

Um die besonderen Eigenschaften einer logistischen Regression zu beschreiben, ist es sinnvoll, diese mit dem Alternativmodell der linearen Regressionsanalyse zu vergleichen, denn gerade im Abgleich mit der linearen Regressionsanalyse können die besonderen Stärken des Logit-Modells prägnant herausgearbeitet werden. Die logistische Regressionsanalyse kann daher am besten als statistische, methodische Antwort auf Schwächen und Probleme der linearen Regressionsanalyse begriffen werden, die sich ergeben, wenn die abhängige Variable dichotomer Natur ist.

Das typische (bivariate) Regressionsmodell besteht aus einer unabhängigen und einer abhängigen Variablen, die beide intervallskaliert sind (und idealerweise kontinuierlich verteilt). Im Regressionsmodell wird unterstellt, dass der (kausale) Zusammenhang zwischen unabhängiger und abhängiger Variablen in Form einer linearen Funktion dargestellt werden kann¹:

$$Y = \beta_0 + \beta_1 X \quad \text{GL (2.1)}$$

Diese lineare Funktion entspricht dem kausalen Prozess, in dem eine Veränderung von X eine Veränderung von Y bewirkt. Der konkrete Wert einer einzelnen Ausprägung von Y allerdings wird in der Regel durch weitere, zusätzliche Einflussfaktoren bestimmt. Des Weiteren bleibt noch ein Messfehler zu berücksichtigen,

¹ Ich begnüge mich hier und im Folgenden mit der Darstellung des bivariaten Zusammenhangs zwischen der abhängigen Variablen und einer einzigen unabhängigen Variablen. Grundsätzlich sind alle Erörterungen, die im Folgenden bezüglich des bivariaten Falls gemacht werden, auf den multivariaten Fall verallgemeinerbar. Ich bevorzuge die Darstellung des bivariaten Falls aus Gründen der Einfachheit und der Didaktik, insbesondere, weil so auch graphische Darstellungen möglich sind. Ich verschiebe die explizite Diskussion der multivariaten Analyse auf das 5. Kapitel, da hier die zu erörternden Konzepte nur sinnvoll diskutiert werden können, wenn man den multivariaten Charakter eines Modells explizit berücksichtigt.

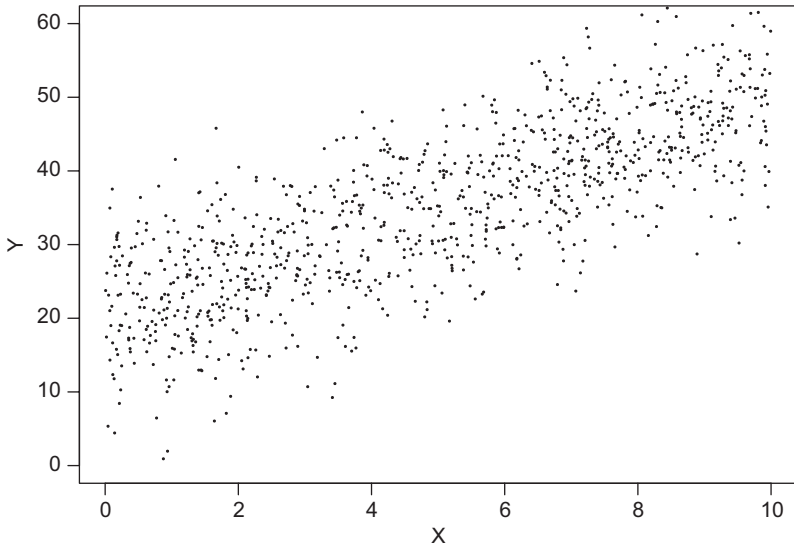


Abb. 2.1 Streudiagramm

so dass der spezifische Y-Wert des i -ten Falls einer Stichprobe mit Hilfe der Regressionsfunktion nur mehr oder weniger ungenau vorausgesagt werden kann. Der Y-Wert eines konkreten Falls y_i kann mit Gl. 2.1 daher aus dem Wert der unabhängigen Variablen x_i nicht exakt bestimmt werden, sondern weicht um den Fehler e_i vom Wert, der durch die Gleichung vorausgesagt wird, ab.

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad \text{GL (2.2)}$$

Die Gl. 2.1 gibt nur den „theoretischen“, d. h. allgemeinen bzw. mittleren Zusammenhang zwischen der unabhängigen Variablen X und der abhängigen Variablen Y wieder. Die tatsächlich beobachteten Werte von Y streuen entsprechend der Verteilung der Fehlerkomponente E um die durch das Modell vorausgesagten Werte von Y . Graphisch lässt sich der Zusammenhang zwischen X und Y in Form eines Streudiagramms abbilden (Abb. 2.1).

Der „wahre“ Zusammenhang zwischen X und Y , wie er sich in Gl. 2.1 in Form der Koeffizienten ausdrückt, ist der „theoretisch korrekte“ Zusammenhang, d. h. der Zusammenhang, wie er in der sogenannten Grundgesamtheit, d. h. der Population, der unser theoretisches Interesse gilt, tatsächlich besteht. Nehmen wir z. B. an, uns interessierte der Zusammenhang zwischen der Selbst-Einstufung auf der

Links-Rechts-Skala und der Sympathie für die CDU bei der Bundestagswahl 2009. Unsere Vermutung wäre dann, dass die Sympathie für die CDU desto höher ausfällt, je weiter rechts man sich selbst ideologisch positioniert. Der „wahre“ Zusammenhang ist uns jedoch nicht bekannt. Um ihn korrekt darzustellen, müssten wir die entsprechenden Daten für alle wahlberechtigten Bürger der BRD erheben, was aus praktischen Gründen unmöglich ist. Der Sinn von inferenzstatistischen Verfahren besteht darin, den wahren Zusammenhang, wie er in der Grundgesamtheit besteht, mit einer Stichprobe von Befragten mehr oder weniger genau zu schätzen. Die geschätzten Werte von Y sollen durch ein „Dach“ gekennzeichnet werden, ebenso die geschätzten Werte für die Koeffizienten.

$$\begin{aligned}\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i \\ &= b_0 + b_1 x_i\end{aligned}\quad \text{GL (2.3)}$$

Die Koeffizienten b_0 und b_1 sind ihrerseits Schätzer der „wahren“ Koeffizienten β_0 und β_1 .² Die in Gl. 2.3 dargestellte Regressionsfunktion bzw. Regressionsgerade wird dabei so bestimmt, dass sie die „bestmögliche“ Schätzung der wahren Geraden darstellt. Angewandt auf das Beispiel in Abb. 2.1 heißt das, dass wir diejenige Gerade suchen, die die „bestmögliche“ Repräsentation der Punktwolke darstellt. Anders ausgedrückt: Wir suchen diejenige lineare Funktion bzw. Geradengleichung, die den wahren Zusammenhang zwischen X und Y am besten repräsentiert, womit in diesem Fall gemeint ist, dass diese Gerade die bestmögliche Schätzung für die wahre Gleichung darstellt. Diese Regressionsgerade wird durch die „Methode der kleinsten Quadrate“ (im Englischen OLS für „ordinary least squares“) gefunden, d. h. wir bestimmen die Parameterwerte der Geraden so, dass die Summe der quadratischen Abweichungen der tatsächlich beobachteten Werte von Y von den aufgrund der Regressionsgeraden geschätzten Werten minimiert wird. Die Logik des OLS-Verfahrens ist eine der Minimierung des Schätzfehlers, da die quadratischen Abweichungen der beobachteten Werte von den prognostizierten als Ausmaß dieses Fehlers interpretiert werden können. Die so gefundene Gerade lässt sich dann graphisch darstellen, wie in Abb. 2.2 zu sehen.

Damit diese Schätzung aber mit einer gewissen Zuverlässigkeit eine „gute“ bzw. sogar die bestmögliche Schätzung des wahren Zusammenhangs darstellt, müssen folgende Bedingungen erfüllt sein (vgl. Berry und Feldman 1985, S. 10 f.; Berry 1993, S. 12; Ohr 2010, S. 645 ff.):

² Ich folge hier der üblichen Konvention, Parameter in der Grundgesamtheit mit griechischen Buchstaben zu versehen und Parameter in der Stichprobe mit lateinischen.

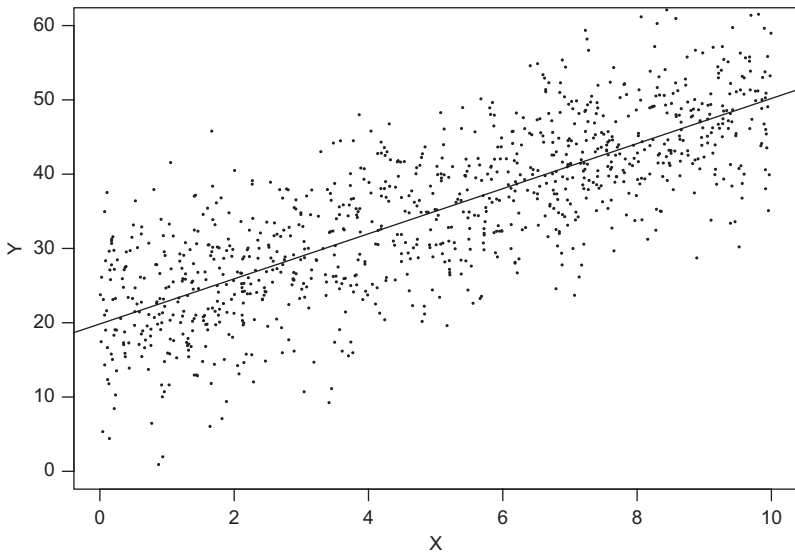


Abb. 2.2 Streudiagramm mit Regressionsgerade

1. *Messung*: Die unabhängigen Variablen müssen mindestens intervallskaliert oder dichotom sein, die abhängige Variable ist intervallskaliert, kontinuierlich und (theoretisch) unbegrenzt. Alle Variablen werden ohne Messfehler gemessen.
2. *Modellspezifikation*: Im Modell sind alle relevanten Variablen und nur relevante Variablen enthalten. Die Form der unterstellten Beziehung zwischen abhängiger und unabhängiger Variable ist linear.
3. *Erwartungswert des Fehlers*: Der Erwartungswert des Fehlers, d. h. der Residuen, ist Null
4. *Homoskedastizität*: Die Varianz des Fehlerterms unterscheidet sich nicht hinsichtlich der verschiedenen Merkmalsausprägungen der unabhängigen Variablen.
5. *Normalverteilung des Fehlers*: Der Fehler ist normalverteilt.
6. *Autokorrelation des Fehlers*: Es gibt keine Korrelation des Fehlers bezüglich verschiedener Merkmalsausprägungen der unabhängigen Variablen.
7. *Keine Korrelation zwischen dem Fehler und der unabhängigen Variablen*
8. *Kollinearität*: Es darf keine perfekte Multikollinearität zwischen den unabhängigen Variablen bestehen.

Die dritte, vierte und sechste Bedingung alleine sind schon hinreichend dafür, dass die mit der OLS-Methode gefundenen Schätzer für die Parameter der Regressions-

gleichung die besten, linearen, unverzerrten Schätzer sind (im Englischen BLUE für „best linear unbiased estimator“). Mit „beste“ ist hiermit gemeint, dass der sogenannte Standardfehler der Schätzung, also die mittlere Abweichung³ des geschätzten Parameterwerts vom wahren der Grundgesamtheit minimiert wird. Die OLS-Methode garantiert damit, dass wir im Mittel mit unserem geschätzten Parameterwert so nahe an dem wahren liegen wie durch eine Schätzung überhaupt möglich.

Liegen die genannten Bedingungen vor, insbesondere die fünfte Annahme, dass die Fehler normalverteilt sind, gilt außerdem, dass die gefundene Regressionsgleichung diejenige ist, die mit der größten Wahrscheinlichkeit die wahre Gleichung ist. Genauer gesagt ist sie diejenige Gleichung, die mit der größten Wahrscheinlichkeit die gemachten Beobachtungen produzieren würde. Tatsächlich können die gemachten Beobachtungen ja unter einer Vielzahl verschiedener „wahrer“ Parameterwerte auftreten. Nehmen wir den einfachen Fall eines Münzwurfs an. Eine beliebige Folge von Kopf- und Zahl-Würfen kann genauso entstehen, wenn wir die Münzwürfe mit einer „fairen“ Münze durchführen, bei der die Wahrscheinlichkeiten von „Kopf“ und „Zahl“ jeweils $\frac{1}{2}$ sind, wie wenn wir sie mit einer gezinkten Münze durchführen würden, bei der „Kopf“ mit Wahrscheinlichkeit $\frac{2}{3}$ und „Zahl“ nur mit Wahrscheinlichkeit $\frac{1}{3}$ auftreten würde. Die bedingte Wahrscheinlichkeit, mit der eine bestimmte Folge von Kopf und Zahl auftritt, je nachdem, um welchen Typus der Münze es sich handelt, ist aber unterschiedlich. Die Folge „Kopf-Zahl-Zahl-Kopf-Zahl“ tritt bei einer fairen Münze mit der Wahrscheinlichkeit $\frac{1}{2}^5 = 1/32 = 0,031$, bei der gezinkten Münze hingegen mit $\frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} = 4/243 = 0,016$ auf. Die a-priori-Wahrscheinlichkeit, dass wir mit einer fairen Münze eine Sequenz wie die tatsächlich beobachtete hervorrufen würden, ist also ungefähr doppelt so groß wie die der gezinkten Münze.

Genauso sind prinzipiell viele verschiedene theoretische „wahre“ Gleichungen vorstellbar, die die beobachtete Punktwolke hervorrufen könnten, so wie in Abb. 2.3 dargestellt. Für jede dieser Gleichungen bzw. Geraden kann eine bedingte Wahrscheinlichkeit berechnet werden, mit der die tatsächlich vorliegende Punktwolke aufgetreten wäre, wenn die jeweilige angenommene „theoretische“ Gleichung auch die wahre gewesen wäre. Die Regressionsgerade bzw. die mit Hilfe der OLS-Schätzmethode ermittelte Gerade ist unter dieser Vielzahl von grundsätzlich vorstellbaren theoretischen Gleichungen eben diejenige, bei der diese bedingte Wahrscheinlichkeit maximiert wird. Allgemein werden solche Verfahren als Maximum-Likelihood-Verfahren bezeichnet. Die OLS-Methode kann daher als Spezialfall eines ML-Verfahrens betrachtet werden.

³ Genauer gesagt handelt es sich beim Standardfehler um die Wurzel aus der mittleren quadratischen Abweichung.

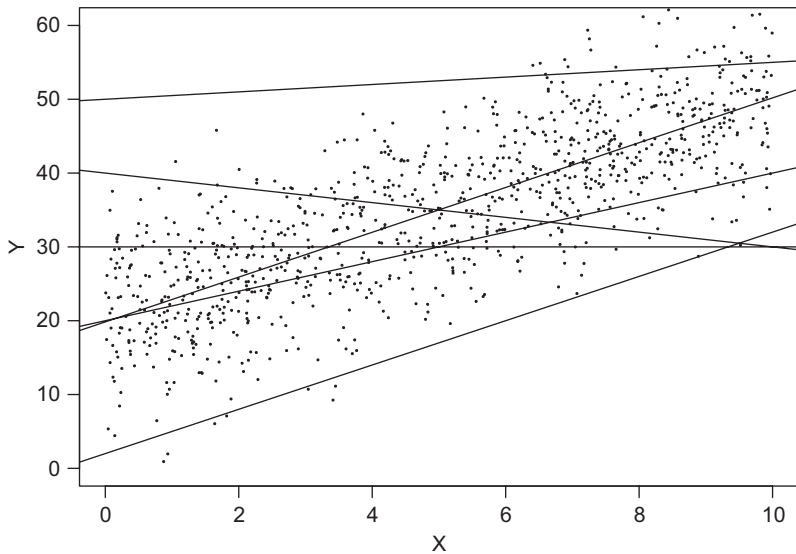


Abb. 2.3 Streudiagramm mit diversen theoretischen Gleichungen

Die lineare Regressionsanalyse ist wie erwähnt das Standardanalyseverfahren schlechthin und findet oft auch Anwendung, wenn die oben genannten Bedingungen nicht ganz und gar erfüllt sind. Das Schätzverfahren der OLS-Methode verhält sich dabei in der Regel relativ robust gegen solche Verletzungen der Annahmen, d. h. selbst wenn diese nicht hundertprozentig erfüllt sind, führt das Verfahren in der überwiegenden Anzahl der Fälle immer noch zu brauchbaren, d. h. in einem pragmatischen Sinne realistischen Schätzwerten.

Zwar verlangt die erste der oben genannten Bedingungen, dass die abhängige Variable kontinuierlicher Natur sein muss, rein technisch gesehen aber bereitet es keinerlei Probleme, die Methode der normalen linearen Regression auch dann anzuwenden, wenn die abhängige Variable dichotomer Natur ist. Geschätzt wird in diesem Fall wieder der repräsentative Wert der abhängigen Variablen in Abhängigkeit von der Ausprägung der unabhängigen Variablen, bzw. der Wert, der die durch einen bestimmten X-Wert bedingte Verteilung der abhängigen Variablen als einzelner Wert am besten repräsentiert. Im Falle einer kontinuierlichen Variablen ist dieser repräsentative Wert der Mittelwert, im Falle einer dichotomen Variablen die relative Häufigkeit der Ausprägung, die das Vorhandensein der mit der abhängigen Variablen gemessenen Eigenschaft ausdrückt. Da die abhängige Variable üblicherweise mit 0 und 1 vercodet ist, fällt diese relative Häufigkeit eleganterweise mit

dem Mittelwert zusammen, was die Anwendbarkeit einer normalen linearen Regression zu unterstützen scheint. Da die relative Häufigkeit bzw. der Mittelwert, wie schon erwähnt, auch als Wahrscheinlichkeit des Auftretens der Ausprägung „1“ interpretiert werden kann, wird die Anwendung der normalen linearen Regression auf Analysen mit einer dichotomen abhängigen Variablen auch als Modell linearer Wahrscheinlichkeit oder „linear probability“ (vgl. Aldrich und Nelson 1985) bezeichnet. Allerdings führt die Anwendung der normalen linearen Regressionsanalyse mit OLS-Schätzung in solchen Fällen zu nicht unproblematischen Ergebnissen, die zum einen konzeptueller Natur, zum anderen statistischer bzw. methodischer Natur sind. Das logistische Regressionsmodell ist sozusagen die angemessene Antwort auf bestimmte Defekte des linearen Modells, wenn man es ohne weiteres auf dichotome abhängige Variablen anwendet. Ich werde daher im Folgenden auch erst diesen „Umweg“ über das Modell der linearen Wahrscheinlichkeit einschlagen, um vor diesem Hintergrund die Stärken des logistischen Modells klarer hervortreten zu lassen.

Als Beispiel wähle ich die Untersuchung des Zusammenhangs zwischen der Wahlteilnahme und der Bewertung des Parteiensystems. Der zugrundeliegende Datensatz ist die so genannte GLES⁴-Studie 2009. Die Studie besteht unter anderem aus mehreren Querschnittstudien, von denen in unserem Zusammenhang vor allem die Vorwahl- und die Nachwahlstudie von Interesse sind. In der Nachwahlbefragung war unter anderem die Frage nach der beabsichtigten Wahlteilnahme enthalten. Während diese Frage bei Studien, die vor einer Wahl stattfinden, häufig mit Antwortvorgaben gestellt werden, die eine Wahrscheinlichkeitsaussage machen („werde sicher zur Wahl gehen“, „werde wahrscheinlich zur Wahl gehen“ etc.), kann sie nach einer Wahl nur in dichotomer Weise gestellt haben, da es hier nur zwei sinnvolle Antwortvorgaben gibt („Ja, habe gewählt“, „Nein, habe nicht gewählt“). Wenn man die Wahlteilnahme erklären will, so könnte man z. B. einen Zusammenhang mit der sogenannten „Politikverdrossenheit“ sehen. Die interessierende Hypothese könnte z. B. lauten: „Je größer die Politikverdrossenheit bei den Bürgern, desto seltener nehmen sie an Wahlen teil.“ Die unabhängige Variable der Politikverdrossenheit operationalisiere ich im Beispiel eher im Sinne einer Parteienverdrossenheit als die maximale Bewertung, die für eine der fünf etablierten Parteien (CDU/CSU, SPD, FDP, Grüne, Linke) auf dem Sympathieskalometer angegeben wurde. Hierbei konnten die Befragten für jede der Parteien auf einer Skala von – 5 bis + 5 angeben, wie viel sie von den Parteien hielten. Es scheint durchaus sinnvoll, Politikverdrossenheit bzw. die Abwesenheit von Politikverdrossenheit als den Maximalwert der Parteienbewertungen zu operationalisieren. Der Mittelwert der Bewertungen wäre weniger angebracht, da z. B. eine große Varianz der wahrge-

⁴ German Longitudinal Election Study.

Tab. 2.1 Kreuztabelle: Zusammenhang zwischen Bewertung der Parteien und Wahlteilnahme

| | Bewertung der am besten bewerteten Partei | | | | | | | | | | |
|-----------|---|----|----|----|----|----|-----|-----|-----|-----|-----|
| | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
| Nichtwahl | 30 | 6 | 9 | 19 | 20 | 70 | 55 | 57 | 46 | 33 | 16 |
| Wahl | 8 | 1 | 8 | 14 | 21 | 83 | 138 | 254 | 400 | 345 | 253 |

Tab. 2.2 Kreuztabelle: Zusammenhang zwischen Bewertung der Parteien und Wahlteilnahme

| | Bewertung der am besten bewerteten Partei | | | | | | | | | | |
|-----------|---|------|------|------|------|------|------|------|------|------|------|
| | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
| Nichtwahl | 78,9 | 85,7 | 52,9 | 57,6 | 48,8 | 45,8 | 28,5 | 18,3 | 10,3 | 8,7 | 5,9 |
| Wahl | 21,1 | 14,3 | 47,1 | 42,4 | 51,2 | 54,2 | 71,5 | 81,7 | 89,7 | 91,3 | 94,1 |

nommenen Sympathie der Parteien keineswegs bedeuten muss, dass man politikverdrossen ist. Ausschlaggebend scheint hierfür eher zu sein, wie gut bzw. schlecht selbst die noch am besten bewertete Partei abschneidet. Wir erwarten also einen positiven Zusammenhang zwischen Nichtwahl und Politikverdrossenheit bzw. einen positiven Zusammenhang zwischen der Wahlteilnahme und dem „Mangel“ an Politikverdrossenheit. Im Folgenden werde ich allgemein von der Parteienbewertung sprechen, ohne jedes Mal darauf hinzuweisen, dass sich dies im Sinne des „ Mangels“ an Politikverdrossenheit auf die Bewertung der am besten bewerteten Partei bezieht.

Die Verteilung der kombinierten Werte der abhängigen und unabhängigen Variablen in Form einer Kreuztabelle ist in Tab. 2.1 zu sehen. Aussagekräftiger als die absoluten Zahlen sind die Spaltenprozentzahlen, bzw. die innerhalb der durch die unabhängige Variable gebildeten Gruppen prozentuierten Zahlen (Tab. 2.2).

Offensichtlich nimmt der Anteil der Wähler wie erwartet mit zunehmender positiver Bewertung des Parteiensystems zu. Analog zur linearen Regressionsanalyse wird ein Streudiagramm angefertigt, das die charakteristische Form des Zusammenhangs auf graphische Weise enthüllen soll (Abb. 2.4).

Ein derartiges „Streudiagramm“ ist allerdings wenig hilfreich, da keine „Streuung“ der Datenpunkte vorhanden ist. Dies liegt daran, dass sowohl die X- als auch die Y-Variable nur wenige diskrete Ausprägungen haben, so dass nur relativ wenige konkrete Kombinationen von Werten insgesamt möglich sind. Da alle Wertekombinationen mindestens einmal besetzt sind und im Streudiagramm nicht erkennbar ist, wie oft die einzelnen Kombinationen besetzt sind, lässt sich so nicht erkennen, welche Kombinationen von Werten besonders häufig auftreten. Dieses

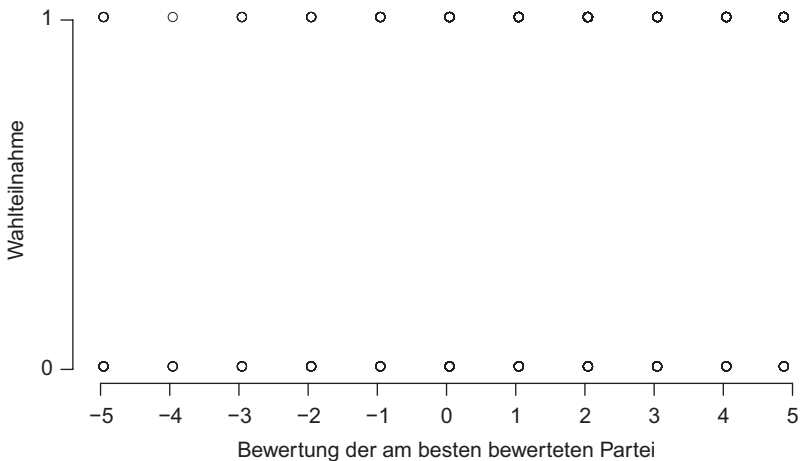


Abb. 2.4 Streudiagramm über den Zusammenhang zwischen Wahlteilnahme und Parteienbewertung

Problem kann man umgehen, indem man die Werte „jittered“, also leicht „verzittert“. In der Graphik erscheinen also nicht die tatsächlichen Werte, sondern die gezeichneten Werte sind zufällig um den wahren Wert herum verteilt⁵. In Abb. 2.5 sind die „gejitterten“ Werte dargestellt.

Die gejitterte Graphik gibt sicherlich schon deutlich besseren Aufschluss über die Natur des Zusammenhangs. Es ist klar zu erkennen, dass es eine massive Häufung von Werten im Bereich rechts oben gibt, d. h. die Absicht zur Wahlteilnahme ist dort besonders stark ausgeprägt, wo eine positive Bewertung des Parteiensystems vorliegt, ganz so, wie es unsere Hypothese in diesem Fall auch vorausgesagt hat.

Noch eindeutiger lässt sich die Verteilung der 0- und 1-Werte der abhängigen Variablen zusammenfassend darstellen, indem man direkt die bedingten Mittelwerte der abhängigen Variablen einträgt, also die gruppenspezifischen Mittelwerte der abhängigen Variablen mit der unabhängigen Variablen als Gruppierungsvariablen. Man kann dann noch zusätzlich die durch eine lineare Regression geschätzte Gerade in die grafische Darstellung legen, wie es in Abb. 2.6 gemacht wird.

⁵ Die „jitter“-Funktion von R, die ich hier verwende, zieht für die Verteilung der gejitterten Werte eine Gleichverteilung in einem spezifizierbaren Intervall heran, so dass die Ansammlungen der um einen Wert gruppierten gejitterten Werte in einem Rechteck um diesen auftreten.

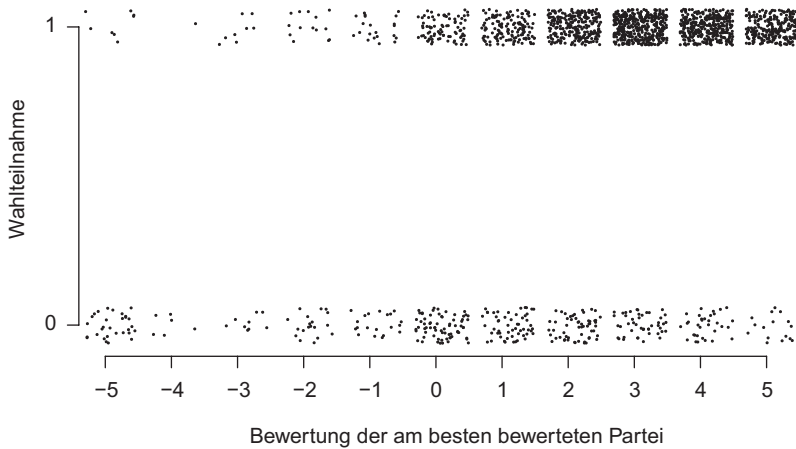


Abb. 2.5 Streudiagramm mit ge jitterten Werten über den Zusammenhang zwischen Wahlteilnahme und Parteienbewertung

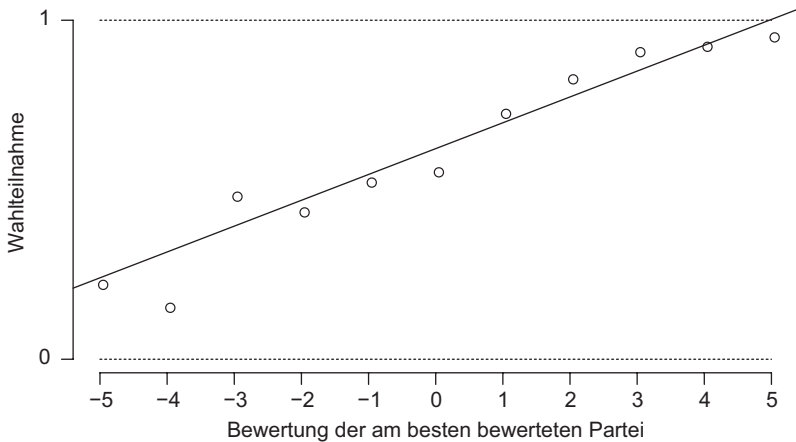


Abb. 2.6 Anteil der Befragten mit Wahlteilnahmeabsicht in Abhängigkeit von Parteienbewertung mit durch lineare Regression geschätzten Regressionslinie

Die konzeptuelle Problematik der Anwendung der linearen Regression lässt sich an einem anderen empirischen Beispiel desselben Datensatzes noch besser illustrieren. Die abhängige Variable ist die Wahl der CDU mit der Zweitstimme, die unabhängige Variable ist die Differenz der Sympathiebewertungen für CDU und SPD. Da es um einen Vergleich der Wahlabsicht zu Gunsten einer der beiden gro-

Tab. 2.3 Kreuztabelle: Zusammenhang zwischen CDU-Wahl und Differenz der Sympathiebewertung von CDU und SPD

| | | Differenz zwischen CDU- und SPD-Sympathiebewertung | | | | | | | | | | | | | | | | | | | | |
|-----|--|--|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| SPD | | 9 | 3 | 18 | 13 | 32 | 46 | 49 | 50 | 65 | 33 | 30 | 8 | 9 | 1 | 3 | 0 | 1 | 0 | 1 | 0 | 0 |
| CDU | | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 3 | 20 | 50 | 93 | 60 | 64 | 52 | 53 | 35 | 23 | 11 | 6 |

Tab. 2.4 Zusammenhang zwischen CDU-Wahl und Differenz der Sympathiebewertung von CDU und SPD, dargestellt in Spaltenprozentzahlen

| | | Differenz zwischen CDU- und SPD-Sympathiebewertung | | | | | | | | | | |
|-----|--|--|-------|-------|-------|-------|-------|------|-------|-------|------|------|
| | | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 |
| SPD | | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 96,1 | 100,0 | 100,0 | 91,7 | 60,0 |
| CDU | | 0 | 0 | 0 | 0 | 0 | 0 | 3,9 | 0 | 0 | 8,3 | 40,0 |

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----|--|------|------|------|------|-------|------|-------|------|-------|-------|
| SPD | | 13,8 | 8,8 | 1,6 | 4,5 | 0,0 | 1,9 | 0,0 | 4,2 | 0,0 | 0,0 |
| CDU | | 86,2 | 91,2 | 98,4 | 95,5 | 100,0 | 98,1 | 100,0 | 95,8 | 100,0 | 100,0 |

ßen Volksparteien geht, gehen in die Analyse nur Fälle ein, die eine Wahlabsicht zu Gunsten einer der beiden Parteien geäußert haben. Der Wert „0“ der abhängigen Variablen bedeutet demnach, dass der Befragte eine Wahlabsicht für die SPD geäußert hat. Wenig überraschend fällt die Wahlabsicht für die CDU gegenüber der für die SPD desto höher aus, je größer die Differenz der Sympathiebewertung der beiden Parteien zu Gunsten der CDU ausfällt (Tab. 2.3, 2.4).

Sowohl das Streudiagramm als auch die Darstellung der durch die Ausprägungen der unabhängigen Variablen bedingten Mittelwerte der abhängigen Variablen zeigen den starken Zusammenhang graphisch anschaulich (Abb. 2.7).

Die Gleichung der Regressionsgerade lautet:

Wahrscheinlichkeit für Wahl der CDU = 0,50 + 0,09 × Differenz

Obwohl die Regressionsgerade ca. 65 % der Varianz der abhängigen Variablen erklärt, ist das konzeptuelle Problem der Analyse auf Anhieb mit dem bloßen Auge erkennbar. Obwohl die Wahrscheinlichkeit auf einen Wertebereich zwischen 0 und 1 beschränkt ist, es sich also um Modelle mit „limited dependent variables“ (Long 1997) handelt, schneidet die Regressionsgerade die obere und untere Begrenzungslinie des Bereichs der zulässigen Werte ungefähr bei - 5,5 und + 5,5. Die mit Hilfe des linearen Wahrscheinlichkeitsmodells geschätzten Werte für die Wahrscheinlichkeit der CDU-Wahl sind größer als 1, wenn die Distanz größer als 5,5 Skalenpunkte auf der Sympathieskala beträgt, und kleiner als 0, wenn die Distanz kleiner

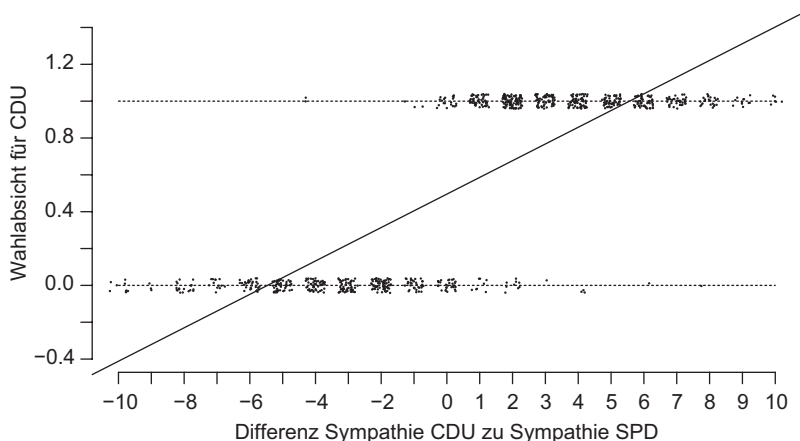


Abb. 2.7 Wahlabsicht für CDU in Abhängigkeit von Differenz der Sympathiebewertung von CDU und SPD

als $-5,5$ Skalenpunkte ist, wenn also die SPD um mindestens 5,5 Punkte besser bewertet wird als die CDU.

Die Idee der Regressionsanalyse besteht darin, den Mittelwert für jede Gruppe, die durch die unabhängige Variable gebildet wird, so gut wie möglich zu schätzen, wobei diese Schätzwerte – im Gegensatz zu einer Varianzanalyse – gleichzeitig dadurch beschränkt sind, dass sie durch eine Funktion in der unabhängigen Variablen abgebildet werden. Auf diese Weise kann die Regressionsanalyse eingesetzt werden, um mit Hilfe der Kenntnis der unabhängigen Variablen eine Prognose hinsichtlich des aufgrund des funktionalen Zusammenhangs erwarteten Wertes der abhängigen Variablen abzugeben. Ist die abhängige Variable beschränkt, d. h. kann sie sich nur innerhalb bestimmter Grenzen bewegen, dann ist eine lineare Funktion für eine Prognose oder Schätzung offensichtlich ungeeignet. Denn wenn man einen nach unten und oben offenen Wertebereich der unabhängigen Variablen annimmt, wenn sich also der Definitionsbereich der Funktion von $-\infty$ bis $+\infty$ erstreckt, dann nimmt auch die abhängige Variable Werte von $-\infty$ bis $+\infty$ an, außer die Steigung der Gerade ist gleich 0. Befinden sich alle geschätzten Werte eines linearen Wahrscheinlichkeitsmodells innerhalb des zulässigen Bereichs, wie es z. B. in Abb. 2.6 der Fall ist, so ist dies lediglich der Beschränkung der realen Werte der unabhängigen Variablen auf einen Bereich geschuldet, in dem auch die Werte der entsprechenden Regressionsgerade nicht die Grenzen von 0 und 1 überschreiten. In diesen Fällen führt zwar auch die lineare Regressionsanalyse zu durchaus brauchbaren Ergebnissen, sobald wir aber den Sättigungsbereich noch innerhalb der Bandbreite von

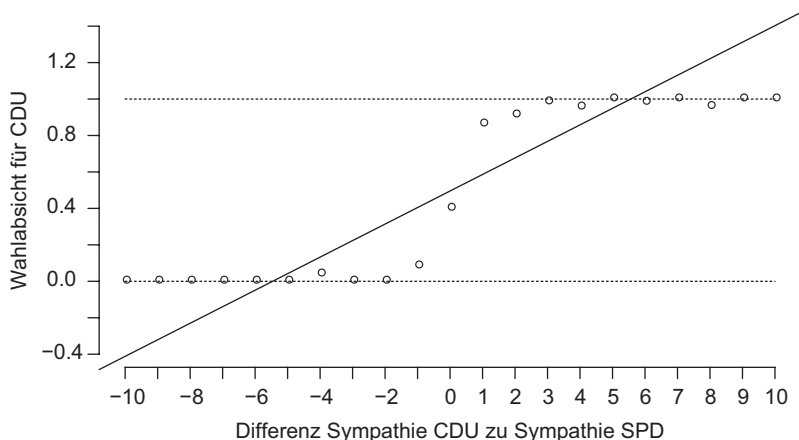


Abb. 2.8 Anteil der Befragten mit Wahlabsicht für CDU in Abhängigkeit von Differenz der Sympathiebewertung von CDU und SPD

realistisch zu erwartenden Werten der unabhängigen Variablen erreichen, stößt die lineare Regression offenkundig an ihre Grenzen und ihr Einsatz zur Schätzung der abhängigen Variablen ist nicht sinnvoll.

Die hervorstechende Eigenschaft einer linearen Funktion besteht darin, dass ihre Steigung konstant ist. Die Steigung einer Funktion aber drückt den Effekt aus, den eine Veränderung der unabhängigen Variablen auf die abhängige Variable ausübt. Im Falle von linearen Funktionen ist dieser Effekt, den X auf Y ausübt, an jeder Stelle immer gleich groß. Dies ist auch der wesentliche Grund für die große Beliebtheit von linearen Modellen, da sie sehr leicht und anschaulich zu interpretieren sind. Wie in Abb. 2.8 jedoch unmittelbar zu erkennen ist, verläuft der Zusammenhang zwischen der Wahlabsicht für die CDU und der Differenz der Sympathiebewertung nicht linear. Einen erkennbaren Effekt übt die Differenz nur im mittleren Bereich zwischen -2 und $+3$ aus. Links und rechts davon tritt eine Art von Sättigungseffekt ein. Bei einem Sympathievorsprung von 3 Punkten der CDU vor der SPD äußern schon annähernd alle Wahlberechtigten eine Wahlabsicht für die CDU, ein noch größerer Sympathievorsprung kann sich daher kaum noch positiv auf die Wahlabsicht für die CDU auswirken. Umgekehrt äußern fast alle Befragten eine Wahlabsicht für die SPD, wenn diese mindestens zwei Punkte besser bewertet wird als die CDU. Dieses Ergebnis ist auch theoretisch durchaus einleuchtend. Mittlere Differenzen zwischen der Sympathie für CDU und SPD, wenn also beide Parteien annähernd gleich gut bewertet werden, determinieren das Wahlverhalten noch nicht eindeutig. Hier können spezifische Faktoren wie die Bewertung der Kandidaten oder die inhaltliche Positionierung der Parteien auf für die Befragten

wichtigen Issues einen Sympathievorsprung aufwiegen, so dass die Wahlabsichten zwischen den Parteien verteilt sind. *Doch gerade weil in diesem mittleren Bereich die Wahlabsicht nicht hundertprozentig durch die Sympathiedifferenz determiniert ist, übt diese hier den größten Effekt aus, d. h. Veränderungen der Sympathiedifferenz bewirken hier die größten Veränderungen hinsichtlich der Verteilung der Wahlabsicht.* Allgemein gilt, dass der Zusammenhang nichtlinearer Art ist, bzw. dass der Effekt, den die unabhängige Variable auf die abhängige Variable ausübt, je nach dem Ort auf der Kurve, den wir betrachten, verschieden ist. Je näher die Kurve ihren Sättigungsgrenzen von 0 und 1 kommt, desto geringer ist der Effekt, am stärksten ist er im Übergangsbereich in der Mitte. Ist also der Y-Wert nahe bei 1, wird eine Änderung von X nur einen geringfügigen Effekt auf Y ausüben. Den stärksten Effekt übt X auf Y an der Stelle der Kurve aus, wenn Y bei 0,5 liegt. Bei einer bivariaten Analyse, wenn nur eine einzige unabhängige Variable X in die Funktion eingeht und somit der Wert von Y ausschließlich durch X erklärt wird, bedeutet dies, dass der Effekt, den X auf Y ausübt, durch den Wert von X an dieser Stelle bestimmt ist.

Die in Abb. 2.8 erkennbare Funktion weist eine spezifische Form von Nichtlinearität auf, die noch weiter beschrieben werden kann. Auch wenn der Effekt, den die unabhängige Variable auf die abhängige ausübt, variiert, so besitzt er doch immer dasselbe Vorzeichen. Die Kurve ist zwar auf einen Bereich zwischen 0 und 1 beschränkt, sie nimmt aber kontinuierlich zu oder ab. Gehen die X-Werte gegen $+\infty$ oder $-\infty$, so konvergiert die Kurve gegen 0 oder 1. Nahe an den Grenzen ist die Steigung der Kurve, die ja den Effekt von X auf Y widerspiegelt, annähernd 0. Formal lassen sich diese Bedingungen folgendermaßen darstellen:

Infobox 2.1: Eigenschaften einer S-förmigen Kurve

Sei $Y = f(X)$

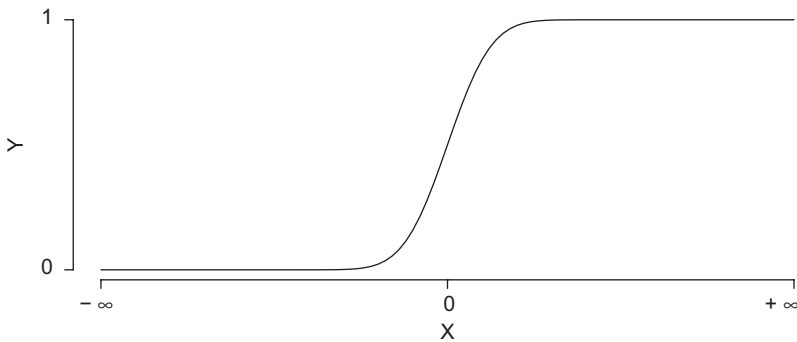
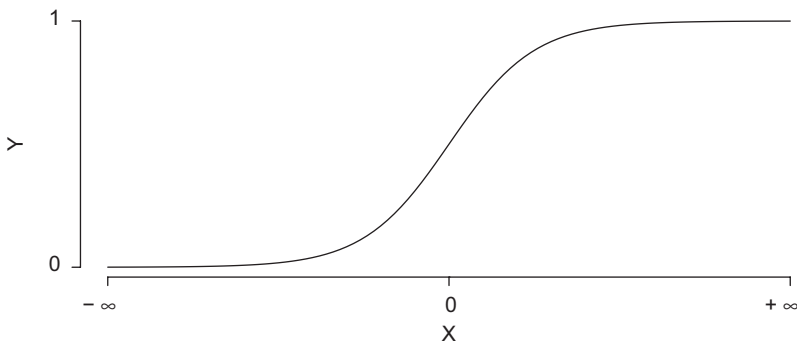
$$f'(X) > 0 \quad \text{und} \quad \lim_{X \rightarrow \infty} f(X) \rightarrow 1 \quad \text{und} \quad \lim_{X \rightarrow -\infty} f(X) \rightarrow 0$$

Oder

$$f'(X) < 0 \quad \text{und} \quad \lim_{X \rightarrow \infty} f(X) \rightarrow 0 \quad \text{und} \quad \lim_{X \rightarrow -\infty} f(X) \rightarrow 1$$

$$\lim_{X \rightarrow \infty} f'(X) \rightarrow 0 \quad \text{und} \quad \lim_{X \rightarrow -\infty} f'(X) \rightarrow 0$$

Es existiert genau ein Wert von X mit: $f''(X) = 0$

**Abb. 2.9** Logitfunktion**Abb. 2.10** Probitfunktion

Die letzte Eigenschaft sagt aus, dass die Funktion nur einen Wendepunkt besitzt. Der Kurventyp, der sich hierdurch ergibt, ist eine sogenannte S-Kurve oder S-förmige (S-shaped) Kurve, da sie entfernt an die Form des Buchstabens „S“ erinnert. Prototypische und auch die bekanntesten Vertreter dieser Kurvenform sind die logistische Funktion und die Probit-Funktion. Die Probit-Funktion ist nichts Anderes als die kumulierte Wahrscheinlichkeitsdichtefunktion bzw. die Verteilungsfunktion der Normalverteilung (Abb. 2.9, 2.10).

Jede S-förmige Kurve ist im rein pragmatischen Sinn offensichtlich besser geeignet, einen Funktionsverlauf wie den in Abb. 2.8 anzunähern und damit geeignete Schätzwerte der abhängigen Variablen auf Grund der Kenntnis des Wertes der unabhängigen Variablen zu liefern, als eine lineare Funktion. Eine lineare Funktion

widerspricht der 2. Bedingung einer korrekten Modellspezifikation, da der untersuchte Zusammenhang in Fällen mit einer begrenzten abhängigen Variablen gar nicht linear sein kann. S-förmige Kurven stellen, wenn die in Kasten 2.1 genannten Eigenschaften für den zwischen unabhängiger und abhängiger Variablen unterstellten Zusammenhang zutreffen, immer eine zumindest gute Annäherung an die Form der Beziehung zwischen unabhängiger und abhängiger Variable dar, selbst dann, wenn ihre funktionale Form der theoretischen Form der Beziehung zwischen unabhängiger und abhängiger Variablen nicht wirklich entsprechen sollte. Allerdings wissen wir dies eigentlich nie bei empirischen Zusammenhängen, also auch nicht bei den ganz normalen Anwendungen linearer Regressionsmodelle. Die einzige Aussage, die wir verlässlich treffen können, ist die, dass wir mit Hilfe des unterstellten statistischen Modells hinreichend gute Schätzungen der abhängigen Variablen auf Basis der Information der unabhängigen Variablen liefern können, dass es gerechtfertigt scheint davon auszugehen, *als ob* der untersuchte Zusammenhang tatsächlich auch die unterstellte Form hätte. Entscheidend ist lediglich, dass die unterstellte Form der Beziehung auch theoretisch Sinn ergibt, sowie dass die auf ihrer Annahme beruhenden Schätzungen nicht auf weiteren Annahmen beruhen, die im klaren Widerspruch zu den gemachten Beobachtungen stehen. Es verbleibt bei der Wahl des angemessenen statistischen Modells daher immer ein nicht unerheblicher Spielraum, der unter pragmatischen Aspekten wie der Einfachheit der Durchführung und der Leichtigkeit und Verständlichkeit der Interpretation der Ergebnisse genutzt werden kann. In diesem Sinne erweisen sich logistische Funktionen als eine besonders attraktive Alternative der Modellwahl bei S-förmigen Kurven und erfreuen sich daher ihrer großen Popularität.

Nur weil die Funktion nicht linear ist, heißt dies nicht zwingend, dass die OLS-Schätzmethode nicht angewandt werden darf. Allerdings sind beim Modell der linearen Wahrscheinlichkeit die hierfür notwendige Bedingungen ebenfalls nicht erfüllt. Dies liegt daran, dass die Werte der abhängigen Variablen lediglich zwei Ausprägungen annehmen. Für welche konkrete Schätzfunktion man sich auch entscheiden mag, der geschätzte Wert für die abhängige Variable ist der einer Wahrscheinlichkeit, während die tatsächlich auftretenden Werte 0 oder 1 sind. Nehmen wir an, der relative Anteil der Y-Werte mit „1“ für einen bestimmten X-Wert sei P und der geschätzte Wert für die Wahrscheinlichkeit von $Y = 1$ für diesen X-Wert sei ebenfalls P , d. h. die Wahrscheinlichkeit bzw. der Mittelwert der 0-1-Werte von Y wird korrekt geschätzt. Dann beträgt der Fehler bezüglich des konkret beobachteten Wertes von Y , der ja entweder 1 oder 0 ist, entweder $1-P$, nämlich genau dann, wenn der beobachtete Wert tatsächlich gleich 1 ist, oder P , was genau dann der Fall ist, wenn der tatsächliche Wert gleich 0 ist. Außerdem tritt der erste Fehler mit einer Wahrscheinlichkeit von P auf, der zweite mit einer von $1-P$. Es gibt also

lediglich zwei konkrete Werte des Fehlers, die jeweils mit einer bestimmten Wahrscheinlichkeit auftreten. Der Fehler ist daher binomialverteilt und nicht normalverteilt. Damit genügt jedes Schätzmodell für eine dichotome Variable nicht der fünften Bedingung, dass die Fehler normalverteilt sein müssen, um das OLS-Verfahren anzuwenden. Des Weiteren gilt für die Varianz der Fehler:

$$\sigma_e^2 = P(1-P)^2 + (1-P)P^2 = P(1-P) \quad \text{GL (2.4)}$$

Die Varianz des Fehlers der Schätzung hängt also vom Wert von P ab. Ist P z. B. 0,5, dann beträgt sie 0,25, ist P hingegen 0,9, beträgt sie 0,09. Da verschiedene P -Werte die Funktionswerte verschiedener Werte der unabhängigen Variablen sein müssen, ist somit auch die vierte Bedingung des Vorliegens von Homoskedastizität verletzt.

Fazit: Die Anwendung des einfachen linearen Regressionsmodells in Form des Modells der linearen Wahrscheinlichkeit auf Untersuchungen, bei denen die Y -Variable dichotomer Natur ist, ist ungeeignet. Dies betrifft sowohl die konkrete Form der geschätzten Funktion als auch das angewandte OLS-Verfahren zur Bestimmung der Koeffizienten der „wahren“ Gleichung. Angemessen ist hingegen die Annahme einer S-förmigen Kurve, wenn der Einfluss, den die unabhängige Variable auf die begrenzte abhängige Variable ausübt, der Form von Abb. 2.8 ist. Die Schätzung dieser S-förmigen Kurve muss dann mit einem anderen Verfahren als OLS durchgeführt werden.



<http://www.springer.com/978-3-658-05081-8>

Logistische Regressionsanalyse

Eine Einführung

Behnke, J.

2015, VII, 131 S., Softcover

ISBN: 978-3-658-05081-8