

Kapitel 1

Notiz zur logistischen Regression

1.1 Grundlagen

Bei dichotomen abhängigen Variablen ergeben sich bei einer normalen linearen Regression Probleme. Während man die Ausprägungen einer dichotomen Variable mit 0 und 1 noch als Unter- bzw. Obergrenze für die Wahrscheinlichkeit des Auftretens eines Merkmals interpretieren kann, kann die Regressionsgerade bei solchen Daten durchaus Werte < 0 und > 1 annehmen. Diese sind dann nicht mehr sinnvoll als Wahrscheinlichkeiten interpretierbar.

Statistisch kommt das Problem hinzu, dass die Annahme, die Fehler seien normalverteilt für eine dichotome abhängige Variable nicht sinnvoll ist. Ebenso wenig ist Homoskedastizität, also eine konstante Varianz, gegeben [2, vgl. S. 9].

Bei einer dichotomen Abhängigen (z.B. mit den Ausprägungen 0 und 1) ist ihr Mittelwert eine Funktion darüber, mit welcher Wahrscheinlichkeit ein Fall zur höheren Kategorie (hier: 1) gehört, er gibt den Anteil der Fälle an, für die das zutrifft. Eine Regression gibt dann die geschätzte Wahrscheinlichkeit an, bei Kenntnis der Unabhängigen zur höheren Kategorie zu gehören [1, vgl. S. 6].

Deshalb geht man davon aus, dass in solchen Fällen die Verteilung der Auftretswahrscheinlichkeiten der Ereignisse der abhängigen Variablen nicht linear ist, sondern eher einer „S“-förmigen Kurve gleicht. Eine Darstellung einer solchen Verteilung zeigt Abbildung 1.1. Sie zeigt, dass für X_i mit einer Wahrscheinlichkeit von $P(Y = 1) \hat{Y}_i$ den Wert 1 annehmen wird.

Die Vorteile der in Abbildung 1.1 gezeigten Verteilung sind anhand der Grafik leicht zu erkennen. Bei einem linearen Verlauf würde eine Änderung in

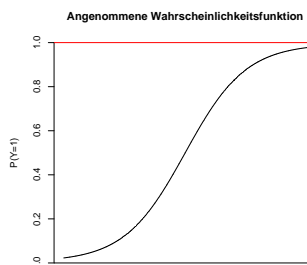


Abbildung 1.1: Angenommene Verteilung der Auftretswahrscheinlichkeiten bei einer dichotomen Variablen.

der erklärenden Variable (X) eine immer gleich große Änderung der erklärten Variable hervorrufen. Das ist aber bei einem Wertebereich von $Y \in \{0, 1\}$ eine nicht sinnvolle Annahme. Wahrscheinlicher ist die Annahme, dass nahe der tatsächlich annehmbaren Werte die theoretische Verteilung sich asymptotisch verhält, während es zwischen den beiden einen Bereich geben muss, in dem der „Umschwung“ von einem Extrem zum anderen statt findet. Für den Einfluß von Änderungen in der erklärenden Variable auf die Erklärte heißt das, dass im Bereich des asymptotischen Verlaufes große Änderungen in X nötig sind, um kleine Änderungen in Y hervorzurufen, während im Bereich des Umschwunges kleine Änderungen von X ausreichen, um große Änderungen in Y hervorzurufen. Genau das bildet eine „S“ – förmige Kurve ab.

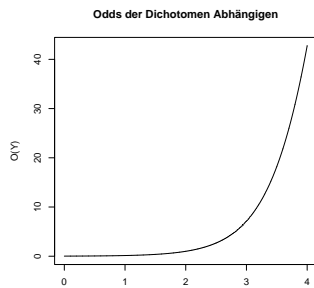


Abbildung 1.2: Odds der angenommenen Verteilung einer dichotomen Variablen.

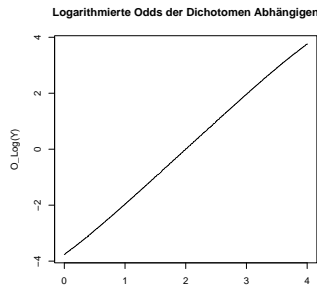


Abbildung 1.3: Logarithmierte Odds der angenommenen Verteilung einer dichotomen Variablen.

Bei einem solchen Verlauf ist nun eine lineare Analyse nicht ohne weiteres möglich. Es ist aber relativ leicht möglich, die Daten so zu modifizieren, dass sie einem linearen Verlauf folgen.

Dabei ist zu beachten, dass die inhaltliche Beziehung zwischen Abhängiger und Unabhängigen nicht linear bleibt, lediglich die formale Beziehung (bezogen auf die Parameter und die Verteilungen) wird in eine lineare Beziehung umgewandelt [1, S. 10]. Die Fragestellung wird damit gewandelt in die Frage danach, mit welcher Wahrscheinlichkeit gilt, unter Kenntnis von X , $\hat{Y}_i = 1$, also: $P(Y = 1) = \alpha + \beta X$ [1, S. 12].

Zwei Modifikationen nutzt man bei der logistischen Regressionsanalyse. Zunächst werden Wahrscheinlichkeitsverhältnisse (englisch: „odds“) für $Y = 1$ gebildet. Odds berechnen sich als Wahrscheinlichkeit dividiert durch die Gegenwahrscheinlichkeit: $O = \frac{P}{1-P}$. Da bei der gezeigten angenommenen Verteilung davon ausgegangen wird, dass jede Ausprägung einer Wahrscheinlichkeit entspricht, ist diese Umrechnung hier gerechtfertigt. Bei realen Daten wird die jeweils angenommene Wahrscheinlichkeit geschätzt. Die Berechnung von Odds hat den Effekt, dass der Wertebereich nach $+\infty$ erweitert wird. Linear ist der Verlauf aber immer noch nicht, wie die Abbildung

1.2 zeigt.

Das erreicht man erst mit der zweiten Umrechnung, dem Logarithmieren der Odds: $O_{Log} = \ln O$. Diese Werte werden auch Logits genannt. Abbildung 1.3 zeigt den Effekt: der Verlauf wird linear und der Wertebereich umfaßt nun auch negative Zahlen. Die Werte der umgeformten Variable haben nun metrisches Skalenniveau. Damit ist eine lineare Regressionsanalyse wieder möglich.

Odds als Wahrscheinlichkeitsverhältnis werden meist als einfache Zahl angegeben.

Ein $O = \frac{0.4}{0.6}$ ist also $O = 0.\bar{6}$. Interpretiert wird es als das Auftreten des Ereignisses in $0.\bar{6}$ Fällen zu 1 (also als $O = \frac{0.6}{1}$), oder, intuitiver, das Auftreten des Ereignisses in 66.6 gegen 100 Fälle, in denen es nicht auftritt. Durch den jeweils gleichen Nenner können Odds miteinander verglichen werden: ein Odd von 9 (zu 1) ist dreimal höher als ein Odd von 3 (zu 1) [2, vgl. S. 12].

Wichtig ist auch das Verhalten der Logits und Wahrscheinlichkeiten. Wie oben bereits beschrieben, wird ein Zusammenhang zwischen Wahrscheinlichkeit des Auftretens eines der beiden Merkmale der Abhängigen und der Unabhängigen postuliert, bei dem nahe der Werte Null und Eins eine Änderung von $\Delta_x = 1$ in X nur geringe Änderungen in Y hervorruft, während in der Mitte der Verteilung bei gleichem Δ_x größere Änderungen in Y auftreten (vgl. wieder Abb. 1.1, S. 1). Logarithmiert man die Odds dieser Verteilung, so gelangt man zu einer linearen Verteilung, bei der eine Änderung von $\Delta_x = 1$ an jeder Stelle die gleiche Änderung in $O_{Log}(Y)$ hervorruft (vgl. Abb. 1.3). Daraus ergibt sich:

„[L]inearity in logits defines a theoretically meaningful nonlinear relationship with the probabilities.“ [2, S. 15]

Die Werte ergeben sich jeweils als:

$$O = \frac{P}{1-P} \quad (1.1)$$

$$O_{Log} = \ln O \quad (1.2)$$

$$\hookrightarrow P = \frac{O}{1+O} \hat{=} \frac{e^{O_{Log}}}{1+e^{O_{Log}}} \quad (1.3)$$

Das Grundprinzip der logistischen Regression. *Die logistische Regression „modelliert den Wahrscheinlichkeitsübergang eine kategorial (hier: binär) ausgeprägten Variablen (unter der Annahme der logistischen Verteilung der Residuen).“ [3, S. 111]*

1.2 Schätzung der Koeffizienten

Zur Schätzung der Koeffizienten wird die **Maximum-Likelihood-Methode** verwendet, wobei für die logistische Regression eine logarithmierte Likelihood (LL) berechnet wird. Es wird dann für eine dichotome Abhängige jeweils die Auftrittswahrscheinlichkeit gewählt, die die logarithmierte Likelihood maximiert. Dies ist dann ein Maß darüber, mit welchen Parametern es am wahrscheinlichsten ist, die gegebene Verteilung der Daten so wie sie ist zu beobachten. Es wird also getestet, mit welchen Parametern die gegebenen Daten am besten reproduziert werden. Als Kriterium wird dabei die LL gewählt, die die geringste Devianz hat.

„Die [...] Devianz ergibt sich aus der Multiplikation des LL – Wertes des geschätzten Modells mit -2 . Je kleiner ihr Wert ausfällt, desto besser ist das Gesamtmodell zu beurteilen. [Die Devianz ist nun] asymptotisch χ^2 – verteilt mit $I - k - 1$ Freiheitsgraden ($I =$ Zahl der Beobachtungen, $k =$ Zahl der unabhängigen Variablen). Damit ermöglicht die Devianz einen Test der Nullhypothese, daß das Modell eine perfekte Anpassung liefert. Eine hohe Irrtumswahrscheinlichkeit

für die Ablehnung der Nullhypothese (hohes Signifikanzniveaus) spricht insoweit für eine gute Anpassung.“ [3, S. 114f.]

Der Abstand des LL – Wertes von 0 ist dabei sowohl von der „Trennfähigkeit der Variablen“ also der Güte des Modells, als auch „von der Verteilung der Beobachtungen auf die Kategorien der abhängigen Variablen“ [3, S. 115] abhängig. Ist die Abhängige sehr schief verteilt, wird die LL immer kleiner sein als bei gleich stark besetzten Gruppen [3, vgl. S. 115].

Deshalb hat man weitere Maße eingeführt, wie zum Beispiel den **Likelihood Ratio-Test**. Wenn man für ein Modell ohne Unabhängige eine LL berechnet (L_0) und von dieser die LL eines Modelles mit einer oder mehreren Unabhängigen (L_1) subtrahiert, dann ist das Ergebnis multipliziert mit -2 ein χ^2 – verteilter Wert mit so vielen Freiheitsgraden wie Unabhängigen im Modell. Die Gleichung lautet [3, vgl. S. 116]:

$$-2LL = \text{LR-Test} = -2 \ln \frac{L_0}{L_1} = -2(L_0 - L_1) \quad (1.4)$$

Damit kann, wie bei linearen Regressionen, überprüft werden, ob und auf welchem Niveau die Verbesserung der Vorhersage durch die Kenntniss von X zufällig ist oder nicht [2, S. 45ff.]. $-2LL$ ist also eine Größe, mit der geprüft wird, ob die Nullhypothese von $\beta_1 = \beta_2 = \dots = \beta_k = 0$ gilt.

Dabei ist zu beachten, dass L_1 angibt, wie *schlecht* das Modell mit den unabhängigen Variablen die Daten abbildet. Es wird damit also mit der statistischen Signifikanz von L_1 die *nicht erklärte* Varianz ausgedrückt [1, S. 20]. Ein ideales Modell würde also am Ende ein nicht signifikantes L_1 und ein signifikantes $-2LL$ aufweisen.¹

Dieses Vorgehen kann auch zur Prüfung des Einflusses zusätzlicher Unabhängiger genutzt werden. Nimmt man die Differenz der LL der Modelle mit allen interessierenden Variablen und ohne die zusätzlichen Variablen, kann auch sie mit -2 multipliziert in der gleichen Weise wie oben interpretiert werden.

Neben $-2LL$ gibt es noch **MacFadden's R^2** , das auch auf L_0 und L_1 zurückgreift. Es wird berechnet als [3, vgl. S. 116]:

$$\text{McF-}R^2 = 1 - \frac{L_1}{L_0} \quad (1.5)$$

„Liefert der LR – Test eine Antwort auf die Frage nach der Signifikanz des Modelles und damit nach der Übertragbarkeit der Ergebnisse auf die Grundgesamtheit, stellt **McF- R^2** ein Maß dar, mit dem die Trennkraft der unabhängigen Variablen insgesamt mit einem Wert benannt und damit vergleichbar (zwischen verschiedenen Modellen) gemacht werden kann.“ [3, S. 116f.]

1.3 Signifikanztests

Die Tests auf Signifikanz, die bei einer logistischen Regression angewendet werden, haben die gleiche Bedeutung wie immer. Sie geben an, mit welcher Wahrscheinlichkeit

¹[1] weicht von der bislang vorgestellten Benennung ab. Bei ihm sind L_0 und L_1 D_0 und D_M respektive, für „deviation“, „with [...] zero [...] of the independent variables in the equation“, bzw. „deviation χ^2 for the full model“ [1, S. 20] und G_M entspricht $-2LL$. Zur besseren Erinnerung fügt er noch hinzu: „(For mnemonic fans, a statistically significant G_M is Good; a statistically significant D_M is Dreadful.)“ [1, S. 21]

man davon ausgehen kann, dass die untersuchten Zusammenhänge tatsächlich so existieren und nicht nur zufällig zustande gekommen sind. Das Programmpaket SPSS gibt beispielsweise standardmäßig die (zweiseitige) Wald-Statistik aus. Sie ist der quadrierte Regressionskoeffizient dividiert durch seinen Standardfehler [2, vgl. S. 30]. Die Signifikanztests können aber nur darüber Auskunft geben, ob ein Zusammenhang besteht. Über seine Stärke geben sie keine Auskunft.

Genau dazu dient das „Bayesian Information Criterion“ oder *BIC*:

$$BIC = z^2 - \ln n \quad (1.6)$$

Dabei sind z der Regressionskoeffizient dividiert durch seinen Standardfehler und n ist die Stichprobengröße [2, nach S. 31]. Als Daumenregel werden *BIC* – Werte von

$0 \leq BIC < 2$ als schwache,

$2 \leq BIC < 6$ als positive,

$6 \leq BIC < 10$ als starke und

$BIC \geq 10$ als sehr starke

Zusammenhänge interpretiert.

Es gibt verschiedene Möglichkeiten, die Koeffizienten zu standardisieren, aber dabei tritt das Problem auf, das eine Standardisierung von dichotomen Variablen nur die beiden möglichen Werte der Variable auf andere Niveaus hebt und damit der Unterschied der standardisierten Werte zu den unstandardisierten so gering ist, dass ein konkreter Nutzen fehlt [2, vgl. S. 32]. Zudem ergeben die verschiedenen Möglichkeiten zur Berechnung von standardisierten Koeffizienten auch unterschiedliche Werte.

Bei [1] erfährt man über standardisierte Koeffizienten, dass sie in der logistischen Regression analog zur linearen Regression gebildet werden. Allerdings beziehen sie sich dann auf Änderungen in Logits der Wahrscheinlichkeiten, dass Y einen bestimmten Wert annimmt. Die beobachtbaren Werte des $O_{Log}(Y)$ von $-\infty \rightarrow +\infty$ erlauben jedoch nicht die Berechnung von Mittelwert oder Standardabweichung [1, S. 46]. Mit den vorhergesagten Werten und der erklärten Varianz kann aber die Standardabweichung berechnet werden und damit wird eine Schätzung der standardisierten Regressionskoeffizienten möglich. Menard schlägt dafür folgende Formel vor: $b_{YX}^* = (b_{YX})(s_X)(R)/s_{logit(\hat{Y})}$. Dabei sind b_{YX}^* der standardisierte Koeffizient, b_{YX} der nicht standardisierte Koeffizient, s_X die Standardabweichung der Unabhängigen X und $s_{logit(\hat{Y})}$ die Standardabweichung des Logits von \hat{Y} [1, vgl. S. 46]. Menard selbst nutzt diese standardisierten Koeffizienten, „to compare the magnitude of the effects of the predictors by converting them to a common scale of measurement.“ [1, S.49]

1.4 Modellgüte

Erlaubt die Kenntnis der Unabhängigen bessere Schätzer der Abhängigen? Wenn ja, wie viel besser? Diese Fragen stellen sich sowohl bei linearen als auch bei logistischen Regressionsmodellen. Bei der logistischen Regression kommt, je nach Forschungsinteresse, noch eine weitere Frage hinzu: Wie gut werden exakte Werte der Abhängigen (also z.B. das Eintreten des Falles $Y = 1$) vorhergesagt?

Analog zu den R^2 in der linearen Regression kann man in der logistischen Regression ein Maß zur Modellgüte bestimmen. Beim linearen Modell wird ja die Fehlerquadratsumme eines Modells ohne die Unabhängigen mit der Fehlerquadratsumme eines Modells mit Unabhängigen verglichen. Diesen Fehlerquadratsummen entsprechen bei der logistischen Regression die oben benannte $L0$ multipliziert mit -2 und $L1$ multipliziert mit -2 . Damit ergibt sich ein Pseudo $-R^2$ von

$$R^2 = \frac{(-2 \ln L0) - (-2 \ln L1)}{(-2 \ln L0)} \quad (1.7)$$

Diese Größe gibt zwar die Verbesserung der Vorhersagekraft durch Kenntnis der Unabhängigen im Vergleich zu ihrer Nichtbeachtung an, aber sie gibt nicht die erklärte Varianz an. Ihr Wertebereich liegt zwischen 0 (keine Verbesserung) und unter 1 [2, S. 49]. Die LL hängt auch von der Anzahl der Fälle ab, was ein anderes Pseudo $-R^2$ berücksichtigt. $-2 * LL$ entspricht einem χ^2 , woraus sich ergibt:

$$R^2 = \frac{\chi^2}{\chi^2 + N} \quad (1.8)$$

dessen Interpretation ganz ähnlich der des Pseudo $-R^2$ aus Gleichung 1.7 ist.

Es gibt aber auch **Nagelkerkes- R^2** , der standardmäßig von SPSS ausgegeben wird. Es gibt an, „wieviel Varianz der abhängigen Variablen durch die betrachteten unabhängigen Variablen erklärt wird“ [3, S. 133] und wird damit analog zum R^2 in der linearen Regression interpretiert. Ebenfalls standardmäßig gibt SPSS die Werte des **Wald - Testes** an.

„Das Funktionsprinzip des Wald - Tests ist eng angelehnt an die Überprüfung der Signifikanz einzelner Koeffizienten innerhalb der linearen Regressionsanalyse (t-Test). Auch hier wird die Null-Hypothese getestet, daß ein bestimmtes β_j Null ist, d.h. die zugehörige unabhängige Variable keinen Einfluß auf die Trennung der Gruppen hat. Die Formel der Wald-Teststatistik W lautet:

$$W = \left(\frac{\beta_j}{s_{\beta_j}} \right)^2 \quad (1.9)$$

mit:

s_{β_j} = Standardfehler von β_j ($j = 0, 1, 2, \dots, k$)

W ist wiederum asymptotisch χ^2 - verteilt[...] bei einem Freiheitsgrad von Eins.“ [3, S. 119f.]

Neben diesen Maßen gibt es noch die Möglichkeit, Gruppenzugehörigkeiten bei dichotomen Abhängigen zu untersuchen. Denn das

„Ziel der logistischen Regression ist es, die Koeffizienten β_j derart zu schätzen, daß eine optimale Trennung der zwei Ausprägungen der abhängigen Variablen erreicht wird. [...] Für den Fall standardisierter unabhängiger Variablen gilt dabei, daß die Höhe des β - Wertes anzeigt, wie gut es um deren Trennkraft bestellt ist (je größer desto besser!). Hohe β - Werte bedeuten eine große Steigung der logistischen Funktion und begrenzen damit den Wahrscheinlichkeitsübergang auf einen engen Wertebereich.“ [3, S. 112f.]

Die durch das Modell vorhergesagte Gruppenzugehörigkeit kann mit der tatsächlichen in einer 2×2 – Tabelle untersucht werden. Denn bei einem perfekten Modell ist zu erwarten, dass alle Fälle entlang der Diagonalen liegen.

Um diese Tabellen zu untersuchen und eine Aussage über die Vorhersagekraft eines Modelles machen zu können, schlägt Menard ein Maß der proportionalen Fehlerveränderung („*proportional change in error*“ [1, S. 24]) vor. Er benutzt nicht den Begriff der Fehlerreduktion, weil es sein kann, dass ein Modell die Daten tatsächlich schlechter vorhersagt, als man sie ohne Kenntnis der Unabhängigen schätzen würde.

Menard stellt drei mögliche Formeln zur Berechnung eines solchen Maßes vor, die sich darin unterscheiden, wie man den Vorhersagefehler definiert. Grundsätzlich berechnet Menard aber alle mit der gleichen Formel [1, S. 24]:

$$\text{Predictive efficiency} = \frac{\text{Errors without model} - \text{Errors with model}}{\text{Errors without model}} \quad (1.10)$$

Die „Errors with model“ sind dabei die falschen Schätzer der Abhängigen mit Kenntnis der Unabhängigen. Die „Errors without model“ werden auf unterschiedliche Weise berechnet, von denen hier zwei vorgestellt werden.

Bei einem einfachen Vorhersagemodell („prediction model“ [1, S. 25]) wird der Modus der Abhängigen als der Vorhersagewert ohne Kenntnis der Unabhängigen gewählt [1, S. 28].

„This method of defining errors without the model is the same as the one used in defining Goodman and Kruskal’s λ for contingency tables with nominal variables [. . .] Because of the similarity to Goodman and Kruskal’s λ , the index is here referred to as λ_p . [It] is a proportional reduction in error (PRE) measure like R^2 when it is positive but, when the model does worse than predicting the mode, λ_p may be negative, indicating the proportional *increase* in error. [. . .] In general, the full range of possible values for λ_p in all tables with N cases is from $1 - N$ to 1.“ [1, S. 28, Hervorhebung i.O.]

Für ein „klassifizierendes Modell“ („classification model“ [1, S. 28]) wird untersucht, inwieweit eine Vorhersage der Fälle für die Zugehörigkeit distinkter Kategorien korrekt ist. Formalisiert ist das:

$$\text{Errors without model} = \sum_{i=1}^N f_i \left(\frac{N - f_i}{N} \right) \quad (1.11)$$

Dabei ist N die Stichprobengröße und f_i die Zahl der in Kategorie i beobachteten Fälle.

„This is the same formula for error without model as is used for Goodman and Kruskal’s τ . [. . .] Parallel to λ_p , [this] index will be referred to as tau – p (τ_p), or tau for prediction tables. [. . .]

τ_p requires that, even in the estimation of error without the model, cases must be separated into distinct groups or categories, and not all placed in the same category. [. . .] For tables with equal marginal distribution, τ_p varies between -1 and $+1$ [but] with extremely skewed and inconsistent marginal distributions, the minimum value of τ_p is equal to $1 - [N^2/(2N - 2)]$.“ [1, S. 29]

Diese Maße sind allerdings nicht durchgängig gebräuchlich und es gibt viele weitere Vorschläge, solche Maße zu bestimmen. Bei einem logistischen Regressionsmodell ist die Genauigkeit der Vorhersage der Gruppenzugehörigkeiten unter Umständen sehr interessant. Dennoch:

„Especially for theory testing, goodness of fit is simply more important than accuracy of classification. [...] Often the two approaches, goodness of fit and accuracy of predictions, will produce consistent results. It is entirely possible, however, to have a model that fits well but does a poor job of predicting category membership.“ [1, S. 32]

In Fällen, in denen tatsächlich die Modellgüte hoch aber die Vorhersage schwach ist, ist eine wichtige Frage sicherlich die nach der Definition der „Errors without model“. Sollten hier nicht sinnvolle Annahmen gemacht worden sein, ist evtl. durch eine Veränderung dieses Fehlertermes die Konsistenz zwischen Modellgüte und Vorhersage wieder herstellbar. Gelingt das nicht, erscheint mir das Modell eher nicht angemessen.

Eine weitere Möglichkeit, die tatsächlichen und die geschätzten Werte der Abhängigen zu nutzen, stellt die Suche nach Ausreißern dar. [3, S. 118] stellt das „Pearson Residuum r “ vor, das für jeden Fall berechnet wird als $r = (y_i - p(y_i = 1)) / \sqrt{p(y_i = 1) * (1 - p(y_i = 1))}$. Ist der Wert von $r > 0.5$ kann man von einem Ausreißer in dem Sinne sprechen, dass die „Diskrepanz zwischen der tatsächlichen Gruppenzugehörigkeit und der ermittelten Gruppenzugehörigkeit groß“ [3, S. 118] ist.

1.5 Interpretation der Ergebnisse

Durch die formale Linearisierung des nicht-linearen Zusammenhanges ergeben sich allerdings bei der Interpretation Schwierigkeiten. Man kann nun nicht mehr analog zur linearen Regression von einem Einfluß von X auf Y sprechen, da ja eine Veränderung in X von $\Delta_x = 1$ bei verschiedenen X – Werten unterschiedliche Änderungen in Y hervorrufen wird.

Bei den Logits der angenommenen Wahrscheinlichkeiten der Abhängigen kann man zwar eine lineare Regression vornehmen. Aber wie interpretiert man die Koeffizienten? Man weiß, dass eine Änderung in X eine bestimmte Änderung in $O_{Log}(P(Y_i))$ bedeutet. Aber eine einfach nachvollziehbare Rückbeziehung auf Y ist damit nicht möglich.

Eine weitere Interpretationsmöglichkeit stellen die Odds dar. Die Koeffizienten einer logistischen Regression stehen mit den Logits in folgender Beziehung, über die sie durch Umformen in Bezug zu den Odds gebracht werden können:

$$\begin{aligned} \ln O &= \alpha + \beta X & (1.12) \\ \Leftrightarrow \ln\left(\frac{P}{1-P}\right) &= \alpha + \beta X \\ \Leftrightarrow e^{\ln\left(\frac{P}{1-P}\right)} &= e^{\alpha + \beta X} \\ \Leftrightarrow \frac{P}{1-P} &= e^{\alpha} * e^{\beta X} & (1.13) \end{aligned}$$

In Gleichung 1.13 gibt der Term e^{β} das Verhältnis zwischen zwei Odds an (deshalb wird er auch „odds ratio“ genannt). Dabei gilt, dass $e^{\beta} = \frac{O(x_i)}{O(x_i+1)}$. Eine

Veränderung von einer Einheit in der erklärenden Variable sorgt also für eine e^β entsprechende Veränderung in den zugehörigen Odds [1, vgl. S. 49], auch [3, S. 121]. In Prozenten ausgedrückt ist die Veränderung des Wahrscheinlichkeitsverhältnisses der Abhängigen bei einer Änderung von $\Delta_x = 1$ der Unabhängigen: $\% \Delta = (e^\beta - 1) * 100$ [2, vgl. S. 23]. Weiterhin muß beachtet werden, dass der exponierte Koeffizient durch das multiplikative Modell seinen neutralen Punkt bei $e^\beta = 1$ hat und nie unter 0 fallen kann. Werte $0 < e^\beta < 1$ bezeichnen also negative Zusammenhänge, Werte $e^\beta > 1$ positive.

Die Umwandlung der Koeffizienten in Bezug auf Odds macht bereits deutlich, dass der Zusammenhang zwischen ihnen nicht nur nicht linear, sondern auch nicht additiv ist. Daher kann bei der Umwandlung der Einflüsse auf die Wahrscheinlichkeiten nicht mehr nur ein Koeffizient angegeben werden, der die Verteilung beschreibt, sondern man muss für die jeweils interessierenden Werte X_i diese einzeln berechnen. Eine Möglichkeit ist, die Steigung der logistischen Kurve (wie in Abbildung 1.1) an einem bestimmten Punkt zu berechnen.² Sie kann interpretiert werden als die Veränderung von Y durch eine Änderung von X an dieser Stelle um eine Einheit [2, vgl. S. 25].³ Dieser Effekt wird auch „Randeffekt“, auch „partial derivative [...] marginal or instantaneous effect“ [2, vgl. S. 25] genannt. Berechnet man diese Größe für $P = 0.5$, also an der Stelle, an der theoretisch der Umschwung vom Auftreten des Ereignisses zum Nichtauftreten statt findet, bzw. an der Stelle, an der X seinen Mittelwert hat, dann kann man anhand der Steigung Aussagen darüber treffen, ob dieser Umschwung eher „plötzlich“ oder „gemächlich“ von statten geht. Damit wird sozusagen die „Trennschärfe“ der Unabhängigen untersucht: ist die Steigung hoch, ist der Einfluß auf das Eintreten des Ereignisses bei der Abhängigen deutlicher oder eindeutiger als bei kleineren Steigungen, wobei dann möglicherweise der Einfluss der Abhängigen nicht so eindeutig ist. Das gilt für metrische Unabhängige.

Bei dichotomen Unabhängigen machen diese Berechnungen wenig Sinn. Statt dessen kann man bei ihnen die Differenz der vorhergesagten Wahrscheinlichkeiten des Eintretens des Ereignisses für $X = 0$ und $X = 1$ berechnen. Dabei kann der Mittelwert von Y als Startpunkt gewählt werden, der die Gruppe repräsentiert, bei denen das Ereignis nicht eintritt. Die Differenz wird dann zu der vorhergesagten Wahrscheinlichkeit für $X = 1$ gebildet [2, vgl. S. 27].

Bei allen vorgestellten Berechnungen für Kennwerte zur Interpretation darf nicht vergessen werden, dass durch den nichtlinearen und nichtadditiven Zusammenhang verschiedene Werte von X und Y auch je unterschiedliche Kennwerte zur Folge haben. Das erschwert die Interpretation zusätzlich. Durch die vorgestellten Möglichkeiten der Umrechnung bieten sich andererseits viele Möglichkeiten, Zusammenhänge zu beschreiben oder auch nur für interessante „typische“ Fälle Wahrscheinlichkeiten zu berechnen.

²Das entspricht der Steigung einer Tangente durch den interessierenden Punkt (x;y).

³Wobei es sich hier streng genommen nur um einen Punkt handelt. Würde man tatsächlich eine Einheit in X weitergehen, bekäme man eine andere Steigung für die Tangente an dieser zweiten Stelle. Die Steigung bezieht sich also nicht auf eine gut interpretierbare Veränderung von X, denn streng genommen geht es hier um einen Intervall, der unendlich klein ist und deshalb als Punkt aufgefasst wird.

Literaturverzeichnis

- [1] MENARD, SCOTT: *Applied Logistic Regression Analysis*. Nummer 106 in *Sage University Papers series on Quantitative Applications in the Social Sciences*. Sage Publications, Thousand Oaks, CA, 1. Auflage, 1995.
- [2] PAMPEL, FRED C.: *Logistic Regression. A Primer*. Nummer 132 in *Sage University Papers series on Quantitative Applications in the Social Sciences*. Sage Publications, Thousand Oaks, CA, 1. Auflage, 2000.
- [3] RESE, MARIO: *Logistische Regression*. In: BACKHAUS, KLAUS, BERND ERICHSON, WULFF PLINKE und ROLF WEIBER (Herausgeber): *Multivariate Analysemethoden*, Seiten 104–144. Springer, Berlin et al., 9. Auflage, 2000.