

## 20 Ordinale Regression

Fortsetzung der Seiten in der 9. Auflage

### Eine andere Modellbegründung

**Alternative Modellbegründungen.** In einem Modell der ordinalen Regression wird versucht, die Wahrscheinlichkeit für das Auftreten der Werte (Kategorien)  $j$  ( $j = 1, 2, 3$  bis zum letzten Wert  $J$ ) einer ordinalskalierten Variable  $y$  durch Einflussvariablen  $x_1, x_2$  und  $x_3$  vorherzusagen.<sup>1</sup> Dazu wird die (bedingte) kumulierte Wahrscheinlichkeit  $P(y \leq j/x)$  für das Auftreten eines Wertes von  $y$  gleich bzw. kleiner als ein bestimmter Wert  $j$  in nichtlinearer Abhängigkeit von den Einflussvariablen modelliert.<sup>2</sup> Als Einflussvariablen kommen metrische und kategoriale Variablen in Frage. Kategoriale Variable müssen vor Aufnahme in das Modell in Binärvariable gewandelt werden ( $\Rightarrow$  Kap. 18.3 im Buch).<sup>3</sup>

Nach einer Schätzung der Parameter (Koeffizienten) des Modells können die Wahrscheinlichkeiten für das Auftreten der Kategorien  $j$  von  $y$  ermittelt werden. Die Vorzeichen und die Höhe der geschätzten Koeffizienten erlauben Aussagen über die Wirkungsrichtung und Wirkungsstärke der Einflussvariablen.

Das Modell der ordinalen Regression kann in alternativer Weise begründet werden:

- $P(y \leq j/x)$  wird als eine nichtlineare Funktion der Variablen  $x_1, x_2, x_3$  postuliert. Diese Modellbegründung wird im Buch erläutert.
- In einer zweiten Modellbegründung wird die ordinalskalierte Variable  $y$  als eine grobe (kategorisierte) Messung einer nicht direkt messbaren latenten metrischen (stetigen) Variable  $y^*$  interpretiert. Aus dieser Konzeption ergeben sich die gleichen Modellgleichungen für  $P(y \leq j/x)$  wie in der Darstellung im Buch. Dieser Modellansatz soll hier erläutert werden.

**Das Messmodell für die Variable  $y$ .** Mit dem im Buch gewählten Beispiel zum politischen Interesse von Befragten (Variable POL2 in ALLBUS90.SAV) soll der Modellansatz erklärt werden.

Das politische Interesse der Bürger kann man sich als eine metrische, nicht direkt messbare Variable  $y^*$  vorstellen, die in der Befragung nur grob (kategorisiert) mittels vorgegebener Antworten erfasst wird. Ergebnis der Befragung ist eine ordinalskalierte Variable  $y$ . Als Antwortkategorien für die

---

<sup>1</sup> Zur einfacheren Darstellung beschränken wir uns hier auf maximal drei Einflussvariablen.

<sup>2</sup> Mit  $/x$  in  $P(y \leq j/x)$  soll ausgedrückt werden, dass es sich um eine bedingte Wahrscheinlichkeit handelt: z.B. der Wahrscheinlichkeit  $\Pr(y \leq 1)$  unter der Bedingung eines gegebenen Wertes von  $x$ . Für z.B.  $j = 3$  ist die kumulierte Wahrscheinlichkeit  $P(y \leq 3/x) = P(y = 1/x) + P(y = 2/x) + P(y = 3/x)$ .

<sup>3</sup> Der SPSS-Anwender muss sich aber darum nicht kümmern, weil das Programm dafür automatisch sorgt.

Befragung zum politischen Interesse war im ALLBUS vorgegeben: „überhaupt nicht“, „wenig“, „mittel“, „stark“ und „sehr stark“. Diese Antwortkategorien sind in der Variable  $y$  (= POL2) mit  $y = 1, 2, 3, 4$  und  $5$  kodiert. Der Zusammenhang zwischen der gemessenen Variable  $y$  zu der latenten Variable  $y^*$  ist im nachfolgenden Schema dargestellt. Schwellenwerte<sup>4</sup>  $\tau$  (tau) trennen die Bereiche der Variablen  $y^*$ , die durch die vorgegebenen Antwortkategorien zu verschiedenen Messwerten für  $y$  führen. Aus modelltheoretischen Gründen wird unterstellt, dass  $y^*$  den gesamten realen Zahlenbereich von  $-\infty$  bis  $+\infty$  abdecken kann. Daher wird für die Schwellenwerte  $\tau$  angenommen, dass  $\tau_0 = -\infty$  und  $\tau_5 = +\infty$  ist.

$$y = 1 \quad (\text{überhaupt nicht}) \quad \text{falls } \tau_0 \leq y^* < \tau_1 \quad (1a)$$

$$y = 2 \quad (\text{wenig}) \quad \text{falls } \tau_1 \leq y^* < \tau_2 \quad (1b)$$

$$y = 3 \quad (\text{mittel}) \quad \text{falls } \tau_2 \leq y^* < \tau_3 \quad (1c)$$

$$y = 4 \quad (\text{stark}) \quad \text{falls } \tau_3 \leq y^* < \tau_4 \quad (1d)$$

$$y = 5 \quad (\text{sehr stark}) \quad \text{falls } \tau_4 \leq y^* < \tau_5 \quad (1e)$$

Dieses Messmodell wird in der Abb. 20.6 veranschaulicht. Die Abstände zwischen den Schwellenwerten können verschieden groß sein.

In allgemeiner Form kann man das in den Gleichungen 1a bis 1e dargestellte Messmodellbeispiel in folgender Gleichung formulieren:

$$y_i = j \text{ wenn } \tau_{j-1} \leq y_i^* \leq \tau_j \text{ für } j = 1 \text{ bis } J^5 \quad (2)$$

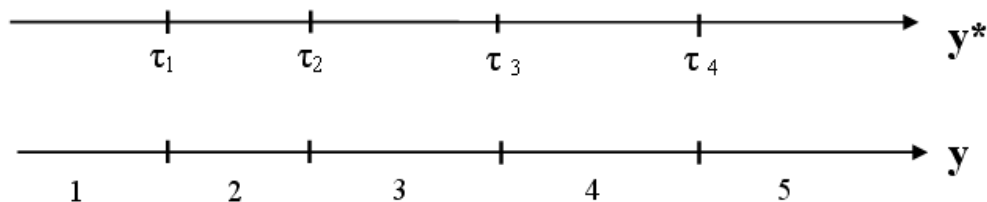


Abb. 20.6. Das Messmodell in grafischer Darstellung

**Das Regressionsmodell für die latente Variable  $y^*$ .** Für die latente Variable  $y^*$  wird wie bei einer linearen Regression ( $\Rightarrow$  Gleichung 18.10 in Kap. 18) angenommen, dass sie linear abhängig ist von den Einflussvariablen  $x_1$ ,  $x_2$  und  $x_3$  sowie einer zufälligen Fehlervariable  $\varepsilon$  (epsilon):

$$y_i^* = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \varepsilon_i \quad (3)$$

Um das Modell auch grafisch zu veranschaulichen, wird es vereinfacht. Im Folgenden wird nur eine metrische Einflussvariable  $x$  (z.B. die Dauer der Schuljahre) betrachtet. Die lineare Modellgleichung lautet nun

<sup>4</sup> Diese werden im Englischen cut points oder thresholds genannt.

<sup>5</sup> Der Index  $i$  bezeichnet den Datenfall  $i$ .

$$y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (4)$$

In Abb. 20.7 ist auf der senkrechten Achse die latente metrische Variable  $y^*$  und auf der waagerechten Achse die metrische Einflussvariable  $x$  abgebildet. Die Grafik zeigt ein Streudiagramm, in dem sich die Streupunkte zu einer „Punktwolke“ formieren ( $\Rightarrow$  Abb. 18.1 und 18.2 für Streudiagrammbeispiele in Kap.18 sowie Abb. 17.1 a in Kap 17). Jeder Streupunkt in der Grafik entspricht einer Wertekombination von  $y_i^*$  und  $x_i$  für eine Person  $i$  (einen Fall). Aus der Lage der „Punktwolke“, die sich von links unten nach rechts oben erstreckt, wird der positive Zusammenhang zwischen dem politischen Interesse der Befragten und der Einflussvariable  $x$  sichtbar. In die „Punktwolke“ ist die Regressionsgerade gelegt. Wie im Modell der linearen Regression, sind die senkrechten Abweichungen der Streupunkte von der Regressionsgerade in der Fehlervariablen  $\varepsilon$  erfasst.

Die Schwellenwerte  $\tau_1$ ,  $\tau_2$ ,  $\tau_3$  und  $\tau_4$  sind in Abb. 20.7 durch unterbrochene waagerechte Linien dargestellt. Sie teilen  $y^*$  in die fünf Messwerte  $y = 1$ ,  $y = 2$ ,  $y = 3$ ,  $y = 4$  und  $y = 5$ .

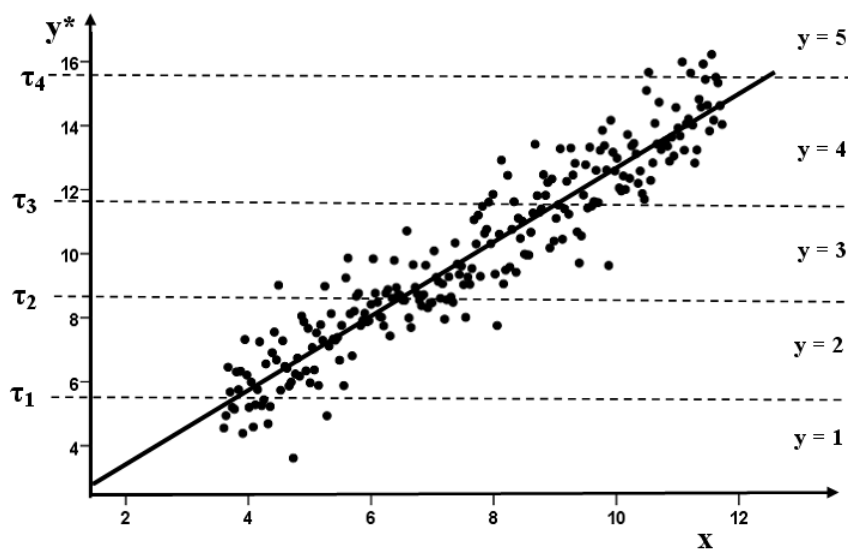


Abb. 20.7. Das Regressionsmodell der latenten Variable  $y^*$

**Das Regressionsmodell für die gemessene Variable  $y$ .** In Abb. 20.8 ist der Zusammenhang zwischen der gemessenen Variable  $y$  und der Einflussvariable  $x$  in einem Streudiagramm zu sehen. Alle Personen (Fälle) mit  $y^*$ -Werten oberhalb des Schwellenwertes  $\tau_4$  werden mit  $y = 5$  erfasst, alle Personen (Fälle) mit  $y^*$ -Werten zwischen  $\tau_3$  und  $\tau_4$  mit  $y = 4$  usw. Nun erscheinen die Streupunkte als horizontale Linien. Aus dieser Anordnung der Streupunkte der Fälle wird deutlich, dass eine unverzerrte Schätzung der Koeffizienten der Regressionsgleichung mit Hilfe der Methode der kleinsten Quadrate wie bei der linearen Regression nicht gelingen kann. Durch den Übergang von  $y^*$  zu  $y$  verändern sich die Werte auf der senkrechten Achse beträchtlich. In Abb. 20.7 liegen die Werte etwa im Zahlenbereich 3 bis 16 und in Abb. 20.8 im Zahlenbereich 1 bis 5. Zudem entspricht die Verteilung der Fehlervariablen  $\varepsilon$  nicht den Modellvoraussetzungen für eine lineare Regres-

sion. Die Fehlervariable  $\varepsilon$  in Abb. 20.8 hat nicht eine konstante Varianz und entspricht auch keiner Normalverteilung ( $\Rightarrow$  Kap. 18.1.2). Eine Schätzung der Koeffizienten mit Hilfe der Methode der kleinsten Quadrate verbietet sich auch allein, weil  $y$  keine metrische Variable ist.

Auch folgender anhand der Abbildungen gut zu erläuternder Aspekt stützt die Argumentation. Verringert man das Messmodell um einen Schwellenwert [z.B. um den Schwellenwert  $\tau_4$  und fasst damit die Kategorien  $y = 4$  („stark“) und  $y = 5$  („sehr stark“) zur Kategorie  $y = 4$  („stark und sehr stark“) zusammen], so hat dieses – wie man aus Abb. 20.7 sehen kann – keinen Einfluss auf die Lage der „Streupunktwolke“ sowie die Lage der Regressionsgrade. Aber die „Streupunktwolke“ in Abb. 20.8 verändert sich. Die Verringerung um eine Kategorie führt zum Verschwinden einer horizontal gelegenen Reihe von Beobachtungen und dieses würde die Schätzung der Koeffizienten beeinflussen. Dieses Argument gilt sinngemäß auch für die Erweiterung des Modells um einen Schwellenwert und damit zu einer Erhöhung der Anzahl der Kategorien von  $y$ .

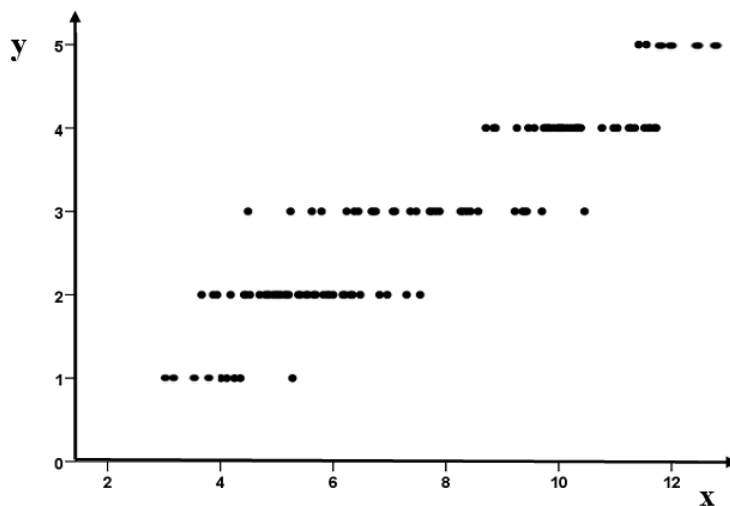


Abb. 20.8. Das Regressionsmodell der gemessenen Variable  $y$

**Annahmen zur Verteilung der stochastischen Fehlervariable  $\varepsilon$ .** In Abb. 20.9 ist das Modell in einer dreidimensionalen Perspektive zu sehen. Für drei beispielhaft gezeigte Werte  $x = x_1$ ,  $x = x_2$  und  $x = x_3$  der Einflussvariable überlagert die Verteilung der Fehlervariable  $\varepsilon$  die Regressionsgerade. Auf der senkrechten Achse ist mit  $f(y^*/x)$  die (bedingte) Häufigkeitsverteilung (Dichtefunktion) der latenten Variable  $y^*$  abgebildet.

Für die Verteilung von  $\varepsilon$  werden je nach Modellvariante verschiedene Annahmen gemacht. Für alle hier betrachteten Modellvarianten gilt aber, dass die Varianz der Fehlervariable  $\sigma_\varepsilon^2$  für verschiedene  $x$ -Werte konstant (gleich) ist.<sup>6</sup>

<sup>6</sup> Im Buch wird zum Abschluss des Kapitels 18 darauf hingewiesen, dass man mit SPSS auch (allgemeine) Modelle mit ungleicher Varianz schätzen kann.

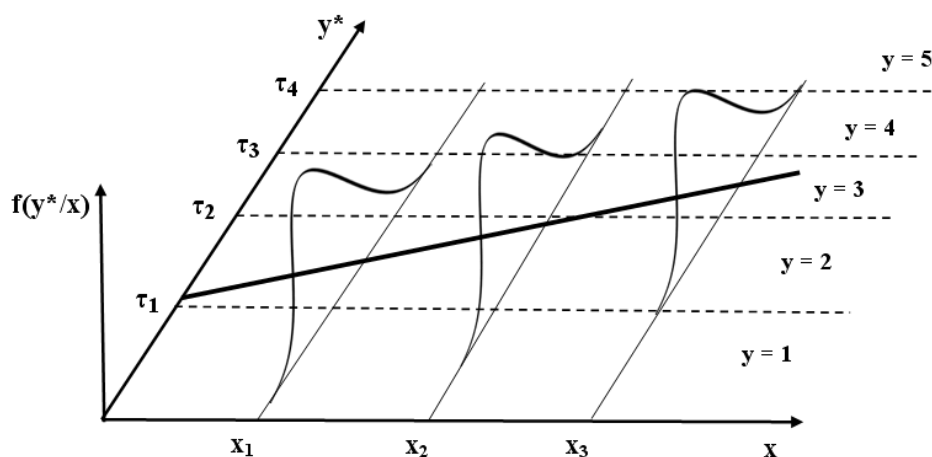
Wird für  $\varepsilon$  eine Normalverteilung unterstellt, so handelt es sich um das Probit-Modell der ordinalen Regression. Wird für  $\varepsilon$  die logistische Verteilung angenommen, so hat man es mit dem Logit-Modell der ordinalen Regression zu tun. Die Abb. 20.9 entspricht für den Fall der Normalverteilung von  $\varepsilon$  der Abb. 18.4 a in Kap. 18 des Buches.

Im Fall des Logit-Modells hat die Verteilung der Fehlervariablen  $\varepsilon$  eine logistische Verteilung mit der Dichtefunktion

$$f_{\text{Log}}(\varepsilon) = \frac{e^\varepsilon}{(1+e^\varepsilon)^2} \quad (5a)$$

und einem Mittelwert  $\mu_\varepsilon = 0$  und einer Varianz  $\sigma_\varepsilon^2 = \pi^2/3 \approx 3,29$ .<sup>7</sup> Die Verteilungsfunktion lautet:

$$F_{\text{Log}}(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{e^t}{(1+e^t)^2} dt = \frac{e^\varepsilon}{1+e^\varepsilon} = \frac{1}{1+e^{-\varepsilon}} \quad (5b)$$



**Abb. 20.9.** Das Regressionsmodell der latenten Variable  $y^*$  in dreidimensionaler Darstellung

Im Fall eines Probit-Modells hat die Fehlervariable  $\varepsilon$  eine Standardnormalverteilung mit der Dichtefunktion:

$$f_{\text{NV}}(\varepsilon) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\varepsilon^2} \quad (6a)$$

Diese hat einen Mittelwert  $\mu_\varepsilon = 0$  und eine Varianz  $\sigma_\varepsilon^2 = 1$ . Die Verteilungsfunktion lautet:

<sup>7</sup>  $\pi \approx 3,14$  ist die Kreiszahl.

$$F_{\text{NV}}(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt. \quad (6b)$$

Im Buch wird der S-förmige Verlauf der Verteilungsfunktion für das Logit- und Probit-Modell in einer Grafik veranschaulicht ( $\Rightarrow$  Abb. 20.2).

Neben den hier betrachteten Verteilungsfunktionen können weitere Verteilungen angenommen werden ( $\Rightarrow$  McCullagh, 1980, S. 115 und die Linkfunktionen in Tabelle 20.3 im Buch).

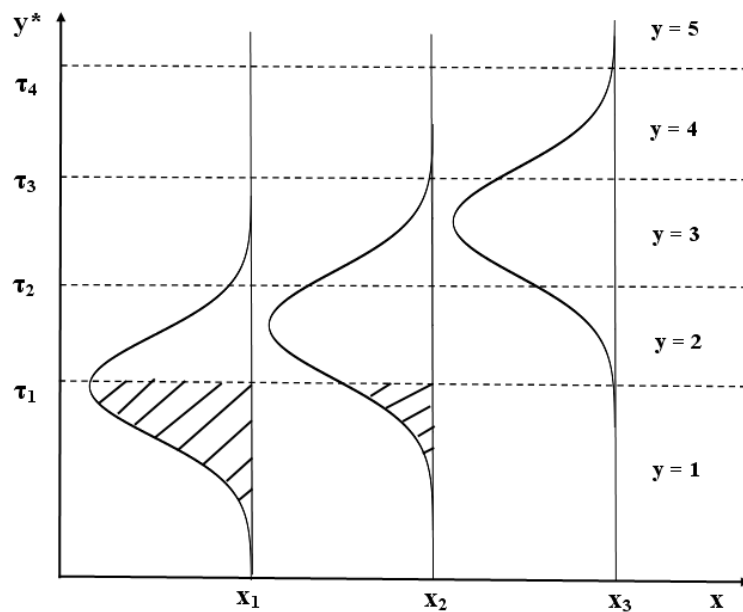
**Modellierung der Wahrscheinlichkeiten und ihre Darstellung in Grafiken.** Bei Annahme einer Verteilung für die Fehlervariable  $\varepsilon$  lässt sich bei Kenntnis der Parameter des Modells die (bedingte) Wahrscheinlichkeit für das Auftreten von  $y \leq j$  bei gegebenem Wert der Einflussvariable  $x$  ermitteln.

Für z.B.  $y = 1$  (dieser Wert wird gemessen, wenn  $y^*$  in den Grenzen zwischen  $\tau_0 = -\infty$  und  $\tau_1$  liegt) gilt für die Wahrscheinlichkeit für  $y \leq 1$  bei gegebenem Wert von  $x$ :

$$P(y_i \leq 1/x) = P(y_i^* \leq \tau_1/x) \quad (7a)$$

In Abb. 20.10 wird dieses veranschaulicht. Abb. 20.10 ist eine Darstellungsvariante der Abb. 20.9. Die Kurvenverläufe der Verteilungen der Fehlervariable  $\varepsilon$  sind um 90 Grad nach links in die Ebene der  $y^*/x$ -Achsen gedreht<sup>8</sup>. Mit dieser zweidimensionalen Grafik wird die grafische Abbildung der Wahrscheinlichkeiten vereinfacht. Die Wahrscheinlichkeit für  $y \leq 1$  (bei gegebenem  $x$ -Wert) entspricht dem schraffierten Flächenanteil der Verteilung von  $y^*$ , der durch die Obergrenze  $\tau_1$  abgesteckt wird ( $\Rightarrow$  Abb. 20.10). Man sieht, dass dieser (schraffierte) Flächenanteil (und damit die Wahrscheinlichkeit) mit größer werdenden  $x$ -Werten kleiner wird. Für  $x = x_3$  ist die Wahrscheinlichkeit nahe Null.

<sup>8</sup> Um die Grafik nicht zu überladen, ist die Regressionsgerade nicht abgebildet.



**Abb. 20.10.** Die Verteilung der Fehlervariablen  $\varepsilon$  für verschiedene Werte von  $x$

Einsetzen von  $y^*$  gemäß Gleichung (4) in Gleichung (7a) ergibt:

$$P(y_i \leq 1/x) = P(\beta_0 + \beta_1 x_i + \varepsilon_i \leq \tau_1) \quad (7b)$$

Subtrahiert man  $\beta_0 + \beta_1 x_i$  innerhalb der Klammer in 7b, so ergibt sich:

$$P(y_i \leq 1/x) = P(\varepsilon_i \leq \tau_1 - \beta_0 - \beta_1 x_i) \quad (7c)$$

Die Wahrscheinlichkeit, dass die Zufallsvariable  $\varepsilon$  unterhalb der Obergrenze  $\tau_1 - \beta_0 - \beta_1 x_i$  liegt, entspricht dem Anteil der durch die Grenze abgesteckten Fläche unter der Verteilung von  $\varepsilon$ .

Dieser Flächenanteil entspricht dem Wert der Verteilungsfunktion  $F(\varepsilon)$  an der Stelle der Obergrenze  $\tau_1 - \beta_0 - \beta_1 x_i$ . Daher ist:

$$P(y_i \leq 1/x) = F(\tau_1 - \beta_0 - \beta_1 x_i) \quad (7d)$$

Analog ist:

$$P(y_i \leq 2/x) = F(\tau_2 - \beta_0 - \beta_1 x_i) \quad (7e)$$

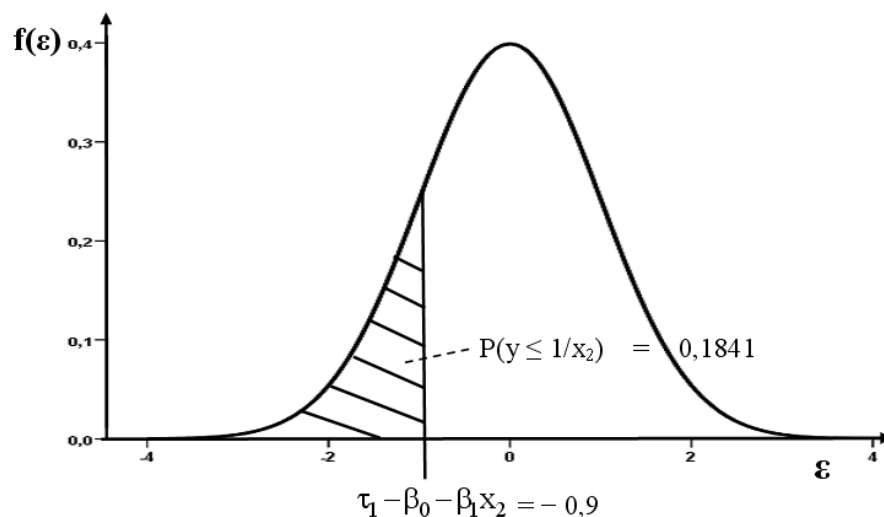
Gleichungen (7d) und (7e) lautet verallgemeinert für die Kategorien  $j = 1$  bis  $J$ :

$$P(y_i \leq j/x) = F(\tau_j - \beta_0 - \beta_1 x_i) \quad (8)$$

Die Wahrscheinlichkeit  $P(y_i \leq 1/x) = P(y_i^* \leq \tau_1/x) = P(\varepsilon_i \leq \tau_1 - \beta_0 - \beta_1 x_i) = F(\tau_1 - \beta_0 - \beta_1 x_i)$  gemäß Gleichungen (7a), (7c) bzw. (7d) soll auch mit einem Zahlenbeispiel veranschaulicht werden. Der Einfachheit halber haben wir dazu für die Verteilung von  $\varepsilon$  einer Standardnormalverteilung angenommen (Probit-Modell).

Unter Annahme fiktiver Werte für die Parameter ( $\tau_1 = 5,7$ ,  $\beta_0 = 2,6$  und  $\beta_1 = 0,8$ )<sup>9</sup> und für  $x_2 = 5$  ergibt sich (für die Obergrenze von  $\varepsilon$  bzw. für die Stelle der Verteilungsfunktion)  $\varepsilon = \tau_1 - \beta_0 - \beta_1 x_i = 5,7 - 2,6 - 0,8 \cdot 5 = -0,9$ . Die Wahrscheinlichkeit  $P(y_i^* \leq \tau_1 = 5,7/x)$  ist also bei  $x = 5$  gleich der Wahrscheinlichkeit  $P(\varepsilon \leq -0,9/x)$ . In Abb. 20.11 entspricht  $P(y_i \leq 1/x_2 = 5)$  dem schraffierten Flächenanteil unterhalb der Kurve der Dichtefunktion, der durch den Wert  $\varepsilon = \tau_1 - \beta_0 - \beta_1 x_2 = -0,9$  abgegrenzt wird. In Abb. 20.12 entspricht  $P(y_i \leq 1/x_2 = 5)$  dem Wert der Verteilungsfunktion an der Stelle  $\varepsilon = \tau_1 - \beta_0 - \beta_1 x_i = -0,9$ . Aus einer Tabelle der standardisierten Normalverteilung (verfügbar auf den Internetseiten zum Buch) ergibt sich  $P(\varepsilon \leq -0,9) = P(\varepsilon \geq 0,9) = 0,1841$ .

Steigt der Wert der Einflussvariable auf z.B.  $x_3 = 10$ , so ergibt sich  $\varepsilon = \tau_1 - \beta_0 - \beta_1 x_i = -4,9$ . In der Tabelle einer Standardnormalverteilung wird die Wahrscheinlichkeit  $P(\varepsilon \leq -4,9)$  nicht mehr ausgewiesen, da sie nur sehr gering ist.<sup>10</sup> Ein höherer  $x$ -Wert führt – wie schon in Abb. 20.10 zu sehen – bei einem positiven Zusammenhang zwischen  $y^*$  und  $x$  zu einer kleineren Wahrscheinlichkeit  $P(y_i \leq 1/x)$ .

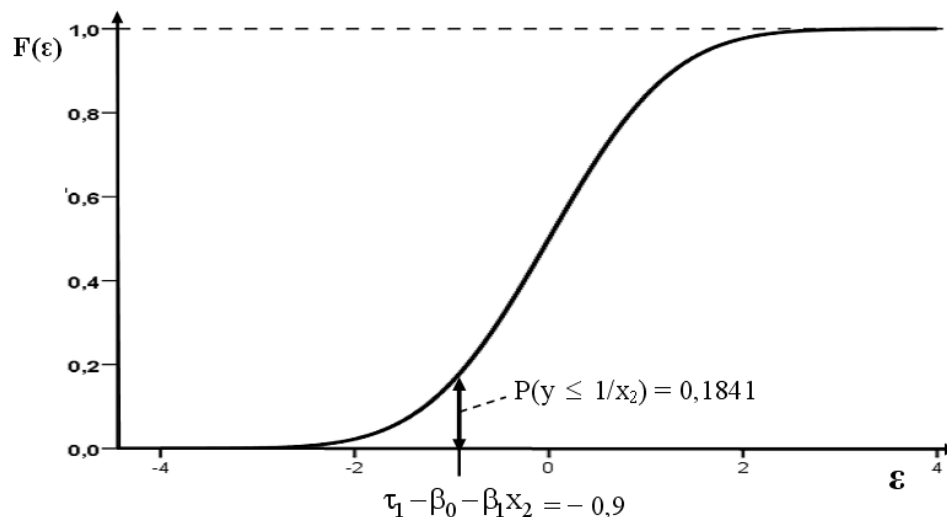


**Abb. 20.11.** Die Wahrscheinlichkeit  $P(y_i \leq 1/x_2)$  als Fläche unter der Dichtefunktion

<sup>9</sup> Der Wert für  $\tau_1$  wurde so gewählt, dass er Abb. 20.7 entspricht.

<sup>10</sup> In Abb. 20.9 und 20.10 berührt die unterbrochene Linie  $\tau_1$  den linken Rand der Verteilung der Fehlervariable bei  $x = x_3$  und in Abb. 20.11 rückt die Obergrenze an den linken Rand der Verteilung.





**Abb. 20.12.** Die Wahrscheinlichkeit  $P(y_i \leq 1/x_2)$  als Wert der Verteilungsfunktion

Aus den aus Gleichung (8) bestimmbaren Wahrscheinlichkeiten lässt sich durch Differenzbildung die Wahrscheinlichkeit für das Auftreten einer bestimmten Kategorie  $j$  ermitteln. So ist z.B. für  $j = 2$

$$P(y_i = 2/x) = P(y_i \leq 2/x) - P(y_i \leq 1/x) = F(\tau_2 - \beta_0 - \beta_1 x_i) - F(\tau_1 - \beta_0 - \beta_1 x_i) \quad (9)$$

Gleichung (9) lautet verallgemeinert für  $y = j$  (für  $j = 1$  bis  $J$ ):

$$P(y_i = j/x) = P(y_i \leq j/x) - P(y_i \leq j-1/x) = F(\tau_j - \beta_0 - \beta_1 x_i) - F(\tau_{j-1} - \beta_0 - \beta_1 x_i) \quad (10)$$

Für den Fall  $P(y_i = 1/x)$  fällt der zweite Term auf der rechten Seite der Gleichung (10) weg, weil wegen  $\tau_0 = -\infty$

$$P(y_i \leq 0/x) = P(\varepsilon_i \leq \tau_0 - \beta_0 - \beta_1 x_i) = F(\tau_0 - \beta_0 - \beta_1 x_i) = F(-\infty - \beta_0 - \beta_1 x_i) = 0 \quad (11)$$

ist. Daher ist:

$$P(y_i = 1/x) = P(y_i \leq 1/x) = F(\tau_1 - \beta_0 - \beta_1 x_i) \quad (12)$$

Für die letzte Kategorie  $J$  der Variable  $y$  (mit  $J = 5$  in unserem Beispiel) ist der erste Term auf der rechten Seite von Gleichung (10) gleich 1, weil wegen  $\tau_5 = \infty$

$$P(y_i \leq 5/x) = P(\varepsilon_i \leq \tau_5 - \beta_0 - \beta_1 x_i) = F(\tau_5 - \beta_0 - \beta_1 x_i) = F(\infty - \beta_0 - \beta_1 x_i) = 1. \text{ Es ist daher:}$$

$$P(y_i = 5/x) = 1 - F(\tau_4 - \beta_0 - \beta_1 x_i) \quad (13)$$

Der im obigen Zahlenbeispiel schon sichtbare gegenläufige Zusammenhang zwischen  $P(y_i \leq j/x)$  und der Einflussvariable  $x$  wird in Abb. 20.13 veranschaulicht. Auf der senkrechten Achse sind die Wahrscheinlichkeiten  $P(y_i \leq j/x)$  und auf der waagerechten Achse die Werte der Einflussvariable  $x$  abgebildet. Exemplarisch für das Logit-Modell ist der Verlauf von  $P(y_i \leq 1/x)$  und  $P(y_i \leq 3/x)$  in Abhängigkeit von der Höhe der Einflussvariable  $x$  zu sehen: die Wahrscheinlichkeit  $P(y_i \leq j/x)$

sinkt mit wachsendem  $x$ . Dieses gilt für den hier angenommenen positiven Zusammenhang zwischen  $y^*$  und  $x$  (d.h. bei positivem Koeffizienten  $\beta_1$ ). Besteht ein negativer Zusammenhang zwischen  $y^*$  und  $x$  (d.h. bei negativem Koeffizienten  $\beta_1$ ), so steigt  $P(y_i \leq j/x)$  mit wachsendem  $x$ .

Auch der nichtlineare Zusammenhang zwischen der Wahrscheinlichkeit  $P(y_i \leq j/x)$  und  $x$  wird aus der Abb. 19.20 ersichtlich. Der S-förmige Verlauf der logistischen Verteilungsfunktion ( $\Rightarrow$  Abb. 20.2 im Buch) ist in Abb. 20.13 als Spiegelbild zu sehen. Im Fall eines negativen Koeffizienten  $\beta_1$  entspräche der Verlauf dem der logistischen Funktion.

Für die Darstellung in Abb. 20.13 wurden die oben angeführten hypothetischen Werte der Parameter genutzt:  $\tau_1 = 5,7$ ,  $\beta_0 = 2,6$  und  $\beta_1 = 0,8$ , ergänzt um  $\tau_3 = 11,8$ .<sup>11</sup>

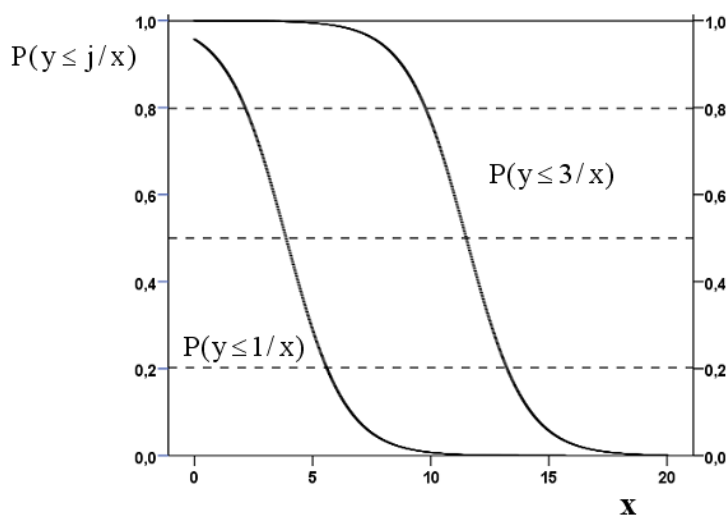


Abb. 20.13. Der Zusammenhang zwischen  $P(y \leq j/x)$  und der Einflussvariable  $x$

**Die Wahrscheinlichkeiten im Logit-Modell.** Für das in Gleichungen (10) dargestellte Modell in der Logit-Modellvariante ergeben sich unter Verwendung von Gleichung (5b) folgende Wahrscheinlichkeiten für die fünf Kategorien der ordinalskalierten Variable  $y$ :

$$P(y_i = 1/x) = F_{\text{Log}}(\tau_1 - \beta_0 - \beta_1 x_i) = \frac{1}{1 + e^{-(\tau_1 - \beta_0 - \beta_1 x_i)}} \quad (14a)$$

$$P(y_i = 2/x) = F_{\text{Log}}(\tau_2 - \beta_0 - \beta_1 x_i) - F_{\text{Log}}(\tau_1 - \beta_0 - \beta_1 x_i) = \frac{1}{1 + e^{-(\tau_2 - \beta_0 - \beta_1 x_i)}} - \frac{1}{1 + e^{-(\tau_1 - \beta_0 - \beta_1 x_i)}} \quad (14b)$$

<sup>11</sup> Der Wert für  $\tau_3$  wurde so gewählt, dass er Abb. 20.7 entspricht.

$$P(y_i = 3/x) = F_{\text{Log}}(\tau_3 - \beta_0 - \beta_1 x_i) - F_{\text{Log}}(\tau_2 - \beta_0 - \beta_1 x_i) = \frac{1}{1 + e^{-(\tau_3 - \beta_0 - \beta_1 x_i)}} - \frac{1}{1 + e^{-(\tau_2 - \beta_0 - \beta_1 x_i)}} \quad (14c)$$

$$P(y_i = 4/x) = F_{\text{Log}}(\tau_4 - \beta_0 - \beta_1 x_i) - F_{\text{Log}}(\tau_3 - \beta_0 - \beta_1 x_i) = \frac{1}{1 + e^{-(\tau_4 - \beta_0 - \beta_1 x_i)}} - \frac{1}{1 + e^{-(\tau_3 - \beta_0 - \beta_1 x_i)}} \quad (14d)$$

$$P(y_i = 5/x) = 1 - F_{\text{Log}}(\tau_4 - \beta_0 - \beta_1 x_i) = 1 - \frac{1}{1 + e^{-(\tau_4 - \beta_0 - \beta_1 x_i)}} \quad (14e)$$

Im Unterschied zur Wahrscheinlichkeit  $P(y_i \leq j/x)$  ( $\Leftrightarrow$  Abb. 20.13) ist der Verlauf von  $P(y_i = j/x)$  bei wachsendem  $x$  unterschiedlich.  $P(y_i = j/x)$  kann auch zunehmen.

Die Parameter der Gleichungen (14a) bis (14e) müssen aus Daten geschätzt werden, um Schätzwerte für die Wahrscheinlichkeiten zu gewinnen. Wie im nächsten Abschnitt dargestellt, ist dieses aber nicht ohne weiteres möglich.

**Das Identifikationsproblem.** Weil die latente Variable  $y^*$  nicht messbar ist, ist der unbekannte Mittelwert  $\mu_\varepsilon$  und die unbekannte Varianz  $\sigma_\varepsilon^2$  der Fehlervariable  $\varepsilon$  nicht durch Schätzung aus den Daten bestimmbar.<sup>12</sup> Daher gibt es hinsichtlich der zu schätzenden Parameter des Modells so genannte Identifikationsprobleme. Die Parameter des Modells sind nicht vollständig bestimmt: unterschiedliche Parameterkonstellationen führen zu gleichen Modellergebnissen. Das bedeutet, dass nicht alle Koeffizienten des Modells durch Parameterschätzung bestimmt werden können. Die Lösung besteht darin, Annahmen über die Varianz  $\sigma_\varepsilon^2$  bzw. über die Höhe von Parametern zu machen.

Das Identifikationsproblem mit Bezug zur unbekanntem Varianz  $\sigma_\varepsilon^2$  kann man sich mit Hilfe der Abb. 20.10 veranschaulichen. Angenommen ein alternatives Modell unterscheide sich lediglich in der Höhe der Varianz  $\sigma_\varepsilon^2$ : diese sei für alle  $x$ -Werte um einen konstanten Faktor größer als in Abb. 20.10 (d.h. die Verteilungskurven wären breiter). Nun stelle man sich neue Trennwerte  $\tau_1^*$ ,  $\tau_2^*$ ,  $\tau_3^*$  und  $\tau_4^*$  vor. Diese neuen Trennwerte seien so gelegen, dass die zugehörigen Flächenanteile (die den Wahrscheinlichkeiten für das Auftreten von  $y \leq 1, 2, 3, 4$  und 5 entsprechen) den in Abb. 20.10 zu sehenden Flächenanteilen für die Trennwerte  $\tau_1$ ,  $\tau_2$ ,  $\tau_3$  und  $\tau_4$  entsprechen. Mit dieser Überlegung soll verdeutlicht werden, dass unterschiedlichen Parametersets  $\tau_j$  und  $\tau_j^*$  ( $j = 1, 2, 3, 4$ ) zu gleichen Wahrscheinlichkeiten für die Werte von  $y$  führen. Zur Lösung dieses Identifikationsproblems wird für das Logit-Modell  $\sigma_\varepsilon^2 = \pi^2/3$  ( $\sigma_\varepsilon = \sqrt{\pi^2/3} \approx 1,81$ ) und für

<sup>12</sup> Im Unterschied dazu ist im Modell der linearen Regression  $\sigma_\varepsilon^2$  schätzbar ( $\Leftrightarrow$  Kap. 18.2.1 im Buch).

das Probit-Modell  $\sigma_\varepsilon^2 = 1$  ( $\sigma_\varepsilon = 1$ ) angenommen.<sup>13</sup> Es wären auch anderen Annahmen möglich. Aber diese Annahmen führen jeweils zu den einfachsten Modellen.

Es gibt ein weiteres Identifikationsproblem. Dieses soll durch folgende Überlegungen erläutert werden. Es sei angenommen, dass das in Gleichung (4) formulierte Modell  $y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i$  mit dem Parameter  $\beta_0$  sowie den Schwellenwerten  $\tau_1$ ,  $\tau_2$ ,  $\tau_3$  und  $\tau_4$  (allgemein  $\tau_j$  für  $j = 1, 2, 3, 4$ ) das wahre Modell sei in dem Sinne, dass dieses Modell die Daten erzeuge. Nun stellen wir uns vor, dass ein alternatives Modell  $y_i^* = \beta_0^* + \beta_1 x_i + \varepsilon_i$  bestehe mit dem Parameter  $\beta_0^* = \beta_0 - \delta$  sowie den Schwellenwertparametern  $\tau_1^* = \tau_1 - \delta$ ,  $\tau_2^* = \tau_2 - \delta$ ,  $\tau_3^* = \tau_3 - \delta$  und  $\tau_4^* = \tau_4 - \delta$ . Dabei sei  $\delta$  eine beliebige Konstante. In allgemeiner Formulierung gelte also für die Schwellenwerte des Alternativmodells:

$$\tau_j^* = \tau_j - \delta \text{ für } j = 1, 2, 3, 4 \quad (15)$$

Gemäß Gleichung (8) berechnet sich die Wahrscheinlichkeit für  $y \leq j$  für das alternative Modell als

$$\Pr(y_i \leq j/x) = F(\tau_j^* - \beta_0^* - \beta_1 x_i) = F((\tau_j - \delta) - (\beta_0 - \delta) - \beta_1 x_i) = F(\tau_j - \beta_0 - \beta_1 x_i) \quad (16)$$

Aus dieser Darstellung ergibt sich, dass die beiden hier betrachteten Modelle, die sich durch eine Konstante  $\delta$  in den Parametern  $\beta_0$  und  $\beta_0^*$  sowie in den Schwellenwertparametersets  $\tau_j$  und  $\tau_j^*$  unterscheiden, zu gleichen Wahrscheinlichkeiten für das Auftreten der Werte von  $y$  führen.<sup>14</sup> Dieser Sachverhalt beschreibt das Identifikationsproblem: verschiedene Kombinationen von Werten von  $\beta_0$  und  $\tau_j$  führen zu gleichen Wahrscheinlichkeiten, d.h. zu gleichen Modellergebnissen. Die Koeffizienten  $\beta_0$  und  $\tau_j$  für  $j = 1$  bis 4 des Modells sind unbestimmt, sie lassen sich nicht durch Schätzung bestimmen.

Zur Lösung dieses Identifikationsproblems sind unendlich viele Lösungen durch Restriktionsannahmen für einen der Parameter möglich. Aber zwei Lösungsvarianten sind bei der Einbettung in Software zur Berechnung von Modellen der ordinalen Regression üblich:

Es wird  $\tau_1 = 0$  angenommen und  $\beta_0$  geschätzt.<sup>15</sup>

Es wird  $\beta_0 = 0$  angenommen und  $\tau_1$  geschätzt.<sup>16</sup>

SPSS hat zur Lösung dieses Identifikationsproblems die Restriktion  $\beta_0 = 0$  gewählt. Gleichung (8) wird zu:

<sup>13</sup> Bedingt durch diese unterschiedlichen Annahmen sind die Schätzwerte der  $\beta$ -Koeffizienten der Einflussvariablen des Logit-Modells ca. 1,8-fach so groß wie die des Probit-Modells.

<sup>14</sup> Bezogen auf das obige Beispiel mit den fiktiven Parametern: die z.B. um den konstanten Wert 1 geringeren Parameter  $\tau_1^* = 4,7$  und  $\beta_0^* = 1,6$  führen für jeden  $x$ -Messwert zur gleichen Stelle  $\varepsilon = \tau_1 - \beta_0 - \beta_1 x_i$  der Verteilungsfunktion und somit zu gleichen Wahrscheinlichkeiten.

<sup>15</sup> Diese Restriktion impliziert für Gleichung (15)  $\delta = \tau_1$ .

<sup>16</sup> Diese Restriktion impliziert für Gleichung (15)  $\delta = \beta_0$ .

$$P(y_i \leq j/x) = F(\tau_j - \beta_1 x_i) \quad (17)$$

Zu beachten ist, dass unabhängig davon, welche Annahmen man zur Lösung der Identifikationsprobleme macht, die mit Hilfe eines Modells berechneten Wahrscheinlichkeiten für  $y$  davon unberührt bleiben. Die Schätzwerte für die  $\beta$ -Koeffizienten der Einflussvariablen  $x$  sind durch unterschiedliche Annahmen zur Lösung des Identifikationsprobleme nicht betroffen. Dieses gilt auch für die Signifikanztests. Unterschiede von Berechnungsergebnissen verschiedener Software zeigen sich in unterschiedlichen Schätzwerten für die Schwellenwertkoeffizienten und diese sind für die Interpretation berechneter Ergebnisse unbedeutend [vergleichbar mit dem Schätzwert  $b_0$  (der Konstanten) im Modell der linearen Regression].

**Parallele Modellgleichungen.** Aus der Herleitung der Modellgleichungen (8) ergibt sich, dass der Koeffizient  $\beta_1$  der Einflussvariablen  $x$  für alle Kategorien  $j$  gleich ist („parallel regression model“) und damit zur Grundlage des Modell gehört.<sup>17</sup>

In Abb. 20.13 ist dieser Sachverhalt veranschaulicht. Die Steigung der beiden Kurven für  $P(y_i \leq 1/x)$  und  $P(y_i \leq 3/x)$  ist für z.B.  $P = 0,8$ ,  $P = 0,5$  oder  $P = 0,2$  (hervorgehoben durch unterbrochene waagerechte Linien) die gleiche. Die Kurvenverläufe sind lediglich wegen der unterschiedlichen Werte für  $\tau_1$  und  $\tau_3$  parallel verschoben. Wie im Buch ausgeführt, sollte die Annahme paralleler Modellgleichungen mit Hilfe eines Tests geprüft werden. Wird die Hypothese paralleler Gleichungen verworfen, dann verliert das Modell seine Gültigkeit. In dem Fall sollte man prüfen, ob eine andere Verteilungsfunktion zu einem besseren Modell führt. Es könnte auch geprüft werden, ob man durch eine Zusammenfassung von Kategorien von  $y$  ein besseres Modell gewinnt. Schließlich könnte man auch die multinomiale logistische Regression in Erwägung ziehen.

**Gleiche Ergebnisse für beide Modellbegründungen.** Setzen wir wegen des Identifikationsproblems  $\alpha_j = \tau_j$  (damit soll explizit ausgewiesen werden, dass die „wahren“ Schwellenwerte  $\tau_j$  nicht geschätzt werden können) und gehen vom Modell gemäß Gleichung (3) aus [anstelle des vereinfachten Modells in Gleichung (4) mit nur einer Erklärungsvariable], so ergibt sich für  $j = 1$  bis  $J-1$ :

$$P(y_i \leq j/x) = F(\alpha_j - \beta_1 x_{1,i} - \beta_{2,i} x_{2,i} - \beta_{3,i} x_{3,i}) \quad (18)$$

Die Parameter (Koeffizienten) dieser  $J-1$  Modellgleichungen werden mit der Maximum Likelihood-Methode geschätzt. Dazu sei auf die Ausführungen im Buch verwiesen.

Die Gleichung (18) entspricht den Gleichungen (20.3a) bzw. (20.8) in Kap. 20 im Buch für das Logit- bzw. Probit-Modell.

Es ist somit gezeigt worden, dass die im Buch gewählte Modellbegründung und die hier erläuterte zu gleichen Modellgleichungen führen. Welche der beiden Begründungszusammenhänge man für das Modell der ordinalen Regression wählt, ist

<sup>17</sup> Enthält das Modell weitere Einflussvariablen, so gilt dieses auch für deren Koeffizienten.

insofern zweitrangig. Aber diese Aussage bedarf einer kleinen Einschränkung. In manchen Fällen ist die Vorstellung, dass die ordinalskalierte Variable  $y$  als eine grobe (kategorisierte) Messung einer latenten Variable  $y^*$  interpretiert wird, nicht angemessen. In einem Überblicksbeitrag in der Zeitschrift *WiSt*<sup>18</sup> z.B. wird für ein Beispiel zur Veranschaulichung der ordinalen Regression die ordinal skalierte Variable „Flugkategorie“ („Economy Class“, „Business Class“, „First Class“) als aus einer latenten metrischen Variable  $y^*$  („individuelle Neigungen für eine Flugkategorie“) entstehend betrachtet. Für uns ist das kein gutes Beispiel für diese Modellinterpretation. Die Flugkategorie ist nicht überzeugend als stetige Variable vorstellbar. Für ein solches Beispiel ist der im Buch erläuterte Modellbegründung angemessener.

### Literatur

- T. J. Gerpott, I. Mahmudova, Ordinale Regression. Eine anwendungsorientierte Einführung, in: *WiSt* (2006), S. 495-498.
- J. Janssen, W. Laatz, Statistische Datenanalyse mit SPSS für Windows, Berlin u.a. 2007.
- P. MacCullagh, Regression Models for ordinal Data, in: *Journal of the Royal Statistical Society, Series B(Methodological)*, Volume 42, Issue 2 (1980), S. 109-142.
- J. Scott Long, *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks u.a. 1997.

---

<sup>18</sup> T. J. Gerpott, I. Mahmudova (2006).