# 21 Logistische Regressionsanalyse

**Was Sie in diesem Kapitel lernen**

- Wie kann man Unterschiede in einer kategorialen abhängigen Variablen auf Unterschiede in einer oder mehreren unabhängigen Variablen zurückführen?
- Wie lässt sich die Idee der multiplen Regressionsanalyse auf kategoriale abhängige Variablen übertragen?
- Woran erkennt man die Güte einer Klassifikation?
- Inwieweit hängt die Lebenszufriedenheit von positiven und negativen täglichen Ereignissen ab?

Im letzten Kapitel haben wir das log-lineare Modell kennengelernt, das wir heranziehen können, um Zusammenhänge zwischen kategorialen Variablen zu analysieren. Im log-linearen Modell wird keine Unterscheidung zwischen abhängigen und unabhängigen Variablen getroffen; es ist ein Modell für ungerichtete Zusammenhänge. Wir haben darüber hinaus gesehen, dass sich auf der Grundlage eines log-linearen Modells auch ein Logit-Modell definieren lässt. Im Logit-Modell werden Unterschiede in einer abhängigen kategorialen Variablen auf Unterschiede in einer oder mehreren unabhängigen kategorialen Variablen zurückgeführt. Wie geht man aber vor, wenn die abhängige Variable kategorialer Natur ist und die unabhängigen Variablen alle oder teilweise metrischer Natur sind? Für solche Fragestellungen ist die logistische Regressionsanalyse geeignet. Sie stellt das Pendant zur multiplen Regressionsanalyse dar, die für metrische abhängige Variablen entwickelt wurde (Kap. 18). Wir werden die logistische Regressionsanalyse ausführlich für dichotome (zweiwertige) Variablen darstellen und anschließend zeigen, wie dieser Ansatz auf den Fall einer mehrkategorialen abhängigen Variablen mit geordneten oder ungeordneten Kategorien übertragen werden kann.

## 21.1 Grundidee der logistischen Regressionsanalyse für dichotome abhängige Variablen

Dichotome Variablen können nur zwei Werte annehmen. Beispiele für solche Variablen sind die Lösung einer Aufgabe (Kategorien: gelöst, nicht gelöst), die Symptombelastung (Kategorien: Symptom liegt vor, Symptom liegt nicht vor) oder die Antwort auf eine Behauptung, mit der eine Einstellung gemessen wird (Kategorien: trifft zu, trifft nicht zu). Wie in der multiplen Regressionsanalyse gehen wir davon aus, dass interindividuelle Unterschiede in der abhängigen Variablen nicht perfekt aufgrund individueller Unterschiede in der unabhängigen Variablen vorhergesagt bzw. erklärt werden können, sondern dass wir die Ausprägung der abhängigen Variablen nur mit einer bestimmten Unsicherheit vorhersagen können. Diese Unsicherheit wird in der logistischen Regressionsanalyse dadurch berücksichtigt, dass die Wahrscheinlichkeit, ein bestimmtes Verhalten zu zeigen (z. B. eine bestimmte Kategorie zu wählen), betrachtet wird. In der logistischen Regressionsanalyse ist die Wahrscheinlichkeit, mit der eine bestimmte Kategorie gewählt wird, eine Funktion der unabhängigen Variablen. Diese Wahrscheinlichkeit kann mit der Zunahme der Werte der unabhängigen Variablen ansteigen (positiver Zusammenhang), abfallen (negativer Zusammenhang) oder gleich bleiben (Unabhängigkeit).

Wir werden die logistische Regression zunächst für den einfachsten Fall einer unabhängigen Variablen (einfache logistische Regressionsanalyse) und verschiedene Formen der Modelldarstellung behandeln. Wir beziehen uns dabei auf das Populationsmodell und zeigen anschließend, wie die Parameter geschätzt und Hypothesen überprüft werden können.

### 21.1.1 Einfache logistische Regressionsanalyse

Die logistische Regressionsanalyse lässt sich in drei Formen darstellen: In der ersten Darstellungsform betrachtet man die Wahrscheinlichkeit einer Kategorie der abhängigen Variablen als Funktion der unabhängigen Variablen. In der zweiten Darstellungsform bildet man den Wettquotienten (s. Kap. 15.3.3) und modelliert ihn als Funktion der unabhängigen Variablen. In der dritten Darstellungsform wird der Logit (s. Abschn. 20.6) in eine Linearkombination der unabhängigen Variablen zerlegt.

#### Darstellung in Form bedingter Wahrscheinlichkeiten

Im Falle einer dichotomen abhängigen Variablen $Y$ können den Kategorien beliebige Werte zugewiesen werden, sofern sie sich zwischen den beiden Kategorien unterscheiden. Wir wählen daher die einfachste Zuordnung, indem wir einer Kategorie den Wert 0, der anderen Kategorie den Wert 1 zuordnen. In der logistischen Regressionsanalyse mit einer einzigen unabhängigen Variablen $X$ werden die bedingten Wahrscheinlichkeiten $P(Y = 1 \mid X = x)$ und $P(Y = 0 \mid X = x)$ betrachtet. $P(Y = 1 \mid X = x)$ ist also die Wahrscheinlichkeit, mit der eine Person auf der Variablen $Y$ den Wert $y = 1$ erhält, wenn sie auf der Variablen $X$ den Wert $x$ aufweist. $P(Y = 0 \mid X = x)$ ist die entsprechende bedingte Wahrscheinlichkeit für die Kategorie mit dem Wert $y = 0$. Wenn die beiden Kategorien einander ausschließen, was bei der logistischen Regression vorausgesetzt wird, dann ist $P(Y = 0 \mid X = x)$ die Gegenwahrscheinlichkeit zu $P(Y = 1 \mid X = x)$ und es gilt:

$$P(Y = 1 \mid X = x) = 1 - P(Y = 0 \mid X = x)$$

Es reicht daher, nur eine der beiden Wahrscheinlichkeiten zu modellieren. Üblicherweise weist man der Kategorie, die von besonderem Interesse ist (z. B. Lösung einer Aufgabe, Zustimmung zu einer Aussage etc.), den Wert 1 zu und beschränkt sich auf die bedingte Wahrscheinlichkeitsfunktion $P(Y = 1 \mid X)$. Die bedingte Wahrscheinlichkeitsfunktion $P(Y = 1 \mid X)$ weist jedem Wert $x$ von $X$ den Wert $P(Y = 1 \mid X = x)$ der bedingten Wahrscheinlichkeit zu.

In der logistischen Regression wird die Abhängigkeit der bedingten Wahrscheinlichkeitsfunktion $P(Y = 1 \mid X)$ von der unabhängigen Variablen $X$ durch eine nicht-lineare Funktion beschrieben. Der Grund hierfür ist, dass eine lineare Funktion mit einem gravierenden Problem behaftet wäre: Handelt es sich bei der unabhängigen Variablen $X$ um eine metrische Variable, die in ihrem Wertebereich nicht beschränkt ist, so hätte eine lineare Funktion zur Folge, dass die bedingte Wahrscheinlichkeitsfunktion unterhalb eines bestimmten $x$-Wertes und oberhalb eines bestimmten $x$-Wertes den zulässigen Wertebereich der Wahrscheinlichkeit von 0 bis 1 verlassen würde (s. Abb. 21.1). In Abbildung 21.1 hätte z. B. eine Person mit dem Wert x = 5 auf der unabhängigen Variablen eine bedingte Wahrscheinlichkeit, die größer als 1 wäre. Dies ist aber theoretisch nicht möglich, da Wahrscheinlichkeiten nur Werte zwischen 0 und 1 annehmen können (s. Abschn. 7.1.5). Um dieses Problem zu beheben, wird die Funktion nun so definiert, dass sie gegen 0 strebt, wenn die Variable $X$ gegen $-\infty$ strebt, und gegen den Wert 1, wenn die Variable $X$ gegen $+\infty$ strebt. Hierzu greift man auf die Exponentialfunktion zurück, da diese vorteilhafte mathematische Eigenschaften hat. Wie bei der einfachen linearen Regression (s. Kap. 16) wird die Funktion durch zwei Parameter bestimmt: (1) einen Parameter ($\beta_0$), der zum Ausdruck bringt, dass generell die Wahrscheinlichkeit, eine bestimmte Kategorie zu wählen,

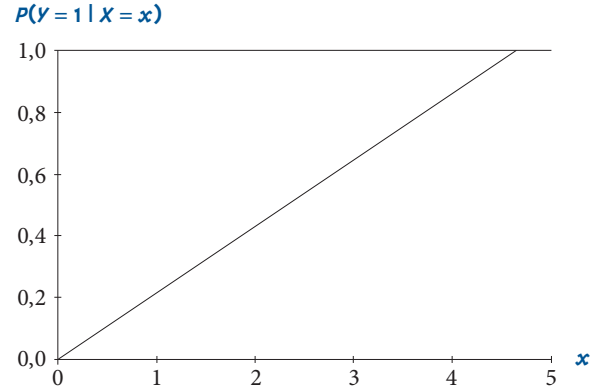


**Abbildung 21.1** Abhängigkeit der bedingten Wahrscheinlichkeit von einer metrischen unabhängigen Variablen im linearen Modell

bei $X = 0$ auf einem höheren oder geringeren Niveau liegen kann; (2) einen Parameter ($\beta_1$), der anzeigt, wie stark die Wahrscheinlichkeit, die Kategorie zu wählen, mit Zunahme der Werte auf der unabhängigen Variablen ansteigt. Die Funktion lautet:

$$P(Y=1\,|\,X) = \frac{e^{\beta_0+\beta_1\cdot X}}{1+e^{\beta_0+\beta_1\cdot X}} \qquad \text{(F 21.1)}$$

Die bedingte Wahrscheinlichkeit ist umso größer, je größer $\beta_0$ und je größer $\beta_1$ ist. In den Abbildungen 21.2 und 21.3 sind die Funktionen für mehrere Parameter $\beta_0$ und $\beta_1$ angegeben. Die Funktion hat in allen Fällen die charakteristische Form einer logistischen Ogive, d. h. einer Funktion, die am einen Ende der Abszisse gegen $-\infty$, am anderen Ende gegen $+\infty$ strebt und einen Wendepunkt aufweist.

**Bedeutung von $\beta_0$.** Der Parameter $\beta_0$ bestimmt die Wahrscheinlichkeit $P(Y=1\,|\,X=0)$ für den Wert 0 der unabhängigen Variablen $X$:

$$P(Y=1\,|\,X=0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$$

Je größer $\beta_0$ ist, umso größer ist diese Wahrscheinlichkeit. In Abbildung 21.2 sind die bedingten Wahrscheinlichkeitsfunktionen für vier verschiedene Parameter $\beta_0$ dargestellt. Der Parameter $\beta_1$ wurde für alle vier Kurven der Einfachheit halber auf den Wert 1 festgelegt. Die Abbildung zeigt, dass die Wahrscheinlichkeit $P(Y=1\,|\,X=0)$ umso größer ist, je größer der Parameter $\beta_0$ ist. Da die Kurven denselben Parameter $\beta_1$ haben, schneiden sich die bedingten Wahrscheinlichkeitsfunktionen nicht und sind anhand einer Verschiebung entlang der $x$-Achse inei-

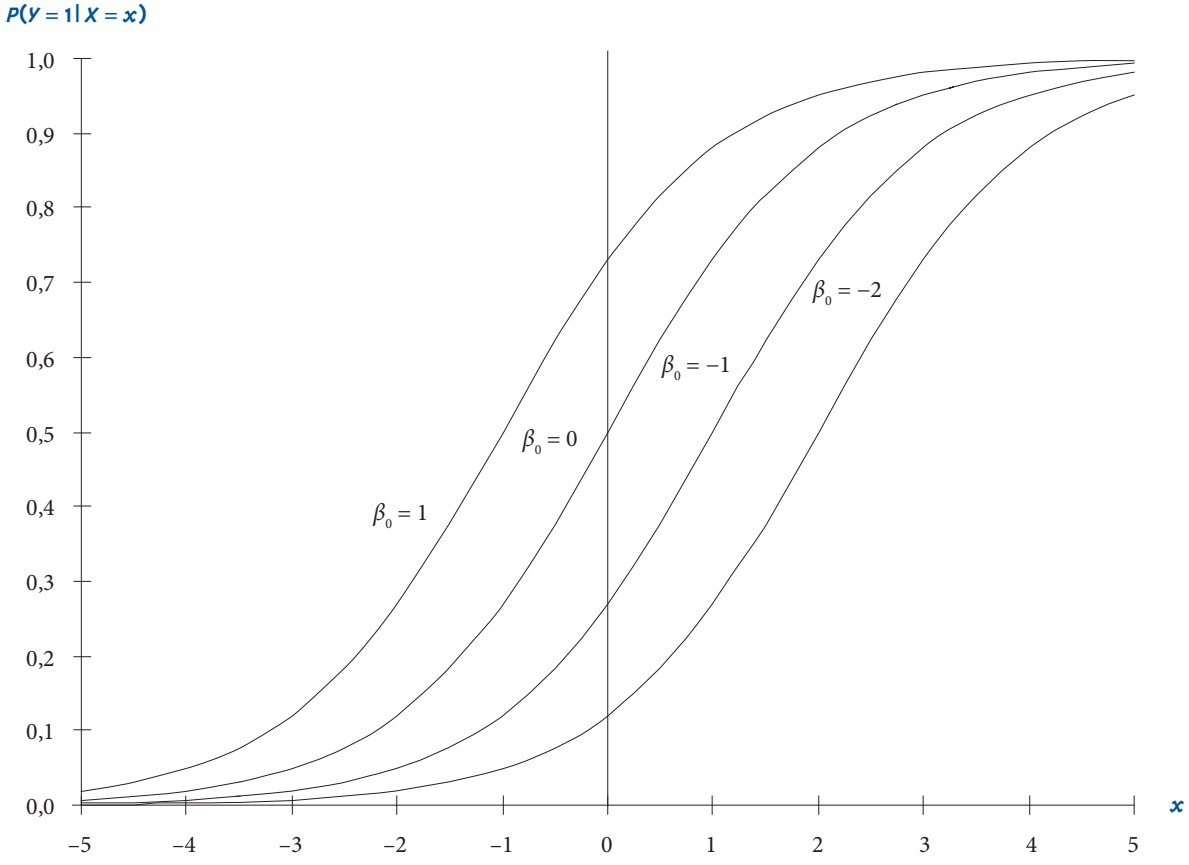


**Abbildung 21.2** Abhängigkeit der bedingten Wahrscheinlichkeit von einer metrischen unabhängigen Variablen im logistischen Regressionsmodell: Auswirkungen verschiedener Regressionskonstanten $\beta_0$

nander überführbar. Je größer der Parameter $\beta_0$ ist, umso weiter links liegt die Kurve. Je nachdem, um welche unabhängige Variable es sich handelt, muss der Wert 0 nicht zwangsläufig ein möglicher oder sinnvoller Wert sein. Will man den Parameter $\beta_0$ als Wert der bedingten Wahrscheinlichkeit an der Stelle $X = 0$ interpretieren, bietet es sich an, die Variable $X$ zu zentrieren. Eine zentrierte Variable erhält man, indem man von jedem $x$-Wert den Mittelwert von $X$ abzieht. Dann hängt der Wahrscheinlichkeitswert, den man für den Mittelwert von $X$ erwartet, nur von $\beta_0$ ab.

**Bedeutung von $\beta_1$.** Der Parameter $\beta_1$ bestimmt die Steigung der Wahrscheinlichkeitsfunktion und damit, wie stark sich Unterschiede auf der unabhängigen Variablen $X$ auf Unterschiede in den bedingten Wahrscheinlichkeiten auswirken. In Abbildung 21.3 sind die bedingten Wahrscheinlichkeitsfunktionen für fünf verschiedene Werte von $\beta_1$ dargestellt. Der Einfachheit halber wurde in allen Fällen $\beta_0 = 0$ gesetzt. Die Abbildung zeigt Folgendes:

- Ist der Parameter $\beta_1$ größer als 0, so ist die bedingte Wahrscheinlichkeit der Kategorie mit dem Wert 1 eine monoton steigende Funktion der Variablen $X$. Die Wahrscheinlichkeit strebt asymptotisch dem Wert 0 zu, je kleiner $X$ wird, ohne dass die Wahrscheinlichkeit den Wert 0 je erreichen kann. Die Wahrscheinlichkeitskurve strebt mit größer werdendem $X$ asymptotisch dem Wahrscheinlichkeitswert 1 zu, ohne ihn je zu erreichen.
- Ist der Parameter $\beta_1$ kleiner als 0, so ist die Wahrscheinlichkeit der Kategorie 1 eine monoton fallende Funktion der Variablen $X$. Die Wahrscheinlichkeit strebt mit kleiner werdendem $X$ asympto-

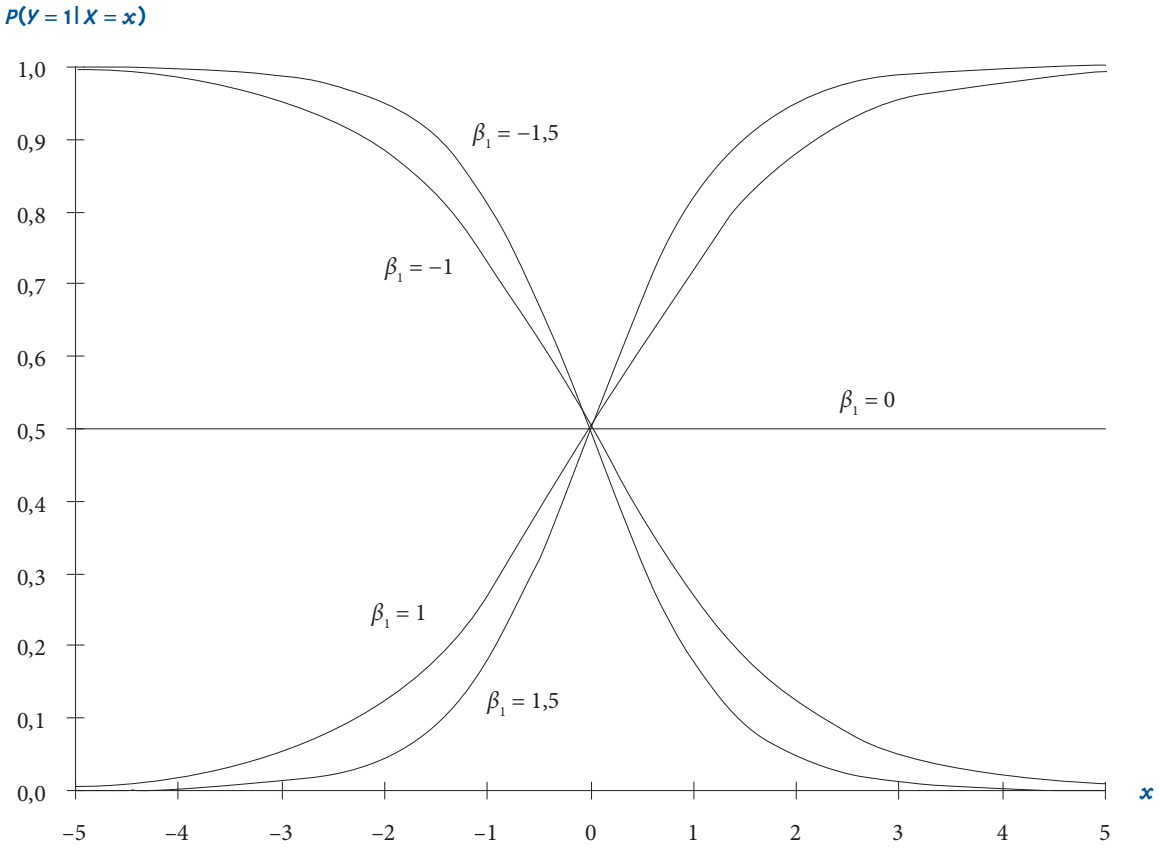


**Abbildung 21.3** Abhängigkeit der bedingten Wahrscheinlichkeit von einer metrischen unabhängigen Variablen im logistischen Regressionsmodell: Auswirkungen verschiedener Regressionsgewichte $\beta_1$

tisch dem Wert 1 und mit größer werdendem $X$ asymptotisch dem Wert 0 zu.

- Ist der Parameter $\beta_1$ gleich 0, dann sind beide Variablen $X$ und $Y$ voneinander unabhängig. Die Wahrscheinlichkeitsfunktion ist dann eine Konstante und verläuft parallel zur $x$-Achse.
- Unterschiede zwischen zwei $x$-Werten wirken sich unterschiedlich stark auf Unterschiede in den bedingten Wahrscheinlichkeiten aus. Liegen die $x$-Werte um den Wendepunkt der Wahrscheinlichkeitsfunktion herum, dann wirkt sich derselbe Unterschied sehr viel deutlicher aus als in den Randbereichen, wo die Wahrscheinlichkeitsfunktion Werte nahe bei 0 oder 1 annimmt.
- Je größer der Betrag von $\beta_1$ ist, umso stärker wirken sich Unterschiede zwischen zwei $x$-Werten,

**Vertiefung**

### Einfache logistische Regression und einfache lineare Regression

Wie bei der einfachen linearen Regressionsanalyse für metrische abhängige Variablen (s. Kap. 16) wird auch bei der logistischen Regression der Zusammenhang zwischen der abhängigen und der unabhängigen Variablen durch zwei Parameter bestimmt, die das Niveau bei $X = 0$ und den Anstieg der Funktion festlegen, die die Abhängigkeit zwischen $X$ und $Y$ beschreibt. Wir haben schon gesehen, dass das lineare Regressionsmodell für dichotome abhängige Variablen nicht geeignet ist, da die Werte der bedingten Wahrscheinlichkeitsfunktion auf den Bereich zwischen 0 und 1 eingeschränkt sind. Darüber hinaus gibt es noch zwei weitere Gründe, warum wir die in Kapitel 16 behandelte einfache lineare Regressionsanalyse nicht auf dichotome Variablen anwenden können.

**Normalverteilung der Residuen.** Um die Parameter auf Signifikanz überprüfen zu können, wird bei der Regressionsanalyse für metrische Variablen angenommen, dass die Residuen bedingt normalverteilt sind. Die Normalverteilung setzt stetige Variablen voraus. Dichotome Variablen können aber nur zwei Werte annehmen, und somit können auch die Residuen nur zwei Werte annehmen. Die Verteilungsvoraussetzungen der Regressionsanalyse für metrische Variablen können bei dichotomen Variablen daher nicht erfüllt sein.

**Homoskedastizität.** Darüber hinaus wird im einfachen linearen Regressionsmodell vorausgesetzt, dass die bedingten Varianzen der Residualvariablen und somit auch der Variablen $Y$ gleich sind und nicht von der Ausprägung der Variablen $X$ abhängen (Annahme der Homoskedastizität). Auch diese Annahme ist bei dichotomen Variablen schon a priori immer falsch. Dies lässt sich leicht veranschaulichen. In Abschnitt 7.2.2 haben wir darauf hingewiesen, dass die Varianz einer dichotomen Variablen berechnet wird, indem die Wahrscheinlichkeit der einen Kategorie mit der Wahrscheinlichkeit der anderen Kategorie multipliziert wird. Für die bedingte Varianz $Var(Y \mid X)$ der Variablen $Y$ gegeben die Variable $X$ ergibt sich daher für den dichotomen Fall:

$$\begin{aligned} Var(Y \mid X) &= P(Y=1 \mid X) \cdot P(Y=0 \mid X) \\ &= P(Y=1 \mid X) \cdot [1 - P(Y=1 \mid X)] \end{aligned} \quad \text{(F 21.2)}$$

Die bedingte Varianz weist daher eine charakteristische Form auf, die in Abbildung 21.4 dargestellt ist. Sie steigt an, bis sie ihren maximalen Wert von 0,25 erreicht hat, um dann wieder abzufallen. Der maximale Wert wird an dem Wert $x$ von $X$ erreicht, an dem die bedingte Wahrscheinlichkeit $P(Y = 1 \mid X)$ den Wert 0,5 annimmt. Je weiter ein Wert der unabhängigen Variablen von dieser Stelle maximaler bedingter Varianz entfernt ist, desto kleiner wird die bedingte Varianz. Sie strebt gegen 0, wenn die bedingte Wahrscheinlichkeit gegen 0 oder 1 strebt. Das heißt, die bedingte Varianz ist eine symmetrische Funktion, die nur von der bedingten Wahrscheinlichkeit abhängt. Die Homoskedastizität kann also nicht erfüllt sein. Die bedingte Varianz gibt die Prognoseunsicherheit an. Sie ist dann klein, wenn die bedingten Wahrscheinlichkeiten Werte nahe 0 und 1 annehmen. Sie ist maximal, wenn die bedingte Wahrscheinlichkeit gleich 0,5 ist.

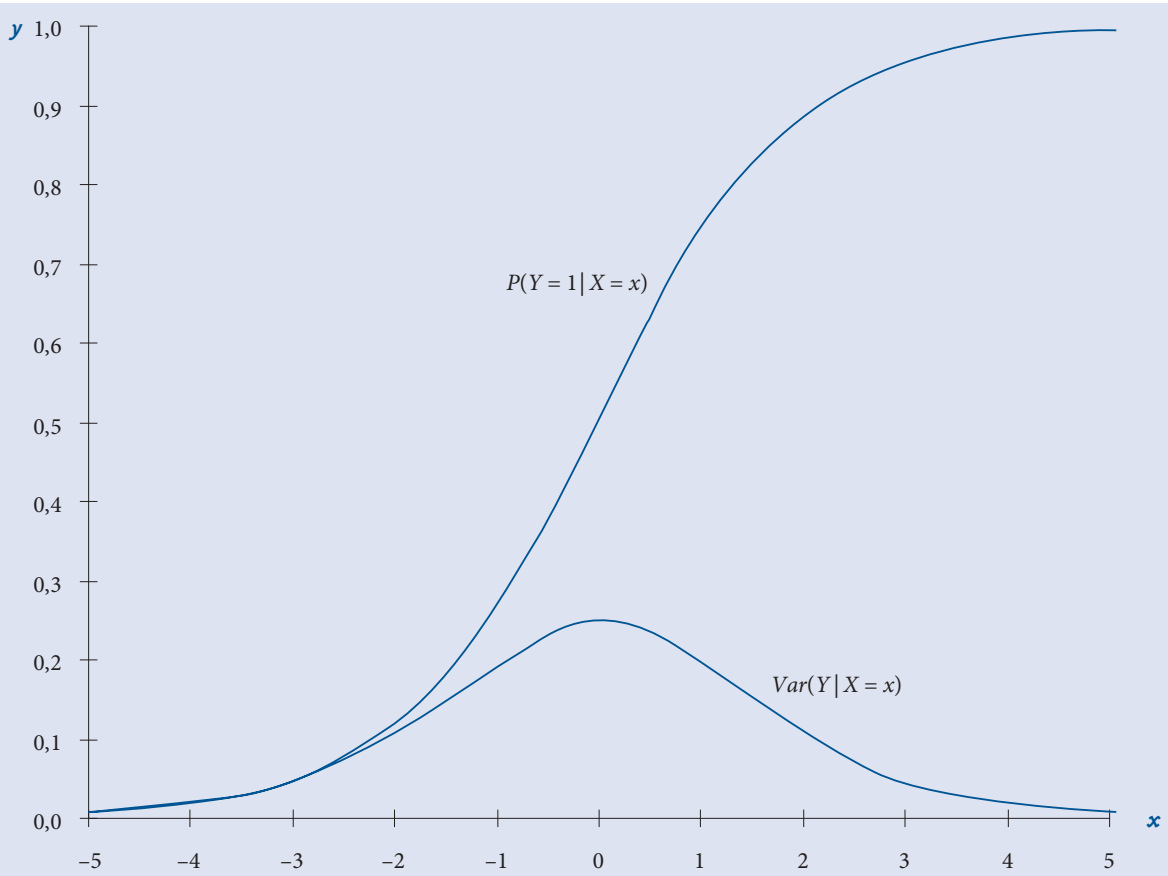


**Abbildung 21.4** Abhängigkeit der bedingten Wahrscheinlichkeit und der bedingten Varianz von einer metrischen unabhängigen Variablen im logistischen Regressionsmodell

die um den Wendepunkt herum liegen, auf Unterschiede in der bedingten Wahrscheinlichkeitsfunktion aus.

- Die bedingte Wahrscheinlichkeit ist gleich 0,5, wenn gilt: $X = -\beta_0/\beta_1$ (s. Übung 3).

### Darstellung in Form der bedingten Wettquotienten

Das Modell der logistischen Regression lässt sich noch in einer zweiten Form darstellen, in der der bedingte Wettquotient in ein Produkt von Parametern zerlegt wird. Den Wettquotienten haben wir bereits in Abschnitt 15.3.3 kennengelernt. Dieser entspricht dem Verhältnis aus der Wahrscheinlichkeit eines Ereignisses und seiner Gegenwahrscheinlichkeit. Im logistischen Regressionsmodell betrachtet man den bedingten Wettquotienten, der das Verhältnis aus der bedingten Wahrscheinlichkeit $P(Y = 1 \mid X = x)$ und der Gegenwahrscheinlichkeit $[1 - P(Y = 1 \mid X = x)]$ darstellt. Die bedingten Wettquotienten $P(Y=1|X=x)/[1-P(Y=1|X=x)]$ sind Werte der bedingten Wettquotientenfunktion $P(Y=1|X)/[1-P(Y=1|X)]$. Aus Gleichung F 21.1 folgt durch Umstellen (s. Übung 4):

$$\frac{P(Y=1|X)}{1-P(Y=1|X)} = e^{\beta_0+\beta_1\cdot X} = e^{\beta_0}\cdot e^{\beta_1\cdot X} = e^{\beta_0}\cdot(e^{\beta_1})^X \qquad \text{(F 21.3)}$$

Der Wettquotient (engl. odds) lässt sich in das Produkt aus $e^{\beta_0}$ und $(e^{\beta_1})^X$ zerlegen.

**Bedeutung von $e^{\beta_0}$.** Der erste Parameter $(e^{\beta_0})$ entspricht dem Wettquotienten an der Stelle $X = 0$. Ist

$\beta_0 = 0$, so ist $e^{\beta_0} = 1$ und der Wettquotient ebenfalls gleich 1, beide Kategorien haben somit die gleiche Wahrscheinlichkeit von 0,5 an der Stelle $X = 0$. Ist $\beta_0 > 0$, dann ist $e^{\beta_0} > 1$ und die Wahrscheinlichkeit der Kategorie mit dem Wert $y = 1$ größer als die Wahrscheinlichkeit der Kategorie mit dem Wert $y = 0$. Ist $\beta_0 < 0$, dann ist $e^{\beta_0} < 1$ und die Wahrscheinlichkeit der Kategorie mit dem Wert $y = 1$ kleiner als die Wahrscheinlichkeit der Kategorie mit dem Wert $y = 0$.

**Bedeutung von $e^{\beta_1}$ (Odds-Ratio).** Der Parameter $e^{\beta_1}$ gibt die Veränderung des Wettquotienten an, wenn die unabhängige Variable um eine Einheit erhöht wird. Machen wir uns dies zunächst an dem Beispiel von $X = 0$ und $X = 1$ deutlich. Für beide Werte gilt nach Gleichung F 21.3:

$$\frac{P(Y=1|X=0)}{1-P(Y=1|X=0)} = e^{\beta_0} \qquad \text{(F 21.4)}$$

$$\frac{P(Y=1|X=1)}{1-P(Y=1|X=1)} = e^{\beta_0} \cdot e^{\beta_1} \qquad \text{(F 21.5)}$$

Setzt man Gleichung F 21.4 in Gleichung F 21.5 ein, erhält man:

$$\frac{P(Y=1|X=1)}{1-P(Y=1|X=1)} = \frac{P(Y=1|X=0)}{1-P(Y=1|X=0)} \cdot e^{\beta_1}$$

und hieraus:

$$e^{\beta_1} = \frac{\dfrac{P(Y=1|X=1)}{1-P(Y=1|X=1)}}{\dfrac{P(Y=1|X=0)}{1-P(Y=1|X=0)}} \qquad \text{(F 21.6)}$$

Der Parameter $e^{\beta_1}$ entspricht also einem Wettquotientenverhältnis (Odds-Ratio), das wir schon in Abschnitt 15.3.3 als Zusammenhangsmaß bei kategorialen Variablen kennengelernt haben. Ist $\beta_1 = 0$ und somit $e^{\beta_1} = 1$, dann hängt der Wettquotient nicht von der Ausprägung der unabhängigen Variablen ab, es gibt somit keinen Zusammenhang zwischen den beiden Variablen $X$ und $Y$. Ist $\beta_1 > 0$, dann ist $e^{\beta_1} > 1$, und es besteht ein positiver Zusammenhang zwischen $X$ und $Y$: Je größer $X$ wird, umso größer wird auch das Odds-Ratio und somit die Chance, dass die Kategorie mit dem Wert $y = 1$ gewählt wird im Vergleich zur Kategorie mit dem Wert $y = 0$. Wenn $\beta_1 < 0$ und somit auch $e^{\beta_1} < 1$, ist der Zusammenhang negativ. Die Chance, die Kategorie mit dem Wert $y = 1$ im Vergleich zur Kategorie mit dem Wert $y = 0$ zu wählen, sinkt mit Zunahme von $X$. Der Parameter $e^{\beta_1}$ gibt also an, mit welchem Faktor das bedingte Wettquotientenverhältnis gewichtet wird, wenn der Wert der Variablen $X$ um eine Messeinheit erhöht wird. Das Wettquotientenverhältnis für $X = 3$ setzt sich daher aus dem Produkt $e^{\beta_0} \cdot e^{\beta_1} \cdot e^{\beta_1} \cdot e^{\beta_1} = e^{\beta_0} \cdot (e^{\beta_1})^3$ zusammen, also aus $e^{\beta_0}$, dem Wettquotienten für $X = 0$, und dem Faktor $e^{\beta_1}$ für jede Erhöhung des Wertes von $X$ um eine Einheit. Beträgt z. B. $e^{\beta_1} = 1{,}2$, ist der Wettquotient für $X = 3$ um das $1{,}2^3 = 1{,}73$-fache größer als der Wettquotient für $X = 0$.

### Darstellung in Form des Logits

In der dritten Darstellungsform der logistischen Regressionsanalyse greift man auf den Logit zurück. Der Logit ist der logarithmierte Wettquotient. Der Wettquotient ist das Verhältnis aus einer Wahrscheinlichkeit und ihrer Gegenwahrscheinlichkeit (s. Abschn. 15.3.3). Logarithmiert man beide Seiten der Gleichung F 21.3, erhält man die bedingte Logit-Funktion $\ln(P(Y=1|X)/[1-P(Y=1|X)])$ als lineare Funktion der unabhängigen Variablen $X$ (s. Übung 5):

$$\ln\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right) = \beta_0 + \beta_1 \cdot X \qquad \text{(F 21.7)}$$

Mit »ln« wird der logarithmus naturalis bezeichnet. Diese Darstellung der logistischen Regression hat den Vorteil, dass die rechte Seite der Gleichung F 21.7 der rechten Seite von Gleichung F 16.28 der einfachen linearen Regression exakt entspricht. Dies zeigt die Analogie zwischen der einfachen logistischen und der einfachen linearen Regression: In der einfachen linearen Regression wird der bedingte Erwartungswert $E(Y|X=x)$ additiv linear zerlegt, in der logistischen Regression der bedingte Logit $\ln(P(Y=1|X=x)/[1-P(Y=1|X=x)])$.

Die additive Zerlegung in F 21.7 folgt direkt aus der Gleichung der logistischen Regression nach F 21.1. Umgekehrt könnte man auch Gleichung F 21.1 aus Gleichung F 21.7 ableiten und Gleichung F 21.7 zur Grundlage der Definition der logistischen Regression

machen. Dem läge die folgende Überlegung zugrunde: Wir haben bereits in Abbildung 21.1 gesehen, dass eine linear-additive Zerlegung der bedingten Wahrscheinlichkeit nicht sinnvoll ist, da die bedingte Wahrscheinlichkeit in ihrem Wertebereich beschränkt ist. Man kann nun die bedingte Wahrscheinlichkeit so transformieren, dass diese Beschränkung aufgehoben wird. Dies bewirkt die Logit-Transformation. Der Logit strebt gegen $-\infty$, wenn der Wettquotient gegen 0 strebt, die Wahrscheinlichkeit der Kategorie mit dem Wert $y = 1$ somit unendlich klein wird. Dabei muss der Fall $P(Y = 1 \mid X) = 0$ ausgeschlossen werden, da der Logarithmus an der Stelle 0 nicht definiert ist. Der Logit strebt gegen $+\infty$, wenn der Wettquotient gegen $\infty$ strebt. Dies ist der Fall, wenn die bedingte Wahrscheinlichkeit $P(Y = 1 \mid X)$ gegen 1 strebt. Dabei muss wiederum der Fall $P(Y = 1 \mid X) = 1$ ausgeschlossen werden, da ansonsten der Nenner des Wettquotienten 0 und somit nicht definiert wäre. Die logistische Regression setzt somit voraus, dass die Zugehörigkeit zu einer Kategorie nie perfekt vorhergesagt werden kann. Sie ließe sich z. B. dann nicht anwenden, wenn alle Personen mit einem Wert $x < x_S$ die erste Kategorie und alle Personen mit einem Wert $x \geq x_S$ die zweite Kategorie wählen würden, wobei $x_S$ ein Schwellenwert wäre, der die Variablen $X$ in zwei Hälften teilt, innerhalb deren die Kategorienzugehörigkeit jeweils perfekt vorhergesagt werden könnte. In diesem Fall wäre eine Bestimmung der Parameter nicht möglich.

**Bedeutung von $\beta_0$.** Die Konstante $\beta_0$ entspricht dem Wert des Logits an der Stelle $X = 0$ und hat somit eine analoge Bedeutung zum Achsenabschnitt in der einfachen linearen Regression.

**Bedeutung von $\beta_1$.** Das Regressionsgewicht $\beta_1$ zeigt an, um welchen Wert der Logit sich ändert, wenn der Wert der Variablen $X$ um eine Einheit erhöht wird. Ist das Regressionsgewicht $\beta_1$ gleich 0, besteht kein Zusammenhang zwischen beiden Variablen.

**Beispiel**

**Lebenszufriedenheit und positive Tagesereignisse**

In einer Studie an $n = 502$ Personen wurde der Zusammenhang zwischen der Lebenszufriedenheit einer Person und der Anzahl ihrer positiven Tagesereignisse in den letzten 24 Stunden untersucht. Personen markierten auf einer Liste von 30 positiven Tagesereignissen alle, die sie in den letzten 24 Stunden erlebt hatten. Der mögliche Wertebereich der unabhängigen Variablen erstreckt sich daher von 0 bis 30. Die abhängige Variable Lebenszufriedenheit wurde anhand des Items »Ich bin mit meinem Leben zufrieden« erfasst. Der Antwort »ja« wurde der Wert 1, der Antwort »nein« der Wert 0 zugeordnet. Eine Analyse des Zusammenhangs beider Variablen auf der Basis einer logistischen Regressionsanalyse ergab folgende Ergebnisse: $b_0 = \hat{\beta}_0 = 0{,}0113$ und $b_1 = \hat{\beta}_1 = 0{,}0728$. Hieraus lässt sich die bedingte Wahrscheinlichkeitsfunktion schätzen, die mit $\hat{P}(Y = 1 \mid X)$ bezeichnet wird, wobei das Symbol ˆ wieder angibt, dass es sich um eine Schätzung handelt.

**Darstellung in Form der bedingten Wahrscheinlichkeiten.** Für die geschätzte bedingte Wahrscheinlichkeitsfunktion ergibt sich somit:

$$\hat{P}(Y = 1 \mid X) = \frac{e^{0{,}0113 + 0{,}0728X}}{1 + e^{0{,}0113 + 0{,}0728X}}$$

Graphisch ist diese Funktion in Abbildung 21.5 dargestellt. Diese Graphik zeigt Folgendes:

- Der Wert $b_0$ ist größer als 0. Dadurch ist die Wahrscheinlichkeit, mit dem Leben zufrieden zu sein, selbst dann größer als 0,50, wenn man kein positives Lebensereignis auf der Liste angeben konnte ($X = 0$).
- Die bedingte Wahrscheinlichkeit liegt für $X = 0$ nahe bei 0,50, da der Wert von $b_0$ nur geringfügig größer als 0 ist.
- Die bedingte Wahrscheinlichkeit, mit dem Leben zufrieden zu sein, nimmt mit der Anzahl der positiven Tagesereignisse zu, da $b_1$ größer als 0 ist. Der Zuwachs ist nicht linear.

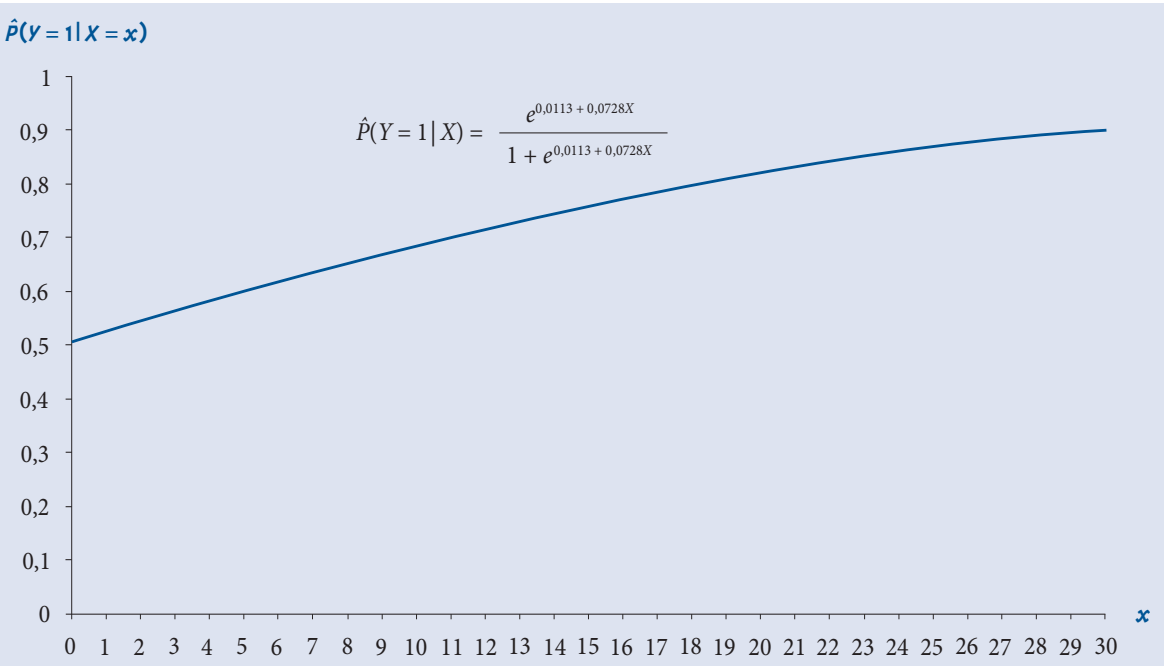


**Abbildung 21.5** Lebenszufriedenheit und positive Tagesereignisse: Bedingte Wahrscheinlichkeitsfunktion

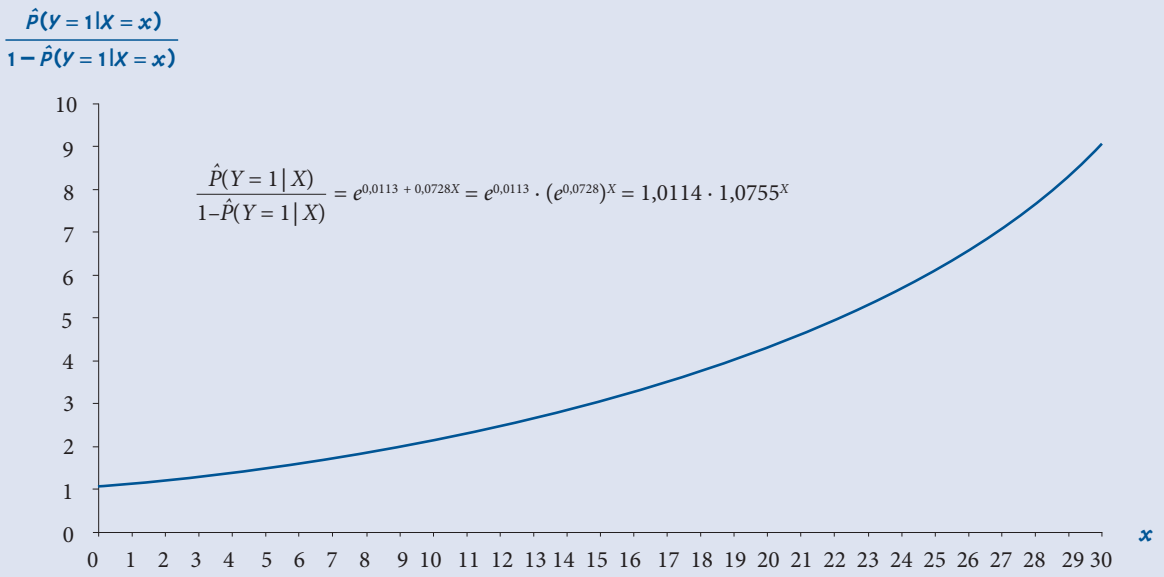


**Abbildung 21.6** Lebenszufriedenheit und positive Tagesereignisse: Bedingte Wettquotientenfunktion

**Darstellung in Form der bedingten Wettquotienten.** Für das Modell in der Form der Wettquotienten ergibt sich folgende Gleichung für die geschätzte bedingte Wettquotientenfunktion:

$$\frac{\hat{P}(Y=1|X)}{1-\hat{P}(Y=1|X)} = e^{0{,}0113+0{,}0728X}$$
$$= e^{0{,}0113}\cdot(e^{0{,}0728})^{X} = 1{,}0114\cdot 1{,}0755^{X}$$

Der Wert $e^{b_0} = e^{0{,}0113} = 1{,}0114$ zeigt an, dass die Chance, mit dem Leben zufrieden zu sein, größer als 1 ist, wenn eine Person kein positives Lebensereignis angeben konnte. Mit jedem positiven Lebensereignis, das eine Person erlebt hat, erhöht sich diese Chance um den Faktor $e^{b_1} = e^{0{,}0728} = 1{,}0755$. Die Chance, mit dem Leben zufrieden zu sein, ist bei 30 positiven Tagesereignissen in den letzten 24 Stunden um den Faktor $(e^{b_1})^{30} = (e^{0{,}0728})^{30} = 1{,}0755^{30} = 8{,}8779$ größer als ohne ein positives Lebensereignis in diesem Zeitraum. Dieser Chancenzuwachs ist in Abbildung 21.6 dargestellt. Man sieht, dass der bedingte Wettquotient mit der Zunahme positiver Tagesereignisse ansteigt.

**Darstellung in Form der bedingten Logits.** Für das Modell in der Form der Logits ergibt sich schließlich folgende Schätzgleichung für die bedingte Logit-Funktion:

$$\ln\left(\frac{\hat{P}(Y=1|X)}{1-\hat{P}(Y=1|X)}\right) = 0{,}0113 + 0{,}0728\cdot X$$

Diese Funktion ist in Abbildung 21.7 dargestellt. Die Abbildung zeigt, dass die bedingten Logits mit zunehmender Zahl positiver Tagesereignisse positiv-linear ansteigen. Der Logit liegt für Personen, die kein positives Tagesereignis erlebt haben, bei 0,0113. Mit jedem positiven Lebensereignis erhöht sich der bedingte Logit um den Wert 0,0728.

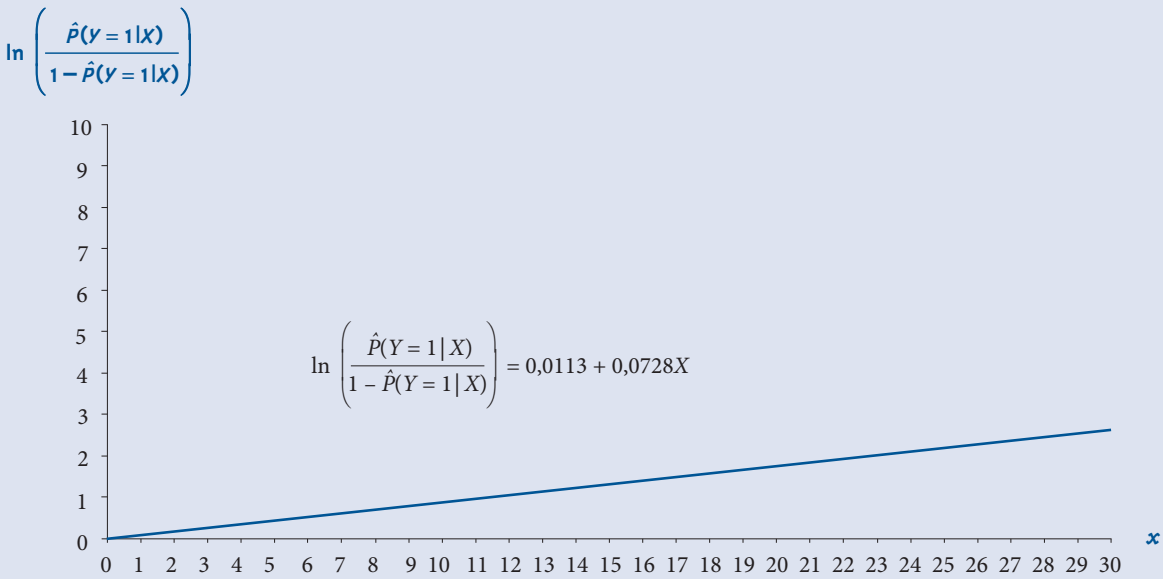


**Abbildung 21.7** Lebenszufriedenheit und positive Tagesereignisse: Bedingte Logitfunktion

Bei der Regressionsanalyse für metrische abhängige Variablen (Kap. 16 und 18) haben wir neben den unstandardisierten Regressionskoeffizienten noch standardisierte Regressionskoeffizienten behandelt, die man erhält, wenn die abhängige und die unabhängigen Variablen standardisiert werden. Dies ist bei der logistischen Regressionsanalyse nicht so einfach möglich. Um die standardisierten Koeffizienten zu erhal-