



Fakultät für Wirtschaftswissenschaften
Institut für Volkswirtschaftslehre (ECON)
Lehrstuhl für Ökonometrie
Prof. Dr. Melanie Schienle

Bachelorthesis

Ökonometrische Analyse der
Behandlungsmethode Recipro[®]
in der Endontologie
im Hinblick auf den Behandlungserfolg

von

Luisa Sophie Werner
Matrikelnummer 1778501
Technisch Volkswirtschaftslehre (Bachelor of Science)

Datum der Abgabe
April 07, 2017

Abstract

Die vorliegende Arbeit widmet sich der zentralen Fragestellung, ob die Anwendung des Systems Reciproc® in der Endontologie zu einer erfolgreicherer Behandlung führt, als die Behandlung mit konventionellen Methoden. Basis der Untersuchung ist ein Datensatz der Akademie für zahnärztliche Fortbildung Karlsruhe (AZFK), der Informationen zu Wurzelkanalbehandlungen im Zeitraum von 1999 bis 2016 betrachtet. Es wird das Verfahren der logistischen Regression angewendet. Dieses eignet sich, um für Patienten die Wahrscheinlichkeiten auf eine erfolgreiche Behandlung zu prognostizieren. Die Arbeit liefert Ergebnisse darüber, welche Patientenmerkmale und auf welche Art und Weise den Behandlungserfolg beeinflussen. Sie kommt zu dem Ergebnis, dass ein jüngeres Patientenalter, der vitale Zustand des Zahnes unmittelbar vor der Behandlung und die Anzahl an medizinischen Zahnreinigungen am behandelten Zahn nach der Behandlung in der Regel zu einer höheren Erfolgswahrscheinlichkeit der Wurzelkanalbehandlung führen. Es wird unter Betrachtung verschiedener Definitionen von Behandlungserfolg herausgestellt, dass der Einsatz von Reciproc® in der Regel erfolgsversprechender ist, als der Einsatz von den vorherigen Methoden (Handinstrumente und BioRaCe®). Zu Beginn werden die angewendeten statistischen Methoden vorgestellt. Im zweiten Teil wird die Vorhergehensweise bei der Modellwahl erläutert und angewendet. Zum Schluss werden die Ergebnisse der Arbeit diskutiert. Daraus ergibt sich insbesondere, dass fehlende Werte zu verfälschten Ergebnissen führen könnten.

Inhaltsverzeichnis

Abkürzungsverzeichnis	iv
Abbildungsverzeichnis	v
Tabellenverzeichnis	vi
1 Einleitung	1
1.1 Medizinischer Hintergrund	1
1.1.1 Recipro [®]	1
1.2 Motivation und Ziel der Untersuchung	2
1.3 Methode	3
1.4 Übersicht über die Ergebnisse	3
2 Methodik	3
2.1 Logistische Regression	3
2.1.1 Odds und Wahrscheinlichkeiten	3
2.1.2 Herleitung und Beschreibung der logistischen Regression	4
2.1.3 Vorteile der logistischen Regression	7
2.1.4 Maximum-Likelihood-Schätzung und numerische Optimierung	8
2.1.5 Interpretation der Koeffizienten	10
2.2 Verfahren zur Modellwahl und Modellbewertung	11
2.2.1 Goodness-of-Fit Maße: Pseudo - R^2	11
2.2.2 Likelihood-Ratio-Test	13
2.2.3 Korrektklassifikationsrate	13
2.2.4 Receiver Operating Characteristic Kurve	15
2.2.5 Wald-Konfidenzintervalle und Wald-Tests	17
2.2.6 Akaike Informationskriterium und Abwärtsverfahren	18
2.3 Zwei-Stichproben Kolmogorov-Smirnov-Test	20
3 Gegenstand der Untersuchung	21
3.1 Beschreibung des Datenmaterials	21
3.1.1 Einteilung in drei Behandlungsgruppen	21
3.1.2 Definition des Behandlungserfolgs	21
3.1.3 Überblick über die Kovariaten im Datensatz	23

4	Vorstellung der Ergebnisse	25
4.1	Anforderungen an den Datensatz und das Modell	25
4.2	Vorgehensweise	27
4.3	Modellwahl	29
4.4	Likelihood-Ratio-Tests	30
4.5	Pseudo- R^2	31
4.6	Einflussnehmende Kovariaten	31
4.7	Stärke und Richtung der einflussnehmenden Kovariaten	32
4.8	Vergleich der Behandlungsgruppen	35
4.8.1	Erfolgswahrscheinlichkeiten im Zeitverlauf	35
4.8.2	Zwei-Stichproben Kolmogorov-Smirnov-Tests	36
5	Diskussion der Ergebnisse	38
6	Zusammenfassung der Ergebnisse	39
	Literatur	40

Abkürzungsverzeichnis

AIC	Akaike Information Criterion (Akaike Informationskriterium)
AUC	Area Under the Curve (Fläche unter der ROC-Kurve)
AZFK	Akademie für zahnärztliche Fortbildung Karlsruhe
bzw	beziehungsweise
d.h.	das heißt
FNR	False Negative Rate
FPR	False Positive Rate
GOF	Goodness of Fit (Güte der Anpassung)
K-L Information	Kullback-Leibler-Information
KS-Test	Kolmogorov-Smirnov-Test
L	Likelihood
LL	Loglikelihood
LR-Test	Likelihood-Ratio-Test
ML	Maximum-Likelihood
MLE	Maximum-Likelihood-Estimation
NiTi	Nickel-Titan Legierung
OLS	Ordinary Least Squares
ROC	Receiver Operating Characteristic
TPR	True Positive Rate
TNR	True Negative Rate
VWF	Vertikale Wurzelfraktion
WK	Wahrscheinlichkeit
WKB	Wurzelkanalbehandlung
WSR	Wurzelspitzenresektion
YI	Youden Index
z.B.	zum Beispiel

Abbildungsverzeichnis

1	Logistische Regressionskurve , rot: beobachtete Ereignisse y_i , blau: angepasste logistische Regressionskurve $\pi(X)$, schwarze Punkte: vorhergesagte Eintrittswahrscheinlichkeiten $\hat{\pi}(X)$	7
2	ROC-Kurve am Beispiel des Modells zu Definition Erfolg3.	15
3	Korrelationsmatrix für den Datensatz,(Korrelationskoeffizient nach Pearson, blau: positiv linearer Zusammenhang, rot: negativ linearer Zusammenhang).	27
4	Kriterien bei der Modellwahl am Beispiel von Definition Erfolg3: AIC, AUC, Accuracy und ROC-Kurve des Modells mit der größten AUC: erfolg4 ~Alter + MedReinigungen + reci + vital.	28
5	Vorhergesagte Erfolgswahrscheinlichkeiten im Zeitverlauf , abhängig vom Termin der WKB, unterteilt in Behandlungen mit (rot) und ohne (blau) Recipro [®] am Beispiel von Definition Erfolg5.	35
6	Ergebnisse der KS-Tests der Modelle für die Erfolgsdefinitionen 4 (links) und 2 (rechts).	36

Tabellenverzeichnis

1	Aufbau einer Klassifikationstabelle , Tutz (2011).	14
2	Die Klassifikationstabelle der Definition Erfolg3.	14
3	Orientierungswerte für das Maß AUC	16
4	Übersicht über die Erfolgsdefinitionen 1-3	22
5	Übersicht über Erfolgsdefinitionen 4-6	23
6	Übersicht über Erfolgsdefinitionen 7-9	23
7	Kreuztabellen für Erfolgsdefinition 9 mit den Variablen molar, praemolar, Alter und MedReinigungen.	25
8	Kreuztabelle für die Kovariaten vital, MedReinigungen und Alter für Definition Erfolg2.	26
9	Pseudo-R² nach McFadden, Nagelkerke und Cox&Snell für die jeweiligen Erfolgsdefinitionen.	29
10	Übersicht über die Regressoren in den jeweiligen Modellen für die Erfolgsdefinitionen 1-9.	30
11	AUC und Korrektklassifikationsrate (Accuracy) der Modelle für die Erfolgsdefinitionen 1-9.	30
12	Ergebnisse der LR-Tests , (Signifikanzcodes: 0 '****', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ' 1).	31
13	Die geschätzten Koeffizienten $\hat{\beta}_{vital}$ und die Effektkoeffizienten $e^{\hat{\beta}_{vital}}$ für die Erfolgsdefinitionen 1-8.	32
14	Die geschätzten Koeffizienten $\hat{\beta}_{Alter}$ und die Effektkoeffizienten $e^{\hat{\beta}_{Alter}}$ für die Erfolgsdefinitionen 1, 2, 3, 6, 7, 8, 9.	32
15	Die geschätzten Koeffizienten $\hat{\beta}_{MedReinigungen}$ und die Effektkoeffizienten $e^{\hat{\beta}_{MedReinigung}}$ für die Erfolgsdefinitionen 1, 2, 3, 7, 8, 9.	32
16	Die geschätzten Koeffizienten $\hat{\beta}_{Erfahrung}$ und die Effektkoeffizienten $e^{\hat{\beta}_{Erfahrung}}$ für die Erfolgsdefinitionen 1, 4, 5, 6, 7, 8.	33
17	Die geschätzten Koeffizienten $\hat{\beta}_{VorherigeWRs}$ und die Effektkoeffizienten $e^{\hat{\beta}_{VorherigeWRs}}$ für die Erfolgsdefinitionen 4-8.	33
18	Die geschätzten Koeffizienten $\hat{\beta}_{VorherigeWRs}$ und die Effektkoeffizienten $e^{\hat{\beta}_{VorherigeWRs}}$ für Erfolgsdefinition 1.	33
19	Die geschätzten Koeffizienten $\hat{\beta}_{canini}$ und die Effektkoeffizienten $e^{\hat{\beta}_{canini}}$ für die Erfolgsdefinitionen 6-8.	33

20	Die geschätzten Koeffizienten $\hat{\beta}_{molar}$ und die Effektkoeffizienten $e^{\hat{\beta}_{molar}}$ für die Erfolgsdefinitionen 5, 6, 9.	34
21	Die geschätzten Koeffizienten $\hat{\beta}_{praemolar}$ und die Effektkoeffizienten $e^{\hat{\beta}_{praemolar}}$ für die Erfolgsdefinitionen 5, 6, 9.	34
22	Übersicht über die P-Werte der KS-Tests für die jeweiligen Modelle der Erfolgsdefinitionen 1-9.	37

1 Einleitung

1.1 Medizinischer Hintergrund

Hat ein Zahn irreversible pathologische Schäden des Markorgans (Pulpa) erlitten, ist eine Wurzelkanalbehandlung (WKB) notwendig. Ziel der Behandlung ist die Entfernung von pulpalen Geweberesten aus dem Wurzelkanalsystem. Im Anschluss wird der entstandene Hohlraum mit plastischem Material bakteriendicht gefüllt, Schreiber (2015). In den letzten Jahren gab es Innovationen, um die WKB zu vereinfachen und bessere Behandlungsergebnisse zu erzielen. In der Vergangenheit erfolgte die Behandlung oft mühsam durch die Anwendung von Handinstrumenten, sogenannten Wurzelkanalinstrumenten. In der Regel wurden Wurzelkanäle dabei mit einer Sequenz aus mehreren aufeinanderfolgenden Wurzelkanalinstrumenten erweitert, Schreiber (2015). Die Wurzelkanaleingänge wurden initial mit Gates-Bohrern und die Wurzelkanäle selbst anschließend mit Stahl-Hand-Wurzelkanalinstrumenten erweitert, um sie für die Desinfektion zugänglich zu machen. Die Bestimmung der Aufbereitungslänge der Wurzelkanäle wurde mit Röntgenbildern durchgeführt. Zur Desinfektion wurden die Wurzelkanäle ausschließlich mit Natriumhypochlorit (NaOCl) gespült. Auf eine Isolierung des Zahnes mit Kofferdam als Infektionsschutz wurde in der Mehrzahl der Fälle verzichtet. Die Einführung maschineller, vollrotierender Wurzelkanalpräparationsmethoden mit dem BioRaCe[®]-System trug zur Modernisierung bei. Im Zuge der Kanaleingangserweiterung wurden weiterhin Gates-Bohrer eingesetzt. Im Anschluss wurde die Instrumentensequenz der maschinell betriebenen BioRaCe[®]-Instrumente verwendet. Ein drehmomentgesteuerter Motor wurde für die Anwendung der vollrotierenden Instrumente genutzt. Außerdem wurde konsequent und während der kompletten Behandlung Kofferdam angewendet. Die Aufbereitungslänge wurde vorwiegend elektrometrisch mit Apexlokatoren gemessen. Röntgenbilder kamen nur konfirmatorisch zum Einsatz. Zudem wurde das Desinfektionsprotokoll um die ultraschallgestützte Spülung erweitert. Einen Meilenstein in der Endodontologie stellt die Entwicklung des Reciproc[®]-Systems durch die Firma VDW (VDW GmbH, München, Deutschland) dar. Seit November 2011 ist die neue Technologie in der Akademie für Zahnärztliche Fortbildung (AZFK) in Karlsruhe im Einsatz, Bartols (2013).

1.1.1 Reciproc[®]

Beim Reciproc[®]-System handelt es sich um ein Ein-Feilen-System. Das System besteht aus drei Instrumenten, R25, R40 und R50, die nach der Größe des Wurzelkanals ausgewählt werden. Das ausgewählte Instrument wird dann als einzelnes für die Präparation des gesamten Kanals verwendet. Die Feilen haben einen S-förmigen Querschnitt mit nicht-schneidender Spitze, Bartols (2013). Die

Technologie ist nach der reziproken Arbeitsbewegung benannt, in die die Instrumente durch einen neu entwickelten Motor versetzt werden. Die Feile schneidet in einem 150° Rotationswinkel gegen den Uhrzeigersinn und führt zur Entlastung eine Bewegung von 30° in die Gegenrichtung durch. Die Feilen bestehen aus einer weiterentwickelten Nickel-Titan Legierung (NiTi) und sind als Einmal-Artikel ausgewiesen. Das bedeutet, dass sie für einen Behandlungsfall geeignet sind und anschließend verworfen werden, Schreiber (2015). Reciproc[®] weist folgende Vorteile auf: Die Aufbereitungszeit wird durch die Instrumente bei gleichwertiger Aufbereitungsqualität deutlich verringert. Reciproc[®] zeichnet sich somit durch eine erhöhte Effizienz aus. Außerdem kann auf eine Gleitpfadpräparation bei geraden oder moderat gekrümmten Wurzelkanälen in vielen Fällen verzichtet werden. Des Weiteren weisen die weiterentwickelte NiTi Legierungen des Reciproc[®]-Systems eine höhere Flexibilität gegenüber Standard NiTi-Legierungen auf. Dadurch wird der Widerstand gegenüber Ermüdungsbrüchen erhöht. Durch die Anwendung als Einmal-Instrumente wird die potentielle Gefahr der Kreuzkontamination durch unzureichende Sterilisationsprozesse vermieden. Außerdem erleichtert das Ein-Feilen-System die Organisation in der Praxis oder Klinik, Schreiber (2015). Laut Bartols (2013) ist die Methode auch in komplizierten Situationen einsetzbar. Es ergaben sich dennoch Bedenken gegenüber Reciproc[®]. Bei In-Vitro Versuchen wurden intraradikulären Mikrofrakturen in den präparierten Wurzeln beobachtet. Insgesamt ist die Studienlage inhomogen und die klinischen Konsequenzen der Beobachtung nicht schlüssig aus dem in-vitro Modell in die Realität übertragbar. Im ungünstigen Fall wird angenommen, dass Mikrofrakturen der Zahnschubstanz zu vertikalen Wurzelfrakturen (VWF) führen, die keine Erhaltungsoption des Zahnes mehr zulassen. In solchen Fällen muss der betroffene Zahn entfernt werden, Schreiber (2015).

1.2 Motivation und Ziel der Untersuchung

Diese Arbeit soll mit statistischen Methoden auf Basis von Patientendaten der AZFK untersuchen, ob Reciproc[®] für den Patienten eine bessere Aussicht auf eine erfolgreiche WKB bietet, als die bisher angewendeten Methoden. Des Weiteren wird untersucht, welche Variablen für eine erfolgreiche WKB ausschlaggebend sind, und ob sich diese positiv oder negativ auf den Behandlungserfolg auswirken. Da bei diesen Untersuchungen der Behandlungserfolg betrachtet wird, stellt sich die Frage, wie man diesen überhaupt definieren kann. In Kapitel 3.1.2 werden verschiedene Definitionen vorgestellt. Diese beziehen sich auf Nachbehandlungen (Revisionen) und Extraktionen in verschiedenen Zeitintervallen nach der WKB. Es sollen also zunächst Modelle und Variablen gefunden werden, die den Behandlungserfolg erklären. Im Anschluss daran werden die Patientengruppen, bei denen Reciproc[®] angewendet bzw. nicht angewendet wurde, hinsichtlich ihrer Aussicht auf Behandlungs-

erfolg verglichen.

1.3 Methode

Um Behandlungserfolg erklären zu können, ist eine binäre Variable Y notwendig. Diese gibt für jeden Patienten an, ob die Behandlung erfolgreich ($Y=1$) oder nicht erfolgreich ($Y=0$) durchgeführt wurde. Die logistische Regression ermöglicht es, abhängig von Variablen, wie z.B. dem Alter des Patienten, Wahrscheinlichkeiten für das Eintreten von Erfolg zu prognostizieren. In Kapitel 2.1 wird das Verfahren im Detail beschrieben. Außerdem ist von Interesse, welche Variablen den erfolgreichen Ausgang der Behandlung beeinflussen. Um diese Variablen zu bestimmen, wird in erster Linie die Receiver Operating Characteristic (ROC) betrachtet, welche in Kapitel 2.2.4 vorgestellt wird. Um die Patientengruppen zu vergleichen, wird ein Zwei-Stichproben Kolmogorov-Smirnov-Test durchgeführt. Sind die Erfolgswahrscheinlichkeiten bei Reciproc[®]-Behandlungen auf höheren Werten konzentriert, kann behauptet werden, dass die Methode generell erfolgsversprechender als die vorher angewendeten Behandlungsmethoden ist. Näheres dazu im Kapitel 2.3.

1.4 Übersicht über die Ergebnisse

Zusammengefasst ergeben die Untersuchungen folgende Ergebnisse: Die wichtigsten Variablen, die den Behandlungserfolg beeinflussen, sind die Vitalität des Zahnes, das Alter des Patienten zum Behandlungszeitpunkt, sowie die Anzahl der nach der Behandlung durchgeführten Zahnreinigungen am behandelten Zahn. Die Behandlung an einem vitalen Zahn, ein jüngeres Patientenalter, sowie die Durchführung von Zahnreinigungen, erhöhen die Aussicht auf eine erfolgreiche Behandlung. Der Vergleich der Patienten, bei denen Reciproc[®] angewendet bzw. nicht angewendet wurde, ergibt, dass Reciproc[®] in der Regel zu höheren Erfolgswahrscheinlichkeiten führt. In Kapitel 4 werden die Ergebnisse im Detail vorgestellt.

2 Methodik

2.1 Logistische Regression

2.1.1 Odds und Wahrscheinlichkeiten

Für das Verständnis des logistischen Regressionmodells ist die Betrachtung von Odds in Abgrenzung zu Wahrscheinlichkeiten notwendig. Eine Wahrscheinlichkeit betrachtet die Häufigkeit eines betrachteten Ereignisses im Verhältnis zur Anzahl der möglichen Ausgänge. Odds hingegen beschreiben die Chance für das Eintreten eines Ereignisses und stellen das Verhältnis aus Wahrscheinlichkeit für

das Eintreten und Wahrscheinlichkeit für das Nichteintreten eines Ereignisses bzw. das Eintreten des Komplementäreignisses dar, Peng et al. (2002). Kodiert man das Eintreten eines Ereignisses mit einer binären Variable Y die beim Eintreten den Wert 1 und beim Nichteintreten den Wert 0 annimmt, ergeben sich die Odds zu

$$Odds(\pi) = \frac{\pi}{1 - \pi} = \frac{P(Y = 1)}{1 - P(Y = 1)}. \quad (2.1)$$

π stellt die Wahrscheinlichkeit $P(Y = 1)$ für das Eintreten von $Y = 1$ dar, Behnke (2015). Der Wertebereich der Odds liegt zwischen 0 und $+\infty$. Sind die Odds > 1 , ist das Eintreten von $Y = 1$ wahrscheinlicher, als das Nichteintreten. Gilt Odds < 1 , ist es umgekehrt. Sind die Odds gleich 1, sind beide Ausgänge gleich wahrscheinlich mit der Wahrscheinlichkeit 0.5, Urban and Mayerl (2006). Möchte man den Einfluss einer Variablen X auf die Eintrittschancen betrachten, wird das Odds-Ratio, das Verhältnis der Gewinnchancen, herangezogen. \hat{P} weist darauf hin, dass die Wahrscheinlichkeit geschätzt wird. Auf die Notation wird in Kapitel 2.1.2 genauer eingegangen. Es ergibt sich für zwei konkrete Werte $x_0 = m$ und $x_1 = x_0 + 1 = m + 1$, die X annehmen kann, mit einem beliebigen Wert m , wenn X von m auf $m + 1$ erhöht wird, Urban and Mayerl (2006).

$$\frac{\frac{\hat{P}_i(Y|X=x_1)}{1 - \hat{P}_i(Y|X=x_1)}}{\frac{\hat{P}_i(Y|X=x_0)}{1 - \hat{P}_i(Y|X=x_0)}} = \frac{Odds(\pi(x_1))}{Odds(\pi(x_0))}. \quad (2.2)$$

2.1.2 Herleitung und Beschreibung der logistischen Regression

Das Ziel der Regression im Allgemeinen besteht darin, den Zusammenhang zwischen einer abhängigen Variable Y und einer oder mehreren unabhängigen Variablen X durch ein Modell zu beschreiben. Die unabhängigen Variablen werden auch als erklärende Variablen, Kovariaten oder Prädiktoren bezeichnet, Hosmer and Lemeshow (2000). Das Verfahren der logistischen Regression eignet sich für binäre, dichotome abhängige Variablen Y . Dichotom beschreibt ein Merkmal, das nur zwei Ausprägungen haben kann, Dreger (2014). Binär bedeutet, dass die Variable nur die Ausprägungen 0 und 1 annehmen kann, Behnke (2015). Zur Vereinfachung wird die logistische Regression am bivariaten Modell mit einem Prädiktor vorgestellt. Das Modell sagt Wahrscheinlichkeiten für das Eintreten eines Ereignisses $Y = 1$ voraus. Das Nichteintreten wird mit $Y = 0$ beschrieben. Die Eintrittswahrscheinlichkeiten werden mithilfe der sogenannten Responsefunktion dargestellt. Diese Funktion liefert den bedingten Mittelwert der Variable Y abhängig von einer bestimmten Linearkombination der unabhängigen Variablen X . Da dieser Erwartungswert ein Anteilswert, bzw. eine theoretische Wahrscheinlichkeit ist, wird er mit π bezeichnet. π ist gleichzusetzen mit $P(Y = 1|X = x)$ oder kurz $P(Y = 1)$, Behnke (2015). Der kleine Buchstabe x steht dabei für einen konkreten, realisierten Wert.

$$E[Y] = \pi = h(\beta_0 + \beta_1 X). \quad (2.3)$$

Die in der Formel enthaltene Funktion h hat den Zweck, den linearen Term $\beta_0 + \beta_1 X$ in eine Wahrscheinlichkeit zu transformieren. Um diese Funktion als lineare Funktion abhängig von X zu erhalten, muss die Umkehrfunktion der Responsefunktion (2.3) gebildet werden. Diese Funktion g wird als Link-Funktion bezeichnet, Behnke (2015).

$$g(\pi) := h^{-1}(\pi) = \beta_0 + \beta_1 X. \quad (2.4)$$

Der Schätzwert $\hat{\pi}$ für π ergibt sich durch die Schätzung der Werte für die Parameter β_0 und β_1 . Da es sich bei π um eine Wahrscheinlichkeit handelt, sollte dessen Wertebereich zwischen 0 und 1 liegen. π sollte in diesem Bereich alle möglichen Werte annehmen können. Um das zu erreichen, werden anstatt der Wahrscheinlichkeiten die $Odds(\pi)$ betrachtet, siehe Kapitel 2.1.1. Damit sich die geschätzten Wahrscheinlichkeiten im Intervall $(0, 1)$ befinden, muss auf die Odds eine Funktion angewendet werden, die den Bereich $(0, \infty)$ auf $(-\infty, \infty)$ abbildet. Eine solche Funktion ist der Logarithmus. Der Logarithmus der Odds wird auch als Logit bezeichnet. Für die Funktion g ergibt sich

$$g(\pi) = \text{Logit}(\pi) = \ln(odds(\pi)) = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X. \quad (2.5)$$

Um π zu erhalten ergibt sich

$$\pi = \text{Logit}^{-1}(\beta_0 + \beta_1 X) = h(\beta_0 + \beta_1 X). \quad (2.6)$$

und in Abhängigkeit von X wird π folgendermaßen definiert

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}. \quad (2.7)$$

Der Term $\beta_0 + \beta_1 X$ wird dabei als linearer Prädiktor w bezeichnet. Die Funktion h ist also die Umkehrfunktion des Logits. Aus der Formel ist nun ersichtlich, dass π wirklich im Wertebereich $(0, 1)$ liegt. Zähler und Nenner sind beide aufgrund der Exponentialfunktion positiv. Da der Nenner immer um 1 größer ist als der Zähler, konvergiert die Funktion $\pi(X)$ gegen 1, wenn X gegen unendlich läuft, Behnke (2015). Formt man die Gleichung um, erhält man die $Odds(\pi)$ als Exponentialfunktion des linearen Prädiktors. Es gilt

$$e^{|\ln(odds(X))|} = odds(\pi) = \frac{\pi}{1-\pi} = e^{\beta_0 + \beta_1 X}, \quad (2.8)$$

Behnke (2015). Für die Wahrscheinlichkeiten des Eintreten eines Ereignisses und dessen Nichteintretens gilt nun Folgendes:

$$P(Y = 1) = \pi(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}, \quad (2.9)$$

$$P(Y = 0) = 1 - \pi(X) = \frac{1}{1 + e^{\beta_0 + \beta_1 X}}. \quad (2.10)$$

Im bisherigen Teil wurde die logistische Regression für nur eine Kovariante betrachtet. Üblicherweise enthält ein Modell mehrere, bzw. K Kovariaten (multivariater Fall). Dann gilt für den linearen Prädiktor im Allgemeinen

$$w = \beta_0 + \beta_1 X_1 + \dots + \beta_j X_j + \dots + \beta_K X_K, \quad (2.11)$$

und in Vektorschreibweise

$$w = \mathbf{X}\beta.$$

Für einen individuellen Fall gilt dann

$$w_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_K x_{iK}. \quad (2.12)$$

Und auch hier gilt in Vektorschreibweise

$$w_i = \mathbf{X}_i \beta.$$

\mathbf{X} und β haben dabei mit N als Stichprobenumfang und $i \in \{1, 2, \dots, N\}$ und $j \in \{1, 2, \dots, K\}$ folgende Gestalt:

$$\mathbf{X} = \begin{pmatrix} x_{10} & x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1K} \\ x_{20} & x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2K} \\ \vdots & & & & & & \\ x_{i0} & x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{iK} \\ \vdots & & & & & & \\ x_{N0} & x_{N1} & x_{N2} & \dots & x_{Nj} & \dots & x_{NK} \end{pmatrix} \quad \text{und} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_K \end{pmatrix}. \quad (2.13)$$

Die Zeilen der Matrix \mathbf{X} stehen dabei für einzelne Beobachtungen und die Spalten für die Werte der jeweiligen Kovariaten. Gleichung 2.7 ändert sich bei K unabhängigen Variablen zu

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_K X_K}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_K X_K}}. \quad (2.14)$$

Aus Gleichung 2.7 wird die logistische Regressionskurve in Abbildung 1 hergeleitet. Der Graph der Funktion $h(X)$ ist S-förmig und besitzt folgende Eigenschaften: Ist die Funktion monoton steigend, konvergiert sie für $X \rightarrow \infty$ gegen 1 und für $X \rightarrow -\infty$ gegen 0. Es gilt also

$$h'(X) > 0 \quad \text{und} \quad \lim_{X \rightarrow \infty} h(X) \rightarrow 1 \quad \text{und} \quad \lim_{X \rightarrow -\infty} h(X) \rightarrow 0. \quad (2.15)$$

Ist die Funktion fallend, ist der Zusammenhang genau umgekehrt und es gilt

$$h'(X) < 0 \quad \text{und} \quad \lim_{X \rightarrow \infty} h(X) \rightarrow 0 \quad \text{und} \quad \lim_{X \rightarrow -\infty} h(X) \rightarrow 1. \quad (2.16)$$

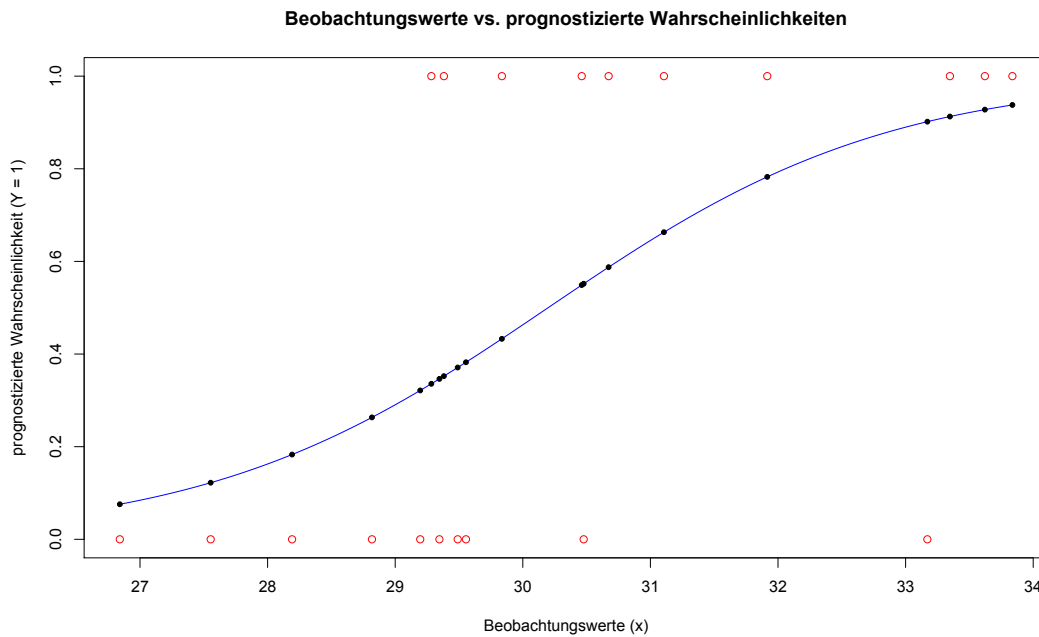


Abbildung 1: Logistische Regressionskurve, rot: beobachtete Ereignisse y_i , blau: angepasste logistische Regressionskurve $\pi(X)$, schwarze Punkte: vorhergesagte Eintrittswahrscheinlichkeiten $\hat{\pi}(X)$.

Außerdem konvergiert die Ableitung $h'(X)$ gegen 0, wenn X an die Grenzen des Definitionsbereich kommt. Dort ist die Änderungsrate geringer als in der Mitte des Definitionsbereichs von X . Es gilt $\lim_{X \rightarrow \infty} h'(X) \rightarrow 0$ und $\lim_{X \rightarrow -\infty} h'(X) \rightarrow 0$. Des Weiteren existiert genau ein Wert X mit $h''(X) = 0$. An dieser Stelle besitzt die Funktion einen Wendepunkt, Behnke (2015).

2.1.3 Vorteile der logistischen Regression

Das Grundkonzept, auf dem die logistische Regression basiert, ist die lineare Regression. Viele Vorteile der logistischen Regression ergeben sich aus den Schwächen der linearen Regression, Behnke (2015). Deshalb ist es sinnvoll, die lineare Regression zunächst kurz vorzustellen. Bei der linearen Regression wird eine abhängige Variable Y durch eine unabhängige Variablen X erklärt. Im bivariaten Fall lautet das theoretische Modell

$$Y = \beta_0 + \beta_1 X. \quad (2.17)$$

Für einen bestimmten Fall $i \in \{1, \dots, N\}$ gilt

$$y_i = \beta_0 + \beta_1 x_i. \quad (2.18)$$

Das theoretische Modell unterscheidet sich von den beobachteten Werten durch einen Fehlerterm e . Im Fall einer bestimmten Beobachtung ergibt sich mit diesem Fehlerterm e_i

$$y_i = \beta_0 + \beta_1 x_i + e_i. \quad (2.19)$$

Um ein bestmögliches Modell zu finden, werden die Fehler e_i für alle N Beobachtungen quadriert aufsummiert. Durch Minimierung dieses Terms wird eine Regressionsgerade gefunden, die den Zusammenhang zwischen X und Y bestmöglich beschreibt. Um ein lineares Modell zu schätzen, sollten einige Bedingungen erfüllt sein, Behnke (2015). Unter Anderem sollte der Fehlerterm normalverteilt mit Erwartungswert 0 und konstanter Varianz σ^2 sein. Es sollte also $e \sim \mathcal{N}(0, \sigma^2)$ gelten. Bei dichotomen Variablen Y ist das nicht der Fall. Da e binomialverteilt ist und sich die Varianz als $\sigma_i^2 = \pi_i(1 - \pi_i)$ errechnet, ist die Annahme konstanter Varianzen für alle Beobachtungen i mit $i \in \{1, \dots, N\}$ nicht erfüllt. Ein weiteres Problem ergibt sich beim Verlauf der Funktion. Aufgrund der Linearität ist der Effekt, den X auf Y ausübt, bei der linearen Regression an jeder Stelle gleich groß. Ist Y dichotom, sind die Effekte für Werte an den Grenzen des Definitionsbereichs von X oft geringer als in dessen Mitte. Durch den S-förmigen Verlauf des Graphen der logistischen Regression wird darauf Rücksicht genommen. Technisch gesehen, kann das Verfahren der linearen Regression auf dichotome abhängige Variablen angewendet werden. Es ist aber meist ungeeignet, um zuverlässige Ergebnisse zu liefern, Behnke (2015).

2.1.4 Maximum-Likelihood-Schätzung und numerische Optimierung

Ein geeignetes Verfahren, um den Koeffizientenvektor $\hat{\beta}$ für β aus 2.13 zu schätzen und damit die Wahrscheinlichkeiten $\hat{\pi}(X)$ für $\pi(X)$ bestimmen zu können, ist das Maximum-Likelihood-Schätzverfahren (MLE). Dieses maximiert die Wahrscheinlichkeit für das Auftreten einer beobachteten Wertekombination der Variablen. Es liefert konsistente und asymptotisch normalverteilte Punktschätzer $\hat{\beta}$ für β , Hastie et al. (2001). Die Likelihood-Funktion (L) gibt die Wahrscheinlichkeit an, mit welcher die empirische Verteilung der Y -Werte richtig geschätzt wird. N ist der Stichprobenumfang, M die Anzahl der Fälle, bei denen $Y = 1$ beobachtet wird und π die Wahrscheinlichkeit $P(Y = 1)$, siehe Formel 2.9 in Kapitel 2.1.2 .

$$L = \prod_{i=1}^N \pi^{y_i} (1 - \pi)^{1-y_i} = \pi^M (1 - \pi)^{(N-M)}. \quad (2.20)$$

Damit die Formel besser handzuhaben ist und einfacher differenziert werden kann, wird sie logarithmiert. Die Logarithmierung ist eine zulässige, monotone Transformation. Es ergibt sich die Loglikelihood-Funktion (LL), Hastie et al. (2001),

$$LL = \ln(L) = \sum_{i=1|Y=1}^N \ln(\pi_i) + \sum_{i=1|Y=0}^N \ln(1 - \pi_i). \quad (2.21)$$

Die Loglikelihood-Funktion bildet den Wertebereich $(0, \infty)$ ab. Durch die Multiplikation mit -2 wird die Funktion positiv und ist χ^2 -verteilt. Diese Umformung liefert die Devianz. Diese entspricht der Varianz im linearen Modell. Steigt die Devianz, ist das Modell schlechter an die Daten angepasst, Behnke (2015). Um ein bestmöglichst angepasstes Modell zu finden, muss die Devianz minimiert werden. Dadurch wird ein $\hat{\pi}$ bestimmt, bei dem alle beobachteten Fälle mit höchster Wahrscheinlichkeit richtig klassifiziert werden, Behnke (2015). Dass diese Maximierung wirklich Wahrscheinlichkeiten liefert, beweist Behnke (2015) in Kapitel 5, Seite 45. Das Maximierungsverfahren führt zu nichtlinearen Gleichungen, Hastie et al. (2001). Aus diesem Grund wird der Newton-Raphson Algorithmus angewendet, um iterativ ein Optimum zu finden. Bei iterativen Verfahren wird zuerst ein Vektor mit Anfangswerten ausgewählt. Dieser wird dann schrittweise angepasst, sodass sich der Wert der Likelihood bei jedem Schritt erhöht. Wird kein ausreichender Anstieg verzeichnet, bricht das Verfahren ab. Numerische Verfahren zur Optimierung haben im Allgemeinen die Form

$$\theta_{t+1} = \theta_t + \lambda_t W_t g_t. \quad (2.22)$$

λ bestimmt die Schrittweite, $W_t g_t$ den Richtungsvektor und θ_t bzw. θ_{t+1} den betrachteten Parameter in Schritt t bzw. $t+1$, Hastie et al. (2001). Möchte man mithilfe des Newton-Raphson Algorithmus einen Parametervektor $\hat{\beta}$ schätzen, muss zunächst der Gradient und die Hessematrix aufgestellt werden. Der Gradient ist ein Vektor, der die ersten Ableitungen in Richtung jeder Parameters enthält. Er hat die Form

$$S = S(\hat{\beta}) = \frac{\partial LL(\hat{\beta})}{\partial \hat{\beta}}, \quad (2.23)$$

mit der Loglikelihood-Funktion LL, siehe 2.21. Die Hessematrix entspricht der zweiten Ableitung und enthält auf der Diagonalen die zweiten Ableitungen und neben der Diagonalen die Kreuzableitungen. Die Hessematrix hat die Gestalt

$$H = H(\hat{\beta}) = \frac{\partial^2 LL(\hat{\beta})}{\partial \hat{\beta}_i \partial \hat{\beta}_j}. \quad (2.24)$$

Der Newton-Raphson Algorithmus optimiert die Devianz schrittweise, indem $\hat{\beta}$ mit folgender Formel bestimmt wird:

$$\hat{\beta}_{n+1} = \hat{\beta}_n - s_n \times H^{-1} \times S. \quad (2.25)$$

$\hat{\beta}_n$ ist das geschätzte $\hat{\beta}$ im n -ten Iterationsschritt, s_n ist in diesem Fall die Schrittweite λ je Iteration aus der allgemeinen Formel 2.22 und H^{-1} die Inverse der Hessematrix. Da der Algorithmus global konvergiert, wird unabhängig vom gewählten Startwert ein Optimum gefunden. Ein geeigneter Startwert reduziert die Anzahl der notwendigen Iterationsschritte, Behnke (2015). Da die Loglikelihood-Funktion konkav ist, liefert das Verfahren eine eindeutige Lösung, Hastie et al. (2001).

2.1.5 Interpretation der Koeffizienten

Nach der Schätzung ist die Interpretation des Modells von Interesse. Im Vergleich zur linearen Regression liefert das logistische Regressionsmodell keinen direkten Zusammenhang zwischen den erklärenden Variablen X_1 bis X_K und der abhängigen Variable Y . Die geschätzte Wahrscheinlichkeit $\hat{\pi}$ lässt aber Rückschlüsse auf den Ausgang von Y zu. Das Modell stellt also einen Zusammenhang zwischen $P(Y = 1)$ bzw. π und den unabhängigen Variablen X dar. Es ist zu beachten, dass dieser Zusammenhang in erster Linie eine Prognose ermöglicht. Ist eine bestimmte Wertekombination x_i der unabhängigen Variablen bekannt, kann das Modell für diesen Fall eine Wahrscheinlichkeit $\hat{\pi}_i$ liefern und damit Rückschlüsse auf das Eintreten $\hat{y}_i = 1$ oder Nichteintreten $\hat{y}_i = 0$ in diesem Fall i ermöglichen, Behnke (2015). Die Koeffizienten der logistischen Regression sind komplizierter und weniger intuitiv zu interpretieren als die der linearen Regression. Es ist insbesondere der Unterschied zwischen einer Wahrscheinlichkeit und einer Chance zu beachten, siehe Kapitel 2.1.1. Die Interpretation setzt an der bereits genannten Gleichung $\text{Logit}(Y = 1) = X\beta = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K$ an. Die Effekte, die die unabhängigen Variablen auf das Logit haben, sind additiv und linear. Erhöht sich eine Variable X_j mit $j \in \{1, \dots, K\}$ ceteris paribus um eine Einheit, nimmt das Logit der Wahrscheinlichkeit π um den geschätzten Wert des Koeffizienten β_j zu. Ceteris paribus bedeutet, dass die übrigen unabhängigen Variablen gleich bleiben. Es kann nur der Effekt einer unabhängigen Variablen betrachtet werden. Alle anderen Variablen gehen mit ihren jeweiligen Mittelwerten in die Funktion ein. Nimmt die betrachtete Variable den Wert null an und alle anderen ihre jeweiligen Mittelwerte, gibt die Konstante für diesen Fall $P(Y = 1)$ an, Behnke (2015). Da die Interpretation von Logits unanschaulich ist, macht es Sinn, sich auf die Odds zu beziehen. Die Logitfunktion ist definiert als $\ln(\text{odds}(\pi))$. Da die Logarithmusfunktion und die Odds streng monoton sind, können die Vorzeichen der Koeffizienten direkt im Bezug auf die Wahrscheinlichkeit interpretiert werden. Ist $\beta_j > 0$, erhöht sich die Wahrscheinlichkeit π , falls X_j ansteigt. Bei $\beta_j < 0$ ist es umgekehrt, Behnke (2015). Gleichung 2.5 kann umgeformt werden zu

$$\text{Odds}(P(Y = 1)) = \text{Odds}(\pi) = \frac{\pi}{1 - \pi} = e^{X\beta}. \quad (2.26)$$

e^{β_j} wird als Effektkoeffizient bezeichnet. Erhöht sich eine Kovariate um eine Einheit, ändern sich die $\text{Odds}(\pi)$ um den Faktor e^{β_j} . Dies wird durch folgende Herleitung am bivariaten Beispiel deutlich. Man betrachte die Formel 2.26. Steigt die unabhängige Variable X vom Wert x_0 auf den Wert $x_1 = x_0 + 1$ um eine Einheit, dann gilt

$$e^{\beta_0 + \beta_1 x_0} = \frac{\pi(x_0)}{1 - \pi(x_0)}, \quad (2.27)$$

$$e^{\beta_0 + \beta_1 x_1} = \frac{\pi(x_1)}{1 - \pi(x_1)}. \quad (2.28)$$

Durch Äquivalenzumformungen können die beiden Gleichungen zu

$$e^{\beta_1} = \frac{\text{Odds}(\pi(x_1))}{\text{Odds}(\pi(x_0))} \quad (2.29)$$

umgeformt werden. Der Effektkoeffizient gleicht also dem Odds-Ratio. Die Effektkoeffizienten geben an, in welchem Verhältnis sich die Odds verändern, wenn sich eine unabhängige Variable *ceteris paribus* um eine Einheit erhöht, Behnke (2015). Zur besseren Veranschaulichung eignet sich ein Zahlenbeispiel. Beträgt ein geschätzter Koeffizient $\hat{\beta}_j = 0.92$, führt das zu dem Effektkoeffizienten $e^{\hat{\beta}_j} = 2.52$. Erhöht sich X_j um eine Einheit, ändert sich die Chance (oder die Odds) um den Wert 2.52. Wie bereits beschrieben, ist die Chance definiert als die Eintrittswahrscheinlichkeit dividiert durch die Gegenwahrscheinlichkeit $\frac{P(Y=1)}{1-P(Y=1)}$. Betrüge diese vorher beispielsweise 50 : 50, verändert sich der Wert durch die Änderung in X_j auf 71 : 29. Der Ausgangspunkt 50 : 50 ist im Beispiel willkürlich gewählt. Die Änderungsrate bei der logistischen Regression ist jedoch nicht konstant für alle Ausgangswerte. Würde man beispielsweise von einer Chance von 71 : 29 ausgehen und anschließend X_j erhöhen, wäre der absolute Effekt geringer. Das wird als inhärent interaktiven Charakter der logistischen Regression bezeichnet, Behnke (2015). Der absolute Effekt auf die Wahrscheinlichkeit, hängt von dem Wert ab, den X_j selbst hat. Dieses Phänomen wird im Kapitel 2.1.2 über den S-förmigen Verlauf des Graphen der logistischen Regression beschrieben und geht aus der Nichtlinearität der logistischen Regression hervor. Während bei der linearen Regression das Verhältnis der Veränderung der abhängigen Variable zur Veränderung der unabhängigen Variable konstant ist, variiert dieses Verhältnis bei der logistischen Regression, Behnke (2015).

2.2 Verfahren zur Modellwahl und Modellbewertung

2.2.1 Goodness-of-Fit Maße: Pseudo - R^2

Goodness-of-Fit Maße (GOF) geben an, wie gut ein Modell an die beobachteten Daten angepasst ist. Solche GOF Maße sind Pseudo- R^2 . Sie sind nicht zu verwechseln mit dem bei der linearen Regression verwendeten R^2 . Während sich das R^2 bei der linearen Regression auf Anteil der Varianz der abhängigen Variable, der durch das Modell erklärt werden kann, bezieht, nehmen Pseudo- R^2 Maße auf die Devianz Bezug, Behnke (2015). Die Devianz ist, wie bereits erwähnt, ein Maß für die Güte der Anpassung an die beobachteten Daten. Diese ist umso besser, je geringer der Vorhersagefehler ist. Der Vorhersagefehler wird in der Regel über die quadratische Summe der Fehler bestimmt. Die Devianz ergibt sich aus der mit -2 multiplizierten und logarithmierten Likelihood, Behnke (2015).

$$\text{Devianz} = -2LL. \quad (2.30)$$

Die Devianz steht immer in Zusammenhang mit einem bestimmten Modell, das mit ML-Schätzung festgelegt wurde, siehe Kapitel 2.1.4. Als Nulldevianz wird die Devianz des Nullmodells bezeichnet. Dieses besteht nur aus einer Konstanten. Da die Devianz von der Stichprobengröße N abhängt, eignet sie sich selbst nicht als GOF Maß. Es gilt aber der Zusammenhang: Je kleiner die Devianz eines bestimmten Modells mit einem bestimmten Parametervektor, desto größer der Anteil der Devianz des Nullmodells, der durch die Hinzunahme der Information, die in unabhängigen Variablen enthalten ist, erklärt werden kann, Behnke (2015). Pseudo- R^2 s vergleichen die Loglikelihood eines betrachteten Modells LL_1 mit der Loglikelihood des Nullmodells LL_0 , Urban and Mayerl (2006). Das McFadden- R^2 wird folgendermaßen definiert:

$$\text{McFadden-}R^2 = 1 - \frac{LL_0}{LL_1}. \quad (2.31)$$

Es handelt sich hierbei um ein modellrelatives Gütemaß, das sich im Wertebereich $[0, 1]$ bewegt. Hat das betrachtete Modell dieselbe Loglikelihood wie das Nullmodell, liefert das McFadden R^2 den Wert 0. Hat das Modell die Likelihood 1 und somit die Loglikelihood 0, nimmt das Pseudo- R^2 den Wert 1 an, Urban (1993). Die Loglikelihood misst den Schätzerfolg eines Modells. Die Kennzahl kann also interpretiert werden als der prozentuale Anteil, um den das Nullmodell (LL_0) durch das Prädiktorenmodell (LL_1) verbessert werden kann, Urban and Mayerl (2006). Basierend darauf kann auch das Cox & Snell- R^2 berechnet werden.

$$\text{Cox \& Snell-}R^2 = 1 - \left(\frac{LL_0}{LL_1} \right)^{\frac{2}{N}}. \quad (2.32)$$

Dieses beschreibt das geometrische Mittel aus den Verhältnissen der individuellen Trefferwahrscheinlichkeiten mit und ohne der Information der im Modell verwendeten unabhängigen Variablen, Behnke (2015). Da das Cox & Snell- R^2 nur den Wertebereich $[0, 1)$ erreicht und somit nie den Wert 1 annimmt, gibt es Nagelkerkes R^2 , das auf einen Wertebereich von $[0, 1]$ standardisiert ist, Urban and Mayerl (2006).

$$\text{Nagelkerkes-}R^2 = \frac{1 - \left(\frac{LL_0}{LL_1} \right)^{\frac{2}{N}}}{1 - \left(\frac{LL_0}{LL_0} \right)^{\frac{2}{N}}}. \quad (2.33)$$

Das Nagelkerke- R^2 ist durch die Standardisierung immer größer als das Cox& Snell- R^2 . Meistens ist es auch größer als das Pseudo- R^2 nach McFadden, Behnke (2015). Die Werte der Pseudo- R^2 Maße hängen stark von der Fragestellung und von den verwendeten Kovariaten ab. Von daher ist es schwierig, Orientierungswerte anzugeben. Generell sind Werte nahe der 1 in der Praxis eher selten zu erwarten, Kopp (2014).

2.2.2 Likelihood-Ratio-Test

Ein weiteres Testverfahren, um zwei konkurrierende Modelle miteinander zu vergleichen, ist der Likelihood-Ratio-Test (LR-Test). Dabei enthält ein Modell M_1 mehr Regressoren als ein anderes Modell M_0 . Man kann M_1 auch als unbeschränktes und M_0 als beschränktes Modell bezeichnen. Das beschränkte Modell ist durch das unbeschränkte eingeschachtelt, Behnke (2015). Grundsätzlich ist ein Modell M_1 mit mehr erklärenden Variablen besser an die Daten angepasst, als eines mit weniger erklärenden Variablen. Der folgende Test prüft die Nullhypothese, ob beide Modelle gleich gut sind. Wird H_0 verworfen, ist der Unterschied zwischen M_1 und M_0 statistisch signifikant, sodass lieber das vollere Modell gewählt werden sollte, Bruin (2011). Für die Teststatistik gilt:

$$lr = -2\ln\left(\frac{L_1}{L_0}\right) = 2\{LL_0 - LL_1\}, \quad (2.34)$$

wobei LL_1 und LL_0 für die Werte der Loglikelihood der beiden Modelle stehen. Unter H_0 ist die Teststatistik mit q als Anzahl der Parameter, in denen sich die Modelle unterscheiden, χ^2 -verteilt.

$$lr \stackrel{H_0}{\sim} \chi_q^2. \quad (2.35)$$

Es ist zu beachten, dass der LR-Test nur geeignet ist, um Modelle zu vergleichen, die aus der gleichen Stichprobe entstehen, Behnke (2015). Oft wird der Vergleich eines Modells mit dem Nullmodell betrachtet. Hier prüft der LR-Test, ob das Prädiktorenmodell besser an die Daten angepasst ist, als das Nullmodell. Gilt H_0 , liefern die Variablen im Prädiktorenmodell keinen Erklärungsgehalt, Urban and Mayerl (2006).

2.2.3 Korrektklassifikationsrate

Ein weiteres Maß, um die Güte der Anpassung zu bewerten, ist die Korrektklassifikationsrate, auch als Accuracy bezeichnet. Das Maß basiert auf den vorhergesagten Wahrscheinlichkeiten und dem Abgleich der tatsächlich beobachteten und prognostizierten Ereignisse, Peng et al. (2002). Mit den geschätzten Wahrscheinlichkeiten werden Gruppenzugehörigkeiten vorhergesagt. Für die Bestimmung der geschätzten Gruppenzugehörigkeiten wird zunächst für jede Beobachtung ($\forall i, i = 1, \dots, N$) die vorhergesagte Wahrscheinlichkeit $\hat{P}(y_i = 1|x_i) = \hat{\pi}_i(x_i)$ für das Eintreten eines Ereignisses abhängig von einer beobachteten Wertekombination der Kovariaten berechnet. Liefert das Modell $\hat{\pi}_i > t$ für einen Threshold-Wert $t \in [0, 1]$ wird für die Beobachtung $\hat{y}_i = 1$ vorhergesagt. Gilt $\hat{\pi}_i \leq t$, wird 0 vorhergesagt. Im Anschluss wird der tatsächliche Ausgangswert von y_i für alle i betrachtet und mit dem geschätzten Wert \hat{y}_i verglichen. Die Korrektklassifikationsrate ist das Verhältnis aus der Anzahl der richtig vorhergesagten Werte (sowohl für 1, als auch für 0), dividiert durch N , Urban and Mayerl

beobachteter Erfolg/ prognostizierter Erfolg	$\hat{Y} = 0$	$\hat{Y} = 1$	Σ
$Y = 0$	Anzahl richtiger Misserfolgsprognosen	Anzahl falscher Erfolgsprognosen	Anzahl Misserfolge
$Y = 1$	Anzahl falscher Misserfolgsprognosen	Anzahl richtiger Erfolgsprognosen	Anzahl Erfolge
Σ	Anzahl Misserfolgsprognosen	Anzahl Erfolgsprognosen	N

Tabelle 1: Aufbau einer **Klassifikationstabelle**, Tutz (2011).

Vorhersage/Beobachtung	$Y = 0$	$Y = 1$	Σ
$\hat{Y} = 0$	92	305	397
$\hat{Y} = 1$	65	710	775
Σ	157	1015	1172

Tabelle 2: Die **Klassifikationstabelle** der Definition Erfolg3.

(2006). Eine Klassifikationstabelle, siehe Tabelle 1, veranschaulicht die Situation. Die Werte auf der Diagonale (Anzahl richtiger Misserfolgsprognosen + Anzahl richtiger Erfolgsprognosen) dividiert durch N ergibt die Korrekturklassifikationsrate. Tabelle 2 zeigt die Klassifikationstabelle für das Modell der Definition Erfolg3. Anhand eines Testdatensatzes wird ermittelt, ob das Modell die Erfolgsereignisse richtig vorhersagt. Aus der Tabelle 1 ergeben sich auch die Maße Sensitivität und Spezifität, Hosmer and Lemeshow (2000). Es gilt:

$$\text{Sensitivität} = P(\hat{y} = 1 | y = 1) = \frac{\# \text{ Richtige Erfolgsprognosen}}{\# \text{ Erfolgsprognosen}}, \quad (2.36)$$

$$\text{Spezifität} = P(\hat{y} = 0 | y = 0) = \frac{\# \text{ Richtige Misserfolgsprognosen}}{\# \text{ Misserfolgsprognosen}}.$$

Damit entspricht $P(\hat{y} = 0 | y = 1)$ dem Verpassen der Vorhersage eines eintretenden Ereignisses (Fehler 2. Art) und $P(\hat{y} = 1 | y = 0)$ einem Falschalarm (Fehler 1. Art), Tutz (2011). Die Sensitivität gibt also den Anteil korrekt klassifizierter Erfolge an und die Spezifität den Anteil korrekt klassifizierter Misserfolge. Zusätzlich existieren die Maße False Positive Rate (FPR) und False Negative Rate (FNR). Die FPR ist der Anteil der Beobachtungen, die fälschlicherweise als Erfolge klassifiziert werden, an allen vorhergesagten Erfolgen, während die FNR der Anteil der Beobachtungen, die fälschlicherweise als Misserfolge klassifiziert werden an allen vorhergesagten Misserfolgen ist, Peng et al. (2002). Bei Tabelle 2 ergeben sich folgende Kennzahlen:

$$\text{Sensitivität} = \frac{710}{1015} = 0.6995,$$

$$\begin{aligned}\text{Spezifizität} &= \frac{92}{157} = 0.586, \\ \text{False positive rate} &= \frac{710}{775} = 0.9161, \\ \text{False negative rate} &= \frac{305}{397} = 0.7683.\end{aligned}$$

Ein Nachteil der Korrektklassifikationsrate ist, dass ein Falschalarm und das Verpassen eines Ereignisses als gleich schlimme Fehler gewichtet werden. Dabei ist es in vielen praktischen Anwendungen wichtiger, das Eintreten eines Ereignisses richtig vorherzusagen, Krzanowski and Hand (2009).

2.2.4 Receiver Operating Characteristic Kurve

Die Receiver Operating Characteristic (ROC) eignet sich, um die Vorhersagegüte bei der Zuordnung von Beobachtungen zu zwei Gruppen zu messen, Tutz (2011). Genauso wie die Korrektklassifikationsrate basiert die ROC-Kurve auf einer Klassifikationstabelle, wie die Tabelle 1. Das Ergebnis der Tabelle ist abhängig von der Wahl eines Thresholds t . In der Regel wird dieser als $t = 0.5$ gewählt.

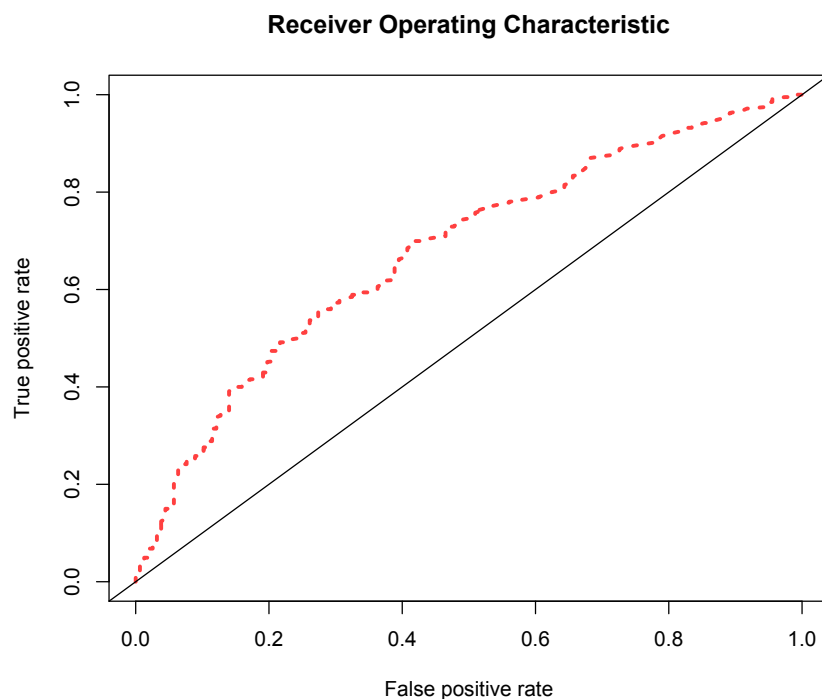


Abbildung 2: ROC-Kurve am Beispiel des Modells zu Definition Erfolg3.

Gegebenfalls macht es Sinn, von dieser Festsetzung abzuweichen. Für eine Beobachtung x_i wird die Gruppenzugehörigkeit $\hat{y}_i = 1$ geschätzt, sofern $\hat{\pi}_i > t$, also der festgelegte Threshold überschritten wird. Ein Modell hat eine hohe Vorhersagegüte, wenn die in Formel 2.36 beschriebenen Maße Sensitivität und Spezifität gleichzeitig möglichst groß werden. Die ROC-Kurve stellt das Verhältnis aus

Sensitivität ($\hat{=}$ True Positive Rate (TPR)) und $1 -$ Spezifität ($\hat{=}$ FPR) für alle möglichen Thresholds gegenüber. Entscheidend ist nicht, welches t als Trennungspunkt gewählt wird, da die ROC Kurve die Ergebnisse für Sensitivität und Spezifität für alle möglichen Thresholds zusammenfasst, Krzanowski and Hand (2009). Die Fläche unter der ROC-Kurve, genannt Area Under The Curve (AUC) gibt an, wie gut das Modell Gruppenzugehörigkeiten vorhersagt. Diese wird beschrieben als

$$\int_0^1 \text{ROC}(t) \partial t \text{ mit } \text{AUC} \in [0.5, 1]. \quad (2.37)$$

Je näher die ROC-Kurve am Punkt $(0, 1)$ liegt, desto erfolgreicher sagt das Modell Ausgangswerte richtig hervor. Das perfekte Modell besteht aus zwei Geraden, die die Punkte $(0, 0)$, $(0, 1)$ und $(1, 1)$ verbinden. Hierfür gilt $\text{AUC} = 1$. Das schlechteste Modell hat eine ROC-Kurve in Form einer Geraden durch den Quadranten. Dieses Modell entspricht zufälligem Raten und es gilt $\text{AUC} = 0.5$. In der Regel liegt die empirische ROC-Kurve zwischen den beiden Extremen, Tutz (2011). Um die Güte der Klassifikation einschätzen zu können, liefern Hosmer and Lemeshow (2000) Orientierungswerte in Tabelle 3. In der Praxis werden selten Werte oberhalb von 0.9 beobachtet. Ein nach anderen

$\text{AUC} = 0.5$	keine Trennung
$\text{AUC} \in [0.7, 0.8]$	akzeptable Trennung
$\text{AUC} \in [0.8, 0.9]$	sehr gute Trennung
$\text{AUC} \geq 0.9$	exzellente Trennung

Tabelle 3: Orientierungswerte für das Maß AUC

Kriterien (bspw. Pseudo- R^2 -Maßen) schlecht angepasstes Modell kann dennoch eine gute Klassifikation bieten, Hosmer and Lemeshow (2000). Krzanowski und Hand beschreiben die ROC-Kurve mit anderen Worten. Es gibt zwei Populationen P (Positiv) und N (Negativ) bezüglich des Eintreten eines Ereignisses. $S(X)$ ist eine sogenannte stetige Score-Funktion. Diese entspricht bei der logistischen Regression der Wahrscheinlichkeit für eine Gruppenzugehörigkeit. Überschreitet $s(x)$ für ein Individuum den Threshold t , wird x der Population P oder N zugeordnet. Beide Gruppen haben die jeweiligen Dichtefunktionen $p(s|P)$ und $p(s|N)$. Je weniger die Dichtefunktionen überlappen, desto besser kann ein Modell klassifizieren. Umso mehr sich die Dichtefunktionen unterscheiden, desto weniger Falschvorhersagen gibt es. Die AUC kann also auch als Ähnlichkeitsmaß für die Scores in den jeweiligen Gruppen interpretiert werden, Krzanowski and Hand (2009). Die ROC-Kurve hat folgende mathematische Eigenschaften, Tutz (2011):

- Sie ist eine monoton steigende Funktion im positiven Quadranten
- Sie ist invariant gegenüber streng monotonen Transformationen

- Sie ist konkav (Ausnahmen: bestes und schlechtestes Modell)

Krzanowski und Hand schlagen folgendes Verfahren vor, um die empirische ROC-Kurve zu schätzen: Zunächst wird eine ROC Kurve definiert als

$$y(t) = h(x(t)), \quad (2.38)$$

mit $x(t) = p(S > t|N)$ und $y(t) = p(S > t|P)$. Eliminiert man t durch Äquivalenzumformungen, erhält man folgende Gleichung für die ROC-Kurve:

$$y = 1 - G[F^{-1}(1 - x)] \text{ mit } x \in [0, 1]. \quad (2.39)$$

f, F bzw. g, G sind dabei die Dichte-/Verteilungsfunktionen in den Populationen N und P . Auf Basis gegebener Daten können Sensitivität und Spezifität geschätzt werden. Setzt man die gegebenen Daten in die Definitionen von Sensitivität und Spezifität aus (2.36) ein, erhält man die empirische ROC Kurve der Form

$$y = 1 - \hat{G}[\hat{F}^{-1}(1 - x)] \text{ mit } x \in [0, 1]. \quad (2.40)$$

Auf Basis der empirischen ROC-Kurve kann über numerische Integration ein Schätzer für die AUC, also \widehat{AUC} mit Varianz $s^2(AUC)$ bestimmt werden, Krzanowski and Hand (2009).

Wahl des optimalen Thresholds

Bei gegebener ROC-Kurve wird ermittelt, bei welchem Threshold die beste Trennung erreicht wird. Hierfür betrachtet man den Youden Index (YI). Dieser gibt an, für welchen Threshold t aus der Menge T der möglichen Thresholds die Summe aus Sensitivität + Spezifität - 1 am größten ist, Krzanowski and Hand (2009), sprich,

$$YI = \max_{t \in T} \{ \text{Sensitivität}(t) + \text{Spezifität}(t) - 1 \}. \quad (2.41)$$

Der YI liefert den maximalen Abstand zwischen der ROC-Kurve und der Diagonale durch den Quadranten.

2.2.5 Wald-Konfidenzintervalle und Wald-Tests

Während die Korrektklassifikationsrate und der Likelihood-Ratio-Test das gesamte Modell betrachten, fokussiert der Wald-Test die Signifikanz einzelner Regressoren, Behnke (2015). Wie bereits in Kapitel 2.1.4 erwähnt, liefert das ML-Schätzverfahren asymptotisch konsistente und normalverteilte Punktschätzer $\hat{\beta}_j$ für β_j mit $j \in 1, \dots, K$. Dabei ist $\hat{\sigma}_{\hat{\beta}_j}^2$ die geschätzte Varianz eines Schätzers $\hat{\beta}_j$, Behnke (2015). Somit gilt für die Verteilung eines geschätzten Koeffizienten

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma_{\hat{\beta}_j}^2). \quad (2.42)$$

Daraus können $(1-\alpha)$ -Konfidenzintervalle konstruiert werden. Das sind Intervalle, in die der unbekannte Parameter mit der Wahrscheinlichkeit $1-\alpha$ fällt. Wenn z_δ das δ -Quantil der Standardnormalverteilung $\mathcal{N}(0, 1)$ ist, gilt

$$P\left(z_{\frac{\alpha}{2}} < \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha. \quad (2.43)$$

Durch Umformungen ergibt sich das Konfidenzintervall

$$\left[\hat{\beta}_j - z_{1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j}, \hat{\beta}_j + z_{1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j}\right]. \quad (2.44)$$

Nimmt eine Variable β_j den Wert 0 an, hat sie keinen Einfluss auf die abhängige Variable Y . Über Konfidenzintervalle und darauf basierende Tests kann überprüft werden, ob ein Koeffizient von 0 verschieden ist. Liegt 0 nicht im $(1-\alpha)$ -Konfidenzintervall, kann mit Irrtumswahrscheinlichkeit α behauptet werden, dass der betrachtete Koeffizient β_j ungleich null ist. Als Test umformuliert lautet das Hypothesenpaar

$$H_0 : \beta_j = 0 \text{ und } H_1 : \beta_j \neq 0. \quad (2.45)$$

Unter H_0 beträgt die Teststatistik w

$$w = \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}}. \quad (2.46)$$

w ist asymptotisch standardnormalverteilt, Woolridge (2002). Somit sind Realisationen der Teststatistik außerhalb des Intervalls $(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}})$ mit Wahrscheinlichkeit α zu erwarten. Liegt die realisierte Teststatistik außerhalb des Intervalls, kann H_0 mit der Irrtumswahrscheinlichkeit α verworfen werden. Das Verwerfen ist gleichbedeutend damit, dass das $(1-\alpha)$ -Konfidenzintervall den Wert 0 nicht enthält. Eine Aussage über die Wahrscheinlichkeit des Verwerfens der Nullhypothese geben die P-Werte an. P-Werte liefern die minimale Irrtumswahrscheinlichkeit α , bei der H_0 gerade noch so verworfen werden kann. Ist der P-Wert sehr klein, ist das Nichtablehnen von H_0 und damit $\beta_j = 0$ unwahrscheinlich. Daraus kann dann geschlussfolgert werden, dass der Koeffizient einen Einfluss auf die abhängige Variable hat, Woolridge (2002)

2.2.6 Akaike Informationskriterium und Abwärtsverfahren

Ein Modell sollte eine Zielvariable möglichst genau erklären und dabei den Trade-Off zwischen Varianz und Verzerrung ausbalancieren. Ein Modell mit vielen Regressoren tendiert in der Regel dazu, erwartungstreu zu sein. Es bringt aber häufig eine höhere Varianz mit sich. Bei einem Modell mit weniger Regressoren ist die Varianz zwar kleiner, dafür kann möglicherweise eine Verzerrung vorliegen. Das Akaike Informationskriterium (AIC) ist eine Kennzahl, die einen Ausgleich zwischen beiden Zielen sucht. Das AIC basiert auf der Kullback-Leibler-Information (K-L-Information), Burnham and

Anderson (2004),

$$I(f, g) = \int_{-\infty}^{\infty} f(X) \ln \left(\frac{f(X)}{g(X|\theta)} \right) dX. \quad (2.47)$$

Diese kann umgeformt werden zu

$$I(f, g) = E_f [\ln(f(X))] - E_f [\ln(g(X|\theta))]. \quad (2.48)$$

Dabei ist $I(f, g)$ ein Distanzmaß zwischen einer wahren Funktion $f(X)$ und einer approximierten Funktion $g(X|\theta)$ mit dem zu schätzenden Parameter θ . In den beschriebenen Fällen ist β der zu schätzende Parameter für θ . Die Minimierung der K-L-Information führt zu einem Modell, das dem wahren Modell möglichst nahe kommt. Akaike hat gezeigt, dass

$$E_Y E_X [\ln(g(X|\hat{\theta}(Y)))] \quad (2.49)$$

geschätzt werden muss, um die K-L-Information bei gegebenen Daten zu minimieren. $\hat{\theta}(Y)$ ist dabei der ML-Schätzer für die Koeffizienten der Variablen, die Y erklären. Außerdem liefert das Maximum der Loglikelihood Funktion $\ln(L(\hat{\theta}|X = x))$ einen verzerrten Schätzer für (2.49). Dessen Verzerrung entspricht der Anzahl der im Modell zu schätzenden Parameter K . Somit ist der asymptotisch unverzerrte Schätzer für die K-L-Information gegeben durch

$$\hat{E}(K - L) = \ln(L(\hat{\theta}|X = x)) - K. \quad (2.50)$$

Daraus kann das AIC hergeleitet werden, Burnham and Anderson (2004),

$$AIC = -2\log(L(\hat{\theta}|X = x)) + 2K = \text{Devianz} + 2K. \quad (2.51)$$

Die Devianz wird in Formel 2.30 beschrieben. Der vordere Term beschreibt die Anpassung des mit dem ML-Verfahren geschätzten Modells, während der hintere Term auf die Komplexität eingeht. Ein Modell ist komplexer, wenn es mehr Regressoren enthält. Der zweite Term der Summe ist also ein Strafterm für die Komplexität, Behnke (2015). Es ist ersichtlich, dass eine gute Anpassung des Modells an die Daten das AIC verringert, und eine höhere Anzahl an Regressoren das AIC erhöht. Ein Modell ist gegenüber eines Konkurrenzmodells zu bevorzugen, wenn es ein kleineres AIC liefert, Konishi and Kitagawa (2008). Der Vergleich von Modellen, die sich auf verschiedene Stichproben beziehen, ist mit (2.51) nur möglich, falls die Stichproben ähnlich groß sind. Ansonsten deutet ein geringeres AIC auch auf einen kleineren Stichprobenumfang und nicht unbedingt auf eine bessere Anpassung hin, Behnke (2015). Bei dem Abwärtsverfahren, das in dieser Arbeit angewendet wird, wird das AIC schrittweise minimiert, um ein bestmögliches Modell zu finden. Dabei wird zunächst ein volles Modell aufgestellt, das alle möglichen Variablen enthält. Pro Iteration wird der Regressor

bestimmt, der am wenigsten Einfluss hat. Dies wird gemessen am P-Wert des Wald-Tests, siehe Kapitel 2.2.5. Das Modell mit einem minimalen Wert des AIC liefert den besten Trade-Off zwischen Anpassungsgüte und Komplexität, Konishi and Kitagawa (2008).

2.3 Zwei-Stichproben Kolmogorov-Smirnov-Test

Der Zwei-Stichproben Kolmogorov-Smirnov-Test (KS-Test) prüft, ob die Verteilung von zwei Grundgesamtheiten verschieden ist. Hierfür wird vorausgesetzt, dass die Stichprobenvariablen X_1, \dots, X_m und Y_1, \dots, Y_n unabhängig sind und aus den Grundgesamtheiten A_1 und A_2 mit den unbekannt, stetigen Verteilungsfunktionen $F(z) = P(X_i \leq z)$ und $G(z) = P(Y_j \leq z)$ mit $i = 1, \dots, m$ und $j = 1, \dots, n$ kommen, Beckmann (1999). Das Hypothesenpaar lautet für einen zweiseitigen Test

$$H_0 : F(z) = G(z) \text{ für } z \in \mathbb{R}, \quad (2.52)$$

$$H_1 : F(z) \neq G(z) \text{ für mindestens ein } z \in \mathbb{R}. \quad (2.53)$$

Die dazugehörige Teststatistik errechnet sich als

$$K_{m,n} = |F_m(z) - G_n(z)|. \quad (2.54)$$

Die Teststatistik gibt die größte, betragsmäßige Differenz zwischen den beiden Verteilungen an. Gilt $K_{m,n} > k_{1-\alpha}$ mit $k_{1-\alpha}$ als $(1 - \alpha)$ -Quantil der Verteilung von $K_{m,n}$, wird H_0 abgelehnt, Beckmann (1999). Abbildung 6 in Kapitel 4.8.2 veranschaulicht die Durchführung von zwei KS-Tests. Die linke Abbildung stellt eine relativ großen maximalen Abstand zwischen den Verteilungsfunktionen dar. Deshalb kann H_0 aus Kapitel 4.8.2 zugunsten von H_1 verworfen werden. Die rechte Abbildung stellt einen KS-Test dar, bei dem der maximale Abstand zwischen den Verteilungsfunktionen nicht groß genug ist, um H_0 zu verwerfen.

3 Gegenstand der Untersuchung

3.1 Beschreibung des Datenmaterials

3.1.1 Einteilung in drei Behandlungsgruppen

Der vorliegende Datensatz umfasst Behandlungen im Zeitraum vom 22.01.1999 bis zum 08.09.2016. Gruppe 1 beinhaltet Behandlungen, bei denen überwiegend Handinstrumente zur endodontischen Therapie zum Einsatz kamen. Behandlungen der Gruppe 2 sind durch die Abrechnungskürzel Phys 2400/240 gekennzeichnet. Bei diesen Behandlungen kamen die maschinenbetriebenen BioRaCe[®]-Instrumente sowie erweiterte Methoden zur Desinfektion von Wurzelkanälen zum Einsatz. Ab November 2011 werden in der AZFK die effizienteren Reciproc[®]-Instrumente verwendet. Die genannten Behandlungsmethoden werden in Kapitel 1.1 beschrieben. Ob die Behandlung mit der fortschrittlichsten Technologie Reciproc[®] die Prognose auf eine erfolgreiche Behandlung verbessert, ist die Hauptfragestellung dieser Arbeit. Deshalb werden die Behandlungen, bei denen Reciproc[®] zum Einsatz kam, mit den übrigen Behandlungen verglichen. Dieser Vergleich wird für verschiedene Definitionen von Behandlungserfolg durchgeführt.

3.1.2 Definition des Behandlungserfolgs

Es gibt verschiedene Möglichkeiten, eine erfolgreiche Behandlung zu definieren. Das härteste und eindeutigste (kurz- und langfristige) Kriterium für Erfolg oder Misserfolg der Behandlung ist die Zahnentfernung. Die Zahnentfernung nach der Therapie ist eindeutig als Misserfolg zu kategorisieren. Bei kürzeren Beobachtungszeiten liegen allerdings keine aussagekräftigen Daten über die längerfristige Entwicklung des Zahnes nach der Behandlung vor. Demzufolge werden die Erfolgsdefinitionen nach folgenden Zeitintervallen betrachtet:

- 1. Erfolgreich, wenn innerhalb der ersten **2 Jahre** nach der Behandlung keine Extraktion erforderlich war und der letzte Besuch mindestens 2 Jahre zurückliegt. (**Erfolg1**)
- 2. Erfolgreich, wenn innerhalb der ersten **3 Jahre** nach der Behandlung keine Extraktion erforderlich war und der letzte Besuch mindestens 3 Jahre zurückliegt. (**Erfolg2**)
- 3. Erfolgreich, wenn innerhalb der ersten **4 Jahre** nach der Behandlung keine Extraktion erforderlich war und der letzte Besuch mindestens 4 Jahre zurückliegt. (**Erfolg3**)

Dabei muss auch berücksichtigt werden, wie lange der letzte Besuch zurückliegt. Sonst kann es zu verfälschten Informationen kommen. Betrachtet man bspw. Erfolgsdefinition 1 und ein Patient hat sich 2014 behandeln lassen, der letzte Eintrag zu diesem Patienten ist aus dem Jahr 2015 und der Zahn wurde nicht extrahiert. Fälschlicherweise könnte der Fall als Erfolg ausgewiesen werden, da

der Behandlungstermin 2 Jahre vor dem Jahr 2016 liegt, auch wenn der Zahn tatsächlich erst ein Jahr überlebt hat. Tabelle 4 fasst zusammen, bei welchen Erfolgsdefinitionen welche Häufigkeiten

Erfolg1	Anzahl der Fälle	Erfolg2	Anzahl	Erfolg3	Anzahl
0	486	0	594	0	678
1	5996	1	4925	1	4013
Σ	6482	Σ	5519	Σ	4691
Σ_{Reci}	1035	Σ_{Reci}	584	Σ_{Reci}	215

Tabelle 4: Übersicht über die **Erfolgsdefinitionen 1-3.**

von erfolgreichen und nicht erfolgreichen Behandlungen vorliegen. Durch die Betrachtung von Zeitintervallen ist es notwendig, kürzer zurückliegende Behandlungen von der Betrachtung auszuschließen. Bei Definition Erfolg1 werden bspw. Behandlungen betrachtet, bei denen innerhalb der ersten beiden nachfolgenden Jahre keine Extraktion stattfand. Somit können Behandlungen, die nach dem 08.09.2014 stattfanden, nicht berücksichtigt werden. Aus Tabelle 4 geht hervor, dass für längere Zeitintervalle weniger Daten vorliegen. Ab einem Zeitintervall von fünf Jahren gibt es zu wenige Ergebnisse aus Gruppe 3 mit Recipro[®], um diese noch sinnvoll vergleichen zu können. Daher ist die Betrachtung größerer Zeitintervalle nicht möglich. Des Weiteren können Revisionsbehandlungen herangezogen werden, um den Behandlungserfolg zu definieren. Auch hier wird der Zeitabstand des revisionsfreien Überlebens nach der Behandlung betrachtet. Ebenso wie bei der Extraktion als Erfolgskriterium, kann maximal ein Zeitintervall von vier Jahren noch sinnvoll ausgewertet werden, siehe Tabelle 5. Eine Revisionsbehandlung umfasst sowohl die erneute endodontische Therapie, sowie den Fall einer WSR. Es ergeben sich folgende Erfolgskategorien:

- 4. Erfolgreich, wenn innerhalb der ersten **2 Jahre** nach der Behandlung keine Revisionsbehandlung erforderlich war und der letzte Besuch mindestens 2 Jahre zurückliegt. (**Erfolg4**)
- 5. Erfolgreich, wenn innerhalb der ersten **3 Jahre** nach der Behandlung keine Revisionsbehandlung erforderlich war und der letzte Besuch mindestens 3 Jahre zurückliegt. (**Erfolg5**)
- 6. Erfolgreich, wenn innerhalb der ersten **4 Jahre** nach der Behandlung keine Revisionsbehandlung erforderlich war und der letzte Besuch mindestens 4 Jahre zurückliegt. (**Erfolg6**)

Als dritte Kategorie von Erfolgsdefinitionen ergibt sich die Zusammenfassung der oben aufgeführten Definitionen, siehe Tabelle 6.

- 7. Erfolgreich, wenn innerhalb der ersten **2 Jahre** nach der Behandlung keine Revisionsbe-

Erfolg4	Anzahl der Fälle	Erfolg5	Anzahl	Erfolg6	Anzahl
0	227	0	256	0	251
1	6255	1	5263	1	4440
Σ	6482	Σ	5519	Σ	4691
Σ_{Reci}	1035	Σ_{Reci}	587	Σ_{Reci}	215

Tabelle 5: Übersicht über **Erfolgsdefinitionen 4-6.**

Erfolg7	Anzahl der Fälle	Erfolg8	Anzahl	Erfolg9	Anzahl
0	688	0	805	0	875
1	5794	1	4714	1	3816
Σ	6482	Σ	5519	Σ	4691
Σ_{Reci}	1035	Σ_{Reci}	587	Σ_{Reci}	215

Tabelle 6: Übersicht über **Erfolgsdefinitionen 7-9.**

handlung oder Extraktion erforderlich war und der letzte Besuch mindestens 2 Jahre zurückliegt. (**Erfolg7**)

-8. Erfolgreich, wenn innerhalb der ersten **3 Jahre** nach der Behandlung keine Revisionsbehandlung oder Extraktion erforderlich war und der letzte Besuch mindestens 3 Jahre zurückliegt. (**Erfolg8**)

-9. Erfolgreich, wenn innerhalb der ersten **4 Jahre** nach der Behandlung keine Revisionsbehandlung oder Extraktion erforderlich war und der letzte Besuch mindestens 4 Jahre zurückliegt. (**Erfolg9**)

Auch hier ist die Betrachtung von Zeitintervallen, die größer als vier Jahre sind, nicht mehr zielführend.

3.1.3 Überblick über die Kovariaten im Datensatz

Im Folgenden wird ein Überblick über die im Datensatz enthaltenen Kovariaten gegeben. Die Variablen **Zahntyp** enthalten Informationen darüber, ob es sich bei dem behandelten Zahn um Molaren (Mahlzähne), Prämolaren (Vormahlzähne), Canini (Eckzähne) oder Incisivi (Schneidezähne) handelt. Im Modell wird diese Zuordnung über drei Dummy-Variablen *molar*, *praemolar* und *canini* realisiert. Über die Zahnnummer kann der Zahn dem jeweiligen Typ zugeordnet werden. Die zusätzliche

Aufnahme einer Dummy-Variable für Incisivi könnte zu einem strikten Multikollinearitätsproblem, siehe Kapitel 4.1, führen. Nehmen alle Variablen den Wert null an, handelt es sich um den Typ Incisivus. Es besteht die Annahme, dass die Behandlung von Molaren aufgrund der erhöhten Anzahl von Wurzelkanälen und der schwierigeren Zugänglichkeit im Mund komplizierter sein könnte. Deshalb könnte die Variable Einfluss auf den Behandlungserfolg haben. Der Datensatz enthält außerdem Informationen über die **Zahnvitalität bei Behandlungsbeginn**. Sie gibt an, ob die WKB am vitalen oder nicht-vitalen Zahn durchgeführt wurde. Die Kürzel VitE (vital) und Trep1 (nicht-vital) kategorisieren dies. Zur einfacheren Handhabung wird die Information in eine binäre Variable übergeführt, die bei der Behandlung am vitalen Zahn den Wert 1 annimmt. Die Variable VorherigeWRs gibt an, ob und wie oft beim selben Patienten in der Vergangenheit eine **WSR** durchgeführt wurde. Gezählt werden deshalb im Zusammenhang nur WSRs, die vor der betrachteten WKB stattfanden. Es werden auch die WSRs an anderen Zähnen miteinbezogen. Eine hohe Anzahl an bereits erfolgten WSRs könnte für ein höheres Grundrisiko für Misserfolg einer WKB stehen. Ebenso wie die Variable VorherigeWRs betrachtet VorherigeWks die **vorherigen WKB**. Hatte ein Patient bereits eine oder mehrere WKB, könnte auch das auf ein erhöhtes Grundrisiko hinweisen. Die Variable MedReinigungen liefert Informationen darüber, wieviele **medizinische Zahnreinigungen** am behandelten Zahn nach der Wurzelkanalfüllung durchgeführt wurden. Eine höhere Anzahl an Zahnreinigungen könnte für eine bessere Hygiene und infolgedessen höhere Erfolgswahrscheinlichkeit der WKB sprechen. Der Datensatz enthält auch Informationen über die **Erfahrung des Behandlers**. Jeder Arzt hat eine Kennziffer, die jeder Behandlung zugeordnet wird. Die Variable erfahrung gibt an, die wievielte WKB des Behandlers die betrachtete WKB ist. Hat ein Behandler viele WKB durchgeführt, enthält die Variable einen hohen Wert. Die Durchführung der Behandlung durch einen erfahrenen Arzt könnte Rückschlüsse auf eine höhere Erfolgswahrscheinlichkeiten zulassen. Die Variable Alter gibt das **Alter des Patienten** bei der WKB an. Da bei älteren Patienten ein höheres Risiko für den Misserfolg der Behandlung bestehen könnte, hat die Variable möglicherweise negativen Einfluss auf die Erfolgswahrscheinlichkeit.

4 Vorstellung der Ergebnisse

4.1 Anforderungen an den Datensatz und das Modell

Um beim logistischen Regressionsmodell sinnvolle Ergebnisse zu erhalten, müssen einige Bedingungen erfüllt sein, Urban and Mayerl (2006). Wie bereits beschrieben, sind ML-Schätzer konsistent und effizient. Somit verbessert sich die Qualität der Schätzung bei einer umfangreichen Stichprobe. Die Tabellen 4, 5 und 6 geben Auskunft über die Gesamtzahl der Fälle. Die Zeitintervalle sind so gewählt, dass die Fallzahl für jede Erfolgsdefinition ausreichend groß ist. Die Erfolgsdefinitionen mit größeren Fallzahlen erreichen in der Regel bessere Schätzergebnisse, Urban and Mayerl (2006). Außerdem

	MedReinigungen				
Erfolg	(0,5]	(5,10]	(11,15]	(16-20]	>20
0	171	128	19	9	2
1	1048	758	151	25	1

	Alter zu Behandlungsbeginn				
Erfolg	(0,20]	(21,40]	(41,60]	(61,80]	>80
0	16	107	359	367	26
1	104	910	1524	1236	42

	praemolar	
Erfolg	0	1
0	641	234
1	2669	1147

	molar	
Erfolg	0	1
0	668	207
1	2584	1232

Tabelle 7: Kreuztabellen für Erfolgsdefinition 9 mit den Variablen molar, praemolar, Alter und MedReinigungen.

sollte die Verteilung der Y-Werte auf die Klassen $Y = 0$ und $Y = 1$ gleichmäßig sein. Die erklärte Variable, hier der Behandlungserfolg, sollte in der Stichprobe die Werte 1 und 0 in ausreichender Anzahl annehmen. Keine Ausprägung sollte in weniger als 10% der Fälle angenommen werden, Urban and Mayerl (2006). Aus den Tabellen 4, 5 und 6 geht hervor, dass dies bei den Erfolgsdefinitionen 2, 3, 7, 8 und 9 erfüllt ist. Bei den anderen Definitionen sind die Misserfolgsereignisse unterrepräsentiert. Durch Betrachtung der Receiver-Operating Characteristic (ROC) wird auf dieses Problem des unbalancierten Datensatzes Rücksicht genommen, Krzanowski and Hand (2009). Außerdem soll-

te der Datensatz keine unvollständigen Informationen enthalten. Alle möglichen Kombinationen aus Kovariaten sollten in ausreichender Anzahl vertreten sein. Aufschluss darüber geben Kreuztabellen. Liegen in diesen keine Nullzellen vor, kommen alle möglichen Kombinationen vom Kovariaten im Modell vor, Urban and Mayerl (2006). In Tabelle 8 wird beispielhaft für das Modell für die Definition Erfolg 2 eine Kreuztabelle berechnet. Es sind kaum Nullzellen vorhanden. Ähnliche Ergebnisse ergeben sich für die anderen Erfolgsdefinitionen. Somit besteht das Problem der unvollständigen Information nicht. Des Weiteren sollte es keine unvollständige Separation geben. Eine unvollständige

	Alter zu Behandlungsbeginn			
MedReinigungen	(0,20]	(21,40]	(41,60]	>60
0-5	27	342	646	450
6-10	7	160	431	382
11-15	1	20	90	62
>15	0	6	23	8

	Alter zu Behandlungsbeginn			
vital	(0,20]	(21,40]	(41,60]	>60
0	70	531	1184	1236
1	79	672	989	758

	MedReinigungen			
vital	0-5	6-10	11-15	>15
0	804	535	108	16
1	661	445	65	21

Tabelle 8: Kreuztabelle für die Kovariaten vital, MedReinigungen und Alter für Definition Erfolg2.

Separation liegt vor, wenn die Werte $Y=1$ und $Y=0$ nur in Kombination mit bestimmten Werten der Kovariaten auftreten. Dies würde zu Problemen bei der ML-Schätzung führen, Urban and Mayerl (2006). Für alle Erfolgsdefinitionen werden Kreuztabellen berechnet, wie für Definition Erfolg 9 in Tabelle 7. Auch bei allen anderen Erfolgsdefinitionen enthalten die Tabellen keine oder wenige Nullzellen. Somit sind Misserfolgsfälle und Erfolgsfälle für alle Ausprägungen der Kovariaten vertreten. Eine weitere Situation, die zu Problemen führen kann, ist Multikollinearität. Diese liegt vor, wenn starke lineare Beziehungen zwischen den Kovariaten bestehen. In diesem Fall würde die ML-

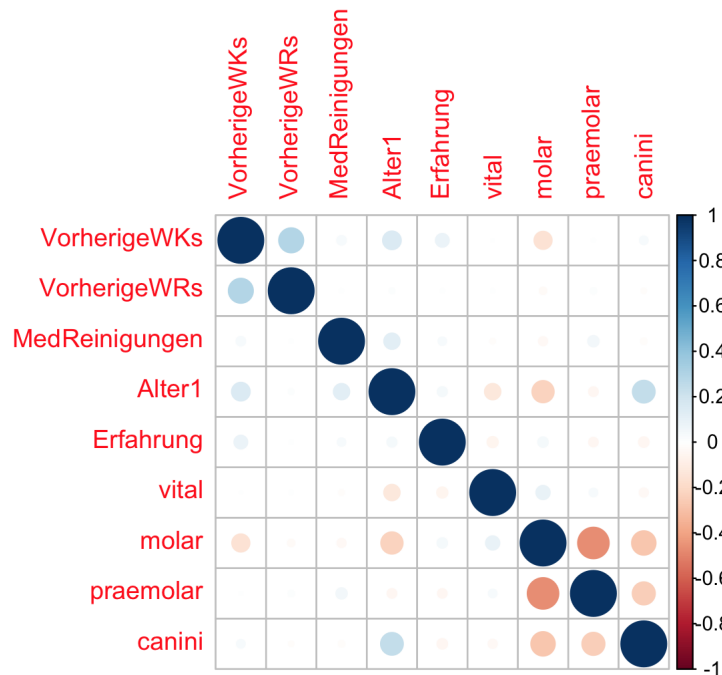


Abbildung 3: Korrelationsmatrix für den Datensatz, (Korrelationskoeffizient nach Pearson, blau: positiv linearer Zusammenhang, rot: negativ linearer Zusammenhang).

Schätzung hohe Standardfehler und eventuell zu hohe Regressionskoeffizienten liefern, Urban and Mayerl (2006). Um zu überprüfen, ob Zusammenhänge zwischen den Kovariaten bestehen, wird eine Korrelationsmatrix bestimmt, siehe Abbildung 3. Basis dafür ist der Korrelationskoeffizient nach Pearson. Große, gesättigte Punkte stehen für starke Zusammenhänge. Kleine, transparente Punkte deuten auf Werte des Korrelationskoeffizienten nahe 0 hin, siehe Legende Abbildung 3. Es ist ersichtlich, dass außerhalb der Diagonalen kaum große, gesättigte Punkte auftreten. Dies spricht dafür, dass die Werte der Korrelationen zwischen den jeweiligen Variablen nahe 0 liegen. Eine Berechnung mithilfe des Spearman'schen Rangkorrelationskoeffizienten ergibt ein ähnliches Ergebnis. Es liegen folglich kaum Zusammenhänge zwischen den Kovariaten vor. Das Problem von Multikollinearität besteht daher nicht.

4.2 Vorgehensweise

Um die Fragestellungen aus Kapitel 1.2 zu beantworten, wird folgende Vorgehensweise angewendet: Zunächst muss für jede Definition des Behandlungserfolgs ein geeignetes Modell gefunden werden, um die Zugehörigkeit zu den Erfolgsgruppen vorherzusagen. Um die Modelle evaluieren zu können, wird der Datensatz zunächst zufällig in einen Trainings- und einen Testdatensatz unterteilt.

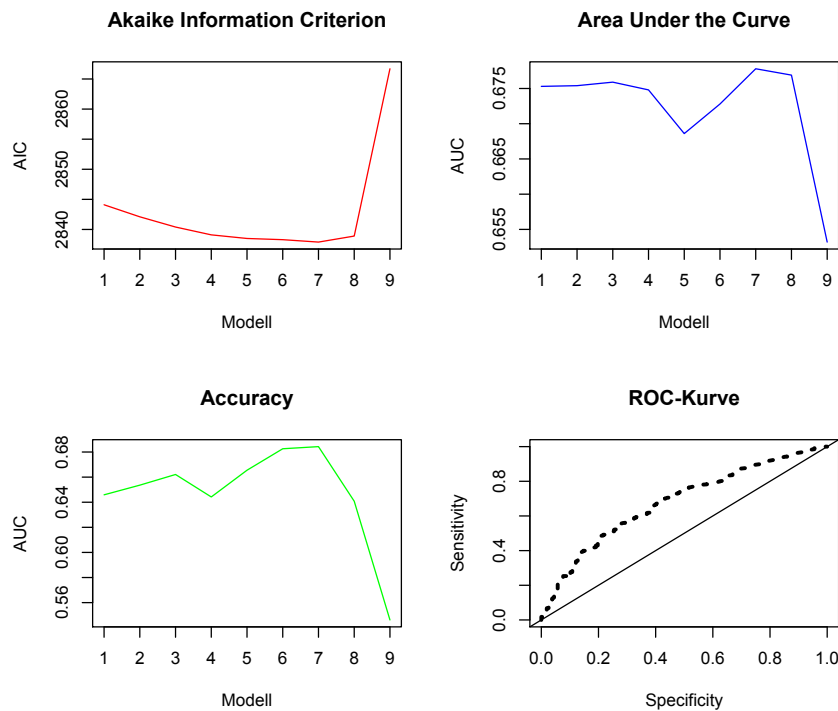


Abbildung 4: Kriterien bei der Modellwahl am Beispiel von Definition Erfolg3: AIC, AUC, Accuracy und ROC-Kurve des Modells mit der größten AUC: `erfolg4 ~ Alter + MedReinigungen + reci + vital`.

Der Trainingsdatensatz enthält 75% und der Testdatensatz 25% der Daten. Der Datensatz ist nicht ausbalanciert, d.h. die Anzahl der erfolgreichen Behandlungen übersteigt die der nicht erfolgreichen. Deshalb wird bei der Modellwahl die AUC der ROC-Kurve herangezogen, siehe Kapitel 2.2.4. Anlässlich der unbalancierten Daten wird ein Threshold abweichend vom Wert 0.5 festgesetzt, bei dem Sensitivität und Spezifität gleichzeitig möglichst hoch sind. Für den Testdatensatz werden mithilfe des im Trainingsdatensatz ermittelten Modells Erfolgswahrscheinlichkeiten und vorhergesagte Erfolge bzw. Misserfolge ausgegeben. Diese werden mit den tatsächlichen Ergebnissen im Testdatensatz verglichen. Daraus wird eine Klassifikationsmatrix und die Korrektklassifikationsrate (Accuracy) bestimmt, siehe Kapitel 2.2.3. In Kombination dazu werden mithilfe des Akaike Information Criterion (AIC), siehe 2.2.6, und dem Abwärtsverfahren der Grad der Anpassung und die Komplexität des Modells berücksichtigt. Gestartet wird beim vollen Modell mit allen verfügbaren Regressoren. In jedem Schritt wird der am wenigsten signifikante Regressor, gemessen am P-Wert, entfernt. Diese P-Werte basieren auf dem Wald-Test, siehe Kapitel 2.2.5. Die binäre Variable, die die Behandlungsgruppenzugehörigkeit kennzeichnet (`reci`), wird dabei nicht aus dem Modell entfernt, da sie für den späteren Vergleich der Behandlungsgruppen von Nöten ist. Anschließend wird das ausgewählte Modell noch-

Erfolgsdefinition	McFadden	Nagelkerke	Cox&Snell
1	0.0507	0.0645	0.0266
2	0.0479	0.065	0.0479
3	0.0459	0.0662	0.0373
4	0.0169	0.0195	0.0051
5	0.0144	0.0173	0.0054
6	0.0088	0.0108	0.0037
7	0.0332	0.0452	0.0222
8	0.0312	0.0453	0.0256
9	0.0296	0.0455	0.028

Tabelle 9: Pseudo- R^2 nach McFadden, Nagelkerke und Cox&Snell für die jeweiligen Erfolgsdefinitionen.

mals auf Basis des ungeteilten Datensatzes geschätzt. Um die Güte der Anpassung an die Daten zu messen, werden die Pseudo- R^2 s von McFadden, Nagelkerke und CoxSnell berechnet, siehe Kapitel 2.2.1. Außerdem wird mithilfe des Likelihood-Ratio-Tests (LR-Test), siehe 2.2.2, geprüft, ob sich das gewählte Modell signifikant vom Nullmodell unterscheidet. Getestet wird sowohl gegen ein Modell aus Interzept und der Dummyvariable *reci*, als auch gegen das Nullmodell. Um zu prüfen, ob die Behandlungsmethode Recipro[®] höhere Erfolge aufweist, werden die kumulierten Verteilungsfunktionen der Erfolgswahrscheinlichkeiten und der prognostizierten Ausgänge für die Behandlungen der Patienten ohne bzw. mit Recipro[®] mit dem Zwei-Stichproben KS-Test verglichen, siehe Kapitel 2.3.

4.3 Modellwahl

Die Fläche unter der ROC-Kurve (AUC) wird als Hauptentscheidungskriterium herangezogen, um für jede Erfolgsdefinition ein Modell zu finden. Abbildung 4 zeigt, wie sich die Kennzahlen AIC, AUC und die Korrektklassifikationsrate (Accuracy) im Abwärtsverfahren entwickeln. Am Beispiel von Erfolgsdefinition 3, wird das AIC durch Modell 7 ($\text{erfolg4} \sim \text{Alter} + \text{MedReinigungen} + \text{reci} + \text{vital}$) minimiert. Gleichzeitig wird bei diesem Modell ein Maximum für die AUC mit einem Wert von 0.6622 gefunden. Auch die Korrektklassifikationsrate wird durch Modell 7 maximiert. Unten rechts in Abbildung 4 wird die ROC-Kurve dieses Modells dargestellt. Wie bereits in Kapitel 2.2.4 erklärt, ist das Modell aussagekräftiger als eine zufällige Zuordnung, da die ROC-Kurve oberhalb der Diagonalen durch den Quadranten verläuft. Somit wird Modell 7 als erklärendes Modell für Erfolgsdefinition 3 ausgewählt und auf Basis der kompletten Daten geschätzt. Nach dem

Regressor	Enthalten in den Modellen zu folgenden Erfolgsdefinitionen	# Modelle in denen Regressor enthalten ist
vital	1, 2, 3, 4, 5, 6, 7, 8, 9	8
Alter	1, 2, 3, 6, 7, 8, 9	7
MedReinigungen	1, 2, 3, 6, 7, 8, 9	7
Erfahrung	1, 2, 4, 5, 6, 7, 8	7
VorherigeWRs	4, 5, 6, 7, 8	5
molar	5, 6, 9	3
praemolar	5, 6, 9	3
canini	6, 7, 8	3
VorherigeWKs	1	1

Tabelle 10: Übersicht über die **Regressoren** in den jeweiligen Modellen für die Erfolgsdefinitionen 1-9.

Erfolgsdefinition	1	2	3	4	5	6	7	8	9
AUC	0.657	0.657	0.662	0.65	0.572	0.564	0.65	0.619	0.629
Accuracy	0.649	0.649	0.627	0.665	0.635	0.595	0.594	0.585	0.584

Tabelle 11: AUC und **Korrektklassifikationsrate (Accuracy)** der Modelle für die Erfolgsdefinitionen 1-9.

gleichen Vorgehen wird für die übrigen Erfolgsdefinitionen ein Modell bestimmt. Die ermittelten Werte der AUC befinden sich im Bereich $[0.564, 0.6622]$, siehe Tabelle 11. Vergleicht man das mit den Richtwerten aus Kapitel 2.2.4, ergibt sich lediglich eine mittelmäßige bis akzeptable Trennung der Erfolgs- und Misserfolgsgruppe durch das Modell. Die Korrektklassifikationsraten bewegen sich zwischen $[0.5836, 0.6654]$, siehe Tabelle 11.

4.4 Likelihood-Ratio-Tests

Mit Likelihood-Ratio-Tests (LR-Tests) wird geprüft, ob sich die gewählten Gesamtmodelle von einem Nullmodell und einem Modell mit der Variable rec_i unterscheiden. Die Hypothesen des Tests sind Folgende:

$$H_0 : \text{Das Modell ist gleich dem Nullmodell/Nullmodell mit } rec_i, \quad (4.1)$$

$$H_1 : \text{Das Modell ist nicht gleich dem Nullmodell/Nullmodell mit } rec_i.$$

Erfolgsdefinition	Likelihood-Ratio-Tests (Signifikanzniveau)	
	LR-Test zwischen Modell x und Nullmodell + Variable reci	LR-Test zwischen Modell x und Nullmodell
1	***	***
2	***	***
3	***	***
4	***	***
5	***	***
6	* (0.034)	0.044
7	***	***
8	***	***
9	***	***

Tabelle 12: Ergebnisse der LR-Tests, (Signifikanzcodes: 0 ‘***’, 0.001 ‘**’, 0.01 ‘*’, 0.05 ‘.’, 0.1 ‘ ’ 1).

Betrachtet werden die ausgewählten Modelle der jeweiligen Erfolgsdefinitionen. Tabelle 12 gibt eine Übersicht über die Ergebnisse dieser Tests. Bei fast allen Modellen liefert der Test extrem kleine P-Werte und lässt damit die Aussage zu, dass sich die Modelle signifikant von den Nullmodellen unterscheiden. Lediglich beim Modell zu Erfolgsdefinition 6 ergibt der Test P-Werte von 0.034 und 0.044. Aber selbst in diesem Fall ist die Nullhypothese bei einem Signifikanzniveau von 5% zu verwerfen.

4.5 Pseudo- R^2

Tabelle 9 stellt die errechneten Pseudo- R^2 -Maße, siehe Kapitel 2.2.1, für die ausgewählten Modelle dar. Die Maßzahlen nehmen keine besonders hohen Werte an, obwohl Pseudo- R^2 s in der Theorie den Bereich $[0, 1]$ bzw. $[0, 1)$ erreichen können. Kopp (2014) behauptet jedoch, dass die Pseudo- R^2 s stark vom Sachverhalt und von den betrachteten Kovariaten abhängen und Werte nahe 1 in der Praxis selten zu erwarten sind. Außerdem indizieren die LR-Tests, die sich genauso wie die Pseudo- R^2 s auf die Loglikelihood beziehen, eine Verbesserung gegenüber dem Nullmodell, siehe Kapitel 4.4.

4.6 Einflussnehmende Kovariaten

Die Modelle geben Auskunft darüber, welche Variablen auf Basis der Daten einen Einfluss auf den Behandlungserfolg haben. Tabelle 10 stellt dar, in welchen Modellen welche Regressoren enthalten

sind. Diese sind nach ihrer Häufigkeit in den Modellen aufgelistet. Die Variable *vital* ist in fast allen Modellen vorhanden, genauso wie die Regressoren *MedReinigungen* und *Alter*. Kommen diese 3 Variablen in den endgültigen Modellen vor, sind sie in der Regel auch signifikant auf einem Signifikanzniveau von 5% nach dem Wald-Test. Demhingegen sind die Variablen *Vorherigewks*, sowie die Dummyvariablen zum Zahntyp *molar*, *praemolar* und *canini* in wenigen Modellen vertreten.

4.7 Stärke und Richtung der einflussnehmenden Kovariaten

Erfolgsdefinition	1	2	3	4	5	6	7	8
$\hat{\beta}_{vital}$	0.2324	0.2103	0.1863	0.4543	0.1867	0.0756	0.3034	0.2147
Signifikanzniveau	*	*	*	***	0.16	0.57	***	**
Effektkoeffizient	1.2616	1.234	1.2048	1.5751	1.2053	1.0785	1.3545	1.2395

Tabelle 13: Die geschätzten **Koeffizienten** $\hat{\beta}_{vital}$ und die **Effektkoeffizienten** $e^{\hat{\beta}_{vital}}$ für die Erfolgsdefinitionen 1-8.

Erfolgsdefinition	1	2	3	6	7	8	9
$\hat{\beta}_{Alter}$	-0.0363	-0.0346	-0.032	0.0081	-0.0248	-0.024	-0.0205
Signifikanzniveau	***	***	***	.	***	***	***
Effektkoeffizient	0.9644	0.966	0.9685	1.0081	0.9755	0.9763	0.9797

Tabelle 14: Die geschätzten **Koeffizienten** $\hat{\beta}_{Alter}$ und die **Effektkoeffizienten** $e^{\hat{\beta}_{Alter}}$ für die Erfolgsdefinitionen 1, 2, 3, 6, 7, 8, 9.

Die geschätzten Koeffizienten in den Modellen können nach dem in Kapitel 2.1.5 beschriebenen Vorgehen interpretiert werden. Positive Koeffizienten $\hat{\beta}_j$ erhöhen die Wahrscheinlichkeit auf Behand-

Erfolgsdefinition	1	2	3	7	8	9
$\hat{\beta}_{MedReinigungen}$	0.0714	0.0731	0.0784	0.0485	0.0619	0.0609
Signifikanzniveau	***	***	***	***	***	***
Effektkoeffizient	1.074	1.0758	1.0816	1.0497	1.0639	1.0628

Tabelle 15: Die geschätzten **Koeffizienten** $\hat{\beta}_{MedReinigungen}$ und die **Effektkoeffizienten** $e^{\hat{\beta}_{MedReinigung}}$ für die Erfolgsdefinitionen 1, 2, 3, 7, 8, 9.

Erfolgsdefinition	1	4	5	6	7	8
$\hat{\beta}_{Erfahrung}$	0.0004	0.0007	0.0011	0.0009	0.0004	0.0008
Signifikanzniveau	***	***	***	**	**	***
Effektkoeffizient	1.0004	1.0007	1.0011	1.0009	1.0004	1.0008

Tabelle 16: Die geschätzten **Koeffizienten** $\hat{\beta}_{Erfahrung}$ und die **Effektkoeffizienten** $e^{\hat{\beta}_{Erfahrung}}$ für die Erfolgsdefinitionen 1, 4, 5, 6, 7, 8.

Erfolgsdefinition	4	5	6	7	8
$\hat{\beta}_{VorherigeWRs}$	-0.3965	-0.3696	-0.2712	-0.2022	-0.2011
Signifikanzniveau	*	*	0.13	0.15	0.13
Effektkoeffizient	0.6727	0.691	0.7625	0.8169	0.8178

Tabelle 17: Die geschätzten **Koeffizienten** $\hat{\beta}_{VorherigeWRs}$ und die **Effektkoeffizienten** $e^{\hat{\beta}_{VorherigeWRs}}$ für die Erfolgsdefinitionen 4-8.

Erfolgsdefinition	1
$\hat{\beta}_{VorherigeWKs}$	-0.0414
Signifikanzniveau	0.1336
Effektkoeffizient	0.9594

Tabelle 18: Die geschätzten **Koeffizienten** $\hat{\beta}_{VorherigeWRs}$ und die **Effektkoeffizienten** $e^{\hat{\beta}_{VorherigeWRs}}$ für Erfolgsdefinition 1.

Erfolgsdefinition	6	7	8
$\hat{\beta}_{canini}$	0.0643	-0.195	-0.1631
Signifikanzniveau	0.75	.	0.11
Effektkoeffizient	1.0664	0.8228	0.8495

Tabelle 19: Die geschätzten **Koeffizienten** $\hat{\beta}_{canini}$ und die **Effektkoeffizienten** $e^{\hat{\beta}_{canini}}$ für die Erfolgsdefinitionen 6-8.

Erfolgsdefinition	5	6	9
$\hat{\beta}_{molar}$	0.3175	0.3257	0.3425
Signifikanzniveau	.	*	***
Effektkoeffizient	1.3737	1.385	1.4085

Tabelle 20: Die geschätzten **Koeffizienten** $\hat{\beta}_{molar}$ und die **Effektkoeffizienten** $e^{\hat{\beta}_{molar}}$ für die Erfolgsdefinitionen 5, 6, 9.

Erfolgsdefinition	5	6	9
$\hat{\beta}_{praemolar}$	0.1909	0.2215	0.2016
Signifikanzniveau	0.21	0.21	*
Effektkoeffizient	1.2103	1.2479	1.2234

Tabelle 21: Die geschätzten Koeffizienten $\hat{\beta}_{praemolar}$ und die **Effektkoeffizienten** $e^{\hat{\beta}_{praemolar}}$ für die Erfolgsdefinitionen 5, 6, 9.

lungserfolg, wenn die dazugehörige unabhängige Variable X_j um eine Einheit ansteigt. Die Chance ändert sich durch die Erhöhung diese Variable um den Faktor $e^{\hat{\beta}_j}$. Tabelle 13 gibt eine Übersicht über die geschätzten Koeffizienten für die Variable `vital`. Außerdem wird der P-Wert des Wald-Tests für den jeweiligen Koeffizienten angegeben. Das Signifikanzniveau richtet sich nach den in Tabelle 12 beschriebenen Signifikanzcodes. Die Variable `vital` kommt in den Modellen zu allen Erfolgsdefinitionen, außer Definition Erfolg 9, vor und ist stets positiv. Daraus kann geschlossen werden, dass der vitale Zustand des Zahnes die Chance auf eine erfolgreiche Behandlung erhöht. Der geschätzte Koeffizient $\hat{\beta}_{vital}$ nimmt Werte zwischen 0.1867 bei Definition 5 und 0.4543 bei Definition 4 an. Hat $\hat{\beta}_{vital}$ bei Definition 4 den Wert 0.4543, ergibt sich der dazugehörige Effektkoeffizient zu $e^{\hat{\beta}_{vital}} = e^{0.4543} = 1.5751$. Betrüge die Chance vorher 50:50, ändert sie sich durch die Änderung der Variable `vital` von 0 zu 1 auf 61:39. Die Tabellen 14 bis 18 geben Auskunft über die geschätzten Koeffizienten $\hat{\beta}_j$ und Effektkoeffizienten $e^{\hat{\beta}_j}$. Zur Übersichtlichkeit wird anstatt nummerierender Indizes der Variablenname für j angegeben. Die Variable `Alter` ist in der Regel negativ (außer bei Definition 6). Das deutet darauf hin, dass ein zunehmendes Alter die Chance auf eine erfolgreiche Behandlung verringert. Die Effektkoeffizienten betragen je nach Modell zwischen 0.9644 und 1.0081. Die geschätzten Koeffizienten für die Variable `MedReinigungen` sind immer positiv. Somit hat die Anzahl der Zahnreinigungen nach der Behandlung einen positiven Effekt auf den Erfolg der Behandlung. Die Effektkoeffizienten betragen je nach Modell zwischen 1.0497 und 1.0816. Ebenso erhöht sich die Wahrscheinlichkeit auf Behandlungserfolg, wenn die Variable `Erfahrung` um eine Einheit steigt,

da auch hier die Koeffizienten $\hat{\beta}_j$ für die jeweiligen Modelle positiv sind. Hat ein Behandler eine WKB mehr durchgeführt, erhöht sich für die nächste WKB in der Regel die Erfolgswahrscheinlichkeit. Die Werte der Effektkoeffizienten sind in Tabelle 16 aufgeführt und bewegen sich im Intervall $[1.0004, 1.0011]$. Die Variablen, die sich auf die Zahntypen beziehen, sind in den Tabellen 19, 20 und 21 aufgeführt. Wider der Erwartung, dass Behandlungen an den Molaren und Praemolaren zu schlechteren Erfolgsaussichten der WKB führen, sind die Koeffizienten $\hat{\beta}_{molar}$ und $\hat{\beta}_{praemolar}$ positiv. Das bedeutet, dass durch die Veränderung der Variablen *molar* und *praemolar* von 0 auf 1 die Chance auf eine erfolgreiche Behandlung steigt. Je nach Modell ändert sich die Chance um einen Faktor zwischen 1.3737 und 1.4085, wenn der Zahn zur Gruppe der Molaren gehört. Gehört der Zahn jedoch zur Gruppe Canini, erhöht sich die Variable *canini* von 0 auf 1. Die Effektkoeffizienten zur Variable *canini* belaufen sich zwischen 0.8228 und 1.0664. Vorherige WKB oder WSR senken die Chance auf eine erfolgreiche Behandlung, da die geschätzten Koeffizienten negativ sind.

4.8 Vergleich der Behandlungsgruppen

4.8.1 Erfolgswahrscheinlichkeiten im Zeitverlauf

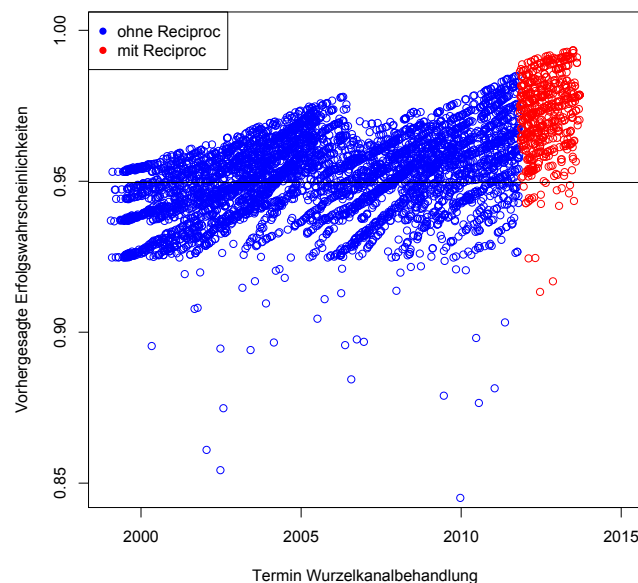


Abbildung 5: Vorhergesagte **Erfolgswahrscheinlichkeiten im Zeitverlauf**, abhängig vom Termin der WKB, unterteilt in Behandlungen mit (rot) und ohne (blau) Reciproc[®] am Beispiel von Definition Erfolg5.

Der Plot in Abbildung 5 stellt die prognostizierten Erfolgswahrscheinlichkeiten im Zeitverlauf am

Beispiel vom Modell zu Erfolgsdefinition 5 dar. Die horizontale Gerade kennzeichnet den optimalen Threshold. Für prognostizierte Wahrscheinlichkeiten oberhalb des Thresholds wird der Ausgangswert Erfolg vorhergesagt. Es wird deutlich, dass der Großteil der Fälle, bei denen Recipro[®] angewendet wurde, oberhalb der Trennlinie liegt und somit der Gruppe der erfolgreichen Behandlungen zuzuordnen ist. Es ist bei keiner Erfolgsdefinition der Fall, dass die prognostizierten Wahrscheinlichkeiten der Behandlungen ohne Recipro[®] die der mit Recipro[®] übersteigen.

4.8.2 Zwei-Stichproben Kolmogorov-Smirnov-Tests

Um den Trend aus Abbildung 5 statistisch zu überprüfen, wird der Zwei-Stichproben KS-Test angewendet, siehe Kapitel 2.3. Hierfür wird der Datensatz inklusive prognostizierter Erfolgswahrscheinlichkeiten in eine Gruppe mit Behandlungen ohne Recipro[®] und eine Gruppe mit Behandlungen mit Recipro[®] unterteilt. Die Verteilung der Wahrscheinlichkeiten in den beiden Gruppen soll nun auf Verschiedenheit getestet werden. Hierfür wird zunächst zweiseitig getestet, um die Nullhypothese der Gleichheit der Wahrscheinlichkeitsverteilungen zu widerlegen. Im Anschluss daran wird

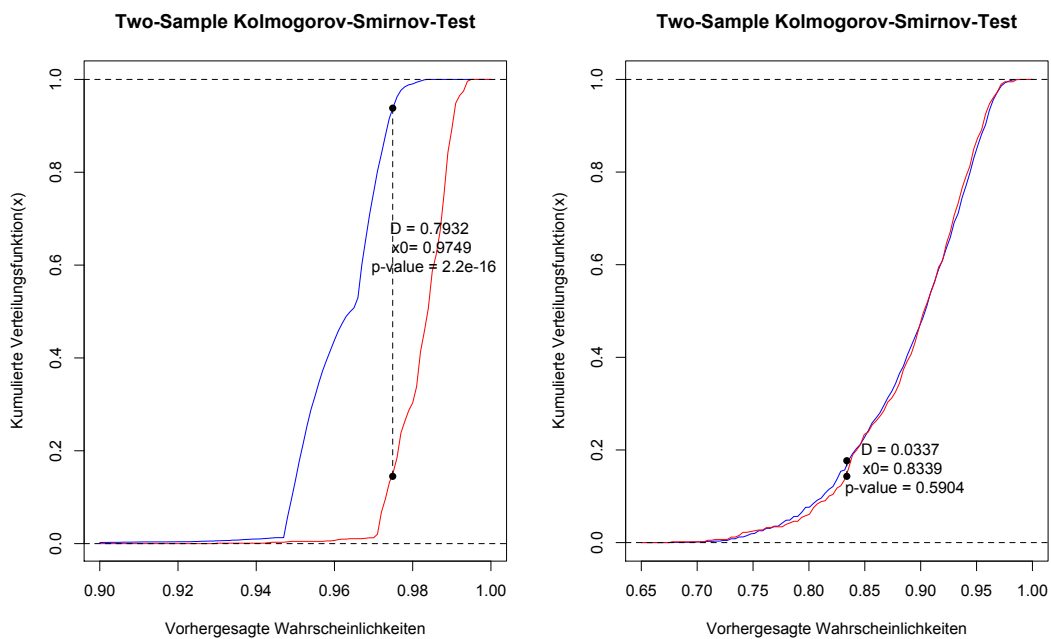


Abbildung 6: Ergebnisse der **KS-Tests** der Modelle für die Erfolgsdefinitionen 4 (links) und 2 (rechts).

ein einseitiger Test durchgeführt, um zu zeigen, dass die Verteilungsfunktion der Behandlungen mit Recipro[®] Methoden auf höheren Werten konzentriert ist. Dies ist der Fall, wenn die Nullhypothese, dass die Verteilungsfunktion der Recipro[®] Behandlungen oberhalb oder gleich der Behandlungen

Erfolgsdefinitionen	1	2	3	4	5	6	7	8	9
P-Wert zweiseitig (Erfolgs-WK)	0.304	0.59	0.0003	2.2e-16	2.2e-16	2.2e-16	2.2e-16	1.4e-06	1.1e-06
P-Wert einseitig (Erfolgs-WK)	0.681	0.304	1.5e-06	2.2e-16	2.2e-16	2.2e-16	2.2e-16	7.2e-07	5.6e-07
P-Wert einseitig (prog. Erfolge)	1	0.784	0.0003	2.2e-16	2.2e-16	8.5e-15	7.7e-05	0.0238	3.1e-06

Tabelle 22: Übersicht über die **P-Werte der KS-Tests** für die jeweiligen Modelle der Erfolgsdefinitionen 1-9.

ohne Recipro[®] liegt, verworfen werden kann. Dann kann auf Basis eines bestimmten Signifikanzniveaus behauptet werden, dass die Recipro[®] erfolgsversprechender bezüglich der beschriebenen Definitionen als die vorherigen Behandlungsmethoden ist. Abbildung 6 veranschaulicht die beiden Verteilungsfunktionen des unterteilten Datensatzes für die Definitionen Erfolg4 und Erfolg2. Dargestellt wird der zweiseitige Test auf Ungleichheit der Verteilungsfunktionen. Als Teststatistik wird der größte betragsmäßige Abstand zwischen den beiden Funktionen verwendet. Die beiden Abbildungen zeigen einen KS-Test, bei dem sich die Verteilungsfunktionen stark unterscheiden, siehe links. Der maximale Abstand der Funktionen ist relativ groß und die Verteilungsfunktion der Recipro[®] Behandlungen ist stets unterhalb der der Verteilungsfunktion der Behandlungen ohne Recipro[®]. Erstere ist somit auf höheren Erfolgswahrscheinlichkeiten konzentriert. Die Nullhypothese der Gleichheit der Verteilungen kann selbst auf einem Signifikanzniveau von $2.2e^{-16}$ gerade noch verworfen werden. Demhingegen stellt die Abbildung rechts, siehe Abbildung 6, einen Test dar, bei dem sich die Verteilungsfunktionen kaum unterscheiden. In diesem Fall beträgt der maximale Abstand der Funktionen lediglich 0.0337 und der P-Wert des Tests liegt bei 0.5904. Tabelle 22 veranschaulicht die Ergebnisse der KS-Tests für alle Erfolgsdefinitionen. Getestet wurden einerseits die Vektoren der prognostizierten Erfolgswahrscheinlichkeiten als auch die prognostizierten Erfolge in Form von binären Werten 0, 1. Abgesehen von den Modellen für die Definitionen Erfolg1 und Erfolg2 kann bei allen Modellen behauptet werden, dass sowohl die Wahrscheinlichkeitsverteilung als auch die Verteilung der prognostizierten Ereignisse signifikant verschieden ist. Außerdem liefern die einseitigen Tests, dass sowohl bezüglich der Wahrscheinlichkeiten als auch bezüglich der prognostizierten Er-

folge die Recipro[®] Behandlungen höhere Erfolgswahrscheinlichkeiten, bzw. eine größere Anzahl an prognostizierten Erfolgen vorweisen. Dies bestätigt den Eindruck, der bereits aus den Plots der Erfolgswahrscheinlichkeiten im Zeitverlauf, in Abbildung 5, entstanden ist.

5 Diskussion der Ergebnisse

Die Analyse des Datensatzes weist einige Schwachstellen auf, auf die in diesem Teil eingegangen wird. Die logistischen Regressionsmodelle werden mit Maximum-Likelihood-Schätzung bestimmt. Ist der Stichprobenumfang N groß, liefert MLE recht zuverlässige Ergebnisse, Urban (1993). Bei den Erfolgsdefinitionen 3, 6 und 9 sind die Stichprobengrößen mit $N = 4691$ und 215 Behandlungen, bei denen Recipro[®]-Instrumente angewendet wurden, am kleinsten. Das Modell für Erfolgsdefinition 6 fällt mehrfach durch Ergebnisse auf, die nicht im Einklang mit den übrigen Modellen stehen. Die geschätzten Koeffizienten fallen in der Regel durch einen hohen P-Wert auf, siehe Tabellen 13 bis 21. Daneben haben die Koeffizienten für die Variablen Canini und Alter andere Vorzeichen als in den Modellen für die übrigen Erfolgsdefinitionen. Das könnte auf eine Verzerrung hinweisen. Auch bei dem LR-Test sind die P-Werte im Vergleich zu den übrigen Modellen am höchsten, siehe Tabelle 4.4. Generell liefern die Modelle relativ niedrige AUC-Werte und Pseudo- R^2 s. Das könnte verschiedene Gründe haben. Möglicherweise gibt es Variablen, die den den Behandlungserfolg beeinflussen, die aber nicht im Datensatz enthalten sind. Die Variation in Y , bzw. π , die durch diese Variablen verursacht wird, kann also keiner bestimmten Variable zugeordnet werden. Weitere Daten, zum Beispiel zur Mundhygiene des Patienten oder zu Vorerkrankungen würden eventuell zu besseren Modellen führen. Ein weiteres Problem bei der durchgeführten Analyse könnte sich durch fehlende Werte ergeben. Der Datensatz beinhaltet nur Behandlungen, die in der AZFK durchgeführt wurden. Die Variablen `VorherigeWks` und `VorherigeWRs` nehmen nur den Wert 1 an, sofern beim Patienten eine WKB oder WSR in der AZFK durchgeführt wurde. Hatte der Patient eine WSR oder WKB, die außerhalb der AFZK stattfand, wird dies nicht im Datensatz berücksichtigt. In diesem Fall liegen bei den Variablen `VorherigeWks` und `VorherigeWRs` falsche Informationen vor. Das gleiche Problem tritt bei Variablen auf, die sich auf den Zeitraum nach der Behandlung beziehen, wie die Anzahl an Zahnreinigungen nach der WKB am behandelten Zahn. Zahnreinigungen, die außerhalb der AZFK vorgenommen wurden, sind nicht protokolliert. Auch bei den Variablen, die sich auf die Erfolgsdefinitionen beziehen (Extraktion und Revision), könnten Falschinformationen auftreten. Für Patienten, die nach der WKB in der AZFK die Zahnklinik gewechselt haben, liegen keine Informationen über Extraktionen oder Revisionen vor. Möglicherweise wurde bei einem Patienten nach der WKB außerhalb der AZFK eine Extraktion oder Revision durchgeführt. Dies wäre dann je nach Definition

als Misserfolg zu werten. Fand die Nachbehandlung aber nicht in der AZFK statt, wird die WKB als erfolgreich gewertet. Auch die Erfahrung des behandelnden Arztes ist ein Maß, das verschieden definiert werden könnte. In der Studie bezieht sich die Variable auf die durchgeführten WKB an der AFZK. Auch hier werden keine Erfahrungen oder Qualifikationen berücksichtigt, die der Arzt außerhalb der AZFK oder durch andere Behandlungen gesammelt hat. Die zusätzliche Erhebung von Patientendaten, auch außerhalb der AZFK, würde vermutlich zu zuverlässigeren Ergebnissen führen.

6 Zusammenfassung der Ergebnisse

Die Arbeit befasst sich mit der Fragestellung, welche Kovariaten den Behandlungserfolg einer WKB beeinflussen und ob die Anwendung der Reciproc[®]-Instrumente zu einer Erhöhung des Behandlungserfolgs führt. Das Verfahren der logistischen Regression liefert, dass in erster Linie die Vitalität des Zahnes, das Patientenalter und die Durchführung von Zahnreinigungen nach der WKB eine Rolle spielen. Der vitale Zustand des Zahnes, medizinische Zahnreinigungen sowie ein jüngeres Patientenalter verbessern die Aussicht auf eine erfolgreiche WKB. Außerdem liefern KS-Tests auf Verteilung, dass Reciproc[®] in der Regel zu einer höheren Erfolgswahrscheinlichkeit führt, als die Anwendung der konventionellen Methoden. Es bleiben aber noch Punkte unbeantwortet. Es sollten noch zusätzliche Kovariaten herangezogen werden, um zuverlässigere Ergebnisse zu erhalten. Außerdem kann es zu Fehlinterpretationen kommen, da sich der Datensatz auf Behandlungen in der AZFK und nicht auf bestimmte Patienten bezieht. Möglicherweise wurden vor oder nach einer WKB in der AZFK eine Revisionsbehandlung oder eine Extraktion durchgeführt. Geschieht das außerhalb der AZFK, wird die Behandlung nicht erfasst.

Literatur

- ANGRIST, J. D. AND J.-S. PISCHKE (2008): *Mostly Harmless Econometrics*, Princeton: Press Princeton.
- BARTOLS, A. (2013): "Clinical Experiences with Reciproc," *Endo*, 7.
- BECKMANN, M. (1999): *Ökonomische Analyse deutscher Auktionen*, Wiesbaden: Springer Fachmedien Wiesbaden GmbH.
- BEHNKE, J. (2015): *Logistische Regressionsanalyse : Eine Einführung*, Methoden der PolitikwissenschaftSpringerLink : Bücher, Wiesbaden: Springer VS.
- BRAMER, M. (2016): *Principles of Data Mining*, Undergraduate Topics in Computer ScienceSpringerLink : Bücher, London: Springer, 3rd ed. 2016 ed.
- BRUIN, J. (2011): "Stata Programs for data analysis," Available at <http://www.ats.ucla.edu/stat/stata/ado/analysis/> (30.12.16).
- BURNHAM, K. P. AND D. R. ANDERSON (2004): "Multimodel Inference Understanding AIC and BIC in Model Selection," *Sage Journals*.
- CALIENDO, M. AND S. KOPEINIG (2005): "Some Practical Guidance for the Implementation of Propensity Score Matching," *Discussion Papers 485*.
- CRONE, S. F. AND S. FINLAY (2012): "Instance sampling in credit scoring: An empirical study of sample size and balancing," *International Journal of Forecasting*, 28, 224–238.
- DREGER, C. (2014): *Ökonometrie : Grundlagen – Methoden – Beispiele*, SpringerLink : Bücher, Wiesbaden: Springer Gabler, 5., überarb. u. akt. Aufl. 2014 ed.
- FAWCETT, T. (2006): "An introduction to ROC analysis," *Pattern Recognition Letters*, 27, 861–874.
- FIELD, A., J. MILES, AND Z. FIELD (2012): *Discovering statistics using R*, Los Angeles, Calif. [u.a.]: Sage, 1. publ. ed.
- GROSS, M. (2015): "Logistische Regression - Modell und Grundlagen," <https://www.inwt-statistics.de/blog-artikel-lesen/LogistischeRegressionKlassifikationsguete.html>.
- GRZEGORCZYK, M. (2003): "Auswertung einer Brustkrebsstudie mit Hilfe von logistischen Regressionsmodellen," .

- HAIJIAN-TILAKI, K. (2013): “Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation,” *Caspian J. Intern Med*, 4, 627–635.
- HAN, D. J. (2009): “Measuring classifier performance: a coherent alternative to the area under the ROC curve,” *MachLearn*, 77, 103–123.
- HANLEY, J. A. AND B. J. MCNEIL (1982): “The meaning and use of the area under the ROC curve,” *Radiology*, 143, 29–34.
- HANLEY, J. A. AND B. M. NEIL (1983): “A Method of Comparing the Areas under Receiver Operating Characteristic Curves derived from the same class,” *Radiology*, 148, 839–843.
- HARRELL, JR., F. E. (2015): “Regression Modeling Strategies : With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis,” .
- HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2001): *The Elements of Statistical Learning*, Stanford: Springer.
- HEDDERICH, J. (2016): *Angewandte Statistik : Methodensammlung mit R*, Berlin: Springer Spektrum, 15., überarb. u. erweiterte Aufl. ed.
- HEERINGA, S., B. T. WEST, AND P. A. BERGLUND (2010): *Applied survey data analysis*, Statistics in the social and behavioral sciences series A Chapman Hall book, Boca Raton [u.a.]: CRC Press.
- HOSMER, D. W. AND S. LEMESHOW (2000): *Applied logistic regression*, Wiley series in probability and statistics, New York [u.a.]: Wiley, 2. ed. ed.
- HUJER, R. (2005): “Vorlesung Mikroökonomie,” Available at http://www.wiwi.uni-frankfurt.de/Professoren/hujer/Lehre/0ek_2/skript.pdf(25.12.16).
- IMAI, K. AND M. RATKOVIC (2013): “Covariate balancing propensity score,” *Journal of the Royal Statistical Society*, 243–263.
- KING, G. AND L. ZENG (2001): “Logistic Regression in Rare Events Data,” *Political Analysis*, 9, 137–163.
- KONISHI, S. AND G. KITAGAWA (2008): *Information criteria and statistical modeling*, Springer series in statistics, New York: Springer.
- KOPP, J. (2014): *Sozialwissenschaftliche Datenanalyse : Eine Einführung*, SpringerLink : Bücher, Wiesbaden: Springer VS, 2., überarb. u. aktualisierte Aufl. 2014 ed.

- KRZANOWSKI, W. J. AND D. J. HAND (2009): *ROC curves for continuous data*, Monographs on statistics and applied probability ; 111, Boca Raton [u.a.]: CRC Press.
- KUHN, M. (2008): “Building Predictive Models in R Using the caret Package,” *Journal of Statistical Software*, 08, 1–26.
- LEGEWIE, J. (2012): “Die Schätzung von kausalen Effekten: Überlegungen von Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule,” *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 123–153.
- LIN, P.-C. AND B. WU (2010): “Kolmogorov-Smirnov Two Sample Test with Continuous Fuzzy Data,” *ResearchGate*.
- MERK, A. (2011): *Optionsbewertung in Theorie und Praxis*, Wiesbaden: Gabler Verlag.
- PENG, C.-Y. J., K. L. LEE, AND G. M. INGERSOLL (2002): “An Introduction to Logistic Regression Analysis and Reporting,” *The Journal of Educational Research*, 3–14.
- ROBIN, X., N. TURCK, A. HAINARD, N. TIBERTI, F. LISACEK, J.-C. SANCHEZ, AND M. MÜLLER (2015): “Package ‘pRoc’,” Available at <https://cran.r-project.org/web/packages/pROC/pROC.pdf> (30.12.16).
- SCHREIBER, M. (2015): “Einfluss der Aufarbeitungsmethode auf den Erfolg der endodontischen Behandlung - vollrotierende vs. reziproke Instrumentation,” .
- STAHLBOCK, R. H., S. F. H. CRONE, AND S. H. LESSMANN (2010): *Data Mining : Special Issue in Annals of Information Systems*, Annals of Information Systems: SpringerLink : Bücher, Boston, MA: Springer US, 1 ed.
- STEIN, P. D. (2014): “Multivariate Analyseverfahren - Logistische Regression,” Available at https://www.uni-due.de/imperia/md/content/soziologie/stein/logistische_regression_14-1-14.pdf(25.12.16).
- TUTZ, G. (2011): *Regression for Categorical Data*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
- URBAN, D. (1993): *Logit-Analyse: statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen*, Fischer.
- URBAN, D. AND J. MAYERL (2006): *Regressionsanalyse: Theorie, Technik und Anwendung*, Studien-skripten zur Soziologie, VS Verlag für Sozialwissenschaften.

WAGNER, S. (2015): “Logistische Regression - Beurteilung der Klassifikationsgüte,” Available at https://www.inwt-statistics.de/blog-artikel-lesen/Logistische_Regression.html (30.12.16).

WOLLSCHLÄGER, D. (2012): “Grundlagen der Datenanalyse mit R : Eine anwendungsorientierte Einführung,” .

WOOLRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Massachusetts: The MIT PRESS.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorstehende Bachelorarbeit selbstständig angefertigt und die benutzten Hilfsmittel, Quellen sowie die befragten Personen und Institutionen vollständig angegeben habe. Darüber hinaus versichere ich, dass ich die Regeln zur Sicherung guter wissenschaftlicher Praxis im Karlsruher Institut für Technologie (KIT) in der gültigen Fassung beachtet habe.

Karlsruhe, 17.03.2017

Luisa Werner