eingegebene Variablen: alter, öffentlich, brwechsel.

Anwendungsbeispiel

- Pro Anstieg des Alters um 1 Jahr reduziert sich die logarithmierte Chance einer Weiterbildungsteilnahme um b = -0,040
- Ebenfalls nicht falsch ist die Aussage, dass sich die Wahrscheinlichkeit einer Weiterbildungsteilnahme mit steigendem Alter reduziert
- Da die logistische Funktionskurve s-förmig verläuft, reduziert sich die Teilnahmewahrscheinlichkeit allerdings nicht proportional, sondern je nach Altersbereich unterschiedlich stark (!)
- Pro Anstieg des Alters um 1 Jahr reduziert sich die Chance einer Weiterbildungsteilnahme um Exp(b) = 0,961, d.h. um 3,9%
- Wird z.B. eine 25jährige mit einer 50jährigen Person verglichen, ist die Teilnahmechance der 50jährigen Person etwa 0,37-mal so groß wie die Teilnahmechance der 25jährigen Person (-0,04 * 25 = -1,0; e^{-1} = 0,368)

Anwendungsbeispiel

- Die Nullhypothese, wonach Alter und Weiterbildungsbeteiligung in der Grundgesamtheit nicht zusammenhängen, kann abgelehnt werden, da ein Wald-Wert von 215,96 (= $(-0{,}04 / 0{,}003)^2$) hochsignifikant ist
- Die Teilnahmechance von Personen, die im öffentlichen Dienst arbeiten, ist 1,965 mal so groß oder beträgt das 1,965-fache oder ist 96,5% größer als bei Personen, die nicht im öffentlichen Dienst arbeiten
- Die Teilnahmechance von Personen mit Berufswechsel ist 1,37mal so groß wie die Chance von Personen ohne Berufswechsel
- Alle Effekte gelten jeweils bei Kontrolle der anderen Prädiktoren

Anwendungsbeispiel

Variablen in der Gleichung

		Regressions koeffizientB	Standardf ehler	Wald	df	Sig.	Exp(B)
Schritt 1[a]	alter	-,040	,003	215,960	1	,000	,961
	öffentlich(1)	,675	,063	115,284	1	,000	1,965
	brwechsel(1)	,313	,102	9,504	1	,002	1,368
	Konstante	-1,426	,107	179,299	1	,000	,240

a. In Schritt 1 eingegebene Variablen: alter, öffentlich, brwechsel.

- Beispiel: Logit einer Person, die 30 Jahre alt ist, im öffentlichen Dienst arbeitet und in den letzten Jahren nicht ihren Beruf gewechselt hat:
  -1,426 - (30*-0,040) + (0,675*1) + (0,313*0) = -1,951
- Die Teilnahmewahrscheinlichkeit beträgt 12,4%:

$$P = \frac{1}{1 + e^{1,951}} = 0,124$$

Anwendungsempfehlungen

- In schrittweisen (hierarchischen) Regressionen können sich Logit-Koeffizienten und Odds-Ratios zwischen verschiedenen Modellen auch dann verändern, wenn keine Korrelationen zwischen den unabhänigen Variablen bestehen! (Mood 2010)
- Bei schrittweisen Modellen sollten daher die „average marginal effects“ (AME, durchschnittliche Marginaleffekte auf die Wahrscheinlichkeit) verwendet werden (berechenbar z.B. in STATA)
- In SPSS können die AME durch ein lineares Wahrscheinlichkeitsmodell approximiert werden

Anwendungsempfehlungen

- Die Interpretation von Interaktionseffekten ist im Logit-Modell nur in Bezug auf die logarithmierten Odds äquivalent zu linearen Regressionsmodellen
- Im Hinblick auf die Wahrscheinlichkeiten, auf die sich die Hypothesen in der Regel beziehen, können die Wechselwirkungen, aufgrund von Nicht-Additivität und Nicht-Linearität im Logit-Modell, ihre statistische Signifikanz verlieren und auch ihr Vorzeichen wechseln (Ai & Norton 2003)
- Interaktionseffekte auf Wahrscheinlichkeitsebene lassen sich mit dem STATA-Befehl „inteff“ vertiefend analysieren (in SPSS gibt es derzeit kein Äquivalent)

Anwendungsempfehlungen

- Treten exorbitant hohe b-Koeffizienten auf, hat dies meist zwei Ursachen:
  - Kollinearität zwischen zwei Prädiktoren (Abhilfe: Prädiktoren zu Skala zusammenfassen, Prädiktor aus Modell entfernen)
  - Kombinationen von abhängiger Variabler und mindestens einer Ausprägung einer unabhängigen Variablen sind zu schwach (d.h. nur mit einem Fall oder gar nicht) besetzt (Abhilfe: Kategorien zusammenfassen, Prädiktor entfernen)

Anwendungsempfehlungen

- Bei schief verteilten AV (d.h. stark überwiegenden Einsen oder Nullen) werden geringere Pseudo-R^2-Werte erzielt als bei ausgeglichener Verteilung
- Einige Pseudo-R^2-Varianten (z.B. nach Cox & Snell, Nagelkerke) haben nicht die PRE-Eigenschaft und sind daher nur bedingt sinnvoll
- Der Mehrwert der Odds-Ratios ist fraglich. „Chancenverhältnis“ ist ein wenig intuitives Konzept
- Da Odds-Ratios kein obere Grenze haben, stellen sie zudem keine standardisierten Effekte dar

Ausgewählte Literatur

- Kopp, J. & Lois, D. (2014): Sozialwissenschaftliche Datenanalyse. Eine Einführung. 2. Auflage. Wiesbaden: Springer VS (Kapitel 7).

- Backhaus et al. (2011): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Berlin: Springer (Kapitel 5).

- Wolf, C. & Best, H. (Hg.) (2010): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: Springer VS (Kapitel 31).

Zitierte Literatur:

- Mood, C. (2010): Logistic Regression: Why we cannot do what we think we can do, and what we can do about it. European Sociological Review 26: 67-82.

- Ai, C.R. & Norton, E.C. (2003): Interaction terms in logit and probit models. Economics Letters 80: 123-129.