TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Verkehrswissenschaften „Friedrich List“ Institut für Wirtschaft und Verkehr

Professur für Ökonometrie und Statistik, insb. im Verkehrswesen

Masterarbeit

Modellierung des Unfallgeschehens im Radverkehr am Beispiel der Stadt Dresden

Jacqueline Martin

Geboren am: 13. Dezember 1996 in Wurzen
Matrikelnummer: 4527707

zur Erlangung des akademischen Grades

Master of Science (M. Sc.)

Betreuer

Dr. rer. pol. Iryna Okhrin
Dipl.-Ing. Sebastian Hantschel

Verantwortliche Hochschullehrer

Prof. Dr. rer. pol. Ostap Okhrin
Univ.-Prof. Dr.-Ing. Regine Gerike

Dresden, 22. Dezember 2020

Zusammenfassung

Das Radverkehrsaufkommen in Deutschland verzeichnete in den letzten Jahren einen Zuwachs, was sich im Umkehrschluss ebenfalls im Anstieg des Unfallgeschehens mit Radfahrendenbeteiligung widerspiegelt. Um den steigenden Unfallzahlen entgegenzuwirken, empfehlen Politik und Verbände v.a. Infrastrukturmaßnahmen zu ergreifen. Davon ausgehend untersucht die vorliegende Arbeit beispielhaft für die Stadt Dresden, wie sich einzelne Infrastrukturmerkmale auf das Unfallgeschehen zwischen Rad- und motorisiertem Verkehr auswirken. Die Datengrundlage der Untersuchung stellen dabei 548 Unfälle mit Radfahrendenbeteiligung aus den Jahren 2015 bis 2019 sowie die Merkmale von 484 Knotenpunktzufahrten dar. Da die Infrastruktur das Unfallgeschehen nicht allein determiniert, werden zudem Kenngrößen des Verkehrsaufkommens einbezogen. Um das Unfallgeschehen zu untersuchen, kommen das Random Forest-Verfahren sowie die Negative Binomialregression in Form von *Accident Prediction Models* mit vorheriger Variablenselektion anhand des LASSO-Verfahrens zum Einsatz. Die Verfahren werden jeweils auf zwei spezielle Unfalltypen für Knotenpunkte angewandt, um differenzierte Ergebnisse zu erlangen. Der erste Unfalltyp *Abbiege-Unfall* umfasst dabei Kollisionen zwischen einem rechtsabbiegenden und einem in gleicher oder entgegengesetzter Richtung geradeausfahrenden Beteiligten, während der zweite Unfalltyp *Einbiegen-/Kreuzen-Unfall* Kollisionen zwischen einem vorfahrtsberechtigten Verkehrsteilnehmenden und einem einbiegenden oder kreuzenden Wartepflichtigen beinhaltet. Für den Unfalltyp *Abbiege-Unfall* zeigen die Verfahren bspw., dass eine über den Knotenpunkt komplett oder teilweise rot eingefärbte Radfahrfurt sowie eine indirekte Führung des linksabbiegenden Radverkehrs anstelle dessen Führung im Mischverkehr höhere Unfallzahlen erwarten lässt, wobei letzteres für den untersuchten Sachverhalt irrelevant erscheint und damit auf eine Schwäche bei der Variableneinbeziehung hindeutet. Im Gegensatz dazu schätzen die Verfahren für den Unfalltyp *Einbiegen-/Kreuzen-Unfall* bspw. höhere Unfallzahlen, wenn die Anzahl der Geradeausfahrstreifen einer Zufahrt zunimmt und wenn der Knotenpunkt durch das Verkehrszeichen Z205 bzw. eine Teil-Lichtsignalanlage anstelle der Vorschrift *Rechts-vor-Links* geregelt wird. Zudem zeigen die Verfahren bei beiden Unfalltypen zumeist, dass die Zahl der Unfälle ab einem bestimmten Verkehrsaufkommen weniger stark ansteigt. Dieses Phänomen ist in der Wissenschaft unter dem Namen *Safety in Numbers-Effekt* bekannt. Ein Vergleich der Modellgüten zwischen den Unfalltypen zeigt zudem, dass beide Verfahren mit ihrem Modell des Unfalltyps *Abbiege-Unfall* bessere Vorhersagen generieren als mit ihrem Modell des Unfalltyps *Einbiegen-/Kreuzen-Unfall*. Weiterhin unterscheiden sich die Modellgüten nach Unfalltyp nur geringfügig zwischen beiden Verfahren, weshalb davon ausgegangen werden kann, dass beide Verfahren qualitativ ähnliche Modelle des entsprechenden Unfalltyps liefern.

Abstract

The volume of bicycle traffic in Germany has increased in recent years, which in turn is also reflected in an increase in the occurrence of accidents involving cyclists. In order to work against the rising accident rates, politics and associations recommend to take especially infrastructure interventions. On this basis, the present study examines how single infrastructure features affect the occurrence of accidents between cyclists and motorized traffic using the example of the city of Dresden. Therefore, the data base of the study consists of 548 accidents involving cyclists from 2015 to 2019 and the characteristics of 484 intersection approaches. Since the infrastructure is not the only factor determining the occurrence of accidents, traffic volume parameters are also included. To investigate the occurrence of accidents, the random forest and the negative binomial regression method in the form of *Accident Prediction Models* with prior variable selection by the LASSO technique are used. Each method is applied to two specific accident types for intersections in order to obtain differentiated results. Thereby, the first accident type *Turning from major road-Collisions* includes collisions between a participant on the major road turning right and a participant driving straight ahead in the same or opposite direction, while the second accident type *Turning from minor road/Crossing-Collisions* includes collisions between a participant with right of way and a participant on the minor road wanting to turn off or to cross. For the accident type *Turning from major road-Collisions*, the methods estimate, for example, that a cycling ford completely or partially colored red over the intersection as well as an indirect guidance of the left-turning bicycle traffic instead of its guidance in mixed traffic leads to more accidents, whereby the latter seems to be irrelevant for the examined situation and thus suggests a weakness in the inclusion of variables. In contrast, for the accident type *Turning from minor road/Crossing-Collisions* the methods estimate, for example, more accidents if the number of straight-ahead lanes of an approach increases and if the intersection is controlled by the traffic sign Z205 or a partial light signal system instead of the *Right has right-of-way* regulation. In addition, the methods mostly show for both accident types that the number of accidents increases less strongly if a certain traffic volume is reached. In science this phenomenon is known as the *Safety in Numbers effect*. A comparison of the model accuracies between the accident types also shows that both methods generate better predictions with their model of the accident type *Turning from major road-Collisions* than with their model of the accident type *Turning from minor road/Crossing-Collisions*. Furthermore, the model accuracies by accident type differ only slightly between the two methods, so it can be assumed that both methods provide qualitatively similar models of the corresponding accident type.

Inhaltsverzeichnis

Abbildungsverzeichnis **VII**

Tabellenverzeichnis **IX**

Abkürzungs- und Symbolverzeichnis **XI**

1 Einleitung **1**

2 Literaturüberblick **3**
2.1 Safety in Numbers-Effekt 3
2.2 Einflussfaktoren von Radverkehrsunfällen 5

3 Grundlagen der Unfallforschung **13**
3.1 Unfallkategorien 13
3.2 Unfalltypen 13

4 Datengrundlage **17**
4.1 Unfalldaten 17
4.2 Infrastrukturmerkmale 25
4.3 Überblick über verwendete Variablen 40

5 Methodik **43**
5.1 Korrelationsbetrachtung 43
5.2 Random Forest 44
5.2.1 Grundlagen 45
5.2.2 Random Forest-Verfahren 47
5.2.3 Modellgütekriterien 48
5.2.4 Variablenbedeutsamkeit 49
5.3 Negative Binomialregression 53
5.3.1 Grundlagen 53
5.3.2 Accident Prediction Models 56
5.3.3 Variablenselektion 58
5.3.4 Modellgütekriterien 59
5.3.5 Variablenbedeutsamkeit 61
5.3.6 Modelldiagnostik 62

6 Durchführung und Ergebnisse **67**
6.1 Korrelationsbetrachtung 67
6.2 Random Forest 68
6.2.1 Modellgütekriterien 69
6.2.2 Variablenbedeutsamkeit 70
6.3 Negative Binomialregression 79
6.3.1 Variablenselektion 80
6.3.2 Modellgütekriterien 81
6.3.3 Variablenbedeutsamkeit 82
6.3.4 Modelldiagnostik 86
6.4 Vergleich beider Verfahren 89
6.4.1 Modellgütekriterien 89
6.4.2 Variablenbedeutsamkeit und Handlungsempfehlungen 91
6.5 Vergleich mit Literaturerkenntnissen 94

7 Kritische Würdigung **97**

8 Zusammenfassung und Ausblick **99**

Literaturverzeichnis **XIII**

Anhang **XVII**

A Grundlagen der Unfallforschung **XVII**

B Datengrundlage **XIX**
B.1 Unfalldaten XIX
B.2 Infrastrukturmerkmale XX

C Methodik **XXI**

D Durchführung und Ergebnisse **XXIII**
D.1 Korrelationsbetrachtung XXIII
D.2 Random Forest XXIV
D.3 Negative Binomialregression XXVIII

Abbildungsverzeichnis

2.1 Safety in Numbers-Effekt: Zusammenhang zwischen DTV und Unfallzahl 5

3.1 Untergliederung des Unfalltyps 2 für Rechtsabbiegende 15
3.2 Untergliederung des Unfalltyps 3 nach Herkunft des bevorrechtigten Fahrzeugs 15

4.1 Häufigkeitsverteilung: Unfallkategorien aller Unfälle 17
4.2 Häufigkeitsverteilung: Unfalltypen aller Unfälle 18
4.3 Repräsentativität des eingegrenzten Datensatzes bzgl. der Unfallkategorien 19
4.4 Repräsentativität des eingegrenzten Datensatzes bzgl. der Unfalltypen . 19
4.5 Häufigkeitsverteilung: Unfallkategorien des eingegrenzten Datensatzes . 20
4.6 Häufigkeitsverteilung: Unfalltypen des eingegrenzten Datensatzes 20
4.7 Tageszeitliche Verteilung: Unfallgeschehen im eingegrenzten Datensatz . 21
4.8 Häufigkeitsverteilung: Beteiligtenkonstellation im eingegrenzten Datensatz 22
4.9 Anteil des Unfallverursachers Rad nach Beteiligtenkonstellation im eingegrenzten Datensatz 22
4.10 Boxplots: Alter der Unfallverursacher und -gegner im eingegrenzten Datensatz 23
4.11 Häufigkeitsverteilung: dreistellige Unfalltypen für Unfälle vom Typ AB im eingegrenzten Datensatz 24
4.12 Häufigkeitsverteilung: dreistellige Unfalltypen für Unfälle vom Typ EK im eingegrenzten Datensatz 25
4.13 Häufigkeitsverteilung: Anzahl der Unfälle vom Typ 23/24 pro Zufahrt . 26
4.14 Häufigkeitsverteilung: Anzahl der Unfälle vom Typ 30/32/34 pro Zufahrt 26
4.15 Boxplots: Ausrundungsradius für Rechtsabbiegende 31
4.16 Häufigkeitsverteilung: Radverkehrsführung im Knotenpunkt 34
4.17 Häufigkeitsverteilung: Absetzung der Radfahrfurt im Knotenpunkt . . . 35
4.18 Häufigkeitsverteilung: Signalgeber des Radverkehrs 36
4.19 Häufigkeitsverteilung: Radverkehrsführung der linksabbiegenden Radfahrenden 37
4.20 Häufigkeitsverteilung: Roteinfärbung der Radfahrfurt 38
4.21 Häufigkeitsverteilung: richtungsbezogener DTV für Kfz und Radverkehr sowie Schwerverkehrsanteil 39

5.1 Schematischer Aufbau eines CART 47

5.2 Schematische Intervalleinteilung der Accumulated Local Effects 52

6.1 Grafische Darstellung der Korrelationen 68
6.2 Optimale Anzahl der Entscheidungsbäume nach Unfalltyp 69
6.3 Variable Importance nach Unfalltyp 70
6.4 Unfalltyp 23/24: PDP und ALE der bedeutsamsten DTV-Größe 72
6.5 Unfalltyp 23/24: PDP und ALE bedeutsamer Infrastrukturgrößen . . . 74
6.6 Unfalltyp 30/32/34: PDP und ALE der bedeutsamsten DTV-Größe . . . 76
6.7 Unfalltyp 30/32/34: PDP und ALE bedeutsamer Infrastrukturgrößen . . 77
6.8 Modelldiagnostikplots der Negativen Binomialregression 87
6.9 Gegenüberstellung realer und vorhergesagter Unfallzahl beider Verfahren nach Unfalltyp 90

A.1 Kurzbeschreibung der Unfalltypen nach M Uko XVII

B.1 Räumliche Verteilung der Radverkehrsunfälle in Dresden XIX
B.2 Ausprägungen des Merkmals *Roteinfärbung der Radfahrfurt* am Beispiel eines Knotenpunktes XX

D.1 Unfalltyp 23/24: PDP der übrigen Variablen XXIV
D.2 Unfalltyp 23/24: ALE der übrigen Variablen XXV
D.3 Unfalltyp 30/32/34: PDP der übrigen Variablen XXVI
D.4 Unfalltyp 30/32/34: ALE der übrigen Variablen XXVII
D.5 Verteilung der Expositionsvariablen nach Logarithmierung XXVIII
D.6 Verteilung der Variable x_{10} vor und nach Logarithmierung XXVIII
D.7 Quantilresiduen vs. geschätzte Werte ohne varianz-stabilisierende Transformation der geschätzten Werte XXXIV

Tabellenverzeichnis

2.1 Koeffizienten der Verkehrsstärken nach Autor 10

3.1 Unfallkategorien nach M Uko 13
3.2 Unfalltypen nach M Uko 14

4.1 Häufigkeitsverteilung: Knotenpunktform 27
4.2 Häufigkeitsverteilung: Verkehrsregelungsart 28
4.3 Häufigkeitsverteilung: Anzahl der Fahrstreifen 29
4.4 Häufigkeitsverteilung: Signalisierung von Links- und Rechtsabbiegenden 30
4.5 Häufigkeitsverteilung: Abstand ruhender Verkehr zum Knotenpunkt . . 32
4.6 Häufigkeitsverteilung: Abbiegeverbote 33
4.7 Deskriptive Statistik der verwendeten Variablen 40

6.1 Modellgüte nach Unfalltyp: Random Forest-Verfahren 70
6.2 Modellgüte nach Unfalltyp: Negative Binomialregression 81
6.3 Ergebnisse nach Unfalltyp: Negative Binomialregression (Auszug) . . . 82
6.4 Modellgütevergleich beider Verfahren nach Unfalltyp 89
6.5 Vergleich bedeutsamer infrastruktureller Variablen beider Verfahren nach Unfalltyp 93

C.1 Schematischer Aufbau einer Kontingenztabelle XXI

D.1 Kategorisierung der metrisch vorliegenden Variablen XXIII
D.2 Variablenselektion mit Hilfe des LASSO-Verfahrens nach Unfalltyp . . XXIX
D.3 Ergebnisse nach Unfalltyp: Negative Binomialregression (vollständig) . XXXI

Abkürzungs- und Symbolverzeichnis

AB	Abbiege-Unfall
ALE	Accumulated Local Effects
C_{korr}	Korrigierter Kontingenzkoeffizient nach Pearson
CART	Classification and Regression Tree
DTV	Durchschnittlich tägliches Verkehrsaufkommen
EK	Einbiegen-/Kreuzen-Unfall
$E(y_k)/\widehat{E(y_k)}$	wahrer/geschätzter Erwartungswert der abhängigen Variable (auch $\mu_k/\hat{\mu}_k$)
F	Fahrunfall
GLM	Generalized Linear Model/Verallgemeinertes Lineares Modell
i	Zählindex für unabhängige Variablen, i=1...p
k	Zählindex für Beobachtungen, k=1...n
L_0	Likelihood-Wert des Nullmodells
L_V	Likelihood-Wert des vollständigen Modells
LASSO	Least absolute shrinkage and selection operator
LR-Test	Likelihood-Ratio-Test
LSA	Lichtsignalanlage
LV	Unfall im Längsverkehr
M Uko	Merkblatt zur Örtlichen Unfalluntersuchung in Unfallkommissionen
McF-R^2	McFaddens-R^2
MIV	Motorisierter Individualverkehr
ML	Maximum-Likelihood
MSE	Mean Squared Error

OOB	Out-of-bag
p	Zählindex für Expositionsvariablen, p=1...a
PDP	Partial Dependence Plot
q	Zählindex für Risikovariablen, q=1...b
Q-Q-Plot	Quantil-Quantil-Plot
RFV	Random Forest-Verfahren
RMSE	Root Mean Squared Error
RV	Unfall durch ruhenden Verkehr
SO	Sonstiger Unfall
s_{pk}	Ausprägung Expositionsvariable
t_{qk}	Ausprägung Risikovariable
ÜS	Überschreiten-Unfall
VI	Variable Importance
x_{ik}	Ausprägung unabhängige Variable
$y_k/\hat{y}_k$	wahre/geschätzte Ausprägung abhängige Variable
$\beta_0/\hat{\beta}_0$	wahres/geschätztes Absolutglied
$\beta_i/\hat{\beta}_i$	wahrer/geschätzter Koeffizient unabhängige Variable
$\gamma_p/\hat{\gamma}_p$	wahrer/geschätzter Koeffizient Expositionsvariable
$\delta_q/\hat{\delta}_q$	wahrer/geschätzter Koeffizient Risikovariable
$\theta/\hat{\theta}$	wahrer/geschätzter Dispersionsparameter
λ	Regularisierungsparameter
$\mu_k/\hat{\mu}_k$	wahrer/geschätzter Erwartungswert der abhängigen Variable (auch $E(y_k)/\widehat{E(y_k)}$)
χ^2	Chi-Quadrat

1 Einleitung

Die Nutzung des Fahrrads bewirkt im Vergleich zum motorisierten Verkehr vielfältige positive Effekte, welche bereits durch zahlreiche wissenschaftliche Studien bestätigt wurden. Demnach geht die zunehmende Nutzung des Fahrrads bspw. mit gesundheitlichen Vorteilen für die Individuen sowie einer Verringerung der gesellschaftlich relevanten negativen Auswirkungen des motorisierten Verkehrs, wie z.B. Staus, CO_2-Emissionen oder Lärmbelästigung, einher. (vgl. EESC (2016), S. 22) Die Studie *Mobilität in Deutschland - MiD* aus dem Jahr 2017 unterstreicht die Relevanz dieser Vorzüge. Demzufolge steigerte sich das Radverkehrsaufkommen in Deutschland zwischen 2002 und 2017 von etwa 25 Millionen auf etwa 28 Millionen Wege pro Tag, was einem Anstieg von etwa zwölf Prozent entspricht. Bezogen auf den Modal Split des Verkehrsaufkommens zeigte die Studie zudem, dass der Anteil der Fahrradwege an den insgesamt etwa 270 Millionen zurückgelegten Wegen pro Tag im Jahr 2002 etwa neun Prozent betrug, während dieser bei den insgesamt etwa 257 Millionen zurückgelegten Wegen pro Tag im Jahr 2017 bei etwa elf Prozent lag, was einem Zuwachs von immerhin zwei Prozentpunkten in 15 Jahren entspricht. Des Weiteren zeigte sich im städtischen Raum eine stärkere Ausprägung des Wachstums als in ländlichen Gebieten. (vgl. Nobis (2019), S. 19ff.) Es ist denkbar, dass sich dieser Trend auch in Zukunft abzeichnet, da bspw. Bikesharing-Systeme an Bedeutung gewinnen. Dieses Wachstum resultiert im Umkehrschluss jedoch ebenfalls in einer erhöhten Zahl der Unfälle mit Radfahrendenbeteiligung. Dafür ermittelte das *European Transport Safety Council* bspw. für die Jahre 2010 bis 2018 einen etwa einprozentigen Anstieg der durchschnittlich jährlich getöteten Radfahrenden sowie einen etwa zweiprozentigen Anstieg der durchschnittlich jährlich schwerverletzten Radfahrenden in Deutschland. (vgl. Adminaité-Fodor und Jost (2020). S. 22ff.) Einen möglichen Grund dafür sieht der *European Transport Safety Council* in der Infrastruktur, welche die Interaktion zwischen Verkehrsteilnehmenden und deren Sicherheit aktiv beinflusse. Daher sollte der Fokus der Verantwortlichen darin liegen, potenzielle Konfliktpunkte zwischen motorisiertem Verkehr und schwächeren Verkehrsteilnehmenden, wie bspw. Radfahrenden, zu reduzieren, indem potenzielle Defizite der vorhandenen Infrastruktur ausfindig gemacht und passende Maßnahmen ergriffen werden. (vgl. Adminaité-Fodor und Jost (2020). S. 41) Auf Grundlage dessen ergibt sich die Frage, ob und wie sich einzelne Infrastrukturmerkmale auf das Unfallgeschehen auswirken. Autoren, die sich mit der Thematik befassen, gelangen dabei zu keinem einheitlichen Konsens, da die betrachteten Merkmale zumeist auf unterschiedliche Weise in die Untersuchungen einbezogen werden. Dennoch sind sich die Autoren

einig, dass das Verkehrsaufkommen einen größeren Erklärungsanteil am Unfallgeschehen aufweist als die Infrastrukturmerkmale. Dabei zeigen sie zumeist, dass die Anzahl der Unfälle mit zunehmendem Verkehrsaufkommen weniger stark ansteigt, was die Autoren als *Safety in Numbers-Effekt* bezeichnen.
Das Ziel der Arbeit liegt daher in der Untersuchung der Einflussfaktoren des Unfallgeschehens zwischen dem motorisiertem Individualverkehr (MIV) und dem Radverkehr am Beispiel der Stadt Dresden, welche in den letzten Jahren zahlreiche Verkehrsinfrastrukturmaßnahmen, wie bspw. den Ausbau von Radwegen sowie die Markierung von Rad- und Schutzstreifen, umsetzte. (vgl. Ohm et al. (2016), S. 13) Die relevante Datengrundlage stellen dabei 548 zwischen 2015 und 2019 polizeilich erfasste Verkehrsunfälle mit Radfahrendenbeteiligung in Dresden dar. Da das *Radverkehrskonzept der Landeshauptstadt Dresden* eine „Erhöhung der Sicherheit aller Verkehrsteilnehmerinnen und Verkehrsteilnehmer durch [den] Umbau bzw. [die] Umgestaltung sicherheitskritischer Verkehrsanlagen“ (ISUP (2017), S. 6) anstrebt, liegt der Fokus dabei - neben Kenngrößen des Verkehrsaufkommens - insbesondere auf den infrastrukturellen Einflussfaktoren der Unfälle, weshalb Merkmale von 484 Zufahrten an 133 Knotenpunkten in die Untersuchungen einbezogen werden. Um differenzierte Ergebnisse zu erlangen, separiert die vorliegende Arbeit zudem zwischen zusammengefassten Untertypen des Unfalltyps Abbiege-Unfall (AB) sowie des Unfalltyps Einbiegen-/Kreuzen-Unfall (EK), welche im *Merkblatt zur Örtlichen Unfalluntersuchung in Unfallkommissionen - M Uko* (FGSV (2012)) festgelegt sind. Als ökonometrische Verfahren kommen dabei das Random Forest-Verfahren (RFV) sowie die Negative Binomialregression in Form von *Accident Prediction Models* mit vorheriger Variablenselektion anhand des Least absolute shrinkage and selection operator (LASSO)-Verfahrens zur Anwendung.
Die vorliegende Arbeit ist wie folgt aufgebaut. Nach einem Literaturüberblick in Kapitel 2 gibt Kapitel 3 einen Einblick in grundlegende Kenngrößen der Unfallforschung, wobei insbesondere auf die verschiedenen Unfalltypen eingegangen wird. Im Anschluss stellt Kapitel 4 die verwendeten Unfalldaten und Infrastrukturmerkmale vor. Weiterhin erläutert Kapitel 5 die verwendete Methodik, wobei zunächst auf ein Maß für die Korrelationsbetrachtung eingegangen wird. Anschließend erfolgt die theoretische Vorstellung des Vorgehens im RFV sowie bei der Negativen Binomialregression. Nachdem Kapitel 6 die Ergebnisse der Korrelationsbetrachtung sowie der beiden Verfahren dargelegt hat, werden die Ergebnisse der Verfahren zudem einerseits untereinander und andererseits mit den Literaturerkenntnissen zusammenfassend verglichen. Zum Abschluss erfolgt eine kritische Würdigung in Kapitel 7.

2 Literaturüberblick

Zunächst stellt das vorliegende Kapitel zwei Studien zum sogenannten *Safety in Numbers-Effekt* vor, welcher in der Literatur eine wesentliche Erkenntnis in der Beschreibung des Unfallgeschehens zwischen MIV und Radverkehr darstellt. Im zweiten Schritt werden weitere wissenschaftliche Arbeiten vorgestellt, die die infrastrukturellen Einflussfaktoren von Radverkehrsunfällen näher beleuchten. Neben den Ergebnissen liegt der Fokus dabei auf dem jeweiligen Untersuchungsgegenstand sowie der verwendeten Methodik der Autoren.

2.1 Safety in Numbers-Effekt

Jacobsen (2015) untersucht den Zusammenhang zwischen der Anzahl der Radfahrenden und dem Unfallgeschehen zwischen dem Kfz- und Radverkehr. Dabei analysiert er anhand von fünf unabhängigen Datensätzen Knotenpunkte verschiedener Länder und Städte. Während der Autor dabei auf Stadtebene bspw. 68 Städte in Kalifornien, USA sowie 47 Städte in Dänemark betrachtet, analysiert er auf Länderebene z.B. das Vereinigte Königreich und die Niederlande. Insgesamt reicht der Zeitraum der Datensätze von 1993 bis 2000, wobei jeder einzelne Datensatz einen anderen Zeitraum abdeckt. Mit Hilfe der Kleinste-Quadrate-Methode schätzt Jacobsen (2015) jeweils den Einfluss des Radverkehrsaufkommens auf das Unfallrisiko pro Radfahrenden. Das Radverkehrsaufkommen wird dabei aufgrund der Datenlage einzelner Untersuchungsgebiete durch verschiedene Maße, wie bspw. *Gefahrene Kilometer pro Person und Tag* oder *Anzahl der Fahrten pro Person und Tag*, beschrieben. Die so ermittelte unabhängige Variable enthält im Exponenten den zugehörigen Koeffizienten. Nimmt dieser den Wert eins an, so wird von einem positiven linearen Zusammenhang ausgegangen. Liegt der Koeffizient hingegen in einem Bereich zwischen null und eins, so liegt ein positiver degressiver Zusammenhang vor, während ein Koeffizient kleiner null einen negativen Zusammenhang impliziert. Das Ergebnis zeigt in nahezu allen Schätzungen einen Koeffizienten zwischen 0,3 und 0,5, d.h. einen positiv degressiven Zusammenhang zwischen dem Radverkehrsaufkommen und der Unfallzahl.[1] Somit steigt das Unfallrisiko ab einer bestimmten Anzahl von Radfahrenden mit zunehmender Anzahl derselben weniger stark an. Anders

[1] Eine Ausnahme bildet dabei die Niederlande mit einem Koeffizienten von etwa -2, was eine rapide Senkung der Unfallzahl mit zunehmender Anzahl an Radfahrenden bedeutet. Der Autor begründet dies mit der Tatsache, dass die Niederlande in den letzten 20 Jahren zahlreiche Maßnahmen implementiert hat, die das Radfahren sicherer machen und dementsprechend ebenfalls mehr Menschen zum Radfahren animieren.

ausgedrückt bedeutet dies, dass die Wahrscheinlichkeit einer Kollision zwischen einem Kfz und einem Radfahrenden mit zunehmender Anzahl an Radfahrenden abnimmt, was wiederum als *Safety in Numbers-Effekt* bezeichnet wird. Dieser für verschiedene Zeitpunkte und verschiedene Städte bzw. Länder konsistente Zusammenhang erscheint dem Autor entgegen der Intuition, denn bei mehr Verkehrsteilnehmenden ließe sich normalerweise eine höhere Unfallzahl vermuten. Jacobsen (2015) findet es unwahrscheinlich, dass die Radfahrenden bei zunehmender Anzahl derselben vorsichtiger fahren. Daher vermutet er den Grund des Effekts darin, dass anstelle der Radfahrenden die Kfz-Fahrenden ihr Fahrverhalten anpassen. Der Autor schlussfolgert zudem, dass Maßnahmen, die das Radverkehrsaufkommen erhöhen, ebenfalls zur Sicherheit der Radfahrenden beitragen können. (vgl. Jacobsen (2015), S. 272ff.)

Nordback et al. (2014) entwickeln einen Ansatz für eine *safety performance function* für Unfälle zwischen dem Kfz- und dem Radverkehr in den USA. Dabei untersuchen sie über zwei Zeiträume signalisierte Knotenpunkte in Colorado, USA. Während sich der erste Zeitraum von 2001 bis 2005 erstreckt und 105 Knotenpunkte umfasst, werden im zweiten Zeitraum zwischen 2008 und 2011 106 Knotenpunkte betrachtet. Nordback et al. (2014) modellieren die Anzahl der erwarteten Unfälle pro Knotenpunkt und Jahr des jeweiligen Zeitraums in Abhängigkeit des *durchschnittlich täglichen Verkehrsaufkommens (DTV) im Kfz-Verkehr* und des *DTV der Radfahrenden*. Die so für beide Zeiträume gebildeten Modelle werden unter Verwendung einer Negativen Binomialregression geschätzt. Dabei kommt die Maximum-Likelihood (ML)-Methode zur Schätzung der Koeffizienten, die jeweils als Potenz der beiden Einflussgrößen in die Modelle einfließen, zum Einsatz. Schließlich überprüfen die Autoren die Signifikanz beider Einflussgrößen mit dem Wald-Test. Die Ergebnisse zeigen einen positiv degressiven Zusammenhang zwischen dem *DTV im Kfz-Verkehr* bzw. dem *DTV der Radfahrenden* und dem Unfallgeschehen, wobei die zugehörigen geschätzten Koeffizienten jeweils Werte zwischen 0,5 und 0,7 annehmen. Somit geschehen weniger Unfälle pro Radfahrenden, wenn das Verkehrsaufkommen steigt, was wiederum als *Safety in Numbers-Effekt* bezeichnet wird. Neben den *safety performance functions* ermitteln die Autoren außerdem *risk performance functions*, welche die Wahrscheinlichkeit eines Unfalls pro eine Million Radfahrende, die den Knotenpunkt passieren, angeben. Auch hier kann der *Safety in Numbers-Effekt* nachgewiesen werden, denn die Ergebnisse zeigen bspw., dass die Wahrscheinlichkeit eines solchen Unfalls mit zunehmendem *DTV der Radfahrenden* sinkt. Demnach weisen Knotenpunkte mit durchschnittlich weniger als 200 pro Tag einfahrenden Radfahrenden ein relativ hohes Unfallrisiko auf, während dieses bei mehr als 600 pro Tag einfahrenden Radfahrenden niedriger ausfällt. (vgl. Nordback et al. (2014), S. 120f.)

Zusammenfassend lässt sich also sagen, dass der *Safety in Numbers-Effekt* durch einen positiv degressiven Zusammenhang zwischen dem Verkehrsaufkommen und der Unfallzahl gekennzeichnet ist. Somit verringert sich die Wahrscheinlichkeit eines Unfalls

mit Überschreiten eines bestimmten Verkehrsaufkommens. Abbildung 2.1 stellt diesen Zusammenhang schematisch dar.

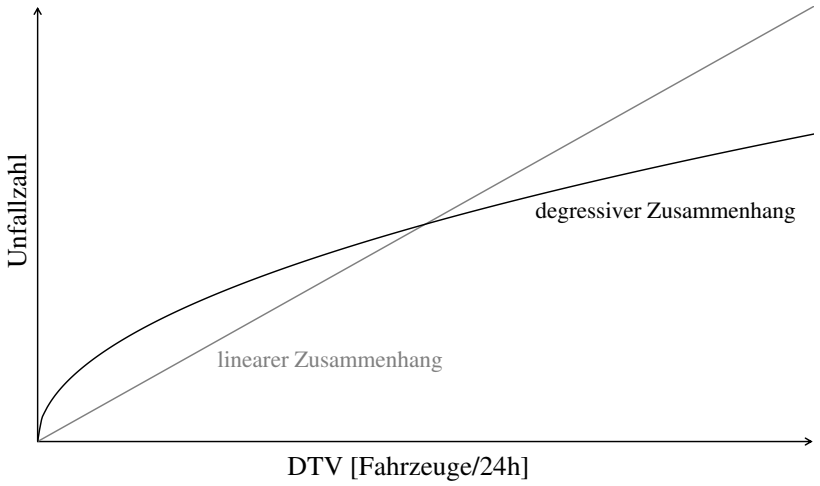


Abbildung 2.1: Safety in Numbers-Effekt: Zusammenhang zwischen DTV und Unfallzahl [Quelle: eigene Darstellung in Anlehnung an Nordback et al. (2014), S. 119]

Neben den eben vorgestellten Autoren gelingt es auch weiteren Autoren, wie bspw. den im Folgenden erwähnten Schepers et al. (2011) sowie Turner et al. (2011), den *Safety in Numbers-Effekt* im Radverkehr nachzuweisen.

2.2 Einflussfaktoren von Radverkehrsunfällen

Das folgende Kapitel gibt einen chronologischen Einblick in wissenschaftliche Arbeiten, die sich mit der Untersuchung infrastruktureller Einflussfaktoren von Radverkehrsunfällen beschäftigen.

Wang und Nihan (2004) betrachten das Unfallgeschehen zwischen Rad- und Kfz-Verkehr an 115 zufällig gewählten vierarmigen signalisierten Knotenpunkten im Ballungsraum von Tokio, Japan. Dabei untersuchen sie die Jahre 1992 bis 1995. Die Autoren unterteilen die Unfälle in drei Klassen, die sich an der Bewegungsrichtung des beteiligten Kfz orientieren. Dabei beinhaltet die erste Klasse Kollisionen mit kreuzenden geradeausfahrenden, die zweite Klasse Kollisionen mit aus gleicher oder entgegengesetzter Richtung fahrenden linksabbiegenden und die dritte Klasse Kollisionen mit aus gleicher oder entgegengesetzter Richtung fahrenden rechtsabbiegenden Kfz. Für jede dieser Klassen schätzen Wang und Nihan (2004) ein Modell zur Vorhersage der Unfallzahl, wobei die Negative Binomialregression zum Einsatz kommt. Die Schätzung der Koeffizienten erfolgt dabei unter Verwendung der ML-Methode. Um das Unfallgeschehen zu untersuchen, stehen den Autoren mehr als 70 Variablen zur Verfügung. Diese sind bspw. *DTV im Kfz-Verkehr* - jeweils für Rechts-, Linksabbiegende und Geradeausfahrende -,

DTV der Radfahrenden, *Anzahl der Fahrspuren* - jeweils für Rechts- und Linksabbiegende -, *Signalsteuerung* oder *Existenz einer Fußgängerbrücke*. Davon beziehen sie zunächst jeweils jene Variablen in die Schätzungen ein, die für die entsprechende Klasse sachlogisch von Relevanz sein könnten. Anschließend entfernen sie Schritt für Schritt nicht-signifikante Variablen, sodass die jeweiligen finalen Modelle lediglich Variablen mit einem Signifikanzniveau von fünf Prozent enthalten. Für die erste Klasse zeigt sich im Ergebnis schließlich, dass die erwartete Unfallzahl bspw. mit zunehmendem Kfz-DTV der Geradeausfahrenden abnimmt, während diese bei der Existenz einer Fußgängerbrücke steigt. Während sich bei der zweiten Klasse bspw. eine niedrigere Unfallzahl bei steigendem Radfahrenden-DTV ergibt, erhöht ein größerer durchschnittlicher Zeitvorsprung der Linksabbiegenden die Unfallzahl. In der dritten Klasse führt z.B. die Änderung der Signalisierung von zwei auf vier Phasen zu einer niedrigeren Unfallzahl, während eine größere Straßenbreite sowie eine größere Anzahl von Rechtsabbiegefahrstreifen diese erhöht. (vgl. Wang und Nihan (2004), S. 318ff.)
Schepers et al. (2011) untersuchen das Unfallgeschehen im Radverkehr an 540 nichtsignalisierten Knotenpunkten innerhalb bebauter Stadtgebiete in den Niederlanden im Zeitraum von 2005 bis 2008. Dabei fokussieren sie sich auf Unfälle zwischen Kfz und Radfahrenden, wobei die Autoren zwischen zwei Unfalltypen unterscheiden. Während der erste Typ Unfälle beschreibt, bei denen der Radfahrende Vorfahrt hatte, bezieht sich der zweite Typ auf Unfälle, bei denen das Kfz Vorfahrt hatte. Die Unterteilung begründen sie damit, dass das jeweilige Unfallgeschehen von verschiedenen Verkehrsflüssen und Infrastrukturmerkmalen beeinflusst wird. Die betrachteten Einflussfaktoren für den ersten Typ sind dabei bspw. *DTV im Kfz-Verkehr*, *DTV der Radfahrenden*, *Art der Radverkehrsinfrastruktur* oder *Abstand Radweg vom Rand der Hauptfahrbahn*. Für den zweiten Typ hingegen beziehen die Autoren neben den Merkmalen *DTV im Kfz-Verkehr* und *DTV der Radfahrenden* bspw. die Faktoren *Anzahl der Spuren auf der Hauptfahrbahn* oder *Existenz einer Mittelinsel* ein. Für beide Typen schätzen Schepers et al. (2011) die jährliche Anzahl der Radverkehrsunfälle pro Knotenpunkt unter Verwendung einer Poisson- sowie Negativen Binomialregression. Dabei werden die Koeffizienten mittels der ML-Methode geschätzt. Anschließend prüfen die Autoren unter Verwendung eines Likelihood-Ratio-Test (LR-Test) und des Akaike-Informationskriterium (AIC), welches der beiden Verfahren für die vorliegenden Daten angemessener ist. Dabei stellt sich heraus, dass die Negative Binomialregression die Daten aufgrund von Überdispersion besser abbildet. Die Signifikanzen der geschätzten Koeffizienten überprüfen die Autoren mit dem Wald-Test. Im Ergebnis zeigt sich, dass bei beiden Typen ein signifikant positiv degressiver Zusammenhang zwischen der Anzahl der Unfälle und dem Kfz- bzw. Radverkehrs-DTV besteht. Das wiederum stellt ein Indiz für den *Safety in Numbers-Effekt* dar. Während sich beim zweiten Typ keines der einbezogenen Infrastrukturmerkmale als signifikant erweist, wird beim ersten Typ ein positiver Zusammenhang zwischen der Anzahl der Unfälle und Knotenpunkten mit gut markierten Zweirichtungsradwe-

gen sowie Knotenpunkten mit rötlich eingefärbter Radverkehrsinfrastruktur ermittelt. Im Gegensatz dazu werden beim ersten Typ weniger Unfälle modelliert, wenn die die Radverkehrsinfrastruktur angehoben - z.B. auf einer Bodenschwelle - geführt wird oder wenn andere geschwindigkeitsreduzierende Maßnahmen vorliegen. Bei einem Abstand zwischen zwei und fünf Metern zwischen dem Radweg und dem Rand der Hauptfahrbahn werden beim ersten Typ ebenfalls weniger Radverkehrsunfälle geschätzt. (vgl. Schepers et al. (2011), S. 857f.)

Turner et al. (2011) untersuchen das Unfallgeschehen im Kfz- und Radverkehr an insgesamt 102 signalisierten Knotenpunkten in Christchurch, Neuseeland und Adelaide, Australien. Die untersuchten Knotenpunkte umfassen dabei 383 Zufahrten, wobei sowohl Zufahrten mit als auch ohne Radverkehrsinfrastruktur betrachtet werden. Die Autoren führen dabei zunächst eine Vorher-Nachher-Studie durch mit dem Ziel, eine erste Beurteilung über die Wirkung der Existenz von Radverkehrsinfrastruktur auf das Unfallgeschehen zu erlangen. Die Vergleichszeitpunkte sind dabei jeweils fünf Jahre bevor und fünf Jahre nachdem eine Infrastrukturmaßnahme umgesetzt wurde. Weiterhin betrachtet die Studie neben der Wirkung auf die Unfallzahl im Gesamten auch die Wirkung auf das Auftreten von vier Hauptunfalltypen, welche die Autoren vorher festgelegt haben. Zwei von ihnen sind dabei z.B. die Konstellation eines den Kfz-Verkehr kreuzenden Radfahrenden sowie die Konstellation eines rechtsabbiegenden Kfz und eines geradeausfahrenden Radfahrenden. Um eine detaillierte Analyse vorzunehmen, entwickeln Turner et al. (2011) für jeden Hauptunfalltyp *safety performance functions*. Diese modellieren die erwartete jährliche Unfallzahl pro Zufahrt in Abhängigkeit bestimmter Einflussfaktoren. Die Einflussfaktoren umfassen dabei neben dem *DTV im Kfz-Verkehr* und dem *DTV der Radfahrenden* ebenfalls Infrastrukturmerkmale wie bspw. die Existenz und Art einer bestimmten Radverkehrsinfrastruktur, die Breite der Zufahrt und des Radstreifens sowie den Abstand zwischen Fahrbahn und Bordsteinkante. Für jeden Hauptunfalltyp entwickeln die Autoren drei solcher *safety performance functions*. Die erste umfasst dabei die Daten beider Städte gemeinsam, wobei für jede Stadt ein eigenes Absolutglied geschätzt wird. Die zweite *safety performance function* modelliert dann nur Daten aus Christchurch, um zu überprüfen, ob sich die Koeffizienten der restlichen Einflussgrößen zwischen den Städten unterscheiden. Im Anschluss betrachten die Autoren jeweils eine *safety performance function*, die lediglich Zufahrten mit vorhandener Radverkehrsinfrastruktur aus beiden Städten enthält, um den Einfluss der einzelnen Merkmale genauer zu untersuchen. Die sich so ergebenden Modelle werden mittels einer Poisson- und einer Negativen Binomialregression geschätzt, wobei die Variablen schrittweise ins Modell aufgenommen werden. Um die finalen Modelle zu ermitteln, nutzen die Autoren bei jeder hinzugefügten Variable das Bayessche Informationskriterium (BIC), um die Sinnhaftigkeit der Aufnahme dieser Variable ins Modell zu überprüfen. Aus der Kombination der Ergebnisse der gesamten Untersuchung schlussfolgern die Autoren bspw., dass der Gesamteffekt eines Radstreifens neutral ist. Während dabei ein

hoher Standard zu einer erhöhten Sicherheit führt, so bewirkt ein niedrigerer Standard eine geringere Sicherheit. Des Weiteren zeigt sich bei Zufahrten mit gemeinsamer Linksabbiege- und Geradeausfahrspur allgemein eine höhere Unfallrate als bei Zufahrten mit separater Linksabbiegespur. Im Gegensatz dazu senken eingefärbte Radstreifen die Unfallraten der Zufahrten im Fall der Konstellation eines rechtsabbiegenden Kfz und eines geradeausfahrenden Radfahrenden, während diese die Unfallraten der Zufahrten im Fall der Konstellation sich kreuzender Beteiligter erhöht. Die Verkehrsstärken weisen zudem in den meisten Modellen einen positiv degressiven Zusammenhang mit der erwarteten jährlichen Unfallzahl pro Zufahrt auf, was auf das Vorliegen des *Safety in Numbers-Effekt* hindeutet. (vgl. Turner et al. (2011), S. 70ff.)

Kim et al. (2012) ermitteln in ihrer Studie kritische Faktoren, die im Jahr 2005 zu Radverkehrsunfällen an 112 städtischen signalisierten vierarmigen Knotenpunkten im Großraum Incheon, Korea führten. Dabei machen sie mittels einer Felduntersuchung 68 potentielle Unfallfaktoren ausfindig. Um nicht alle dieser potentiellen Einflussfaktoren in die Untersuchung einzubeziehen, schätzen die Autoren zunächst einen Korrelations- sowie Kontingenzkoeffizienten und ermitteln so eine Auswahl von 20 signifikanten Einflussfaktoren. Diese umfasst dabei bspw. die Größen *DTV im Kfz-Verkehr*, *DTV der Radfahrenden* sowie die *Existenz einer Bushaltestelle* oder die *Existenz eines Zebrastreifens*. Anhand dieser Variablenauswahl schätzen die Autoren eine Poissonregression, um die erwartete Anzahl von Radverkehrsunfällen pro Knotenpunkt und Jahr zu ermitteln. Für die Schätzung kommt dabei die ML-Methode zum Einsatz. Schließlich führen Kim et al. (2012) eine Rückwärtsselektion der Variablen durch, um die Modellkomplexität zu verringern, wobei die vier unabhängigen Variablen *DTV im Kfz-Verkehr*, *DTV der Radfahrenden*, *Gehwegbreite* und *Existenz einer Verkehrsinsel* übrig bleiben. Im Vergleich dazu schätzen sie eine Negative Binomialregression. Da diese in der Koeffizientenschätzung jedoch keine signifikanten Unterschiede im Vergleich zur Poissonregression aufweist und das Poissonmodell nur eine sehr geringe Überdispersion zeigt, beschließen die Autoren, sich auf die Ergebnisse der Poissonregression zu konzentrieren. Diese zeigen schließlich, dass sowohl ein erhöhter Kfz-DTV als auch ein erhöhter Radverkehrs-DTV zu einer höheren Unfallzahl führt. Im Gegensatz dazu verringern breitere Gehwege die Unfallzahl, während die Existenz von Verkehrsinseln diese ebenfalls erhöht. (vgl. Kim et al. (2012), S. 630f.)

Hamann und Peek-Asa (2013) untersuchen den Einfluss der Radverkehrsinfrastruktur auf die Wahrscheinlichkeit von Unfällen zwischen Rad- und Kfz-Verkehr in den Jahren 2007 bis 2010. Dabei betrachten sie 147 Knotenpunkte mit und 147 Knotenpunkte ohne Unfälle in vier Landkreisen in Iowa, USA. Um die Wahrscheinlichkeiten zu schätzen, verwenden die Autoren eine logistische Regression, wobei die abhängige Variable die Ausprägungen *Knotenpunkt mit Unfall* und *Knotenpunkt ohne Unfall* annehmen kann. Während als mögliche infrastrukturelle Einflussgrößen die Existenz bzw. die Art der Radverkehrsinfrastruktur einbezogen wird, stellen bspw. die Größen *Radverkehrs-*

aufkommen, *Kfz-Verkehrsaufkommen* oder *Straßenbreite* Kontrollvariablen dar. Letztendlich schätzen Hamann und Peek-Asa (2013) zwei Modelle. Das erste Modell enthält neben den Kontrollvariablen die Existenz irgendeiner Radverkehrseinrichtung als binäre infrastrukturelle Einflussgröße, wohingegen das zweite Modell die Art der Radverkehrseinrichtung kategoriell einbezieht und damit näher spezifiziert. Die möglichen Kategorien sind dabei neben der Referenzkategorie *keine Radverkehrsinfrastruktur* die Ausprägungen *Fahrbahnmarkierungen*, *radverkehrsspezifische Beschilderung* sowie die Kombination von beidem. Bei Schätzung des ersten Modells zeigt sich, dass die Existenz irgendeiner Radverkehrsinfrastruktur die Wahrscheinlichkeit eines Unfalls - verglichen mit gar keiner Radverkehrsinfrastruktur - auf 48 Prozent senkt. Im Gegensatz dazu steigt die Unfallwahrscheinlichkeit um 38 Prozent, wenn der Abstand zwischen den Bordsteinen um zehn Fuß zunimmt. Beim zweiten Modell wird eine Reduktion der Unfallwahrscheinlichkeit auf 36 Prozent geschätzt, wenn anstelle von gar keiner Radverkehrsinfrastruktur eine Kombination aus radverkehrsspezifischer Beschilderung und Fahrbahnmarkierungen vorliegt. Während die Wahrscheinlichkeit durch das bloße Vorhandensein von Fahrbahnmarkierungen auf 40 Prozent gesenkt wird, wird diese bei alleinigem Vorhandensein einer radverkehrsspezifischen Beschilderung auf immerhin 62 Prozent reduziert. Ähnlich wie im ersten Modell erhöht sich die Wahrscheinlichkeit eines Unfalls mit einer zehn Fuß breiteren Straße um 37 Prozent. Weiterhin ergibt sich in beiden Modellen ein zehnprozentiger Anstieg der Unfallwahrscheinlichkeit pro fünf zusätzlichen Radfahrenden in 30 Minuten. Im Gegensatz dazu führen zehn zusätzliche Kfz in 30 Minuten zu einer zwei Prozent höheren Unfallwahrscheinlichkeit. (vgl. Hamann und Peek-Asa (2013), S. 106f.)

Klassen et al. (2014) untersuchen das Unfallgeschehen zwischen dem Kfz- und Radverkehr in der Stadt Edmonton, Kanada für die Jahre 2006 bis 2009. Da sie außerdem ein räumliches Verständnis des Auftretens der Unfälle erlangen wollen, schätzen die Autoren für die Stadt, welche 31 verkehrliche Zonen umfasst, jeweils ein Spatial-Mixed-Logit-Modell für Unfälle an Knotenpunkten und Unfälle auf Abschnitten der freien Strecke zwischen Knotenpunkten. Dabei ist die abhängige Variable jeweils binär und enthält die Ausprägungen *Unfall mit schwerer Verletzung* und *Unfall mit leichter Verletzung*. Die räumliche Komponente der Modelle suggeriert, dass Zonen, die räumlich nah beieinander liegen, gemeinsame Faktoren aufweisen könnten, die das Unfallgeschehen beeinflussen. Neben infrastrukturellen Einflussfaktoren beziehen Klassen et al. (2014) auch menschliche Einflussgrößen - wie *Alter* und *Geschlecht* -, zeitliche Faktoren - wie *Tag* und *Jahreszeit* - sowie Umweltfaktoren - wie *Temperatur* und *Windgeschwindigkeit* - in die Untersuchungen ein. Als Merkmale der Infrastruktur der Knotenpunkte hingegen fließen z.B. Größen wie *Anzahl der Zufahrten*, *Vorfahrtsregelung* oder *Radverkehrsbeschilderung* ein. Im Gegensatz dazu enthält das Modell für die Abschnitte der freien Strecke zwischen Knotenpunkten infrastrukturelle Einflussgrößen wie bspw. *Abschnittslänge*, *Parkmöglichkeiten* oder *Radverkehrsbeschilderung*. Bzgl. der infrastruk-

turellen Einflussgrößen zeigt das Ergebnis für Unfälle an Knotenpunkten bspw., dass Knotenpunkte ohne Radverkehrsbeschilderung ein mehr als dreimal so hohes Risiko für schwere Unfälle aufweisen als Knotenpunkte mit einer solchen. Zudem weisen signalisierte Knotenpunkte, an denen alle Zufahrten Zweirichtungsfahrbahnen sind, ein mehr als doppelt so hohes Risiko für schwere Unfälle auf als Knotenpunkte anderer Art. Weiterhin zeigt sich für Abschnitte der freien Strecke zwischen Knotenpunkten bspw., dass schwere Unfälle weniger wahrscheinlich sind an Abschnitten, die ein- bzw. beidseitiges Parken erlauben, im Vergleich mit Abschnitten ohne eine solche Parkmöglichkeit. Für beide Modelle können die Autoren zudem eine signifikante räumliche Korrelation zwischen Zonen, deren Zentren weniger als drei Kilometer voneinander entfernt sind, nachweisen. (vgl. Klassen et al. (2014), S. 301f.)
Zusammenfassend lässt sich sagen, dass die Verkehrsstärke eine wesentliche Größe zur Beschreibung des Unfallgeschehens zwischen Kfz- und Radverkehr darstellt. Während die meisten Autoren einen positiven Zusammenhang zwischen der Verkehrsstärke und der Unfallzahl identifizieren, können einige von ihnen zudem den eingangs beschriebenen *Safety in Numbers-Effekt* nachweisen. Tabelle 2.1 zeigt dafür zusammenfassend die von den Autoren für die Verkehrsstärken ermittelten Koeffizienten, wobei ein Koeffizient zwischen null und eins den *Safety in Numbers-Effekt* repräsentiert. Dabei sind lediglich jene Autoren aufgeführt, deren verwendetes Modell eine Identifikation des *Safety in Numbers-Effekt* zulässt. Außerdem ist für die Autoren, die mehrere Modelle untersuchen, eine Spanne vom minimalen zum maximalen Wert des entsprechenden Koeffizienten angegeben.

Tabelle 2.1: Koeffizienten der Verkehrsstärken nach Autor

Autor	Kenngröße	Koeffizient
Jacobsen (2015, S. 273)	Radverkehrsaufkommen	0,31 bis 0,48[1]
Nordback et al. (2014, S. 120)	Kfz-Verkehrsaufkommen	0,58 bis 0,64
	Radverkehrsaufkommen	0,53 bis 0,65
Wang und Nihan (2004, S. 318f.)	Kfz-Verkehrsaufkommen	-0,13 bis -0,04
	Radverkehrsaufkommen	-0,57 bis -0,25
Schepers et al. (2011, S. 857ff.)	Kfz-Verkehrsaufkommen	0,50 bis 0,73
	Radverkehrsaufkommen	0,48 bis 0,56
Turner et al. (2011, S. 71)	Kfz-Verkehrsaufkommen	-0,05 bis 0,96
	Radverkehrsaufkommen	-0,06 bis 0,69
Kim et al. (2012, S. 630)	Kfz-Verkehrsaufkommen	4,29
	Radverkehrsaufkommen	1,64

[1] *Die Niederlande bildet mit einem Koeffizienten von -2 eine Ausnahme.*
Quelle: eigene Darstellung

Für die infrastrukturellen Einflussfaktoren hingegen lässt sich anhand der aufgeführten Literatur kein umfassender Konsens ableiten, da die Autoren zumeist einerseits verschiedene Merkmale in ihre Untersuchungen einbeziehen und andererseits für die gleichen Einflussfaktoren verschiedene Ergebnisse ermitteln. So bewirken bspw. rot eingefärbte Radfahrstreifen bei Schepers et al. (2011) eine Erhöhung der erwarteten Unfallzahl, während diese bei Turner et al. (2011) sowohl zu Erhöhung als auch zu einer Verringerung dieser führen kann. Weiterhin erlauben die Ergebnisse der Autoren bspw. keine einheitliche Schlussfolgerung bzgl. markierter Radverkehrsinfrastruktur. Während diese bei Schepers et al. (2011) die Unfallzahl erhöht, ermitteln Hamann und Peek-Asa (2013) einen negativen Zusammenhang mit dem Unfallgeschehen. Anhand der verwendeten Methodik fällt zudem auf, dass die Poisson- bzw. die Negative Binomialregression die am häufigsten verwendeten Verfahren der betrachteten Autoren darstellen.

3 Grundlagen der Unfallforschung

Da sich der Untersuchungsgegenstand der Arbeit auf das Radverkehrsunfallgeschehen bezieht, geht das vorliegende Kapitel auf grundlegende Merkmale des Unfallgeschehens, welche im *Merkblatt zur Örtlichen Unfalluntersuchung in Unfallkommissionen - M Uko* (FGSV (2012)) festgelegt sind, ein. Dabei richtet sich das Augenmerk auf die Einteilung von Unfällen in Kategorien und Typen, wobei zwei spezielle Typen wiederum die Grundlage der Untersuchungen der Arbeit darstellen.

3.1 Unfallkategorien

Die Unfallkategorie beschreibt die schwerste Unfallfolge, die ein am Unfall Beteiligter erlitten hat. Dabei lassen sich die Unfallkategorien in Personen- (Kategorie 1 bis 3) und Sachschäden (Kategorie 4 bis 6) aufgliedern, wobei sich die Unfallkategorie bei Sachschadensunfällen nach dem Straftatbestand richtet. Tabelle 3.1 stellt die Hierarchie sowie die Erläuterung der einzelnen Unfallkategorien dar. (vgl. FGSV (2012), S. 7f.)

Tabelle 3.1: Unfallkategorien nach M Uko

Unfallkategorie	Erläuterung (schwerste Unfallfolge)
1	mindestens eine getötete Person
2	mindestens eine schwerverletzte Person
3	mindestens eine leichtverletzte Person
4	Sachschaden, mindestens ein Kfz ist nicht mehr fahrbereit
5	Sachschaden, alle Kfz sind fahrbereit, kein Einfluss berauschender Mittel
6	alle übrigen Sachschäden unter Einfluss berauschender Mittel

Quelle: eigene Darstellung in Anlehnung an FGSV (2012), S. 8

3.2 Unfalltypen

Der Unfalltyp beschreibt die unfallauslösende Konfliktsituation, wobei dessen Bestimmung keine anderen Aspekte, wie z.B. die Schuldfrage, berücksichtigt. Sämtliche Konfliktsituationen können dabei einem der in Tabelle 3.2 dargestellten Unfalltypen, deren Kurzbeschreibungen in Abbildung A.1 in Anhang A zu finden sind, zugeordnet werden.

Tabelle 3.2: Unfalltypen nach M Uko

Unfalltyp	Kürzel	Bezeichnung
1	F	Fahrunfall
2	AB	Abbiege-Unfall
3	EK	Einbiegen-/Kreuzen-Unfall
4	ÜS	Überschreiten-Unfall
5	RV	Unfall durch ruhenden Verkehr
6	LV	Unfall im Längsverkehr
7	SO	Sonstiger Unfall

Quelle: eigene Darstellung in Anlehnung an FGSV (2012), S. 9

Aufgrund der in der Arbeit untersuchten Frage nach dem Einfluss von Infrastrukturmerkmalen auf das Radverkehrsunfallgeschehen empfiehlt sich der Fokus auf spezielle Unfalltypen. Der Grund dafür ist, dass das Auftreten mehrerer Unfälle vom gleichen Unfalltyp, d.h. Unfälle mit einer ähnlichen oder gar gleichen Konfliktsituation, auf mögliche infrastrukturelle Defizite hindeutet. (vgl. FGSV (2012), S. 8f.) Aus diesem Grund und aufgrund der gegebenen knotenpunktbezogenen Daten fokussiert sich die Arbeit daher auf die Unfalltypen 2 (AB) und 3 (EK), welche typischerweise an Knotenpunkten auftreten. Die Eigenschaften und die weitere für die Bearbeitung relevante Unterteilung der beiden Unfalltypen in zwei- bzw. dreistellige Untertypen werden im Folgenden näher erläutert.

Unfalltyp 2: Abbiege-Unfall

Wird ein Unfall durch einen Konflikt zwischen einem Abbiegenden[1] und einem aus gleicher oder entgegengesetzter Richtung kommenden Verkehrsteilnehmenden ausgelöst, so handelt es sich um einen Unfall vom Typ AB. Das gilt ebenfalls für einbiegende Verkehrsteilnehmende[2], die mit entgegenkommenden, nachfolgenden oder sich parallel bewegenden Verkehrsteilnehmenden in Konflikt geraten. Die Untergliederung in zwei- bzw. dreistellige Unfalltypen spezifiziert den Sachverhalt bspw. dahingehend, dass zwischen Rechts- und Linksabbiegenden unterschieden wird. (vgl. FGSV (2012), S. 52f.)
Da sich die vorliegende Arbeit in der folgenden Untersuchung auf Rechtsabbiegende konzentriert, stellt Abbildung 3.1 die Untergliederung des Unfalltyps 2 für Rechtsabbiegende dar. Dabei bezeichnet F einen Fußgänger.
Es ist zu erkennen, dass sich die zweistelligen Unfalltypen 23 und 24 lediglich in der Position der Annäherung des sich in gerader Richtung bewegenden Verkehrsteilneh-

[1] Abbiegen bezeichnet den Vorgang, bei dem ein Verkehrsteilnehmender von einer übergeordneten auf eine untergeordnete Straße auffährt.
[2] Einbiegen bezeichnet den Vorgang, bei dem ein Verkehrsteilnehmender von einer untergeordneten auf eine übergeordnete Straße auffährt.

menden unterscheiden. Dieser nähert sich beim Unfalltyp 23 auf der Fahrbahn, die auch der Rechtsabbiegende nutzt, in gleicher Richtung an. Beim Unfalltyp 24 hingegen bewegt sich dieser versetzt neben der Fahrbahn des Rechtsabbiegenden - z.B. auf einer Radfahrfurt - in gleicher bzw. entgegengesetzter Richtung.

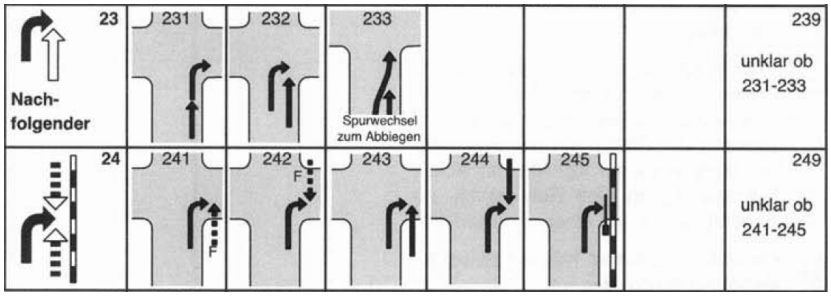


Abbildung 3.1: Untergliederung des Unfalltyps 2 für Rechtsabbiegende [Quelle: FGSV (2012), S. 53]

Unfalltyp 3: Einbiegen-/Kreuzen-Unfall

Kommt es zwischen einem einbiegenden oder kreuzenden Wartepflichtigen und einem vorfahrtsberechtigten Verkehrsteilnehmenden zu einem Unfall, so wird der Unfalltyp EK zugeordnet. (vgl. FGSV (2012), S. 54f.) Abbildung 3.2 zeigt die entsprechenden Untertypen, die in der vorliegenden Arbeit thematisiert werden.
Die dargestellten Untertypen unterscheiden sich in der Herkunft des bevorrechtigten Verkehrsteilnehmenden. Während sich dieser im Unfalltyp 30 von links nähert, erfolgt die Annäherung beim Unfalltyp 32 von rechts. Im Falle des Unfalltyps 34 ist der bevorrechtigte Verkehrsteilnehmende explizit als Radfahrender ausgewiesen, welcher sich auf dem Radweg kommend von links bzw. rechts annähert.

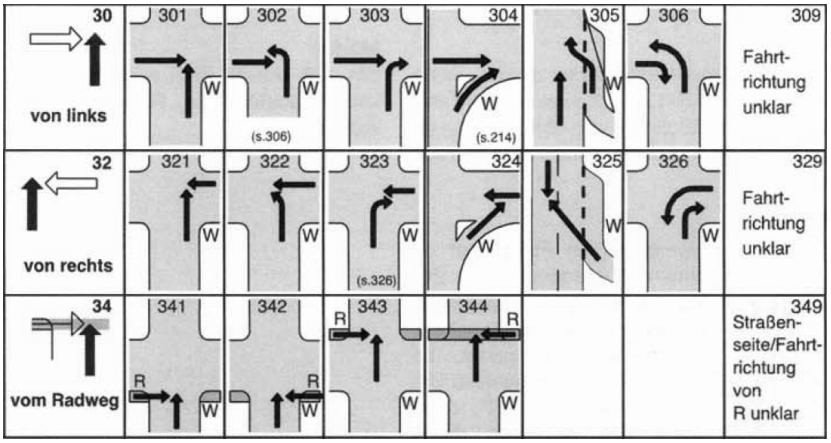


Abbildung 3.2: Untergliederung des Unfalltyps 3 nach Herkunft des bevorrechtigten Fahrzeugs [Quelle: FGSV (2012), S. 55]

4 Datengrundlage

Da der Schwerpunkt der Arbeit in der Untersuchung des Einflusses von Infrastrukturmerkmalen auf das Unfallgeschehen im Radverkehr in Dresden liegt, stellt das folgende Kapitel zunächst den vorliegenden Datensatz der Radverkehrsunfälle vor, welcher dann in den nächsten Schritten auf die für die in der Untersuchung relevanten Unfälle vom Typ AB und EK eingegrenzt wird. Im Anschluss erfolgt eine Vorstellung der Infrastrukturmerkmale, welche letztendlich als Eingangsgrößen für die Schätzungen dienen. Im letzten Teil findet sich ein Überblick über die Ausprägungen der in den Schätzungen verwendeten Variablen in Form einer deskriptiven Statistik.

4.1 Unfalldaten

Grundlage für die Untersuchungen bilden die von der Polizei erhobenen Daten zu Radverkehrsunfällen in der Stadt Dresden für die Jahre 2015 bis 2019. In diesen fünf Jahren kam es zu 6.569 Unfällen mit Radfahrendenbeteiligung, wovon 3.133 Unfälle (47,7 Prozent) an Knotenpunkten und 3.436 Unfälle (52,3 Prozent) auf der freien Strecke stattfanden. Da die Unfallkategorie und der Unfalltyp, wie in Kapitel 3 beschrieben, wesentliche Merkmale der Unfallforschung darstellen, zeigen die Abbildungen 4.1 und 4.2 überblicksartig deren Verteilungen bei den 6.569 Unfällen auf.

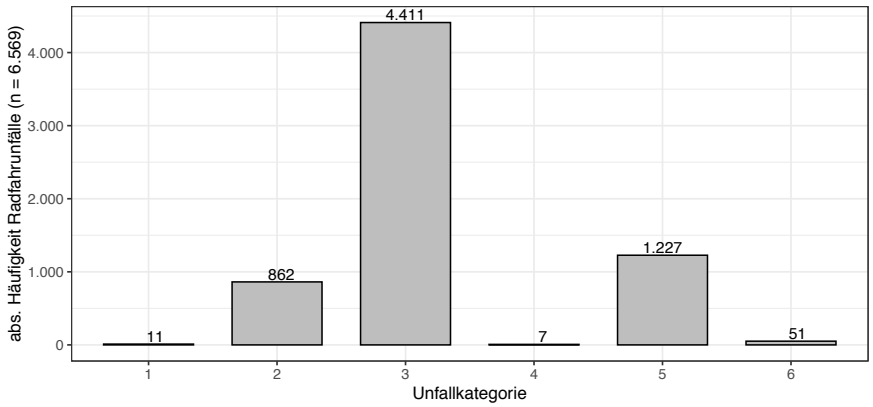


Abbildung 4.1: Häufigkeitsverteilung: Unfallkategorien aller Unfälle [Quelle: eigene Darstellung]

Nach Abbildung 4.1 ereigneten sich von den 6.569 Unfällen 5.284 Unfälle mit Personenschaden (80,4 Prozent, Kategorie 1 bis 3) und 1.285 Unfälle mit Sachschaden (19,6 Prozent, Kategorie 4 bis 6). Dabei trat die Unfallkategorie 3 mit 4.411 Unfällen

am häufigsten auf, gefolgt von Kategorie 5 mit 1.227 Unfällen.
Bei den Unfalltypen hingegen traten laut Abbildung 4.2 die Typen EK mit 2.367 Unfällen sowie AB mit 1.112 Unfällen am häufigsten auf, gefolgt von 1.070 Unfällen vom Typ F.

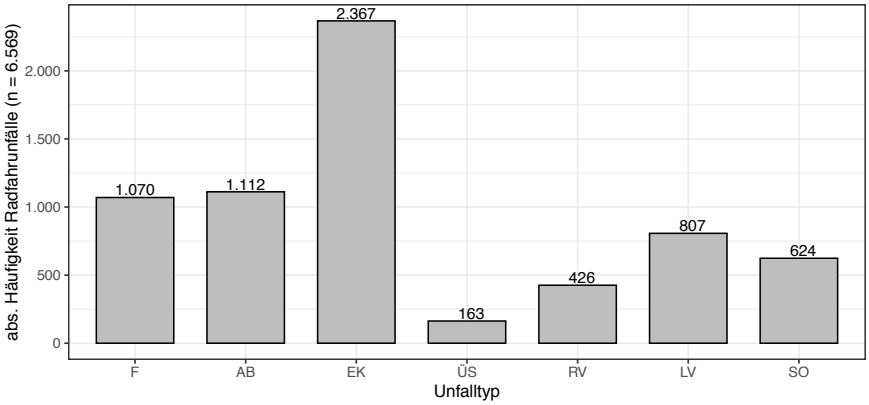


Abbildung 4.2: Häufigkeitsverteilung: Unfalltypen aller Unfälle [Quelle: eigene Darstellung]

Da nicht für jeden dieser Unfälle Angaben zum DTV vorliegen und der DTV eine wesentliche Größe bei der Beschreibung des Unfallgeschehens darstellt, wird der Datensatz auf die Unfälle eingegrenzt, die an Zufahrten von Knotenpunkten stattfanden, an denen Angaben zum DTV vorliegen. Dadurch verringert sich die Anzahl der Unfälle von anfangs 6.569 Unfällen auf 548 Unfälle. Die räumliche Verteilung dieser Unfälle in Dresden ist in Abbildung B.1 in Anhang B dargestellt.
Wie in den Abbildungen 4.3 und 4.4 ersichtlich wird, repräsentiert der eingegrenzte Unfalldatensatz den ursprünglichen Unfalldatensatz hinsichtlich der Unfallkategorien und -typen gut, sodass dieser als angemessenes Abbild des ursprünglichen Unfalldatensatzes angesehen werden kann. Abbildung 4.3 stellt dabei die Anteile der einzelnen Unfallkategorien an den jeweiligen Unfallzahlen im ursprünglichen und im eingegrenzten Unfalldatensatz gegenüber. Es ist zu erkennen, dass die Anteile der einzelnen Unfallkategorien im eingegrenzten Unfalldatensatz nur minimal von denen im ursprünglichen Unfalldatensatz abweichen. Im Falle der Unfalltypen in Abbildung 4.4 zeigt sich ein ähnliches Bild, wobei die Diskrepanz der einzelnen Anteile im eingegrenzten und im ursprünglichen Unfalldatensatz etwas höher ist als bei den Unfallkategorien. Dennoch lässt sich im eingegrenzten Unfalldatensatz ebenfalls erkennen, dass die Unfalltypen EK und AB am häufigsten auftraten.
Nachdem die Eingrenzung des ursprünglichen Datensatzes zu einer angemessenen Datengrundlage führte, werden die entsprechenden Radverkehrsunfälle des eingegrenzten Datensatzes im Folgenden anhand einzelner Merkmale näher charakterisiert, um einen Eindruck über das Wesen der Unfälle zu vermitteln.

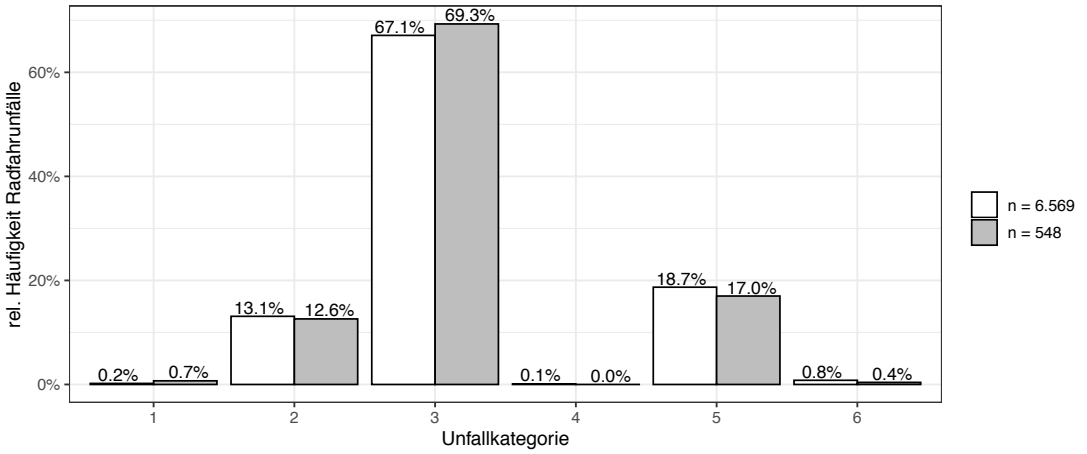


Abbildung 4.3: Repräsentativität des eingegrenzten Datensatzes bzgl. der Unfallkategorien [Quelle: eigene Darstellung]

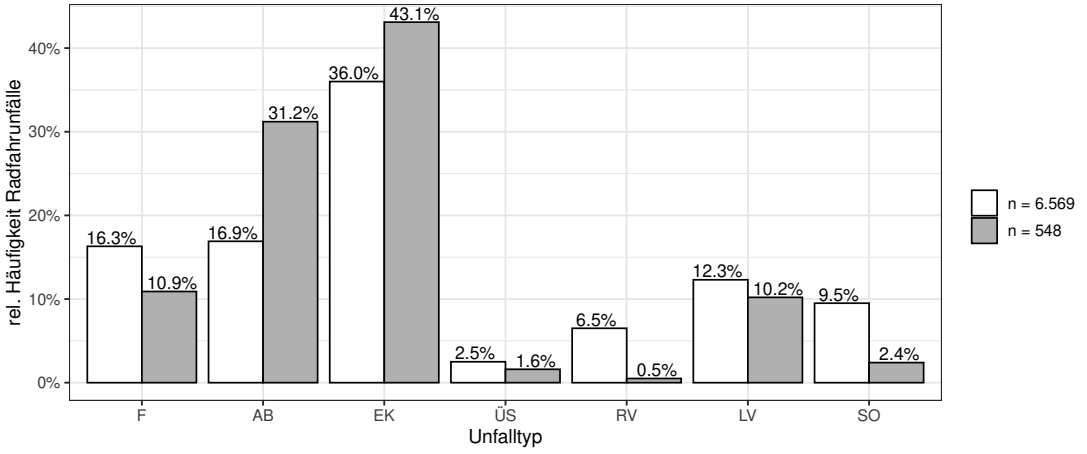


Abbildung 4.4: Repräsentativität des eingegrenzten Datensatzes bzgl. der Unfalltypen [Quelle: eigene Darstellung]

Merkmale der Radverkehrsunfälle im eingegrenzten Datensatz

Die zuvor erlangten Erkenntnisse bzgl. der Unfallkategorien und -typen im eingegrenzten Datensatz werden in den Abbildungen 4.5 und 4.6 noch weiter verdeutlicht. Demnach nehmen Unfälle mit Personenschaden (Unfallkategorie 1 bis 3, 453 Unfälle) einen Anteil von 82,7 Prozent ein, während Unfälle mit Sachschaden (Unfallkategorie 4 bis 6, 95 Unfälle) einen Anteil von 17,3 Prozent erlangen. Zuden dominiert die Unfallkategorie 3 mit 380 Unfällen. (vgl. Abbildung 4.5) Weiterhin ereigneten sich 171 Unfälle vom Typ AB und 236 Unfälle vom Typ EK, wodurch diese die beiden am häufigsten auftretenden Unfalltypen des eingegrenzten Datensatzes darstellen. (vgl. Abbildung 4.6)

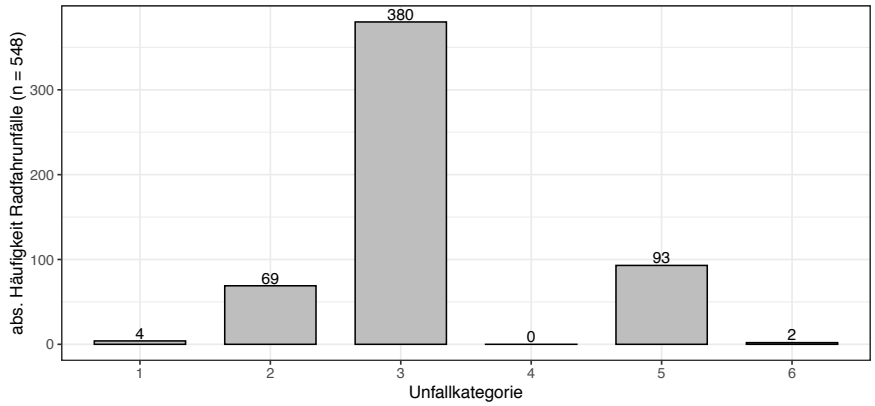


Abbildung 4.5: Häufigkeitsverteilung: Unfallkategorien des eingegrenzten Datensatzes [Quelle: eigene Darstellung]

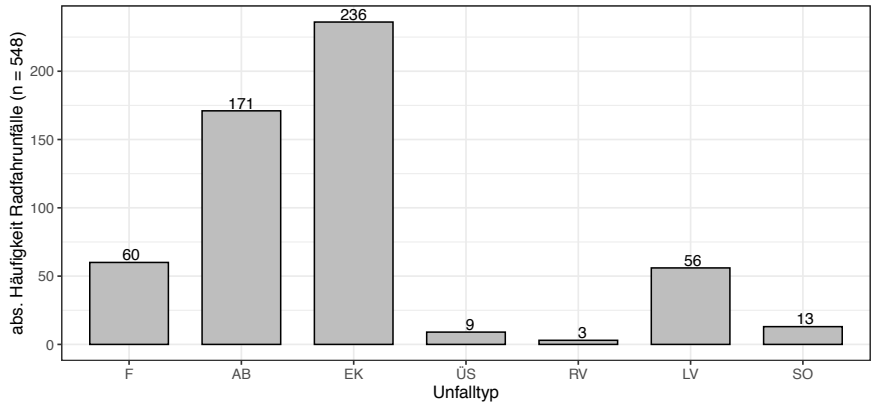


Abbildung 4.6: Häufigkeitsverteilung: Unfalltypen des eingegrenzten Datensatzes [Quelle: eigene Darstellung]

Wie bereits eingangs erwähnt, traten die betrachteten Radverkehrsunfälle in den Jahren 2015 bis 2019 auf. Demzufolge ergibt sich bei einer Grundgesamtheit von 548 Unfällen in fünf Jahren ein Mittelwert von 109,6 Unfällen pro Jahr. Die Verteilung der Unfälle auf die einzelnen Jahre ist im Folgenden dargestellt, wobei sich eine relativ gleichmäßige Verteilung über die Jahre erkennen lässt.

2015:	107 Unfälle
2016:	107 Unfälle
2017:	99 Unfälle
2018:	111 Unfälle
2019:	124 Unfälle
Summe:	548 Unfälle

Weiterhin gibt Abbildung 4.7 einen Überblick über die tageszeitliche Verteilung des Unfallgeschehens, wobei sich die Anzahl der Unfälle auf den betrachteten Zeitraum von fünf Jahren bezieht. Dabei bezeichnet bspw. Stunde 10 den Zeitraum zwischen 10:00 Uhr und 10:59 Uhr. Anhand der Unfallzahlen lassen sich Spitzenstunden ableiten, welche in der Abbildung zwischen 07:00 Uhr und 07:59 Uhr sowie zwischen 16:00 Uhr und 16:59 Uhr zu erkennen sind. Diese spiegeln annähernd die typischen Früh- und Nachmittagsspitzenstunden des allgemeinen Verkehrsaufkommens eines mittleren Werktags in Dresden wider, welche bspw. mit Berufsverkehr assoziiert werden. (vgl. Gerike et al. (2020), S. 31) Demzufolge scheint eine Beziehung zwischen allgemeinem Verkehrsaufkommen und dem Unfallgeschehen im Radverkehr zu bestehen.

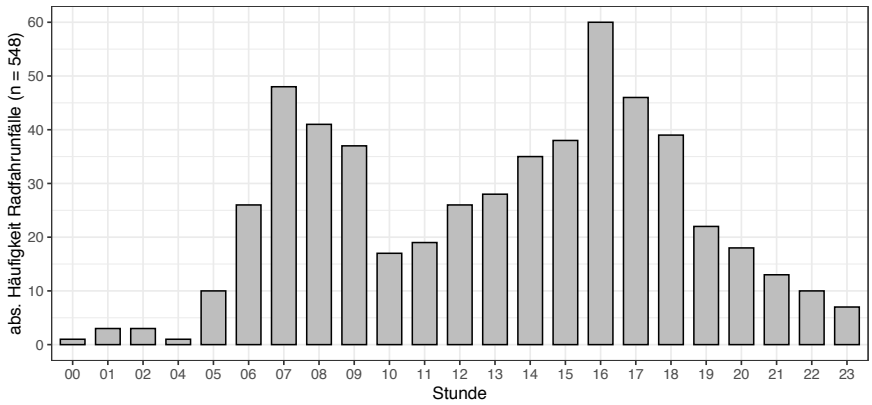


Abbildung 4.7: Tageszeitliche Verteilung: Unfallgeschehen im eingegrenzten Datensatz [Quelle: eigene Darstellung]

Abbildung 4.8 stellt zudem die Beteiligtenkonstellationen der 548 Radverkehrsunfälle dar, wobei *Fg* einen Fußgänger bezeichnet. Anzumerken ist dabei, dass die Abbildung die Konstellationen unabhängig davon aufzeigt, welcher der Beteiligten der Unfallverursacher und welcher der Unfallgegner ist. So beinhaltet bspw. die Konstellation *Rad-PKW* sowohl Unfälle, bei denen der Radfahrende der Verursachende ist als auch Unfälle, bei denen der PKW der Verursachende ist. Es ist zu erkennen, dass die Konstellation *Rad-PKW* mit Abstand am häufigsten auftritt (384 Unfälle), gefolgt von Unfällen, bei denen lediglich ein Radfahrender beteiligt ist (Konstellation *Rad*, 56 Unfälle) sowie der Konstellation *Rad-Rad* (32 Unfälle). Die übrigen Konstellationen treten in einem ähnlichen bzw. geringeren Umfang auf.
Da Abbildung 4.8 keine Aussage über den Unfallverursacher zulässt, stellt Abbildung 4.9

den normierten Anteil dar, bei dem der Radfahrende der Unfallverursachende der entsprechenden Unfallkonstellation ist. Dabei kennzeichnet *Fg* abermals einen Fußgänger. Es lässt sich bspw. erkennen, dass der Radfahrende den Unfall in 25 Prozent aller Unfälle der Konstellation *Rad-PKW* verursacht. Anders ausgedrückt bedeutet dies, dass der Radfahrende lediglich in einem von vier *Rad-PKW*-Unfällen der Verursachende ist.

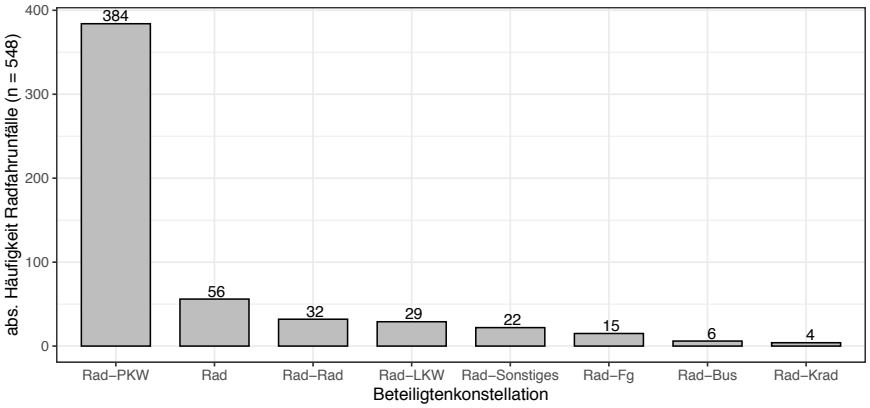


Abbildung 4.8: Häufigkeitsverteilung: Beteiligtenkonstellation im eingegrenzten Datensatz [Quelle: eigene Darstellung]

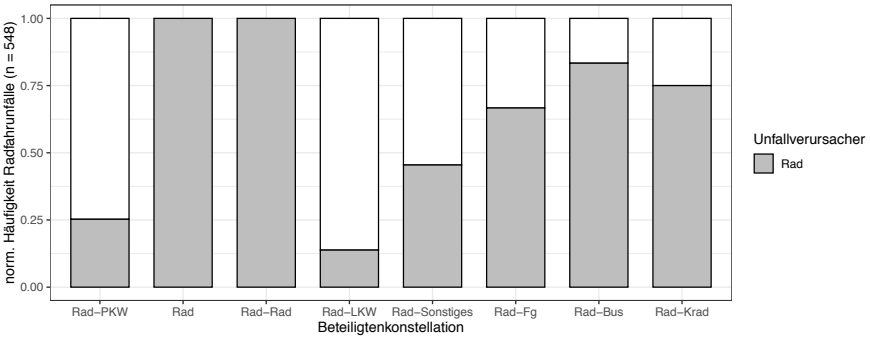


Abbildung 4.9: Anteil des Unfallverursachers Rad nach Beteiligtenkonstellation im eingegrenzten Datensatz [Quelle: eigene Darstellung]

Das Alter der Unfallverursachenden und -gegner ist in Abbildung 4.10 in Form von Boxplots[1] näher charakterisiert. Dabei beträgt der Median des Alters der Unfallverursachenden bspw. 40,00 Jahre und der Mittelwert, welcher durch das rote Kreuz gekennzeichnet ist, 41,03 Jahre. Im Gegensatz dazu beträgt der Median der Unfallgegner 35,00 Jahre und der Mittelwert derselben 38,63 Jahre. Somit sind die Unfallverursachenden im Median und im Mittel etwas älter als die Unfallgegner. Der Grund für die Abweichungen der Mediane vom entsprechenden Mittelwert liegt an den Ausreißern, welche bspw. im Boxplot der Unfallgegner auf der rechten Seite als Punkte gekennzeichnet sind. Anzumerken ist außerdem, dass die Anzahl der Beobachtungen der

[1] Ein Boxplot kennzeichnet das Minimum, das 25-Prozent-Quantil (=unteres Quantil), den Median, das 75-Prozent-Quantil (=oberes Quantil), das Maximum sowie Ausreißerwerte einer Variable grafisch. (vgl. Pruscha (2006), S. 13ff.)

Unfallverursachenden trotz der 548 Unfälle lediglich 519 beträgt, da das Alter der Unfallverursachenden in 29 Fällen unbekannt ist. Die Abweichung der 489 Beobachtungen der Unfallgegner von den insgesamt 548 Unfällen hingegen ist auf die 56 Radverkehrsunfälle ohne Unfallgegner (vgl. Abbildung 4.8) sowie auf das unbekannte Alter dreier Unfallverursacher zurückzuführen.

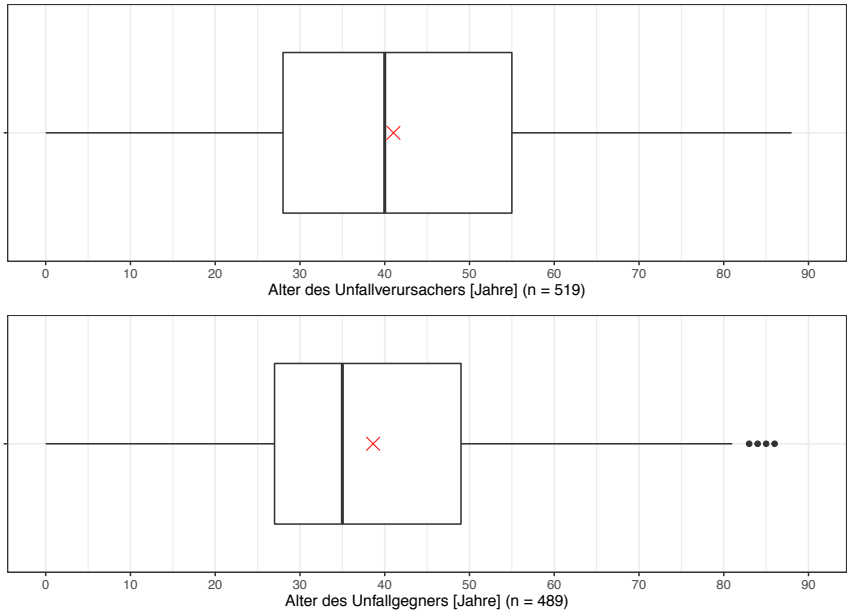


Abbildung 4.10: Boxplots: Alter der Unfallverursacher und -gegner im eingegrenzten Datensatz [Quelle: eigene Darstellung]

Charakterisierung der Unfälle vom Typ AB und EK

Da sich der Fokus auf Unfälle des gleichen Unfalltyps aufgrund der in Kapitel 3.2 angeführten Aspekte anbietet, betrachtet der folgende Abschnitt spezielle Untertypen der Unfalltypen AB und EK, welche die beiden am häufigsten auftretenden Unfalltypen des eingegrenzten Datensatzes (vgl. Abbildung 4.6) und damit die Grundlage der Untersuchungen darstellen, genauer. Eine separate Betrachtung des Unfallgeschehens nach Unfallkategorien hingegen ist nicht erforderlich, da die Schwere der Unfallfolge für die untersuchte Fragestellung irrelevant ist.
Die Untersuchung des Unfallgeschehens vom Typ AB beschränkt sich auf die Untertypen 23 und 24, da diese eine ähnliche unfallauslösende Konfliktsituation beschreiben. Wie bereits in Abbildung 3.1 ersichtlich wurde, unterscheiden sich diese Untertypen lediglich darin, ob sich der Geradeausfahrende in gleicher oder entgegengesetzter Richtung zum Rechtsabbiegenden bewegt und ob für diesen eine Radfahrfurt versetzt zur Fahrbahn vorhanden ist oder nicht. Da die Furtverfügbarkeit, welche in Kapitel 4.2 thematisiert wird, in den Untersuchungen als Einflussgröße einbezogen wird, bietet sich die gemeinsame Betrachtung der Untertypen 23 und 24 an. Diese gemeinsame Betrachtung

wird im Folgenden mit 23/24 bezeichnet.
Zwischen 2015 und 2019 ereigneten sich von den insgesamt 548 Unfällen 111 Unfälle vom Typ 23/24, was einem Anteil von 20,3 Prozent entspricht. Bezogen auf die 171 Unfälle vom Typ AB im eingegrenzten Datensatz (vgl. Abbildung 4.6) beträgt der Anteil 64,9 Prozent.
Die Unterteilung der Unfälle vom Typ 23/24 in dreistellige Unfalltypen ist in Abbildung 4.11 dargestellt. Dabei beschreiben die dreistelligen Typen 231, 232 und 243 eine Bewegung beider Beteiligter in gleiche Richtung und die Typen 242 und 244 eine Bewegung derselben in entgegengesetzte Richtung (vgl. Abbildung 3.1). Während ersteres in 98 Fällen auftrat, ereignete sich zweiteres lediglich in 13 Fällen, sodass in der Mehrheit von einer Bewegung beider Beteiligter in gleicher Richtung ausgegangen werden kann.

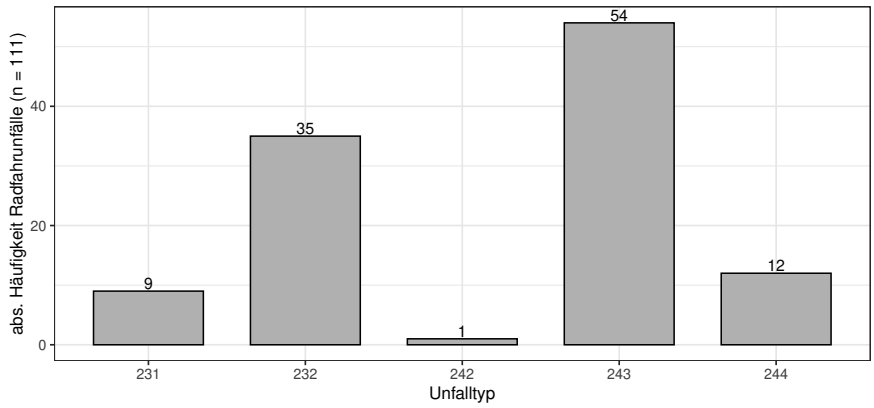


Abbildung 4.11: Häufigkeitsverteilung: dreistellige Unfalltypen für Unfälle vom Typ AB im eingegrenzten Datensatz [Quelle: eigene Darstellung]

Bei Unfällen vom Typ EK beschränkt sich die Untersuchung auf die Untertypen 30, 32 und 34, da auch diese eine ähnliche unfallauslösende Situation aufweisen. Wie in Abbildung 3.2 verdeutlicht wurde, unterscheiden sich diese Untertypen in der Herkunft des bevorrechtigten Verkehrsteilnehmenden von rechts oder links und in der Tatsache, ob sich der Verkehrsteilnehmende auf einer Radfahrfurt bewegt. Da die Furtverfügbarkeit - wie bereits zuvor beschrieben - in den Untersuchungen als Einflussgröße einbezogen wird, können auch diese Untertypen gemeinsam betrachtet werden. Im Folgenden wird die Zusammenfassung dieser Untertypen mit 30/32/34 bezeichnet.
Im betrachteten Zeitraum traten von den insgesamt 548 Unfällen 189 Unfälle vom Typ 30/32/34 auf. Das entspricht einem Anteil von 34,5 Prozent. Gemessen an den 236 Unfällen vom Typ EK (vgl. Abbildung 4.6) nehmen Unfälle vom Typ 30/32/34 einen Anteil von 80,1 Prozent ein.
Abbildung 4.12 stellt die Unterteilung der Unfälle vom Typ 30/32/34 in dreistellige Unfalltypen dar. Die Unfalltypen 301, 302, 303, 341 und 343 charakterisieren dabei den bevorrechtigten Verkehrsteilnehmer als von links kommend, während die Unfalltypen 321, 322, 342 und 344 diesen als von rechts kommend beschreiben (vgl. Abbil-

dung 3.2). Dabei taucht ein von links kommender Vorfahrtsberechtigter in 127 Fällen auf, wohingegen ein von rechts kommender Vorfahrtsberechtigter in 62 Fällen auftritt. Somit kann in der Mehrheit von einem von links kommenden Vorfahrtsberechtigten ausgegangen werden.

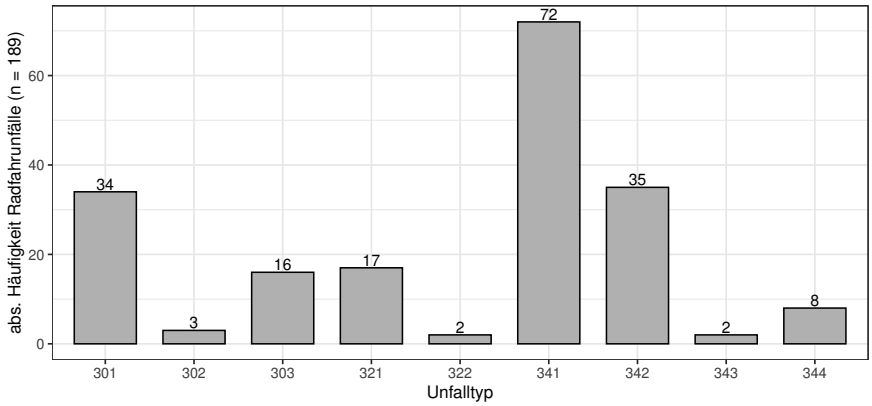


Abbildung 4.12: Häufigkeitsverteilung: dreistellige Unfalltypen für Unfälle vom Typ EK im eingegrenzten Datensatz [Quelle: eigene Darstellung]

Nachdem die Radfahrunfälle anhand einiger Merkmale näher charakterisiert wurden, gibt das folgende Kapitel Aufschluss über die Merkmale der Infrastruktur, die letztendlich als Einflussgrößen des Unfallgeschehens in die Untersuchungen einfließen.

4.2 Infrastrukturmerkmale

Die Untersuchung des Unfallgeschehens erfolgt hauptsächlich auf Basis von infrastrukturellen Merkmalen von Knotenpunktzufahrten, für welche Angaben zum DTV vorliegen. Der Grund, warum die Knotenpunkte nicht im Ganzen einbezogen werden, liegt darin, dass sich die einzelnen Zufahrten eines Knotenpunktes hinsichtlich ihrer infrastrukturellen Eigenschaften unterscheiden können. Somit ermöglicht eine Schätzung auf Basis von Zufahrten anstelle von Knotenpunkten genauere Ergebnisse. Eine Voraussetzung für eine genaue Schätzung liegt zudem darin, dass sich die infrastrukturellen Merkmale der Zufahrten im betrachteten Zeitraum nicht verändert haben. Alles in Allem stehen für die Untersuchung Merkmale von 484 Zufahrten in Dresden zur Verfügung, welche sich auf 133 Knotenpunkte verteilen. Anzumerken ist außerdem, dass die betrachtete Zufahrt jeweils jene ist, aus der sich der Unfallverursachende dem Knotenpunkt nähert.

Verteilung der Unfälle vom Typ 23/24 und 30/32/34 auf die Zufahrten

Wie zuvor erläutert, beschränkt sich die Untersuchung der Radverkehrsunfälle auf Unfälle der zusammengefassten Typen 23/24 sowie 30/32/34. Somit ergeben sich die abhängigen Variablen aus der Anzahl der Unfälle vom Typ 23/24 bzw. 30/32/34 pro Zufahrt im betrachteten Zeitraum. Die Anzahl der Unfälle beläuft sich dabei auf die zuvor erwähnten 111 Unfälle des Typs 23/24 sowie 189 Unfälle des Typs 30/32/34.
Abbildung 4.13 zeigt die Verteilung der Anzahl der Unfälle vom Typ 23/24 auf die Zufahrten. Das Minimum beträgt dabei null Unfälle und das Maximum sieben Unfälle. Weiterhin wird ersichtlich, dass es von den 484 Zufahrten lediglich 63 Zufahrten gibt, die mindestens einen Unfall vom Typ 23/24 aufweisen. Im Gegensatz dazu treten bei 421 Zufahrten keine Unfälle vom betrachteten Typ auf. Der Mittelwert ergibt sich demnach zu 0,23 Unfällen vom Typ 23/24 pro Zufahrt im betrachteten Zeitraum von fünf Jahren.

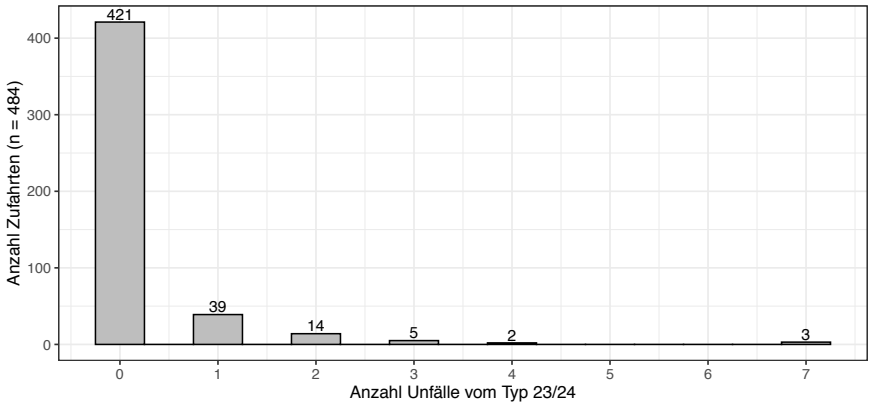


Abbildung 4.13: Häufigkeitsverteilung: Anzahl der Unfälle vom Typ 23/24 pro Zufahrt [Quelle: eigene Darstellung]

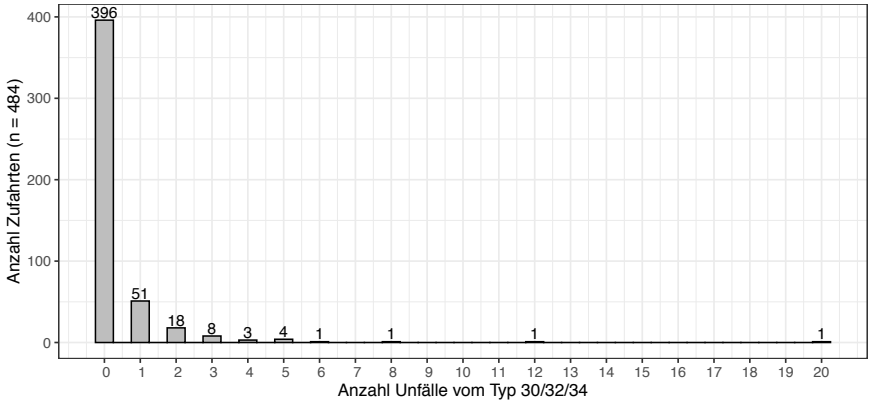


Abbildung 4.14: Häufigkeitsverteilung: Anzahl der Unfälle vom Typ 30/32/34 pro Zufahrt [Quelle: eigene Darstellung]

Für die Unfälle vom Typ 30/32/34 zeigt sich eine ähnliche Verteilung, die in Abbil-

dung 4.14 dargestellt ist. Hierbei beträgt das Minimum null Unfälle und das Maximum 20 Unfälle. Im Gegensatz zum Unfalltyp 23/24 gibt es beim Unfalltyp 30/32/34 88 Zufahrten mit mindestens einem Unfall vom entsprechenden Typ sowie 396 Zufahrten ohne Unfälle vom Typ 30/32/34. Dementsprechend ergibt sich der Mittelwert zu 0,39 Unfällen pro Zufahrt im betrachteten Zeitraum von fünf Jahren.
Der Grund für die geringen Mittelwerte beider betrachteter Unfalltypen liegt in den rechtsschiefen Verteilungen der abhängigen Variablen, welche dadurch zustande kommen, dass die meisten Beobachtungen bzw. Zufahrten sehr kleine Ausprägungen der entsprechenden Unfallzahlen aufweisen.

Infrastrukturelle Eigenschaften der Zufahrten

Im Folgenden werden die infrastrukturellen Merkmale der Zufahrten vorgestellt, die die möglichen unabhängigen Variablen der Schätzungen darstellen. Dabei werden die Verteilungen der einzelnen Merkmalsausprägungen auf die Zufahrten absolut dargestellt, während die Verteilungen der einzelnen Ausprägungen auf die Anzahl der Unfälle vom Typ 23/24 und 30/32/34 relativ angegeben werden. Der Grund dafür ist, dass sich die Anzahl der Beobachtungen für Unfälle vom Typ 23/24 und 30/32/34 unterscheidet, sodass ein Vergleich absoluter Zahlen nicht aussagekräftig wäre.
Die Knotenpunktform wird anhand der Anzahl der Knotenpunktarme gemessen, wodurch sich die Kategorien *Kreuzung*, *Einmündung* und *anderes* ergeben. *Kreuzungen* weisen dabei vier Knotenpunktarme auf, während *Einmündungen* durch drei Knotenpunktarme gekennzeichnet sind. Die Kategorie *anderes* hingegen beschreibt Knotenpunkte mit mehr als vier Knotenpunktarmen. Tabelle 4.1 gibt Auskunft über die Verteilung der einzelnen Ausprägungen auf die Zufahrten sowie auf die Anzahl der Unfälle vom Typ 23/24 und 30/32/34. Dabei umfasst der Datensatz mehr Zufahrten von Kreuzungen als von Einmündungen bzw. der Kategorie *anderes*. Zudem geschahen die meisten Unfälle beider betrachteter Unfalltypen an Zufahrten von *Kreuzungen*, gefolgt von Zufahrten an *Einmündungen* und *anderes*.

Tabelle 4.1: Häufigkeitsverteilung: Knotenpunktform

Knotenpunktform	Zufahrten	Unfalltyp 23/24 (n = 111)	Unfalltyp 30/32/34 (n = 189)
Kreuzung	317	71,2%	66,1%
Einmündung	133	19,8%	26,5%
anderes[1]	34	9,0%	7,4%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$

[1] *Knotenpunkte mit mehr als vier Knotenpunktarmen*
Quelle: eigene Darstellung

Ein weiteres Merkmal, das die infrastrukturellen Gegebenheiten beschreibt, ist die Art der Verkehrsregelung, deren Verteilung in Tabelle 4.2 dargestellt ist. Neben den Ausprägungen *Rechts-vor-Links*, *Verkehrszeichen Z205*, *Verkehrszeichen Z206* kann diese ebenfalls die Ausprägungen *Teil-Lichtsignalanlage (LSA)* sowie *LSA* annehmen. Dabei ist das *Verkehrszeichen Z205* allgemein unter dem Begriff *Vorfahrt gewähren* bekannt, während das *Verkehrszeichen Z206* als *Halt. Vorfahrt gewähren* bekannt ist. Die Verkehrsregelung *LSA* ist sowohl in Bezug auf die Zufahrten als auch in Bezug auf die Unfälle vom Typ 23/24 und 30/32/34 am häufigsten vertreten. Während die anderen Ausprägungen im Falle der Zufahrten und der Unfälle vom Typ 23/24 lediglich eine untergeordnete Rolle spielen, tritt das *Verkehrszeichen Z205* bei Unfällen vom Typ 30/32/34 häufiger auf.

Tabelle 4.2: Häufigkeitsverteilung: Verkehrsregelungsart

Verkehrsregelung	Zufahrten	Unfalltyp 23/24 (n = 111)	Unfalltyp 30/32/34 (n = 189)
Rechts-vor-Links	23	0,9%	1,6%
Verkehrszeichen Z205	72	5,4%	29,6%
Verkehrszeichen Z206	20	0,9%	2,1%
Teil-LSA	54	9,9%	15,9%
LSA	315	82,9%	50,8%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$

Quelle: eigene Darstellung

Weiterhin bildet das *Verkehrszeichen Z720*, welches im Allgemeinen auch als Grünpfeil bekannt ist, ein zu betrachtendes Zufahrtsmerkmal. Dieses findet jedoch bei 410 von 484 Zufahrten keine Anwendung, während es bei lediglich 74 Zufahrten vorzufinden ist. Bei den Unfällen vom Typ 23/24 und 30/32/34 zeichnet sich ein ähnlicher Trend ab. Hierbei spielt es für 73,9 Prozent aller Unfälle vom Typ 23/24 keine Rolle, während es für 26,1 Prozent der Unfälle relevant ist. Für Unfälle vom Typ 30/32/34 ist es zudem in 79,9 Prozent der Fälle irrelevant und lediglich in 20,1 Prozent der Fälle von Bedeutung.

Weitere infrastrukturelle Merkmale sind die Anzahl der Fahrstreifen der Linksabbiegenden, Geradeausfahrenden und Rechtsabbiegenden sowie die Anzahl der Mischfahrstreifen einer Zufahrt. Tabelle 4.3 stellt einen Überblick der entsprechenden Verteilungen der Ausprägungen dar. Dabei weisen alle vier Merkmale einen Minimalwert von null und einen Maximalwert von zwei Fahrstreifen auf, wobei im Falle der Linksabbiege- und Rechtsabbiegefahrstreifen jeweils die Ausprägung null dominiert. Bei den Geradeausfahrstreifen dominiert im Falle der Zufahrten und der Unfälle vom Typ 30/32/34 ebenfalls die Ausprägung null, während bei den Unfällen vom Typ 23/24 alle Ausprägungen einen ähnlichen Anteil aufweisen. Bei der Anzahl der Mischfahrstreifen dominieren zudem jeweils die Ausprägungen null und eins über die Ausprägung zwei.

Tabelle 4.3: Häufigkeitsverteilung: Anzahl der Fahrstreifen

Linksabbiegefahrstreifen	Zufahrten	Unfalltyp 23/24 (n = 111)	Unfalltyp 30/32/34 (n = 189)
0	303	55,0%	63,0%
1	166	42,3%	34,4%
2	15	2,7%	2,6%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$
Geradeausfahrstreifen			
0	332	36,9%	68,3%
1	101	37,8%	23,2%
2	51	25,3%	8,5%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$
Rechtsabbiegefahrstreifen			
0	382	64,9%	76,2%
1	97	35,1%	22,7%
2	5	0,0%	1,1%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$
Mischfahrstreifen			
0	147	36,0%	40,7%
1	324	64,0%	51,9%
2	13	0,0%	7,4%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$

Quelle: eigene Darstellung

Des Weiteren wird die Signalisierung der Links- und Rechtsabbiegenden je als Merkmal einbezogen. Dabei unterscheidet sich die Signalisierung der Linksabbiegenden in den Merkmalen *keine separate Signalisierung* und *separate Signalisierung*. Die Signalisierung der Rechtsabbiegenden hingegen weist fünf Ausprägungen auf. Diese sind *keine separate Signalisierung*, *separate Signalisierung über dreifeldigen Signalgeber*, *Dreiecksinsel ohne LSA*, *Dreiecksinsel mit LSA* sowie *separate Signalisierung über zweifeldigen Signalgeber*. Die Verteilungen der einzelnen Ausprägungen sind in Tabelle 4.4 dargestellt. Bei den Linksabbiegenden dominiert in allen Fällen die Ausprägung *keine separate Signalisierung*, jedoch nimmt die Ausprägung *separate Signalisierung* in allen drei Fällen ebenfalls nicht zu vernachlässigende Anteile an. Bei den Rechtsabbiegenden nimmt die Ausprägung *keine separate Signalisierung* sowohl bei den Zufahrten als auch bei den Unfällen vom Typ 23/24 und 30/32/34 Anteile über 90 Prozent ein, weshalb die übrigen Ausprägungen kaum bis keine Bedeutung finden.

Tabelle 4.4: Häufigkeitsverteilung: Signalisierung von Links- und Rechtsabbiegenden

Linksabbiegende	Zufahrten	Unfalltyp 23/24 (n = 111)	Unfalltyp 30/32/34 (n = 189)
keine separate Signalisierung	379	68,5%	83,1%
separate Signalisierung	105	31,5%	16,9%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$
Rechtsabbiegende			
keine separate Signalisierung	450	94,6%	97,8%
separate Signalisierung[1]	13	2,7%	1,1%
Dreiecksinsel ohne LSA	8	0,9%	0,0%
Dreiecksinsel mit LSA	6	0,9%	1,1%
separate Signalisierung[2]	7	0,9%	0,0%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$

[1] *dreifeldig,* [2] *zweifeldig*
Quelle: eigene Darstellung

Der Ausrundungsradius für Rechtsabbiegende ist ein in Metern gemessenes Merkmal der Zufahrten, welches in Abbildung 4.15 in Form von Boxplots sowohl für die Zufahrten als auch für die Anzahl der Unfälle vom Typ 23/24 sowie 30/32/34 charakterisiert wird. Demzufolge sind minimale Ausrundungsradien in Höhe von null Metern möglich. Diese treten bspw. an jenen Zufahrten auf, von denen ein Rechtsabbiegen aufgrund des Vorliegens einer Einmündung unmöglich ist. Die Mittelwerte, welche abermals durch ein rotes Kreuz gekennzeichnet sind, unterscheiden sich kaum zwischen den drei Fällen. Im Gegensatz dazu unterscheiden sich jedoch z.B. die Maximalwerte. Während diese im Fall der Zufahrten und im Fall der Unfälle vom Typ 23/24 etwas mehr als 90 Meter betragen, nimmt der Maximalwert bei den Unfällen vom Typ 30/32/34 einen Wert von etwa 63 Metern an.

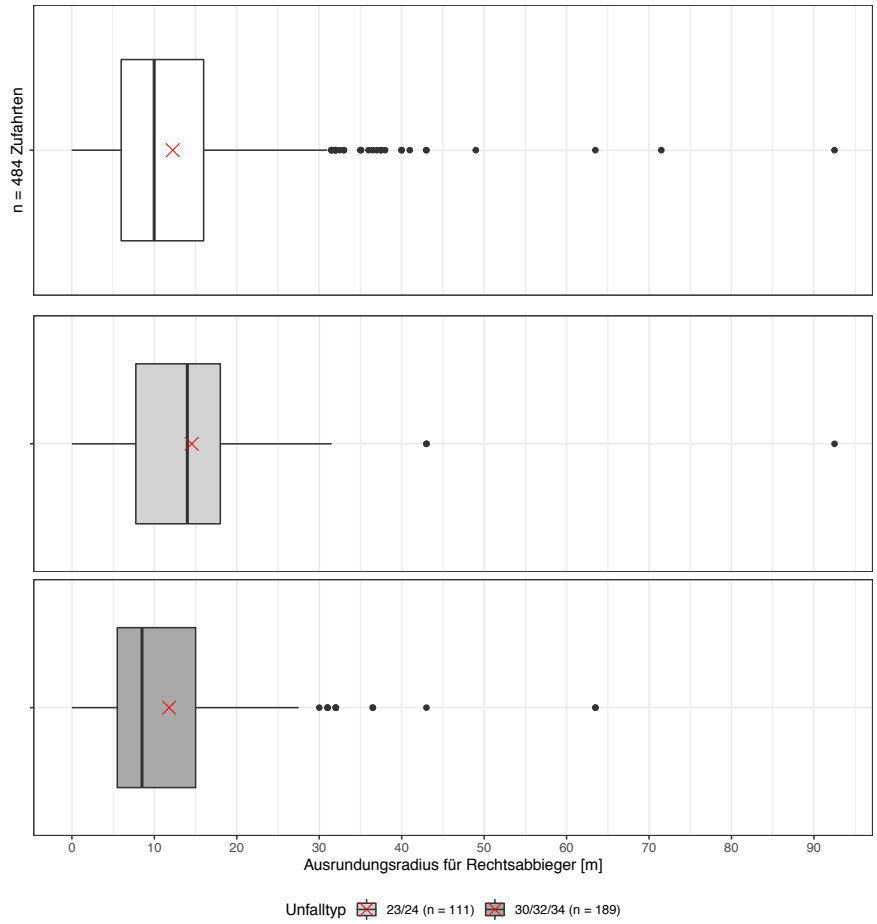


Abbildung 4.15: Boxplots: Ausrundungsradius für Rechtsabbiegende [Quelle: eigene Darstellung]

Der Abstand des ruhenden Verkehrs zum Knotenpunkt kann für die Untersuchung des Unfallgeschehens ebenfalls relevant sein. Dieser wird in sechs Kategorien unterteilt. Die erste Kategorie lautet dabei *Parken im Knotenpunkt* und beschreibt Zufahrten, bei denen kein Mindestabstand zum Knotenpunkt eingehalten werden muss. Neben der Ausprägung *keine Zufahrt vorhanden* gibt es zusätzlich Ausprägungen, die den eigentlichen Abstand widerspiegeln. Dieser wird demnach in die Kategorien *0 Meter bis 5 Meter*, *mehr als 5 Meter bis 15 Meter*, *mehr als 15 Meter bis 25 Meter* sowie *mehr als 25 Meter* unterteilt. Der Trennwert von 25 Metern leitet sich dabei aus der erforderlichen Haltesichtweite gemäß der *Richtlinien für die Anlage von Stadtstraßen - RASt 06* ab. (vgl. FGSV (2006), S. 124) Tabelle 4.5 gibt einen Überblick über die Verteilungen der einzelnen Ausprägungen, wobei die Ausprägung *mehr als 25 Meter* die übrigen Ausprägungen sowohl bei den Zufahrten als auch bei den Unfällen vom Typ 23/24 und 30/32/34 mit Abstand dominiert.

Tabelle 4.5: Häufigkeitsverteilung: Abstand ruhender Verkehr zum Knotenpunkt

Abstand ruhender Verkehr zum Knotenpunkt	Zufahrten	Unfalltyp 23/24 (n = 111)	Unfalltyp 30/32/34 (n = 189)
Parken im Knotenpunkt	3	0,0%	0,0%
keine Zufahrt vorhanden	1	0,0%	0,0%
0 bis 5 m	21	0,0%	4,3%
mehr als 5 m bis 15 m	19	0,9%	2,6%
mehr als 15 m bis 25 m	19	0,0%	3,2%
mehr als 25 m	421	99,1%	89,9%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$

Quelle: eigene Darstellung

Abbiegeverbote stellen ein weiteres zu betrachtendes Merkmal der Zufahrten von Knotenpunkten dar. Diese werden in die Kategorien *keine*, *Linksabbiegeverbot*, *Rechtsabbiegeverbot*, *Geradeausfahrverbot*, *nur Linksabbiegen*, *nur Rechtsabbiegen* sowie *nur Geradeausfahren* unterteilt. Die Verteilungen der einzelnen Ausprägungen sind in Tabelle 4.6 dargestellt, wobei die meisten Zufahrten ohne eine solche Regelung auskommen. Auch Unfälle vom Typ 23/24 und 30/32/34 ereigneten sich in den meisten Fällen an Zufahrten ohne eine solche Regelung. Weiterhin fällt auf, dass sich beim Unfalltyp 23/24, welcher sich auf das Rechtsabbiegen bezieht, dennoch Unfälle an Zufahrten zustande kommen, an denen Rechtsabbiegen aufgrund von Verboten nicht gestattet ist. Das lässt sich bspw. mit der Nichteinhaltung von Vorschriften einiger Verkehrsteilnehmenden erklären.

Tabelle 4.6: Häufigkeitsverteilung: Abbiegeverbote

Abbiegeverbot	Zufahrten	Unfalltyp 23/24 (n = 111)	Unfalltyp 30/32/34 (n = 189)
keine	364	69,4%	80,4%
Linksabbiegeverbot	62	24,3%	9,0%
Rechtsabbiegeverbot	17	2,7%	1,6%
Geradeausfahrverbot	4	0,0%	0,5%
nur Linksabbiegen	20	0,0%	8,5%
nur Rechtsabbiegen	1	0,0%	0,0%
nur Geradeausfahren	16	3,6%	0,0%
	$\sum = 484$	$\sum = 100\%$	$\sum = 100\%$

Quelle: eigene Darstellung

Wie bereits zuvor erwähnt wurde, könnte das Vorhandensein einer Radfahrfurt ein relevantes Merkmal für die Untersuchung des Unfallgeschehens darstellen. Dieses die Radverkehrsführung im Knotenpunkt beschreibende Merkmal weist die Ausprägungen *ohne Furt*, *mit Furt* sowie *Furt irrelevant* in zweifacher Ausführung auf. Dabei bezieht sich die Ausprägung *Furt irrelevant: 1* einerseits auf Zufahrten, an denen kein Konfliktpunkt mit Rechtsabbiegenden zustande kommen kann, da es sich bspw. um eine Einmündung handelt. Andererseits werden mit *Furt irrelevant: 2* Zufahrten beschrieben, an denen lediglich der Radverkehr rechts abbiegen kann, da es sich bspw. bei der rechts von der betrachteten Zufahrt befindlichen Zufahrt um eine ankommenden Einbahnstraße handelt. Diese beiden Ausprägungen könnten bei der Untersuchung des Unfalltyps 23/24 relevant sein, da sich dieser auf Rechtsabbiegende bezieht. Abbildung 4.16 veranschaulicht die Verteilungen der Ausprägungen, wobei die Ausprägung *ohne Furt* in allen drei Fällen den größten Anteil einnimmt. Das Vorhandensein einer Furt scheint dennoch nicht unbedeutend zu sein, denn diese spielt bei etwa je 19 Prozent der Zufahrten und der Unfälle vom Typ 23/24 eine Rolle. Bei den Unfällen vom Typ 30/32/34 ereigneten sich zudem etwa 15 Prozent der Unfälle an Zufahrten mit einer Radfahrfurt.

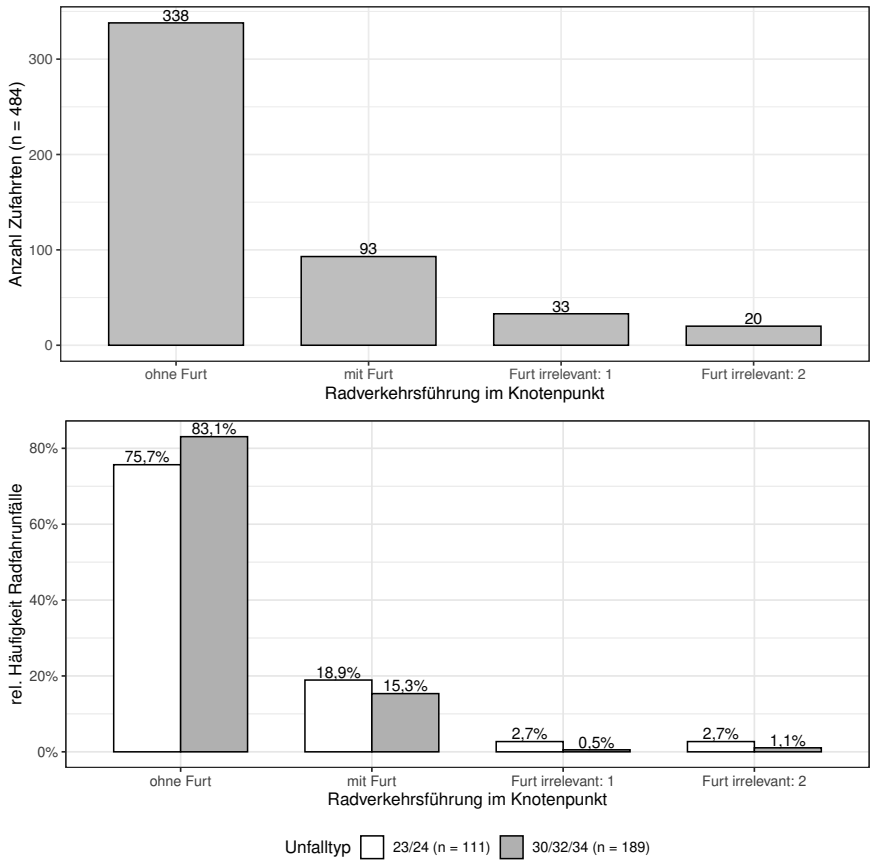


Abbildung 4.16: Häufigkeitsverteilung: Radverkehrsführung im Knotenpunkt [Quelle: eigene Darstellung]

Die Absetzung der Radfahrfurt im Knotenpunkt kann ebenfalls als Infrastrukturmerkmal betrachtet werden. Diese beschreibt den Abstand zwischen Fahrbahnrand und Radfahrfurt. Für die Untersuchungen wird dieses Merkmal kategorial einbezogen, wobei die Ausprägungen *ohne Furt*, *gering*, *mittel* und *weit* zur Verfügung stehen. Um eine geringe Absetzung handelt es sich, wenn der Abstand einen Wert zwischen null und maximal zwei Metern annimmt. Beträgt dieser hingegen mehr als zwei bis maximal vier Meter, so wird eine mittlere Absetzung angenommen. Demzufolge werden Abstände, die mehr als vier Meter groß sind, als weite Absetzung betrachtet. Abbildung 4.17 stellt die jeweiligen Verteilungen der Ausprägungen dar. Da die Radfahrfurt - wie im Abschnitt zuvor beschrieben - nur in wenigen Fällen von Relevanz ist, tritt auch hier die Ausprägung *ohne Furt* in allen drei Fällen am häufigsten auf. Bei Vorhandensein einer Radfahrfurt hingegen ist die Mehrheit der Zufahrten durch eine weite Absetzung der Radfahrfurt charakterisiert. Selbes gilt auch für die Unfälle der beiden betrachteten Typen 23/24 und 30/32/34, denn auch hier ereignet sich die Mehrheit der Unfälle mit Radfahrfurt an Zufahrten mit einer weiten Absetzung der selbigen.

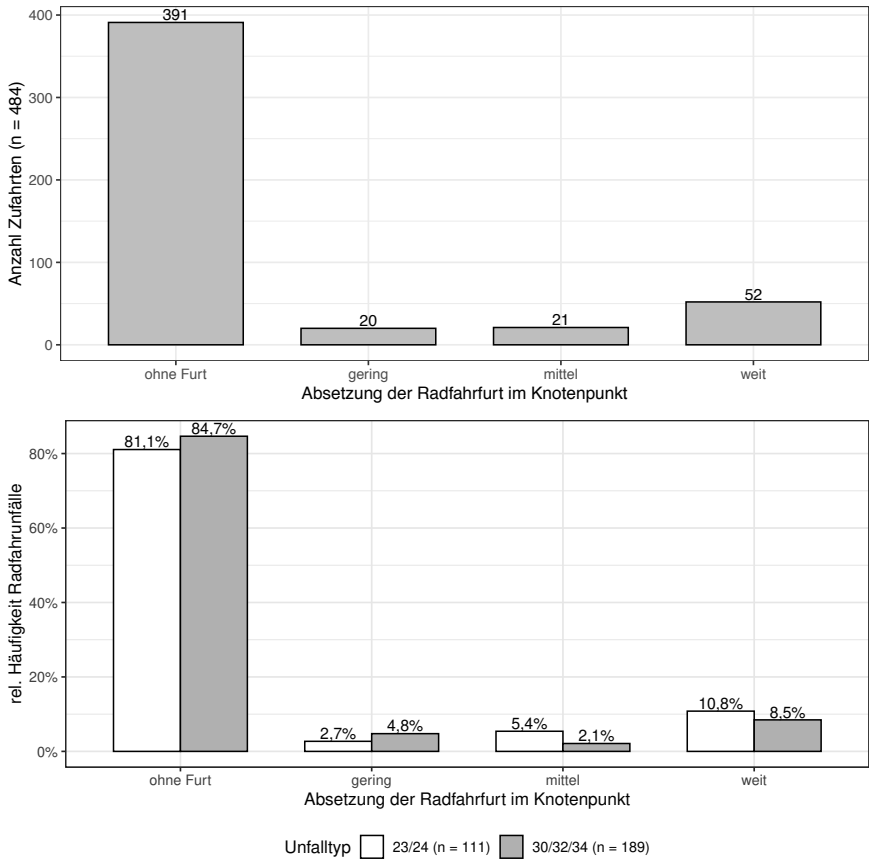


Abbildung 4.17: Häufigkeitsverteilung: Absetzung der Radfahrfurt im Knotenpunkt [Quelle: eigene Darstellung]

Der Signalgeber des Radverkehrs wird in die drei Kategorien *mit MIV*, *eigener Signalgeber* sowie *Fuß/Rad-Kombischeibe* unterteilt, wobei die Verteilungen der einzelnen Ausprägungen in Abbildung 4.18 dargestellt sind. Dabei nutzt der Radverkehr bei den meisten Zufahrten den Signalgeber des MIV. Auch bei den Unfällen vom Typ 30/32/34 sind Zufahrten, bei denen der Radverkehr den Signalgeber des MIV nutzt, mit Abstand am häufigsten vertreten. Bei Unfällen vom Typ 23/24 hingegen treten neben Unfällen an Zufahrten, bei denen der Radverkehr den Signalgeber des MIV nutzt, ebenfalls einige Unfälle an Zufahrten auf, bei denen der Radverkehr seinen eigenen Signalgeber besitzt. Ein Signalgeber in Form einer *Fuß/Rad-Kombischeibe* spielt in allen drei Fällen lediglich eine untergeordnete Rolle.

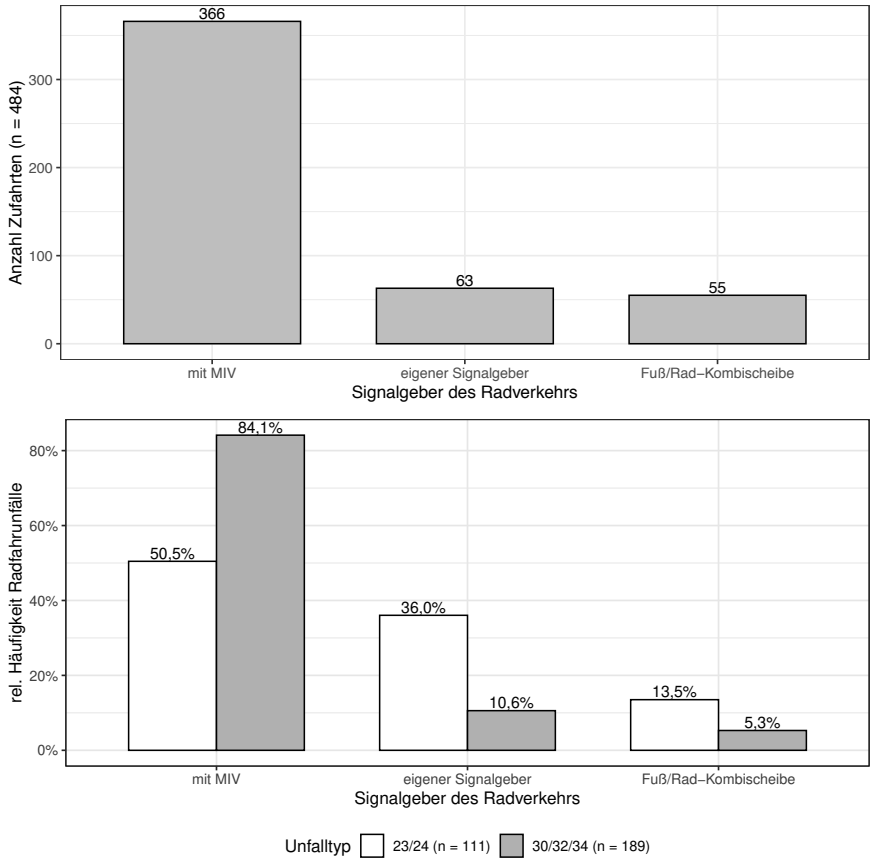


Abbildung 4.18: Häufigkeitsverteilung: Signalgeber des Radverkehrs [Quelle: eigene Darstellung]

Ein weiteres zufahrtsspezifisches Merkmal stellt die Führung des linksabbiegenden Radverkehrs dar. Dieses Merkmal wird in die Kategorien *keine/Mischverkehr*, *direkt aus Radverkehrsanlage*, *indirekt: nicht geführt* sowie *indirekt: geführt* unterteilt. Die Ausprägung *keine/Mischverkehr* beschreibt dabei, dass sich die linksabbiegenden Radfahrenden im Mischverkehr befinden und somit keine separate Radverkehrsführung vorliegt. Des Weiteren bezieht sich das Merkmal *indirekt: nicht geführt* auf eine Beschilderung und das Merkmal *indirekt: geführt* auf einen bspw. durch Markierungen, Aufstellflächen oder Signalgeber geführten Radverkehr. Die Verteilungen der einzelnen Ausprägungen sind in Abbildung 4.19 dargestellt. Dabei tritt die Ausprägung *keine/Mischverkehr* im Fall der Zufahrten und im Fall der Unfälle vom Typ 30/32/34 am häufigsten auf, während die Ausprägung *indirekt: nicht geführt* bei den Unfällen vom Typ 23/24 den größten Anteil aufweist.

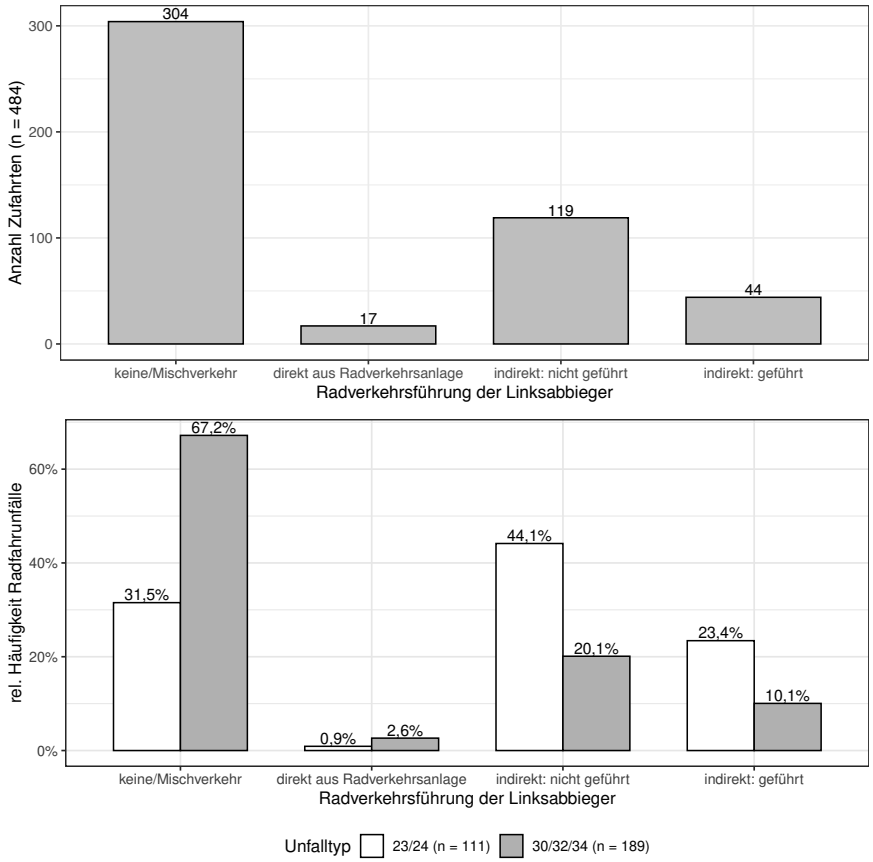


Abbildung 4.19: Häufigkeitsverteilung: Radverkehrsführung der linksabbiegenden Radfahrenden [Quelle: eigene Darstellung]

Die Roteinfärbung der Radfahrfurt stellt ein weiteres Infrastrukturmerkmal dar. Dieses kann neben den Ausprägungen *nein* und *ja, über gesamten Knoten* ebenfalls die Ausprägung *ja, teilweise* annehmen. Dabei beschreibt die Ausprägung *nein* entweder, dass keine zur entsprechenden Zufahrt gehörige Radfahrfurt vorhanden ist oder dass die zu einer Zufahrt gehörige Furt keinerlei Einfärbung enthält. Im Gegensatz dazu liegt bei der Ausprägung *ja, über gesamten Knoten* eine Einfärbung über die gesamte Straßenbreite vor. Die Ausprägung *ja, teilweise* beschreibt demzufolge Furten, bei denen lediglich ein Teilstück rot eingefärbt ist. Zur Verdeutlichung zeigt Abbildung B.2 in Anhang B die möglichen Ausprägungen anhand eines beispielhaften Knotenpunktes auf. Die Verteilung der jeweiligen Ausprägungen ist in Abbildung 4.20 dargestellt, wobei die meisten Zufahrten keine Roteinfärbung der Radfahrfurt aufweisen. Während das gleiche für Unfälle vom Typ 30/32/34 gilt, ist die Radfahrfurt bei den meisten Zufahrten mit Unfällen vom Typ 23/24 rot über den gesamten Knotenpunkt eingefärbt. Dennoch weist auch hier die Ausprägung *nein* einen hohen Anteil auf.

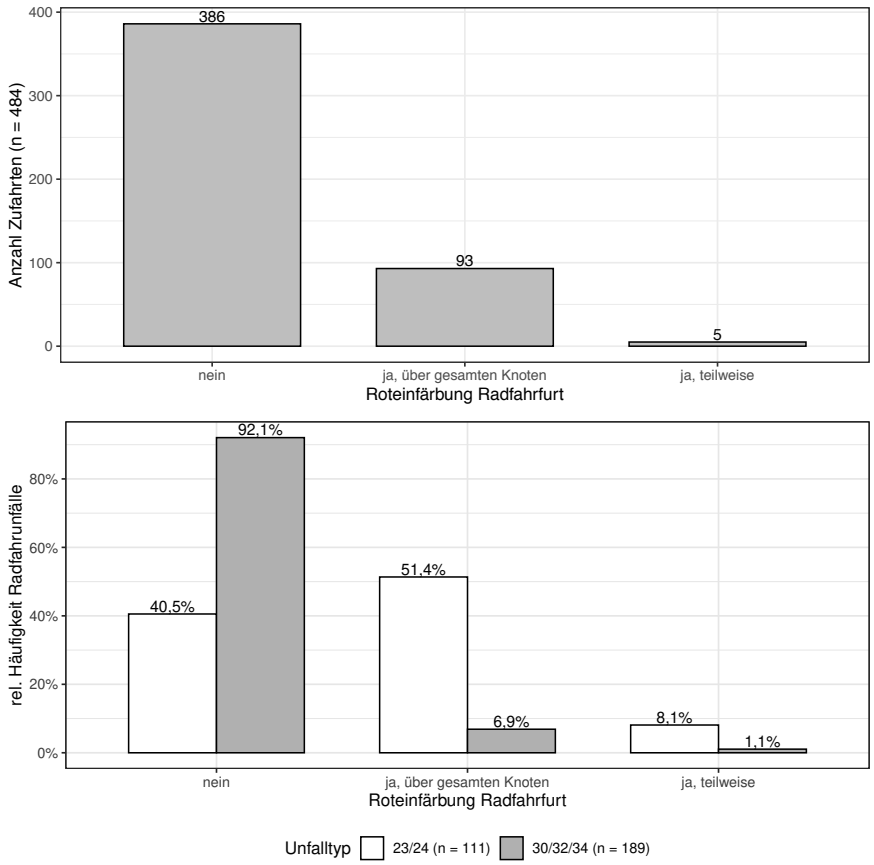


Abbildung 4.20: Häufigkeitsverteilung: Roteinfärbung der Radfahrfurt [Quelle: eigene Darstellung]

Verkehrsnachfrage der Zufahrten

Die Verkehrsnachfrage der einzelnen Zufahrten, welche eine bedeutende Rolle bei der Untersuchung des Unfallgeschehens einnimmt (vgl. Kapitel 2), wird anhand des DTV gemessen, welcher jeweils richtungsfein für den Kfz-, Schwer- sowie Radverkehr vorliegt. Die Schwerverkehrsnachfrage wird dabei als Anteil des Schwerverkehrs am Kfz-DTV einbezogen, wobei der Schwerverkehr alle Fahrzeuge, die mehr als 3,5 Tonnen Gewicht aufweisen - wie z.B. LKW's, Lastzüge, Busse und Straßenbahnen - beinhaltet. Abbildung 4.21 stellt die Ausprägungen der einzelnen Größen der Verkehrsnachfrage überblicksartig dar. Dabei zeigt die Abszisse jeweils die Verkehrsnachfrage und die Ordinate die Anzahl der betroffenen Zufahrten, wobei insgesamt 484 Zufahrten betrachtet werden. Sonderfälle bilden dabei Nachfragen von null sowie Schwerverkehrsanteile von 100 Prozent. Eine Nachfrage von null tritt bspw. auf, wenn die betroffene Zufahrt eine Einbahnstraße ist, welche vom Knotenpunkt abgeht, da in diesem Fall kein Fahrzeug entgegen der Einbahnstraße in den Knotenpunkt einfahren darf. Ein Schwerverkehrsanteil von 100 Prozent hingegen tritt bspw. an jenen Zufahrten auf, an denen das Befahren in Richtung Knotenpunkt lediglich Bussen und Bahnen gestattet ist.

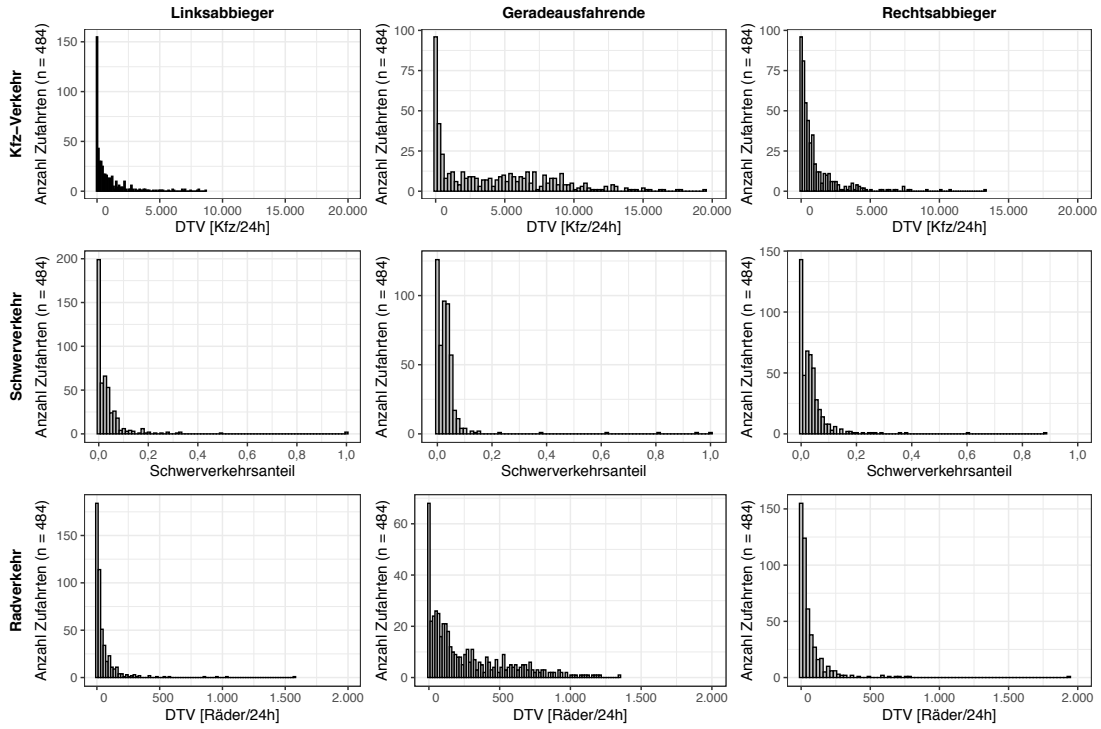


Abbildung 4.21: Häufigkeitsverteilung: richtungsbezogener DTV für Kfz und Radverkehr sowie Schwerverkehrsanteil [Quelle: eigene Darstellung]

4.3 Überblick über verwendete Variablen

Nachdem die Infrastrukturmerkmale der 484 Zufahrten, von denen die meisten kategorial vorliegen, vorgestellt wurden, gibt Tabelle 4.7 einen Überblick über die einzelnen Ausprägungen der verwendeten Variablen in Form einer deskriptiven Statistik. Dabei werden bei kategorialen Merkmalen die möglichen Ausprägungen und bei metrisch skalierten Merkmalen die Kennwerte Minimum, Median, Mittelwert sowie Maximum angegeben. Da jeweils eine Schätzung in Abhängigkeit der Anzahl der Unfälle vom Typ 23/24 sowie vom Typ 30/32/34 pro Zufahrt durchgeführt wird, werden die beiden mit y_1 und y_2 bezeichneten abhängigen Variablen einzeln beleuchtet. Außerdem werden die in den Schätzungen verwendeten Variablenbezeichnungen x_1 bis x_{26} eingeführt.

Tabelle 4.7: Deskriptive Statistik der verwendeten Variablen

	Definition	Minimum	Median	Mittelwert	Maximum
		(Ausprägungen bei kategorialen Variablen)			
y_1	Anzahl der Unfälle vom Typ 23/24 pro Zufahrt	0,00	0,00	0,23	7,00
y_2	Anzahl der Unfälle vom Typ 30/32/34 pro Zufahrt	0,00	0,00	0,39	20,00
x_1	Knotenpunktform	- Kreuzung - Einmündung - anderes			
x_2	Verkehrsregelungsart	- Rechts-vor-Links - Verkehrszeichen Z205: Vorfahrt gewähren - Verkehrszeichen Z206: Halt. Vorfahrt gewähren - Teil-LSA - LSA			
x_3	Grünpfeil (Z720)	- nicht vorhanden - vorhanden			
x_4	Anzahl Linksabbiegefahrstreifen	0,00	0,00	0,41	2,00
x_5	Anzahl Geradeausfahrstreifen	0,00	0,00	0,42	2,00
x_6	Anzahl Rechtsabbiegefahrstreifen	0,00	0,00	0,22	2,00
x_7	Anzahl Mischfahrstreifen	0,00	1,00	0,72	2,00
x_8	Signalisierung Linksabbiegende	- keine separate Signalisierung - separate Signalisierung			

x_9	Signalisierung Rechtsabbiegende	- keine separate Signalisierung - separate Signalisierung: dreifeldig - Dreiecksinsel ohne LSA - Dreiecksinsel mit LSA - separate Signalisierung: zweifeldig			
x_{10}	Ausrundungsradius Rechtsabbiegende [m]	0,00	10,00	12,25	92,50
x_{11}	Abstand ruhender Verkehr zum Knotenpunkt	- Parken im Knotenpunkt - keine Zufahrt vorhanden - 0 m bis 5 m - mehr als 5 m bis 15 m - mehr als 15 m bis 25 m - mehr als 25 m			
x_{12}	Abbiegeverbote	- keine - Linksabbiegeverbot - Rechtsabbiegeverbot - Geradeausfahrverbot - nur Linksabbiegen - nur Rechtsabbiegen - nur Geradeausfahren			
x_{13}	Radverkehrsführung im Knotenpunkt	- ohne Furt - mit Furt - kein Konfliktpunkt mit Rechtsabbiegenden - nur Radverkehr als Rechtsabbiegende			
x_{14}	Absetzung Radfahrfurt	- ohne Furt - gering (0 m bis 2 m) - mittel (mehr als 2 m bis 4 m) - weit (mehr als 4 m)			
x_{15}	Signalgeber Radverkehr	- mit MIV - eigener Signalgeber - Fuß/Rad-Kombischeibe			
x_{16}	Radverkehrsführung Linksabbiegende	- keine/Mischverkehr - direkt aus Radverkehrsanlage - indirekt: nicht geführt - indirekt: geführt			
x_{17}	Roteinfärbung Radfahrfurt	- nein - ja, über gesamten Knoten - ja, teilweise			
x_{18}	DTV Kfz Linksabbiegende [Kfz/24h]	0,00	310,00	880,00	8.660,00
x_{19}	DTV Kfz Geradeausfahrende [Kfz/24h]	0,00	2.900,00	4.068,40	19.450,00
x_{20}	DTV Kfz Rechtsabbiegende [Kfz/24h]	0,00	470,00	1.031,00	13.300,00

x_{21}	Schwerverkehrsanteil Linksabbiegende	0,00	0,02	0,04	1,00
x_{22}	Schwerverkehrsanteil Geradeausfahrende	0,00	0,03	0,04	1,00
x_{23}	Schwerverkehrsanteil Rechtsabbiegende	0,00	0,03	0,04	0,88
x_{24}	DTV Radverkehr Linksabbiegende [Räder/24h]	0,00	20,32	54,61	1.572,87
x_{25}	DTV Radverkehr Geradeausfahrende [Räder/24h]	0,00	145,82	266,13	1.346,40
x_{26}	DTV Radverkehr Rechtsabbiegende [Räder/24h]	0,00	28,03	66,87	1.935,48

Quelle: eigene Darstellung

5 Methodik

Das folgende Kapitel stellt die Methodik vor, mit welcher die Einflussfaktoren der Radfahrunfälle untersucht werden. Nachdem ein Korrelationsmaß vorgestellt wurde, geht das Kapitel dabei zunächst auf das RFV mit seinen Grundlagen, Modellgütekriterien sowie Möglichkeiten zur Ermittlung der Variablenbedeutsamkeit ein. Im Anschluss dazu wird die Negative Binomialregression - deren zugrunde gelegtes Modell ein Generalized Linear Model/Verallgemeinertes Lineares Modell (GLM) ist - als vergleichendes Verfahren vorgestellt. Dabei liegt der Fokus zunächst auf den Grundlagen sowie der Anwendung in Form von Accident Prediction Models. Weiterhin stellt das Kapitel das LASSO-Verfahren zur Variablenselektion vor. Neben Modellgütekriterien und dem Vorgehen bei der Ermittlung der Variablenbedeutsamkeit erfolgt zudem ein Überblick über die Kriterien der Modelldiagnostik, welche nach Durchführung eines linearen Modells zu prüfen sind.

5.1 Korrelationsbetrachtung

Um einen ersten Eindruck vom Zusammenhang der Variablen zu bekommen, kann der 1904 von Pearson eingeführte Kontingenzkoeffizient angewandt werden. Dabei handelt es sich um ein Zusammenhangsmaß für nominale Variablen, welches auf der Größe Chi-Quadrat χ^2 basiert. Um diese zu ermitteln, werden Kontingenztabellen, die die absoluten Häufigkeiten des Auftretens von Ausprägungspaaren zweier nominal skalierter Variablen abbilden, benötigt. Anhang C stellt den schematische Aufbau einer solchen Kontingenztabelle sowie die Erläuterung der zugehörigen Bestandteile dar. Ausgehend davon wird in diesem Kapitel, welches sich auf Benninghaus (2005, S. 116f.) und Cleff (2012, S. 80ff.) stützt, der Kontingenzkoeffizient nach Pearson hergeleitet.
Der χ^2-Wert ergibt sich durch den linken Ausdruck in Gleichung 5.1.1.

$$\chi^2 = \sum_{i=1}^{v} \sum_{j=1}^{w} \frac{(n_{ij} - n_{ij}^e)^2}{n_{ij}^e} \quad mit \quad n_{ij}^e = \frac{n_{i.} \cdot n_{.j}}{n} \tag{5.1.1}$$

Dabei gibt n_{ij} die beobachtete absolute Häufigkeit und n_{ij}^e die unter Unabhängigkeit[1] erwartete absolute Häufigkeit des Auftretens der entsprechenden Ausprägungskombination an. Außerdem bezeichnet v die Anzahl der Ausprägungen der ersten Variable und w analog die Anzahl der Ausprägungen der zweiten Variable. Zudem erfolgt die

[1] Unabhängigkeit unterstellt das Fehlen eines Zusammenhangs zwischen beiden Variablen.

Ermittlung von n_{ij}^e unter Einbeziehung der Randhäufigkeiten $n_{i.}$ bzw. $n_{.j}$ sowie der Gesamtanzahl der Beobachtungen n.
Durch die Bildung der Differenz von n_{ij} und n_{ij}^e wird ein erster Eindruck über die Abweichung der beobachteten Daten von der Unabhängigkeit vermittelt. Je größer die Differenz ist, desto größer ist tendenziell auch die Abhängigkeit der beiden Variablen. Damit sich positive und negative Differenzen jedoch nicht aufheben, werden die Differenzen quadriert und anschließend durch n_{ij}^e dividiert. Generell kann χ^2 Werte größer oder gleich null annehmen. Null ist dieser, wenn beide Variablen unabhängig voneinander sind und somit kein Unterschied zwischen den erwarteten absoluten Häufigkeiten besteht. Je stärker die Abhängigkeit der Variablen ist, desto größer ist tendenziell auch der χ^2-Wert.
Ein Nachteil des χ^2-Wertes liegt jedoch darin, dass dieser mit der Größe der Stichprobe sowie mit der Anzahl der Ausprägungen der beiden Variablen steigt. Um diesem Nachteil entgegen zu wirken, kommen Zusammenhangsmaße, die auf χ^2 basieren, zum Einsatz. Eines dieser Maße ist der eingangs erwähnte Kontingenzkoeffizient nach Pearson C, dessen Berechnungsvorschrift in Gleichung 5.1.2 dargestellt ist.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \in [0; 1[\tag{5.1.2}$$

C liegt im Bereich zwischen null und eins, wobei null wiederum bedeutet, dass beide Variablen unabhängig voneinander sind. Ein größerer Wert spiegelt demzufolge eine größere Abhängigkeit wider. Obwohl sich C bei zunehmender Anzahl der Merkmalsausprägungen dem Wert eins nähern kann, wird dieser nie erreicht. Um einen normierten Kontingenzkoeffizienten zu erhalten, der dennoch den Wert eins annehmen kann, wird Gleichung 5.1.2 um den maximal erreichbaren Wert C_{max}, dessen Berechnungsvorschrift in Gleichung 5.1.3 dargestellt ist, korrigiert. Dabei gibt v die Anzahl der Zeilen und w die Anzahl der Spalten der entsprechenden Kontingenztabelle an.

$$C_{max} = \sqrt{\frac{min(v, w) - 1}{min(v, w)}} \tag{5.1.3}$$

Letztendlich ergibt sich der korrigierte Kontingenzkoeffizient C_{korr} in Gleichung 5.1.4 durch Division der Gleichungen 5.1.2 und 5.1.3.

$$C_{korr} = \frac{C}{C_{max}} = \sqrt{\frac{\chi^2}{\chi^2 + n}} \cdot \sqrt{\frac{min(v, w)}{min(v, w) - 1}} \in [0; 1] \tag{5.1.4}$$

5.2 Random Forest

Das auf Entscheidungsbäumen basierende RFV findet beim maschinellen Lernen sowohl bei Fragestellungen der Klassifikation, falls die abhängige Variable kategoriell ist, als auch bei solchen der Regression, falls die abhängige Variable metrisch skaliert ist,

Anwendung. (vgl. Loh (2011), S. 14) Für ein besseres Verständnis beleuchtet der folgende Abschnitt zunächst Entscheidungsbäume der Klassifikation und Regression sowie den bekannten Classification and Regression Tree (CART)-Algorithmus im Allgemeinen. Die darauf folgenden Ausführungen beziehen sich dann aufgrund der untersuchten Fragestellung auf die Anwendung bei Regressionsfragestellungen. Dabei werden neben dem Bagging-Algorithmus, welcher die Grundlage des RFV darstellt, das Vorgehen im RFV, Modellgütekriterien sowie Kenngrößen zur Ermittlung der Bedeutsamkeit einzelner Variablen vorgestellt.

5.2.1 Grundlagen

Um den Ablauf des RFV zu verstehen, vermittelt das vorliegende Kapitel zunächst ein generelles Verständnis für Entscheidungsbäume und CART's.

Entscheidungsbäume der Klassifikation und Regression

Entscheidungsbäume der Klassifikation und Regression sind Methoden des maschinellen Lernens, um aus einem Trainingsdatensatz ein Vorhersagemodell für neue Daten zu generieren. Die Idee dahinter liegt in der mehrmaligen Aufteilung des Trainingsdatensatzes in Untermengen, sodass am Ende jede Beobachtung genau einer Untermenge zugeordnet werden kann. Dabei erfolgt die Aufteilung des Datensatzes durch die wiederholte Aufsplittung von Variablen an Knoten - beginnend am Wurzelknoten -, wobei zur Aufsplittung ein bestimmtes Kriterium angewandt wird. Dieses entscheidet letztendlich darüber, welche Variable zur Aufsplittung am entsprechenden Knoten herangezogen wird. Weiterhin gibt es ein Kriterium, das darüber entscheidet, nach wie vielen Aufsplittungen pro Pfad keine derselben mehr stattfinden sollen. Im Anschluss werden dann am Ende eines jeden Pfades Blattknoten gebildet, welche die vorhergesagten Werte des entsprechenden Pfades enthalten.
Um einen solchen Entscheidungsbaum zu generieren, gibt es verschiedene Algorithmen, die sich in der möglichen Struktur des Baums unterscheiden. Die Unterschiede liegen dabei bspw. im Kriterium für die Aufsplittung von Variablen an einem Knoten. Der bekannteste dieser Algorithmen ist der CART-Algorithmus, auf dessen Anwendung bei Klassifikations- und Regressionsfragestellungen im Folgenden näher eingegangen wird. (vgl. Loh (2011), S. 14ff.)

CART-Algorithmus

Die sogenannten CART's, welche in den 1980er Jahren von Breiman et al. (1984) entwickelt wurden, sind ein Mittel, um Klassifikations- und Regressionsprobleme zu modellieren. Außerdem stellen sie binäre Entscheidungen dar und arbeiten unabhängig von Annahmen wie der Normalverteilungsannahme der linearen Regression. Weiterhin bilden CART's die Grundlage für andere Algorithmen wie bspw. dem Bagging und dem

RFV. (vgl. Sutton (2005), S. 304ff.)
Wie zuvor beschrieben, gibt es für jeden Algorithmus unterschiedliche Kriterien, die den Aufbau des Entscheidungsbaums beeinflussen. Im Falle der Klassifikation, also bei Vorliegen einer kategorialen abhängigen Variable, nutzt der CART-Algorithmus den „Gini-Impurity-Index“ als Aufsplittungskriterium an den einzelnen Knoten. Dieser stellt ein Maß für die Missklassifikation durch das Aufsplitten am Knoten anhand einer bestimmten Variable dar. Der „Gini-Impurity-Index“ ist umso niedriger, je besser die Kategorien an einem Knoten anhand der gewählten Variable getrennt werden. Letztendlich wird von allen Variablen die Variable zur Aufsplittung verwendet, die die Kategorien am besten separiert bzw. den niedrigsten „Gini-Impurity-Index“ aufweist. (vgl. Raileanu und Stoffel (2004), S. 79f.) Wenn der entsprechende Entscheidungsbaum aufgespannt wurde, wird die vorhergesagte Kategorie für jeden Blattknoten über eine Mehrheitsentscheidung ermittelt.[2] (vgl. Loh (2011), S. 14ff.)
Für Regressionsfragestellungen, bei denen die abhängige Variable metrisch skaliert ist, verwendet das CART-Verfahren ein auf Residuen basiertes Kriterium für die Aufsplittung der Knoten. Das Ziel liegt dabei darin, dass die beobachteten abhängigen Variablen so nah wie möglich an den entsprechenden vorhergesagten Werten liegen. Dem Kriterium zufolge wird also die Variable herangezogen, deren Aufsplittung in zwei Untermengen zur größten Reduktion der gemeinsamen Summe der quadrierten Residuen der für beide Untermengen relevanten Beobachtungen des Trainingsdatensatzes führt. Das Residuum ergibt sich dabei aus der Differenz der jeweiligen abhängigen Variable und dem Mittelwert der abhängigen Variablen der Beobachtungen, welche am aufzusplittenden Knoten relevant sind. Nach Aufspannen des Entscheidungsbaums werden die Blattknoten, die die pfadbezogenen Vorhersagen der abhängigen Variablen angeben, durch Mittelwertbildung der abhängigen Variablen der einzelnen Beobachtungen, die zu einem Pfad gehören, belegt. (vgl. Sutton (2005), S. 315f.)
Abbildung 5.1 zeigt beispielhaft den schematischen Aufbau eines CART. Dabei wird abhängig davon, ob es sich um ein Klassifikations- oder Regressionsproblem handelt, am Wurzelknoten sowie an den internen Knoten eines der oben beschriebenen Aufsplittungskriterien für die Variablen x_i angewandt, um den Entscheidungsbaum aufzuspannen. Weiterhin lässt sich - wie eingangs erwähnt - erkennen, dass an den Knoten binäre Entscheidungen getroffen werden. Falls ein aufzusplittendes x_i kategorial ist, wird dieses ebenfalls durch eine Binärcodierung aufgesplittet. Die Blattknoten enthalten die für den jeweiligen Pfad vohergesagten Werte der abhängigen Variable $\hat{y}$. Diese werden - wie oben beschrieben - bei der Klassifikation über einen Mehrheitsentscheid und bei der Regression über Mittelwertbildung gebildet.
Aufgrund des Untersuchungsgegenstandes der Arbeit beziehen sich die folgenden Erläuterungen auf die Anwendung bei Regressionsfragestellungen.

[2] Gehören zu einem Pfad bspw. drei Beobachtungen, wobei eine Beobachtung der Kategorie A und die beiden anderen Beobachtungen der Kategorie B angehören, so wird dieser Pfad Kategorie B vorhersagen.

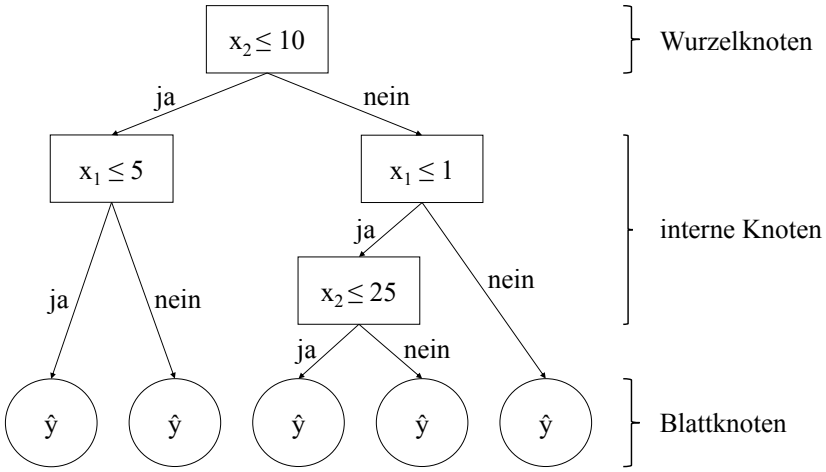


Abbildung 5.1: Schematischer Aufbau eines CART [Quelle: eigene Darstellung in Anlehnung an Sutton (2005), S. 307]

5.2.2 Random Forest-Verfahren

Das RFV basiert auf dem sogenannten Bagging-Algorithmus, welcher auch Bootstrap-Aggregation genannt wird. Dieser kombiniert die Vorhersagen von mehreren auf dem selben Datensatz beruhenden Entscheidungsbäumen, wodurch die Varianz der Vorhersagen reduziert wird. Das wiederum führt zu einer Erhöhung der Genauigkeit der Vorhersagen.
Um das zu erzielen, werden im Bagging in mehreren Iterationen aufeinanderfolgende unabhängige Entscheidungsbäume unter Einbeziehung aller zur Verfügung stehenden unabhängigen Variablen x_i ($i = 1...p$) konstruiert, wobei für jeden Entscheidungsbaum eine zufällige Teilmenge des Datensatzes, welche auch Bootstrap-Stichprobe genannt wird, einbezogen wird. Die Bildung eines einzelnen Entscheidungsbaums erfolgt schließlich nach dem in Kapitel 5.2.1 im Abschnitt *CART-Algorithmus* beschriebenen Vorgehen im Falle der Regression. Das besondere an der Bootstrap-Stichprobe ist einerseits, dass diese den selben Umfang wie der ursprüngliche Datensatz aufweist und andererseits, dass die Beobachtungen k ($k = 1...n$) mehrmals auftauchen können, wodurch die Anzahl der einbezogenen k zwischen den einzelnen Bootstrap-Stichproben unterschiedlich sein kann. Soll nun eine Vorhersage für ein k generiert werden, so durchläuft dieses k zunächst jeden Entscheidungsbaum und bekommt dementsprechend mehrere Vorhersagen $\hat{y}_k$ zugewiesen. Die letztendliche Vorhersage eines bestimmten k erfolgt dann über Mittelwertbildung der einzelnen $\hat{y}_k$. (vgl. Sutton (2005), S. 303ff.)
Durch die Bildung von Bootstrap-Stichproben werden die k nicht in jeder Iteration zur Erstellung des Entscheidungsbaums einbezogen, was diese k wiederum zu Out-of-bag (OOB)-Beobachtungen macht. Tendenziell gehört dabei jedes k in ca. einem Drittel der Iterationen nicht der Bootstrap-Stichprobe an. Anders ausgedrückt bedeutet dies,

dass in jeder Iteration etwa zwei Drittel der k zur Erstellung des Entscheidungsbaums einbezogen werden. Durch dieses Vorgehen separiert der Bagging-Algorithmus automatisch in Trainings- und Testdatensatz. Dabei umfasst der Trainingsdatensatz die jeweiligen k der Bootstrap-Stichprobe und der Testdatensatz die zugehörigen OOB-Beobachtungen k. (vgl. Breiman (2001), S. 11)
Das RFV erweitert den Bagging-Algorithmus um eine zufällige Komponente, wodurch es zur Blackbox wird. Das bedeutet, dass von außen letztendlich nicht mehr nachvollzogen werden kann, wie das Verfahren arbeitet und zum Ergebnis gelangt. Anstatt aus allen p zur Verfügung stehenden x_i das gemäß dem Aufsplittungskriterium für Regressionsfragestellungen beste x_i zur Aufsplittung eines Knotens heranzuziehen, nutzt das RFV hingegen das beste aus einer festgelegten Anzahl an zufällig gewählten x_i. Somit kann der Bagging-Algorithmus als Sonderfall des RFV betrachtet werden, bei dem die Anzahl der Variablen, aus der das beste x_i zur Aufsplittung gewählt wird, gleich p entspricht.
Alles in allem ist das RFV nutzerfreundlich, da nur zwei Parameter festgelegt werden müssen. Diese sind die Anzahl der Variablen für die zufällige Auswahl an jedem Knoten sowie die Anzahl der Entscheidungsbäume. (vgl. Liaw und Wiener (2002), S. 18)

5.2.3 Modellgütekriterien

Die Modellgüte des RFV kann anhand von Kreuzvalidierung[3] - unter Nutzung der zufällig generierten Trainings- (Bootstrap-Stichprobe) und Testdatensätze (OOB-Beobachtungen) - untersucht werden. Als Kriterien für die Modellgüte bei Regressionsfragestellungen können dabei der Mean Squared Error (MSE) sowie der Root Mean Squared Error (RMSE), welche beide auf Residuen basieren, herangezogen werden. Die Ermittlung beider Kenngrößen wird im Folgenden näher erläutert.
In jeder Bootstrap-Iteration durchlaufen die jeweiligen OOB-Beobachtungen k den zugehörigen Entscheidungsbaum und bekommen eine Vorhersage $\hat{y}_{k,OOB}$ zugewiesen. Da jedes k mehrmals bei der Erstellung von Entscheidungsbäumen OOB ist, erlangt es Vorhersagen von verschiedenen Entscheidungsbäumen, deren Mittelwert die eigentliche Vorhersage $\bar{\hat{y}}_{k,OOB}$ darstellt. Der MSE ergibt sich schließlich aus dem Mittelwert der Summe der quadrierten Residuen aller OOB-Beobachtungen k, wobei die Residuen jeweils aus der Differenz zwischen mittlerem vorhergesagtem $\bar{\hat{y}}_{k,OOB}$ und beobachtetem y_k gebildet werden. Gleichung 5.2.1 stellt diesen Zusammenhang dar.

$$MSE = \frac{1}{n}\sum_{k=1}^{n}(\bar{\hat{y}}_{k,OOB} - y_k)^2 \tag{5.2.1}$$

[3] Bei der Kreuzvalidierung wird der Datensatz mehrmals in einen Trainings- und einen Testdatensatz aufgeteilt, wobei sich die Untermengen jeweils gegenseitig ausschließen und jede Beobachtung einmal dem Testdatensatz angehört. Der Trainingsdatensatz dient dann der Erstellung des Modells, während der Testdatensatz auf das Modell angewendet wird und Vorhersagen generiert. Diese werden schließlich mit den tatsächlichen Werten verglichen.

Da der MSE als Einheit eine quadrierte Größe angibt, wird aus Gründen der Interpretationsvereinfachung häufig die Wurzel des MSE als Maß der Modellgüte angewandt. Demnach ergibt sich der RMSE zu Ausdruck 5.2.2.

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\bar{\hat{y}}_{k,OOB} - y_k)^2} \tag{5.2.2}$$

Je geringer der Wert des entsprechenden Maßes ist, desto näher liegen die mittleren vorhergesagten $\bar{\hat{y}}_{k,OOB}$ an den beobachteten y_k, d.h. desto besser bilden die trainierten Entscheidungsbäume die Testdaten ab. (vgl. Breiman (2001), S. 25f. sowie Liaw und Wiener (2002), S. 20)

5.2.4 Variablenbedeutsamkeit

Das folgende Kapitel beschreibt, wie die Bedeutsamkeit und der Einfluss einzelner x_i im RFV untersucht werden können.

Variable Importance

Die Bedeutsamkeit einzelner Variablen im RFV wird in der Literatur auch als Variable Importance (VI) bezeichnet. Die im Folgenden erläuterte Ermittlung der VI für CART's bei Regressionsfragestellungen basiert dabei auf dem zuvor vorgestellten MSE.
Zunächst wird für jeden Entscheidungsbaum b der MSE_b berechnet. Dabei repräsentiert $\hat{y}^b_{k,OOB}$ die Vorhersage der OOB-Beobachtung k des entsprechenden Entscheidungsbaums b und n^b_{OOB} die Anzahl der OOB-Beobachtungen k des selbigen. Gleichung 5.2.3 stellt diesen Zusammenhang dar.

$$MSE_b = \frac{1}{n^b_{OOB}}\sum_{k=1}^{n^b_{OOB}}(\hat{y}^b_{k,OOB} - y_k)^2 \tag{5.2.3}$$

Anschließend werden die Werte des entsprechenden x_i zufällig vertauscht, sodass die ursprüngliche Beziehung zwischen unabhängiger und abhängiger Variable aufgehoben wird. Nachdem die OOB-Beobachtungen k den Entscheidungsbaum b mit den vertauschten Werten von x_i durchlaufen haben, erfolgt eine erneute Berechnung des MSE_b. Die Differenz zwischen diesem neu bestimmten und dem ursprünglichem MSE_b ergibt schließlich die Reduktion des MSE_b durch das Vorhandensein von x_i am entsprechenden Entscheidungsbaum b. Die VI eines x_i ergibt sich letztendlich aus dem Mittelwert der MSE_b-Reduktionen über alle Entscheidungsbäume b durch das Vorhandensein von x_i. Der Ablauf wird für alle x_i wiederholt, sodass diese am Ende nach absteigender VI geordnet werden können. Das x_i, welches dabei die größte mittlere Reduktion des MSE_b aufweist, ist die bedeutsamste Variable und hat damit den größten Einfluss auf die mittleren vorhergesagten Werte $\bar{\hat{y}}_{k,OOB}$. Im Gegensatz dazu weisen x_i, deren Wertevertauschen lediglich zu einer geringen bis gar keinen Verbesserung des MSE_b führt,

nur einen geringen bis gar keinen Einfluss auf $\bar{\hat{y}}_{k,OOB}$ auf. (vgl. Breiman (2002) sowie Grömping (2009), S. 311f.)

Partial Dependence Plots

Eine Möglichkeit, um für Verfahren des maschinellen Lernens den Einfluss einzelner x_i auf die Vorhersage der abhängigen Variable grafisch darzustellen, sind Partial Dependence Plots (PDP's). Diese zeigen, wie sich die über alle Beobachtungen gemittelte marginale Vorhersage $\hat{y}$ ändert, wenn eine bestimmtes x_i all ihre möglichen Ausprägungen annimmt. Des Weiteren lässt sich erkennen, in welcher funktionalen Beziehung x_i und $\hat{y}$ zueinander stehen. Diese Eigenschaft kann sehr von Vorteil sein, da Verfahren wie der CART-Algorithmus unabhängig von Annahmen, wie bspw. der Annahme der linearen Beziehung zwischen der unabhängigen und abhängigen Variable in der linearen Regression, arbeiten. Im Detail erfolgt die Erstellung eines PDP wie folgt.
Für jede mögliche Ausprägung, die das betrachtete x_i annehmen kann, wird $\hat{y}$ unter der Prämisse, dass die Ausprägungen der anderen Variablen jeweils beibehalten werden, gebildet. Dafür werden die Ausprägungen des interessierenden x_i aller k nacheinander durch die jeweils betrachtete Ausprägung ersetzt. Hat x_i bspw. die Ausprägungen A und B, so nimmt x_i im ersten Schritt für alle k den Wert A an, woraufhin die Berechnung von $\hat{y}$ erfolgt. Im zweiten Schritt wird das Prozedere dann für den Wert B wiederholt. Der PDP stellt letztendlich die Ausprägungen von x_i auf der Abszissenachse und die zugehörigen $\hat{y}$ auf der Ordinatenachse gegenüber.
Ein Nachteil dieser Darstellungsform ist jedoch, dass das betrachtete x_i nicht mit anderen Variablen korreliert sein darf. Eine zu hohe Korrelation würde demnach zur Entstehung von unwahrscheinlichen oder gar unmöglichen $\hat{y}$ führen. Der Grund dafür liegt in der Mittelung der Vorhersagen über Beobachtungen, deren „erzwungene“ Ausprägung von x_i in der Realität nicht in Kombination mit Ausprägungen anderer Variablen auftauchen. Soll bspw. der Einfluss der *Größe* untersucht werden, wobei es ebenfalls die Variable *Gewicht* gibt, so liegt im Normalfall eine Korrelation zwischen diesen Variablen vor. Eine „erzwungene“ Ausprägung der Variable *Größe* von bspw. zwei Metern ist in der Realität in Kombination mit einem *Gewicht* von bspw. 40 Kilogramm unwahrscheinlich. Das würde letztendlich dazu führen, dass eine Verzerrung von $\hat{y}$ durch die korrelierte Variable *Größe* auftritt. Somit kann der Effekt eines korrelierten x_i nicht vom Einfluss anderer Variablen isoliert dargestellt werden. (vgl. Apley und Zhu (2016), S. 2ff. und Molnar (2019)) Aufgrund dieses Nachteils wird im Folgenden eine alternative Darstellungsmethode vorgestellt.

Accumulated Local Effects

Das Ziel der Accumulated Local Effects (ALE) liegt ebenfalls in der Ermittlung, wie einzelne x_i die Vorhersage eines Modells des maschinellen Lernens beeinflussen. Im Gegensatz zu den PDP's isolieren sie partiell die Effekte der anderen Variablen durch

Nutzung von Vorhersage-Differenzen, was zu einer Robustheit gegenüber korrelierten Variablen führt.

Um die ALE zu ermitteln, werden die Ausprägungen des interessierenden x_i zunächst in Intervalle unterteilt, wobei jedes Intervall gleich viele k enthalten sollte.[4] Im nächsten Schritt wird für jedes k die Differenz zweier Vorhersagen $\hat{y}$ gebildet, wobei die zum x_i des jeweiligen k gehörenden Intervallgrenzen herangezogen werden. Die Bildung des ersten $\hat{y}$ erfolgt dabei unter Nutzung der oberen Intervallgrenze anstelle der eigentlichen Ausprägung von x_i. Von diesem wird das zweite $\hat{y}$ subtrahiert, welches analog unter Nutzung der unteren Intervallgrenze ermittelt wird.[5] Anzumerken ist dabei, dass für die Ermittlung der $\hat{y}$ lediglich die Ausprägung von x_i substituiert wird, während die Ausprägungen der übrigen Variablen jeweils beibehalten werden. Die lokalen Effekte der einzelnen Intervalle auf $\hat{y}$ ergeben sich dann durch Mittelung der zum jeweiligen Intervall gehörenden Differenzen. Abbildung 5.2 veranschaulicht diese Vorgehensweise am Beispiel von zwei unabhängigen Variablen x_1 und x_2. Dabei kennzeichnet ein Punkt ein k und $z_0, z_1, .., z_6$ die Intervallgrenzen von x_1. Die Mittelwerte werden anschließend in der vorliegenden Intervallreihenfolge akkumuliert, sodass sich der ALE einer bestimmten Ausprägung von x_i letztendlich durch das Aufsummieren der Effekte vom ersten bis zum entsprechenden Intervall der betrachteten Ausprägung von x_i ergibt. Anschließend erfolgt eine Zentrierung der sich so für jedes Intervall ergebenden ALE, sodass sich der mittlere ALE - also der mittlere Effekt von x_i auf $\hat{y}$ - zu null ergibt. Durch die Nutzung von Differenzen werden - wie eingangs erwähnt - die Effekte der anderen Variablen auf $\hat{y}$ ausgeblendet, sodass sich letztendlich der alleinige Effekt von x_i auf $\hat{y}$ ergibt.

Grafisch stellen ALE die Ausprägungen von x_i auf der Abszissenachse und die zugehörigen relativen Effekte auf die Vorhersage $\hat{y}$ auf der Ordinatenachse gegenüber. Der relative Effekt einer Ausprägung von x_i kann dabei ebenfalls als Abweichung zum mittleren Effekt von x_i auf $\hat{y}$ interpretiert werden. (vgl. Apley und Zhu (2016), S. 5ff. und Molnar (2019))

[4] Das kann bspw. durch die Nutzung von Quantilen erzielt werden.

[5] Da die Reihenfolge der Ausprägungen von x_i die Intervallgrenzen beeinflusst, ist es für kategoriale Variablen sinnvoll, die Ausprägungen zunächst in eine Reihenfolge zu bringen. Ergibt sich dabei keine natürliche Reihung, bietet sich die Ordnung der Ausprägungen anhand von Ähnlichkeiten der übrigen x_i an.

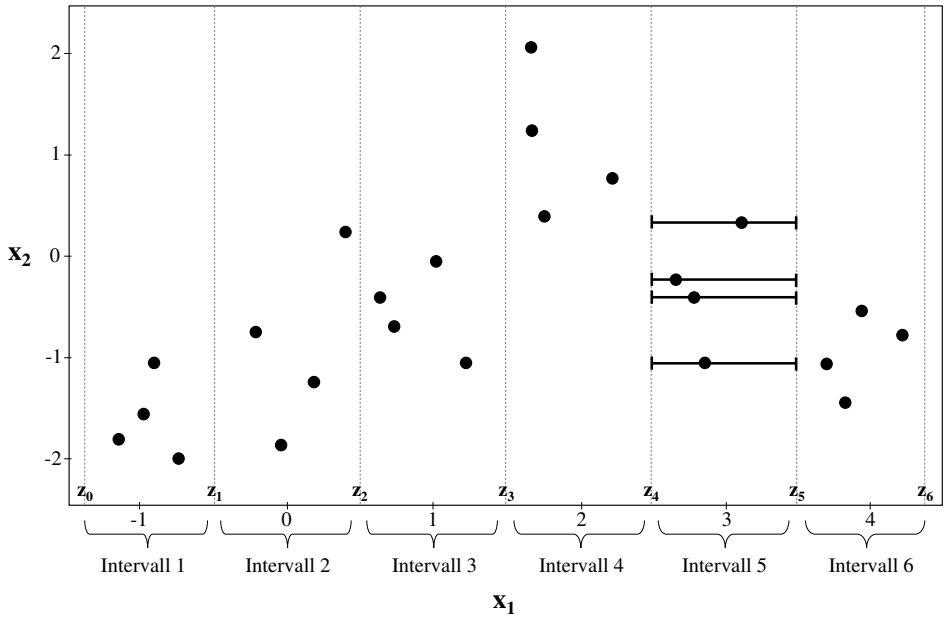


Abbildung 5.2: Schematische Intervalleinteilung der Accumulated Local Effects: Die Vorhersage-Differenz einer Beobachtung im x_1-Intervall 5 ergibt sich zu $\hat{y}(x_1 = z_5, x_2) - \hat{y}(x_1 = z_4, x_2)$. Dabei bleibt x_2 bei der jeweiligen Differenz konstant. Der lokale Effekt des Intervalls 5 ergibt sich letztendlich aus dem Mittelwert der Vorhersage-Differenzen der Beobachtungen dieses Intervalls. [Quelle: eigene Darstellung in Anlehnung an Apley und Zhu (2016), S. 9]

5.3 Negative Binomialregression

Um die Einflussfaktoren des Auftretens von Unfällen an Knotenpunkten zu ermitteln, kommen in der Literatur sogenannte *Accident Prediction Models* zur Anwendung. Diese bedienen sich der Negativen Binomialregression, welche zu den multivariaten Analyseverfahren zählt. Der Negativen Binomialregression liegt dabei ein GLM zugrunde, bei dem die abhängige Variable einer negativen Binomialverteilung folgt. Die negative Binomialverteilung wiederum ergibt sich aus einer Überlagerung von Poisson- und Gammaverteilung. (vgl. Dunn und Smyth (2018), S. 399f.) Aus diesem Grund stellt das folgende Kapitel im ersten Schritt Merkmale eines GLM im Allgemeinen sowie der Poisson- und Gammaverteilung und deren Überlagerung vor. Der nächste Teil zeigt anschließend auf, wie das zu schätzende *Accident Prediction Model* aufgebaut und zu verstehen ist. Um eine angemessene Schätzung durchführen zu können, stellt das Kapitel zudem das LASSO-Verfahren vor, anhand dessen im Vorfeld der Schätzung eine Variablenauswahl getroffen wird. Neben Modellgütekriterien und einem Maß zur Ermittlung der Bedeutsamkeit einzelner Variablen liegt der Fokus zudem auf Kriterien der Modelldiagnostik linearer Modelle.

5.3.1 Grundlagen

Die folgenden Abschnitte vermitteln ein grundlegendes Verständnis für GLM's sowie die Poisson- und Gammaverteilung und deren Überlagerung.

Verallgemeinerte Lineare Modelle (GLM)

Jedes Regressionsmodell besteht aus einer zufälligen und einer systematischen Komponente. Dabei beschreibt die zufällige Komponente die Verteilung der abhängigen Variable y_k. Die systematische Komponente hingegen kennzeichnet die mathematische Beziehung zwischen dem Erwartungswert der abhängigen Variable $E(y_k)$ und den Ausprägungen der unabhängigen Variablen x_{ik}. (vgl. Dunn und Smyth (2018), S. 11)
Beim klassischen linearen Modell unterliegt die zufällige Komponente der Annahme, dass die y_k stochastisch unabhängige, normalverteilte Größen mit dem Erwartungswert $E(y_k) = \mu_k$ und einer konstanten Varianz $Var(y_k) = \sigma^2$ sind. Formal stellt sich das durch den Ausdruck $y_k \sim N(\mu_k, \sigma^2)$ dar.
Die systematische Komponente unterstellt im linearen Modell, dass μ_k in einer direkten linearen Beziehung zu x_{ik} steht. Die x_{ik} wiederum fließen in der funktionalen Form ein, dass sich ein linearer Zusammenhang mit μ_k ergibt. Die sich so ergebende systematische Komponente nimmt die Form in Gleichung 5.3.1 an. Dabei bezeichnet β_0 das Absolutglied und β_i den Koeffizienten von i. (vgl. Dunn und Smyth (2018), S. 31f.)

$$\mu_k = \beta_0 + \sum_{i=1}^{p} \beta_i x_{ik} \tag{5.3.1}$$

Da die Annahme eines normalverteilten y_k - insbesondere einer konstanten Varianz - nicht für alle Daten herangezogen werden kann, kommen die sogenannten GLM's zur Anwendung, welche als parameterlineare Regressionsmodelle durch flexible Verallgemeinerungen der zufälligen und systematischen Komponente Abhilfe schaffen.[6]
Die zufällige Komponente der GLM's unterstellt dabei, dass die stochastisch unabhängigen y_k einer Verteilungsfamilie - nämlich der Exponentialfamilie[7] - angehören. Dabei wird die Varianz von y_k in Abhängigkeit von μ_k modelliert.
Die Verallgemeinerung der systematischen Komponente liegt darin, dass eine Linkfunktion $g(\mu_k)$ mit einem linearen Prädiktor η verbunden wird, sodass der Ausdruck

$$g(\mu_k) = \eta \tag{5.3.2}$$

gilt. Der lineare Prädiktor nimmt dabei folgende Form an.

$$\eta = \beta_0 + \sum_{i=1}^{p} \beta_i x_{ik} \tag{5.3.3}$$

Demnach ergibt sich die allgemeine Beziehung zwischen $g(\mu_k)$ und η - also die systematische Komponente eines GLM der Exponentialfamilie - gemäß Gleichung 5.3.2 und 5.3.3 zu folgendem Ausdruck.

$$g(\mu_k) = \beta_0 + \sum_{i=1}^{p} \beta_i x_{ik} \tag{5.3.4}$$

Dabei werden die x_{ik} in der funktionalen Form einbezogen, dass sie mit $g(\mu_k)$ in linearem Zusammenhang stehen. (vgl. Dunn und Smyth (2018), S. 211ff.)

Poisson-, Gamma- und negative Binomialverteilung

Die Poissonverteilung kommt für GLM's der Exponentialfamilie zum Einsatz, wenn y_k diskrete, nicht-negative Werte aufweist. Dabei verläuft die zugehörige Verteilungsfunktion rechtsschief. Die Poissonverteilung enthält nur einen Parameter μ_k, wodurch sich die zugehörige zufällige Komponente formal zu $y_k \sim Po(\mu_k)$ ergibt.
Als Linkfunktion eines GLM's mit Poissonverteilung kommt üblicherweise eine logarithmische Linkfunktion zum Einsatz, welche dafür sorgt, dass μ_k nicht-negativ wird. Somit entspricht $g(\mu_k)$ dem Ausdruck $ln(\mu_k)$, wodurch sich die systematische Komponente gemäß Gleichung 5.3.4 mit dem Ausdruck 5.3.5 beschreiben lässt.

$$ln(\mu_k) = \beta_0 + \sum_{i=1}^{p} \beta_i x_{ik} \tag{5.3.5}$$

[6] Die Annahme eines normalverteilten y_k kann bspw. nicht angewandt werden, wenn y_k einen relativen Anteil, Zähldaten oder stetige, nicht-negative Werte annimmt. (vgl. Dunn und Smyth (2018), S. 166)
[7] Zur Exponentialfamilie gehören bspw. die stetigen Verteilungen Normal- und Gammaverteilung sowie die diskreten Verteilungen Poisson- und negative Binomialverteilung. (vgl. Dunn und Smyth (2018), S. 212)

Nach Umformung ergibt sich diese letztendlich zu folgendem Ausdruck.

$$\mu_k = exp(\beta_0 + \sum_{i=1}^{p} \beta_i x_{ik}) \tag{5.3.6}$$

Der Erwartungswert sowie die Varianz ergeben sich für ein GLM, welchem eine Poissonverteilung zugrunde liegt, zu $E(y_k) = \mu_k$ bzw. $Var(y_k) = \mu_k$. (vgl. Dunn und Smyth (2018), S. 371f.)
Die Gammaverteilung hingegen kommt für GLM's der Exponentialfamilie zum Einsatz, wenn y_k stetige, nicht-negative Werte aufweist. Diese ist durch die zwei Parameter μ_k und θ gekennzeichnet, wobei θ einen Dispersionsparameter darstellt. Somit ergibt sich die zugehörige zufällige Komponente zu $y_k \sim \Gamma(\mu_k, \theta)$.
Auch in diesem Fall kommt eine logarithmische Linkfunktion zum Einsatz, wodurch die entsprechende systematische Komponente ebenfalls durch den Ausdruck in Gleichung 5.3.5 bzw. 5.3.6 beschrieben wird.
Der Erwartungswert sowie die Varianz ergeben sich für ein GLM mit zugrundeliegender Gammaverteilung zu $E(y_k) = \mu_k$ und $Var(y_k) = \frac{\mu_k^2}{\theta}$. (vgl. Dunn und Smyth (2018), S. 400ff.)
Weist die Verteilung von y_k hingegen eine Überdispersion auf, so kommt die negative Binomialverteilung als zufällige Komponente eines GLM aus der Exponentialfamilie zum Einsatz. Überdispersion bedeutet dabei, dass die Varianz den Erwartungswert übersteigt ($Var(y_k) >> \mu_k$). Das tritt auf, wenn nicht nur y_k selbst, sondern auch μ_k eine Variabilität besitzt.(vgl. Dunn und Smyth (2018), S. 397)
Maher und Summersgill (1996, S. 288) nennen als möglichen Grund für eine Überdispersion den Einfluss nicht beobachteter unabhängiger Variablen auf y_k. Wird für Daten mit Überdispersion ein Modell, das die Überdispersion nicht abbildet, geschätzt, so ist die Folge eine Unterschätzung von Standardfehlern, was wiederum dazu führt, dass Signifikanztests der unabhängigen Variablen einen signifikanteren Einfluss der entsprechenden Variable anzeigen, obwohl die Signifikanz in Wahrheit geringer ausfällt. (vgl. Dunn und Smyth (2018), S. 397)
Die negative Binomialverteilung ergibt sich letztendlich aus der Überlagerung der zuvor vorgestellten Poisson- und Gammaverteilung. Dabei folgt y_k einer Poissonverteilung, während μ_k selbst einer Gammaverteilung folgt. Demnach gilt für die zufällige Komponente des entsprechenden GLM $y_k \sim Po(\mu_k)$ sowie $\mu_k \sim \Gamma(\mu_k, \theta)$.
Die systematische Komponente kann abermals durch den Ausdruck in Gleichung 5.3.5 bzw. 5.3.6 beschrieben werden.
Der Erwartungswert und die Varianz dieser Überlagerung ergeben sich zu $E(y_k) = \mu_k$ und $Var(y_k) = \mu_k + \frac{\mu_k^2}{\theta}$. Um die Überdispersion abzubilden und die Varianz zu korrigieren, enthält $Var(y_k)$ im Term auf der rechten Seite den Parameter θ der Gammaverteilung. Je größer θ ist, desto weiter konvergiert die Varianz gegen den Erwartungswert μ_k. Sehr große Werte von θ bedeuten dementsprechend, dass die negative Binomialverteilung

lung (annähernd) einer Poissonverteilung entspricht.
Mit Hilfe der negativen Binomialverteilung können damit bspw. Zähldaten (Poissonverteilung), deren Erwartungswert selbst von anderen unberücksichtigten Faktoren beeinflusst wird (Gammaverteilung), in Form einer Negativen Binomialregression untersucht werden. (vgl. Dunn und Smyth (2018), S. 399f.)

5.3.2 Accident Prediction Models

Um die erwartete Anzahl von Unfällen $E(y_k)$ in Abhängigkeit von Einflussfaktoren zu ermitteln, kommt häufig die Negative Binomialregression zum Einsatz. Die zugehörigen Modelle - in Form von GLM's mit zugrunde gelegter negativer Binomialverteilung - werden in der Literatur auch als *Accident Prediction Models* bezeichnet. (vgl. bspw. Schepers et al. (2011, S. 855) oder Eenink et al. (2008, S. 7)) Diese Modelle unterteilen die unabhängigen Variablen in Expositionsvariablen s_{pk} sowie Risikovariablen t_{qk}. Dabei bilden die s_{pk} die wichtigsten Determinanten des Unfallgeschehens ab, was zur Folge hat, dass dessen Nichtauftreten zu einer Vorhersage von null Unfällen führt. Denkbare Expositionsgrößen wären dabei Verkehrsstärken oder die Länge des betrachteten Streckenabschnitts. Die t_{qk} hingegen können die Anzahl der Unfälle nur beeinflussen, wenn deren Ausprägungen ungleich null sind. Sind diese hingegen null, so wird das Modell aufgrund der Expositionsgrößen trotzdem Unfälle vorhersagen. Solche das Unfallgeschehen zusätzlich beeinflussenden t_{qk} sind bspw. Merkmale der Infrastruktur.
Gleichung 5.3.7 stellt das ökonometrische Grundmodell eines *Accident Prediction Models* dar. (vgl. Eenink et al. (2008), S. 7)

$$E(y_k) = \beta_0 \cdot \prod_{p=1}^{a} s_{pk}^{\gamma_p} \cdot exp(\sum_{q=1}^{b} \delta_q t_{qk} + \epsilon_k) \tag{5.3.7}$$

Dabei kommen die folgenden Bestandteile zum Einsatz.

$E(y_k)$	erwartete Anzahl Unfälle bei Beobachtung k, k=1...n
β_0	Absolutglied
γ_p	Koeffizient der Expositionsvariable p, p=1...a
s_{pk}	Ausprägung der p.ten Expositionsvariable bei Beobachtung k, p=1...a und k=1...n
δ_q	Koeffizient der Risikovariable q, q=1...b
t_{qk}	Ausprägung der q.ten Risikovariable bei Beobachtung k, q=1...b und k=1...n
ϵ_k	Residuum der Beobachtung k, k=1...n

Der Einfluss von s_{pk} und t_{qk} wird in den *Accident Prediction Models* gemäß Gleichung 5.3.7 multiplikativ verknüpft. Dabei fließen die s_{pk} als Potenzfunktion mit der Basis s_{pk} und dem Exponenten γ_p ein, damit - wie eben beschrieben - das Nichtauftreten eine Unfallzahl von null zur Folge hat. Der Exponent γ_p beschreibt dabei die Form des

Zusammenhangs zwischen dem entsprechenden s_{pk} und $E(y_k)$. Die t_{qk} hingegen werden als Exponentialfunktion in das Modell aufgenommen. Damit wird sichergestellt, dass das Nichtauftreten keinen Einfluss auf $E(y_k)$ hat ($exp(0) = 1$). Der Vorfaktor δ_q eines jeden t_{qk} beschreibt zudem den Einfluss des Auftretens des entsprechenden t_{qk} auf $E(y_k)$, wobei der Einfluss nicht direkt, sondern als Exponent der e-Funktion einfließt. (vgl. Schepers et al. (2011), S. 855)
Durch Nutzung einer logarithmischen Linkfunktion ergibt sich die Grundform der systematischen Komponente der negativen Binomialverteilung gemäß Gleichung 5.3.5, wobei die unabhängigen Variablen in Expositions- und Risikovariablen unterteilt werden. Gleichung 5.3.8 stellt das sich so ergebende zu schätzende Modell dar. Dabei entspricht $ln(\beta_0)$ dem zu schätzenden Absolutglied. Weitere zu schätzende Größen sind außerdem $ln(E(y_k))$, γ_p sowie δ_q.

$$ln(E(y_k)) = ln(\beta_0) + \sum_{p=1}^{a} \gamma_p \, ln(s_{pk}) + \sum_{q=1}^{b} \delta_q t_{qk} + \epsilon_k \tag{5.3.8}$$

Sowohl die Schätzung der eben genannten Größen als auch die des Dispersionsparameters θ der zuvor beschriebenen Gammaverteilung erfolgt unter Verwendung der ML-Methode[8]. Die daraus resultierende Schätzfunktion für die erwartete Anzahl der Unfälle in Gleichung 5.3.9 ergibt sich letztendlich aus der Rücktransformation der zuvor genannten Gleichung 5.3.8.

$$\widehat{E(y_k)} = \hat{\beta}_0 \cdot \prod_{p=1}^{a} s_{pk}^{\hat{\gamma}_p} \cdot exp(\sum_{q=1}^{b} \hat{\delta}_q t_{qk}) \tag{5.3.9}$$

Anzumerken ist dabei, dass sich der Schätzer für die erwartete Anzahl an Unfällen $\widehat{E(y_k)}$ aufgrund der Rücktransformation aus der Berechnungsvorschrift $exp(ln(E(y_k)))$ ergibt. Weiterhin lässt sich der Schätzer des Absolutglieds $\hat{\beta}_0$ analog mit $exp(ln(\hat{\beta}_0))$ bestimmen. Die Schätzer $\hat{\gamma}_p$ sowie $\hat{\delta}_q$ hingegen entsprechen direkt γ_p und δ_q des geschätzten Modells 5.3.8 und bedürfen daher keiner weiteren Berechnungsvorschrift.
Wie zuvor beschrieben, kennzeichnet $\hat{\gamma}_p$ den Zusammenhang des entsprechenden s_{pk} mit $\widehat{E(y_k)}$.[9] Weiterhin beschreibt $\hat{\delta}_q$ die Veränderung von $\widehat{E(y_k)}$, wenn sich ein metrisch skaliertes t_{qk} um eine Einheit erhöht. Im Gegensatz dazu bezeichnet $\hat{\delta}_q$ bei kategorialen t_{qk} die Veränderung von $\widehat{E(y_k)}$, wenn anstelle der Referenzausprägung die entsprechende Ausprägung von t_{qk} auftritt. Der letztendliche Effekt wird in beiden Fällen durch $exp(\hat{\delta}_q)$ beschrieben.

[8] Die ML-Methode versucht die Koeffizienten so zu schätzen, dass die Wahrscheinlichkeit des Zutreffens der beobachteten abhängigen Variablen maximiert wird. (vgl. Backhaus et al. (2008), S. 252)
[9] Dieser kann dabei von negativer ($\hat{\gamma}_p<0$), positiv degressiver ($0<\hat{\gamma}_p<1$), positiv linearer ($\hat{\gamma}_p=1$) oder positiv progressiver Natur ($\hat{\gamma}_p>1$) sein.

5.3.3 Variablenselektion

Aufgrund des nicht-zufälligen Vorgehens der Negativen Binomialregression und der vielen zur Verfügung stehenden unabhängigen Variablen erfolgt vor der eigentlichen Schätzung zunächst eine Variablenselektion, was wiederum die Vorhersagegenauigkeit erhöhen soll. Der Grund dafür liegt im typischen Trade-Off des statistischen Lernens, welcher zwischen Bias und Varianz besteht. Dabei kennzeichnet der Bias die Verzerrung der geschätzten Koeffizienten. Je größer der Bias ist, desto verzerrter sind die geschätzten Koeffizienten und desto schlechter sagt das Modell die Trainingsdaten vorher. Die Varianz hingegen beschreibt die Variabilität der Modellvorhersagen bei verschiedenen Testdaten. Je höher die Varianz im Modell ist, desto sensibler reagieren die Vorhersagen auf Abweichungen in den Testdaten und desto inkonsistenter sind diese bei verschiedenen Testdaten. Demnach sollte ein optimales Modell einen geringen Bias und eine niedrige Varianz aufweisen, was jedoch schwer zu erzielen ist, da eine Senkung des Bias eine Erhöhung der Varianz bedingt und umgekehrt.
Die Einbeziehung aller zur Verfügung stehenden Variablen würde möglicherweise zu einem zu komplizierten Modell führen, was als Overfitting bezeichnet wird. Overfitting kann dabei als Extremfall des Trade-Offs zwischen Bias und Varianz betrachtet werden, da dieses durch einen (sehr) geringen Bias sowie eine (sehr) hohe Varianz gekennzeichnet ist.[10] Neben ungenauen Vorhersagen weisen Modelle mit Overfitting Interpretationsschwierigkeiten der geschätzten Koeffizienten auf, da diese ineffizient sind. Die Ineffizienz kommt dabei daher, dass der Einfluss einer unabhängigen Variable auf die abhängige Variable mit zunehmender Anzahl an unabhängigen Variablen nicht mehr hinreichend präzise ermittelt werden kann. Das wiederum kann dazu führen, dass ein tatsächlicher Einflussfaktor als nicht signifikant erscheint. Umgekehrt ist es genauso möglich, dass eigentlich irrelevante unabhängige Variablen mit zunehmender Anzahl an unabhängigen Variablen als signifikante Einflussgrößen eingestuft werden. Daher empfiehlt es sich, lediglich eine Auswahl an erklärenden Variablen einzubeziehen, die wiederum jeweils starke Effekte aufweisen. (vgl. Backhaus et al. (2008), S. 84) Um diesem Problem entgegenzuwirken, kann das von Tibshirani (1996) eingeführte LASSO-Verfahren - auch bekannt als L1-Regularisierung - angewandt werden.
Das LASSO-Verfahren ist eine Regularisierungsmethode, die simultan zur Koeffizientenschätzung eine Variablenauswahl durchführt. Um die Vorhersagegenauigkeit zu verbessern, wird dabei ein kleiner Bias - also eine geringe Verzerrung der geschätzten Koeffizienten - zugunsten einer starken Verringerung der Varianz der vorhergesagten Werte durch das Schrumpfen bzw. Nullsetzen von Koeffizienten zugelassen.[11]

[10] Im Gegensatz dazu ist Underfittung dadurch gekennzeichnet, dass ein zu einfaches Modell einen (sehr) hohen Bias und eine (sehr) niedrige Varianz aufweist. Demnach gilt es, mit dem gewählten Modell eine Balance zwischen einem einfachen und einem komplizierten Modell zu finden.

[11] Eine andere Regularisierungsmethode stellt die Ridge-Regression (L2-Regularisierung) dar. Ein Nachteil dieses Verfahrens liegt jedoch darin, dass keine Variablenauswahl durchgeführt wird, da die Koeffizienten lediglich geschrumpft werden, aber nie den Wert null erreichen.

Der LASSO-Schätzer der Koeffizienten β_i wird anhand des Optimierungsproblems in Gleichung 5.3.10 ermittelt. Dabei bezeichnet $\hat{\boldsymbol{\beta}}^{\boldsymbol{LASSO}} = (\hat{\beta}_1^{LASSO}, ..., \hat{\beta}_p^{LASSO})^T$ den Vektor der LASSO-Schätzer der Koeffizienten β_i. Weiterhin kann β_0 aufgrund der zugrunde gelegten Annahme zentrierter Daten vernachlässigt werden.

$$\hat{\boldsymbol{\beta}}^{\boldsymbol{LASSO}} = \underset{\beta}{arg\,min} \left(\ell(\beta_0, \beta_i) + \lambda \sum_{i=1}^{p} |\beta_i| \right) \tag{5.3.10}$$

Während der linke Teil der Gleichung die negative Log-Likelihood-Funktion des GLM zeigt, kennzeichnet der rechte Term einen zusätzlichen Bestrafungsterm mit einem nicht-negativen Regularisierungsparameter λ. Dabei werden die Koeffizienten mit zunehmendem λ kontinuierlich in Richtung null geschrumpft. Bei einem hinreichend großen λ kommt es außerdem dazu, dass insignifikante Koeffizienten auf null geschrumpft werden, was zu einem Ausschluss dieser vom Modell führt. Des Weiteren wird das optimale λ üblicherweise mittels Kreuzvalidierung ermittelt. Gleichung 5.3.11 stellt zudem die Zusammensetzung der negativen Log-Likelihood-Funktion der Negativen Binomialregression dar. (vgl. Tibshirani (1996), S. 267ff. und Zhang und Jia (2017), S. 2f.)

$$\ell(\beta_0, \beta_i) = -\frac{1}{n} \sum_{k=1}^{n} \left(y_k \left(\beta_0 + \sum_{i=1}^{p} \beta_i x_{ik} \right) - (\theta + y_k)\, ln \left(\theta + exp \left(\beta_0 + \sum_{i=1}^{p} \beta_i x_{ik} \right) \right) \right) \tag{5.3.11}$$

Neben dem Problem des Overfittings kann das LASSO-Verfahren ebenfalls dem in Kapitel 5.3.6 vorgestellten Problem der Multikollinearität entgegenwirken, da das Verfahren tendenziell eine der korrelierten Variablen im Modell belässt und den Rest der korrelierten Variablen durch Nullsetzen der Koeffizienten ausschließt. (vgl. Friedman et al. (2010), S. 3)
Da die LASSO-Schätzer verzerrt sind und andere Varianzen wie die entsprechenden ML-Schätzer aufweisen, wird die Negative Binomialregression mit den mittels LASSO-Verfahren als relevant erachteten Variablen geschätzt, um unverzerrte ML-Schätzer zu erhalten. Anhand dieser Schätzung erfolgt dann die Beurteilung der Modellgüte sowie der Bedeutsamkeit der Variablen.

5.3.4 Modellgütekriterien

Als Modellgütekriterium kommt neben dem LR-Test ebenfalls das Pseudo-R-Quadrat-Maß McFaddens-R^2 (McF-R^2) zum Einsatz. Des Weiteren werden der MSE und der RMSE anhand von Kreuzvalidierung ermittelt, um die Güte der beiden angewandten Verfahren vergleichbar zu machen. Die Ermittlung der vier genannten Größen ist im Folgenden beschrieben.

Likelihood-Ratio-Test

Dem LR-Test, welcher bei genesteten Modellen zur Anwendung kommt, liegt die Nullhypothese zugrunde, dass die Koeffizienten der betrachteten unabhängigen Variablen null sind und die unabhängigen Variablen somit keinen Erklärungsgehalt für die unterschiedlichen Ausprägungen der abhängigen Variable liefern. (vgl. Backhaus et al. (2008), S. 262) Für die verwendete Teststatistik ergibt sich Gleichung 5.3.12.

$$LR\text{-}Test = -2\,ln\left(\frac{L_0}{L_V}\right) = -2\,(ln(L_0) - ln(L_V)) \tag{5.3.12}$$

Dabei bezeichnet L_0 den Likelihood-Wert des Nullmodells, welches lediglich das Absolutglied enthält, und L_V den Likelihood-Wert des vollständigen Modells. Der Quotient aus L_0 und L_V wird logarithmiert und schließlich mit *-2* multipliziert, wodurch eine asymptotisch χ^2-verteilte Teststatistik generiert wird. Die Anzahl der Freiheitsgrade ergibt sich aus der Differenz der Freiheitsgrade aus dem vollständigen und dem Nullmodell. Dabei weist das Nullmodell einen Freiheitsgrad von eins auf, da es lediglich das Absolutglied enthält. Durch den Vergleich der Teststatistik mit dem Referenzwert der χ^2-Verteilung wird die Signifikanz des Modells beurteilt. Meist wird dabei ein Signifikanzniveau von fünf Prozent zugrunde gelegt. Ein Modell wird als signifikant angesehen, wenn dessen LR-Wert größer als der tabellierte χ^2-Wert ist. In diesem Fall ist das vollständige Modell besser als das Nullmodell, was bedeutet, dass die Nullhypothese verworfen werden kann und die unabhängigen Variablen einen hohen Erklärungsanteil an der abhängigen Variable aufweisen. (vgl. Tabachnick und Fidell (2007), S. 447f.)

McFaddens-R^2

Da das in der linearen Regression verwendete Bestimmtheitsmaß R^2 zur Erklärung der Variation der Daten durch das vorliegende Modell nicht auf GLM's übertragbar ist, kommen Pseudo-R-Quadrat-Maße zur Anwendung. Das bekannteste Pseudo-R-Quadrat-Maß ist dabei McF-R^2, welches ebenfalls eine Gegenüberstellung von L_0 und L_V vorsieht.[12] Dabei berechnet sich die Kenngröße wie folgt.

$$McF\text{-}R^2 = 1 - \frac{ln(L_V)}{ln(L_0)} \tag{5.3.13}$$

Es lässt sich erkennen, dass der Quotient bei geringen Unterschieden zwischen den Modellen nahe eins ist, wodurch McF-R^2 einen Wert nahe null annimmt. Bei großen Unterschieden würde sich ein größeres McF-R^2 ergeben. Dementsprechend kennzeichnet ein großes McF-R^2 eine bessere Modellanpassung. (vgl. Hilbe (2011), S. 65f.) Nach Urban (1993, S. 62) wird bereits bei Werten im Bereich von 0,2 bis 0,4 von einer guten Modellanpassung gesprochen.

[12] Weitere Pseudo-R-Quadrat-Maße sind bspw. Cox und Snell-R^2 sowie Nagelkerke-R^2. Nähere Informationen dazu können Backhaus et al. (2008, S. 263ff.) entnommen werden.

MSE und RMSE

Um die Modellgüte des RFV[13] und der Negativen Binomialregression konsistent vergleichen zu können, wird der MSE bei der Negativen Binomialregression ebenfalls unter Anwendung von Kreuzvalidierung ermittelt. Nach Drittelung des Datensatzes wird das Modell dabei zunächst mit zwei Drittel der k (Trainingsdatensatz) erstellt und anschließend anhand des übrigen Drittels der k (Testdatensatz) in Form des MSE validiert. Das wiederum wird zwei weitere Male wiederholt, sodass am Ende jedes Drittel der k einmal zur Validierung genutzt wird und jedes k einmal dem Testdatensatz angehört. Somit ergeben sich drei verschiedene MSE, wobei sich jeder einzelne MSE aus dem Mittelwert der Summe der quadrierten Residuen ergibt. Die Residuen wiederum werden aus der Differenz zwischen geschätztem $\widehat{E(y_k)}$ und beobachtetem y_k gebildet. Auch hier gilt je geringer der Wert des MSE ist, desto näher liegen die vorhergesagten $\widehat{E(y_k)}$ und beobachteten y_k beieinander, d.h. desto besser bilden die trainierten Modelle die Testdaten ab. Gleichung 5.3.14 zeigt beispielhaft die Berechnung des ersten MSE. Dabei bezeichnet k_1 eine Beobachtung des ersten Testdatensatzes und n_1 die Anzahl der Beobachtungen des selbigen. Die Ermittlung der restlichen MSE erfolgt analog.

$$MSE_1 = \frac{1}{n_1} \sum_{k_1=1}^{n_1} (\widehat{E(y_{k_1})} - y_{k_1})^2 \tag{5.3.14}$$

Der eigentliche MSE der Negativen Binomialregression ergibt sich dann durch Mittelung der drei MSE. Gleichung 5.3.15 stellt diesen Zusammenhang dar. Dabei bezeichnen MSE_1 bis MSE_3 die einzelnen MSE.

$$MSE = \frac{MSE_1 + MSE_2 + MSE_3}{3} \tag{5.3.15}$$

Der RMSE ergibt sich entsprechend zur Wurzel des gemittelten MSE, was Gleichung 5.3.16 formal darstellt.

$$RMSE = \sqrt{\frac{MSE_1 + MSE_2 + MSE_3}{3}} \tag{5.3.16}$$

5.3.5 Variablenbedeutsamkeit

Um die Signifikanz und damit den Einfluss einzelner unabhängiger Variablen i auf die abhängige Variable zu überprüfen, kommt der Wert der Wald-Teststatistik W_i zur Anwendung. Die zugehörige Nullhypothese lautet, dass der entsprechende Koeffizient β_i null ist und somit keinen signifikanten Einfluss auf die abhängige Variable aufweist. Die Teststatistik ergibt sich aus dem quadrierten Quotienten des geschätzten Koeffi-

[13] Im RFV wird jeder Entscheidungsbaum - gemäß Kapitel 5.2.2 und 5.2.3 - zufällig mit zwei Drittel der k (Trainingsdatensatz) erstellt und mit dem übrigen Drittel der k (Testdatensatz) in Form des MSE bzw. RMSE validiert. Bei Betrachtung aller Entscheidungsbäume gehört dabei jedes k mindestens einmal einem Testdatensatz an.

zienten $\hat{\beta}_i$ und dessen geschätztem Standardfehler $s_{\hat{\beta}_i}$. Gleichung 5.3.17 stellt diesen Zusammenhang dar.

$$W_i = \left(\frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \right)^2 \tag{5.3.17}$$

Die Teststatistik folgt einer asymptotischen χ^2-Verteilung. Ist W_i größer als der zugehörige tabellierte Wert der χ^2-Verteilung bei einem vorgegebenem Signifikanzniveau, wird die Nullhypothese abgelehnt und von einem signifikanten Einfluss der unabhängigen Variable i ausgegangen. (vgl. Backhaus et al. (2008), S. 273)

5.3.6 Modelldiagnostik

Da GLM's verallgemeinerte lineare Modelle darstellen, lassen sich die Annahmen der linearen Regression, welche nach Durchführung der Schätzung zu prüfen sind, auf GLM's übertragen. Dabei spielen die Residuen eine Rolle, die sich im Falle der linearen Regression zu $y_k - \hat{\mu}_k$ ergeben. Da die Varianz bei GLM's der Exponentialfamilie - im Gegensatz zu linearen Modellen - jedoch von μ_k abhängt, sind diese Residuen nicht für die Diagnostik von GLM's geeignet. Aus diesem Grund stellt das vorliegende Kapitel zunächst eine geeignete Residuenart vor, die bei Verwendung des korrekten GLM eine Normalverteilung mit einem Erwartungswert von null und einer konstanten Varianz aufweist. Im Anschluss erfolgt dann die Vorstellung der Modellannahmen sowie ein Maß zur Ermittlung einflussreicher Beobachtungen.

GLM-Residuen

Für die Untersuchung diskreter Verteilungen der Exponentialfamilie eignen sich standardisierte Quantilresiduen $r'_{Q,k}$, wobei die Standardisierung dazu führt, dass die Residuen bei Wahl des korrekten Modells eine nahezu konstante Varianz aufweisen. Gleichung 5.3.18 zeigt auf der linken Seite die Berechnungsvorschrift eines $r'_{Q,k}$.

$$r'_{Q,k} = \frac{r_{Q,k}}{\sqrt{1 - h_k}} \quad mit \quad r_{Q,k} = \Phi^{-1}(u_k) \tag{5.3.18}$$

Dabei stellen $r_{Q,k}$ die Quantilresiduen der k dar. Das Besondere dabei ist, dass das $r_{Q,k}$ eines k die selbe kumulative Wahrscheinlichkeit der Standardnormalverteilung aufweist wie das entsprechende y_k bei der gewählten Verteilung der Exponentialfamilie. Weiterhin bezeichnet h_k die Hebelwirkung eines k. Diese beschreibt, wie stark die entsprechende Beobachtung k ihren Erwartungswert beeinflusst. Je größer h_k ist, desto größer ist die Hebelwirkung dieses k. Dementsprechend sagt ein kleines h_k aus, das nicht nur dieses k allein, sondern mehrere k gemeinsam zur Schätzung des Erwartungswertes beitragen. Die $r_{Q,k}$ lassen sich für diskrete abhängige Variablen mit Hilfe der Vorschrift auf der rechten Seite bestimmen. Dabei bezeichnet Φ^{-1} die inverse kumulative Verteilungsfunktion (CDF) der Standardnormalverteilung, wobei die CDF aufgrund der

diskreten Verteilung Sprünge aufweist. u_k bezeichnet einen zufällig gewählten Wert innerhalb dieses Sprungbereichs der CDF. Durch das zufällige Vorgehen wird sichergestellt, dass die Quantilresiduen kontinuierliche Ausprägungen aufweisen, obwohl die untersuchte Verteilungsfunktion durch diskrete Werte gekennzeichnet ist. Des Weiteren folgen die Quantilresiduen einer Standardnormalverteilung, wenn das korrekte Modell aus der Exponentialfamilie gewählt wird.[14] (vgl. Dunn und Smyth (2018), S. 300ff.)
Die im Folgenden im Rahmen der Modellannahmen angesprochenen Residuen meinen die eben vorgestellten standardisierten Quantilresiduen.

Modellannahmen

Die folgenden Ausführungen beziehen sich - bis auf eine separat gekennzeichnete Ausnahme - auf Backhaus et al. (2008, S. 80ff.) sowie Dunn und Smyth (2018, S. 297ff.).
Eine Annahme linearer Modelle liegt in der richtigen Spezifizierung des Modells. Dazu zählt bspw., dass das Modell alle relevanten erklärenden Variablen enthält und dass die Zahl der zu schätzenden Koeffizienten kleiner als die Zahl der Beobachtungen ist. Außerdem ist es bei GLM's notwendig, dass die abhängige Variable der Verteilung der gewählten Exponentialfamilie folgt und dass die korrekte Linkfunktion genutzt wird. Weiterhin muss eine lineare Beziehung zwischen der mittels der Linkfunktion transformierten abhängigen Variable und den unabhängigen Variablen im linearen Prädiktor bestehen. Wird die Transformation von nicht-linearen in lineare Beziehungen vernachlässigt, so kommt es zur Verzerrung der geschätzten Koeffizienten.
Eine weitere Annahme besagt, dass die Residuen einer Normalverteilung mit einem Erwartungswert von null folgen. Somit unterstellt das Modell, dass sich die Schwankungen der Residuen im Mittel ausgleichen. Ein Erwartungswert ungleich null würde sich bspw. bei einem systematischen Messfehler der abhängigen Variable ergeben, was dazu führt, dass das Absolutglied verzerrt wird. Wird hingegen ein Modell ohne Absolutglied spezifiziert, so führt dies zu einer Verzerrung der übrigen Koeffizienten. Um zu überprüfen, ob die Residuen einer Normalverteilung folgen, kommt üblicherweise der Normal Quantile-Quantile-Plot - oder kurz Normal Q-Q-Plot - zum Einsatz. Dieser kommt in der linearen Regression zur Anwendung, um die Beziehung zwischen den theoretischen Quantilen der Normalverteilung und den Quantilen der Residuen der Beobachtungen grafisch darzustellen. Folgen die Residuen einer Normalverteilung, so liegen die Punkte auf einer Linie. Normal Q-Q-Plots können ebenfalls auf die zufällige Komponente eines GLM's angewandt werden, da die Quantilresiduen - wie zuvor beschrieben - bei korrekter Wahl des Modells aus der Exponentialfamilie einer Normalverteilung folgen und demzufolge ebenfalls auf einer Linie liegen. Demnach kann mit Hilfe eines solchen Plots ebenfalls überprüft werden, ob die Annahme einer negativen Binomialverteilung

[14] Pearson- und Devianzresiduen stellen weitere Residuenarten dar, die üblicherweise bei der Untersuchung von GLM's herangezogen werden. Da deren Auftreten bei diskreten Verteilungen jedoch zu einer Abweichung von einer Normalverteilung führen kann, sind diese bei der vorliegenden Untersuchung weniger gut geeignet.

für die abhängige Variable y_k angemessen ist. (vgl. Dunn und Smyth (2018), S. 105ff.) Das Vorliegen von Homoskedastizität beschreibt eine weitere Annahme. Diese besagt, dass die Varianz der Residuen annähernd konstant ist. Ist dies nicht der Fall, so liegt Heteroskedastizität vor. Bei Wahl des korrekten Modells aus der Exponentialfamilie ist die Varianz der Residuen annähernd konstant, sodass von einer Einhaltung der Annahme ausgegangen werden kann. Die Folgen von Heteroskedastizität sind eine ineffiziente Schätzung sowie verzerrte Standardfehler der Koeffizienten. Mit Hilfe eines Plots der Residuen gegen die geschätzten Werte von y_k kann diese Annahme überprüft werden, wobei der Plot kein Trend bzw. Muster aufweisen sollte.
Die nächste Annahme geht davon aus, dass die Residuen in der Grundgesamtheit unkorreliert sind, d.h. dass die Beobachtungen unabhängig voneinander und die Residuen zufällig sind. Ist dies nicht der Fall, liegt Autokorrelation vor. Die Folge von Autokorrelation, welche häufig bei Zeitreihen auftritt, liegt dabei in verzerrten Standardfehlern. Um eine mögliche Verletzung der Annahme zu überprüfen, können die Residuen ebenfalls gegen die geschätzten Werte von y_k geplottet werden. Im Optimalfall, also der Abwesenheit von Autokorrelation, lässt sich dabei abermals kein Trend bzw. Muster erkennen. Liegen aufeinanderfolgende Werte der Residuen hingegen nahe beieinander, so ist das ein Indiz für positive Autokorrelation. Im Gegensatz dazu würden die Residuen bei negativer Autokorrelation stark schwanken.[15]
Eine weitere Annahme besagt, dass zwischen den Residuen und den unabhängigen Variablen keine Korrelation besteht. Eine solche Korrelation tritt bspw. auf, wenn im Modell relevante Variablen, die mit einer der aufgenommenen Variablen korrelieren, vernachlässigt werden. Da die Residuen den Einfluss von vernachlässigten Variablen enthalten, kommt es bei Vorliegen einer solchen Korrelation dazu, dass der Teil der Variation der abhängigen Variable, der den Residuen zugeschrieben wird, fälschlicherweise der aufgenommenen unabhängigen Variable zugeordnet wird. Eine solche Verletzung führt schließlich zu einer Verzerrung des geschätzten Koeffizienten der betroffenen unabhängigen Variable.
Die letzte Annahme liegt in der Abwesenheit von perfekter Multikollinearität zwischen den unabhängigen Variablen. Dabei beschreibt perfekte Multikollinearität einen exakt linearen Zusammenhang zwischen diesen. Generell zeigt sich Multikollinearität in einer zu hohen Korrelation zwischen den entsprechenden Variablen, was bedeutet, dass diese (annähernd) den gleichen Erklärungsgehalt für die abhängige Variable liefern. Anders ausgedrückt bedeutet dies, dass sich die Streuungen der betroffenen Variablen überschneiden, wodurch sich die vorhandene Information nicht mehr eindeutig den Variablen zuordnen lässt. Eine Folge zunehmender Multikollinearität liegt darin, dass die geschätzten Koeffizienten unzuverlässiger werden. Eine weitere Folge ist die Ermittlung zu großer Standardfehler.

[15] Eine weitere Möglichkeit der Identifikation von Autokorrelation besteht darin, die Residuen gegen die eine Zeiteinheit zuvor liegenden Residuen zu plotten und auf Trends zu überprüfen.

Einflussreiche Beobachtungen

Als einflussreiche Beobachtungen werden Beobachtungen mit einem großen Residuum sowie einer großen Hebelwirkung aufgefasst. Ist der Einfluss einer solchen Beobachtung hinreichend groß, so kann das Ausschließen dieser aus dem Modell zu einer substanziellen Änderung der geschätzten Koeffizienten führen. Die Cook-Distanz D_k, deren Berechnungsvorschrift in Gleichung 5.3.19 abgebildet ist, quantifiziert diesen Effekt.

$$D_k = \frac{(r'_{P,k})^2}{p} \frac{h_k}{1 - h_k} \tag{5.3.19}$$

Während p die Anzahl der Koeffizienten - inklusive dem Absolutglied - darstellt, kennzeichnet $r'_{P,k}$ das standardisierte Pearson-Residuum von k, dessen Berechnungsvorschrift im Folgenden auf der linken Seite abgebildet ist.

$$r'_{P,k} = \frac{r_{P,k}}{\sqrt{\theta(1 - h_k)}} \quad mit \quad r_{P,k} = \frac{y_k - \hat{\mu}_k}{\sqrt{Var(\hat{\mu}_k)/w_k}} \tag{5.3.20}$$

Diese enthält neben einem Dispersionsparameter θ abermals die Hebelwirkung h_k. Die Berechnung des nicht-standardisierten Pearson-Residuums $r_{P,k}$ erfolgt gemäß der Berechnungsvorschrift auf der rechten Seite, wobei $Var(\hat{\mu}_k)$ die nicht-konstante Varianz des GLM's beschreibt. Weiterhin bezeichnet w_k die Gewichtung eines k.
Eine Daumenregel besagt zudem, dass Beobachtungen mit D_k größer eins einen potenziell hohen Einfluss aufweisen und daher näher untersucht werden sollten. (vgl. Dunn und Smyth (2018), S. 110f. sowie S. 299f.)

6 Durchführung und Ergebnisse

Nach einer kurzen Betrachtung der Korrelationen zwischen den einbezogenen Variablen erfolgt die Vorstellung der Ergebnisse des RFV sowie der Negativen Binomialregression, wobei die Ergebnisse jeweils getrennt nach Unfalltyp beleuchtet werden. Im Anschluss werden die Ergebnisse beider Verfahren miteinander verglichen. Dabei liegt der Fokus einerseits auf den erzielten Modellgütekriterien und andererseits auf den ermittelten Einflüssen der bedeutsamsten Variablen. Im letzten Schritt erfolgt zudem eine Gegenüberstellung der erzielten Ergebnisse mit den Erkenntnissen der wissenschaftlichen Autoren. Anzumerken sei hierbei, dass alle dargestellten Ergebnisse mit der Statistiksoftware *R* erzielt wurden.

6.1 Korrelationsbetrachtung

Da die Korrelationen zwischen den Variablen mit dem in Kapitel 5.1 beschriebenen Maß C_{korr} bestimmt werden, müssen alle Variablen in kategorialer Form vorliegen. Um das zu erreichen, werden die Variablen, die nicht kategorial vorliegen, kategorisiert. Das betrifft die Variablen y_1, y_2, x_4 bis x_7, x_{10} sowie x_{18} bis x_{26}, deren Bedeutung Kapitel 4.3 zu entnehmen ist. Da die Variablen y_1, y_2 sowie x_4 bis x_7 ganzzahlige Ausprägungen aufweisen, werden diese Ausprägungen jeweils als eigene Kategorie der entsprechenden Variable einbezogen. Die übrigen genannten Variablen, die diese Eigenschaft nicht aufweisen, werden anhand von Intervallen kategorisiert. Die so gebildeten Ausprägungen der kategorisierten Variablen sind in Tabelle D.1 in Anhang D dargestellt.
Nachdem für alle Variablen eine kategoriale Form vorliegt, können die jeweiligen C_{korr} gemäß Gleichung 5.1.4 bestimmt werden, um einen ersten Eindruck von den Zusammenhängen der Variablen zu erlangen. Die sich so ergebenden Korrelationen zwischen den einzelnen Variablen werden in Abbildung 6.1 grafisch veranschaulicht. Dabei kennzeichnet ein weißes Feld einen Kontingenzkoeffizienten von null, was bedeutet, dass zwischen den Ausprägungen der beiden betrachteten Variablen kein Zusammenhang besteht. Je größer der Kontingenzkoeffizient ausfällt, d.h. je roter die Farbe des Feldes ist, desto stärker ist der Zusammenhang zwischen den Ausprägungen der Variablen. Da es sich um eine symmetrische Korrelationsmatrix handelt, zeigt die Abbildung lediglich eine Hälfte der Korrelationen auf.
Aus der Roteinfärbung in Abbildung 6.1 ergibt sich demzufolge bspw. ein stärkerer Zusammenhang zwischen den Variablen *Anzahl der Unfälle vom Typ 23/24 pro Zufahrt* (y_1) und *Roteinfärbung der Radfahrfurt* (x_{17}). Weitere stärkere Zusammenhänge

finden sich zwischen den Variablen *Anzahl Linksabbiegefahrstreifen* (x_4) und *Signalisierung Linksabbiegende* (x_8) sowie zwischen den Variablen *Signalisierung Linksabbiegende* (x_8) und *Radverkehrsführung Linksabbiegende* (x_{16}). Ein ebenfalls stärkerer Zusammenhang lässt sich einerseits zwischen den Variablen *Anzahl Geradeausfahrstreifen* (x_5) und *DTV Kfz Geradeausfahrende* (x_{19}) und andererseits zwischen den Variablen *Radverkehrsführung im Knotenpunkt* (x_{13}) und *Absetzung Radfahrfurt* (x_{14}) erkennen.

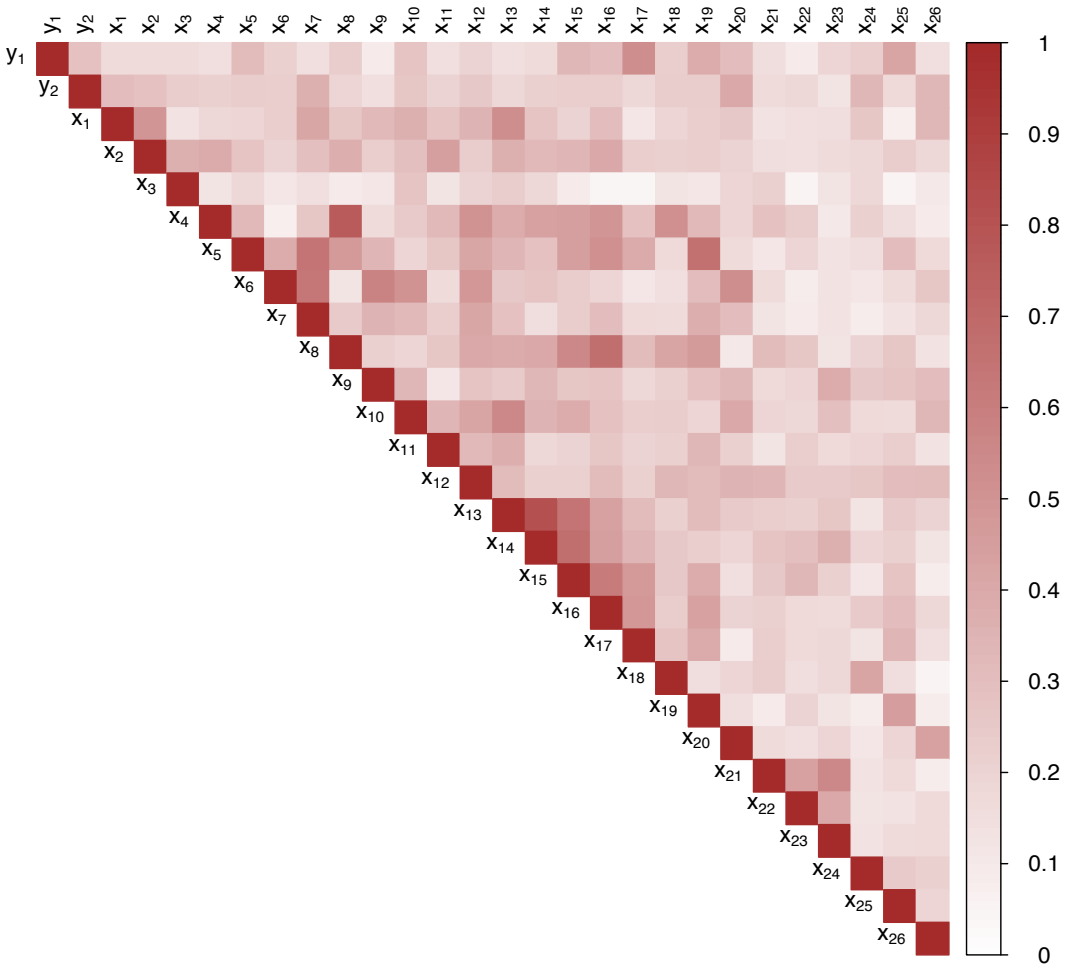


Abbildung 6.1: Grafische Darstellung der Korrelationen [Quelle: eigene Darstellung]

6.2 Random Forest

Das RFV wird mit Hilfe der Software *R* für die betrachteten Unfalltypen 23/24 mit der abhängigen Variable y_1 sowie 30/32/34 mit der abhängigen Variable y_2 durchgeführt. Aufgrund des zufälligen Vorgehens im RFV können dabei jeweils alle zur Verfügung stehenden Variablen (p=*26*) einbezogen werden, ohne vorher eine Auswahl treffen zu müssen. Zudem zeigte sich in einer eigens durchgeführten Simulation, dass das Ergebnis des RFV durch die Einbeziehung (hoch-)korrelierter Variablen nicht beeinflusst wird,

wodurch eine vorherige Variablenselektion ebenfalls überflüssig erscheint.
Die beiden festzulegenden Parameter des RFV sind gemäß Kapitel 5.2.2 die Anzahl der zufällig gewählten Variablen am Knoten sowie die Anzahl der Entscheidungsbäume. Für die Anzahl der zufällig gewählten Variablen, aus denen die beste zur Aufsplittung am jeweiligen Knoten herangezogen wird, kommt für beide betrachtete Unfalltypen die Standardeinstellung der Software *R* zum Einsatz, wodurch sich diese jeweils zu acht ergibt.[1] Die optimale Anzahl der Entscheidungsbäume hingegen hängt vom MSE ab, welcher bei Hinzukommen weiterer Entscheidungsbäume variiert. Um den Zusammenhang grafisch zu veranschaulichen, stellt Abbildung 6.2 die Höhe des MSE in Abhängigkeit der Anzahl der Entscheidungsbäume für die beiden Unfalltypen dar. Die optimale Anzahl der Entscheidungsbäume findet sich dabei jeweils an der Stelle des minimalen MSE. Während dieser beim Unfalltyp 23/24 bei 290 Entscheidungsbäumen erreicht wird, beträgt die optimale Anzahl beim Unfalltyp 30/32/34 313 Entscheidungsbäume. Die gestrichelten Linien veranschaulichen die optimale Anzahl der Entscheidungsbäume grafisch.

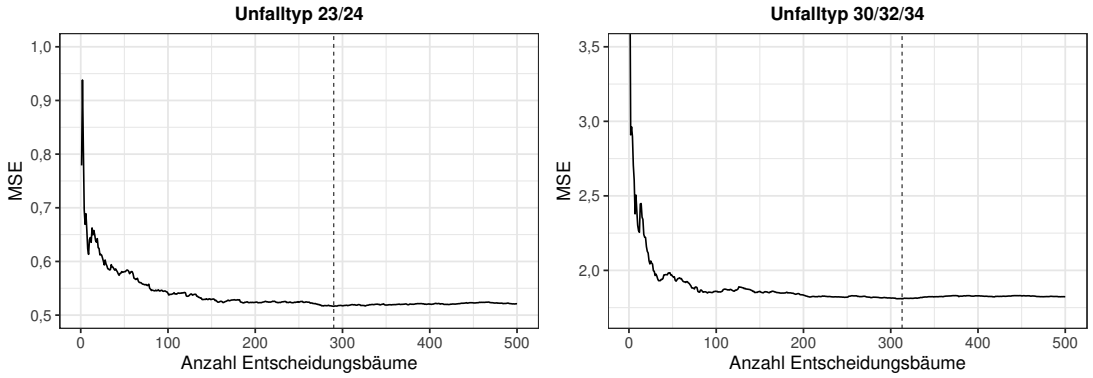


Abbildung 6.2: Optimale Anzahl der Entscheidungsbäume nach Unfalltyp [Quelle: eigene Darstellung]

6.2.1 Modellgütekriterien

Nach Durchführung des RFV mit den eben beschriebenen Parametern ergeben sich die in Tabelle 6.1 dargestellten Modellgütekriterien gemäß Kapitel 5.2.3. Dabei ist der MSE ein Maß dafür, wie gut die mit den Bootstrap-Stichproben trainierten Entscheidungsbäume die Vorhersagen der OOB-Beobachtungen k - in Form der mittleren quadrierten Abweichung der geschätzten und beobachteten y_k - abbilden. Während der MSE für Unfälle vom Typ 23/24 0,53 (Unfälle)^2 beträgt, beläuft sich dieser für Unfälle vom Typ 30/32/34 auf 1,89 (Unfälle)^2. Der RMSE ergibt sich als Wurzel des MSE für den Unfalltyp 23/24 zu 0,73 Unfällen und für den Unfalltyp 30/32/34 zu 1,37 Unfällen.

[1] Standardmäßig ermittelt die Software *R* die Anzahl der zufällig gewählten Variablen mit der Berechnungsvorschrift $p/3$, wobei das Ergebnis abgerundet wird.

Dementsprechend generiert das RFV beim Unfalltyp 23/24 genauere OOB-Vorhersagen als beim Unfalltyp 30/32/34.

Tabelle 6.1: Modellgüte nach Unfalltyp: Random Forest-Verfahren

	Unfalltyp 23/24	Unfalltyp 30/32/34
MSE	0,53	1,89
RMSE	0,73	1,37

6.2.2 Variablenbedeutsamkeit

Um die Bedeutsamkeit einzelner unabhängiger Variablen zu beurteilen, kommen die in Kapitel 5.2.4 beschriebenen Konstrukte VI, PDP sowie ALE zum Einsatz. Abbildung 6.3 stellt dafür zunächst die VI beider betrachteter Unfalltypen dar.

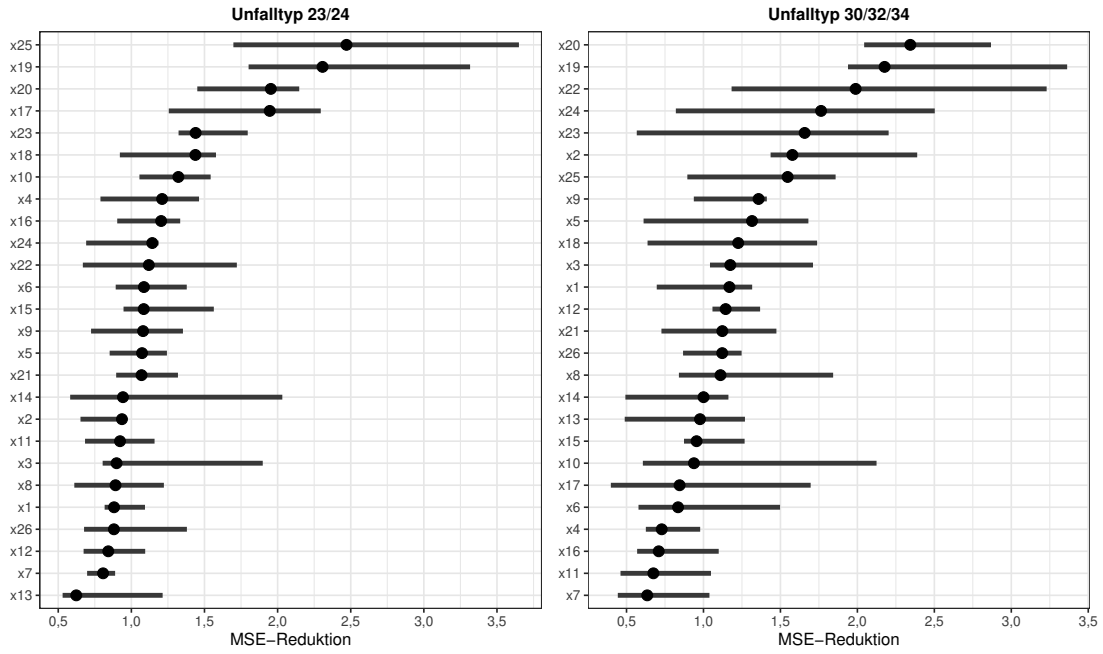


Abbildung 6.3: Variable Importance nach Unfalltyp [Quelle: eigene Darstellung]

Die Abszisse zeigt dabei jeweils die MSE-Reduktion der x_i, während die Ordinate die einzelnen x_i nach absteigendem VI abbildet. Um ein VI zu ermitteln, wird das in Kapitel 5.2.4 beschriebene Vertauschen der Werte eines x_i und die darauffolgende Berechnung der MSE-Reduktion wiederholt durchgeführt, um stabilere Ergebnisse zu erlangen. (vgl. Molnar (2019)) Die horizontalen Linien kennzeichnen dabei jeweils ein Intervall, dessen Anfangspunkt das fünf-Prozent-Quantil der MSE-Reduktionen darstellt, welches sich aus den Wiederholungen ergibt. Das Ende dieser horizontalen Linien hingegen bildet das 95-Prozent-Quantil dieser wiederholt ermittelten MSE-Reduktionen

ab. Weiterhin kennzeichnet der Punkt jeweils den Median der so ermittelten MSE-Reduktionen, anhand dessen die einzelnen x_i nach absteigendem VI geordnet werden.
Bei Betrachtung des Unfalltyps 23/24 fällt auf, dass die drei bedeutsamsten Variablen DTV-Kenngrößen sind. Dabei kennzeichnet x_{25} gemäß Tabelle 4.7 den DTV der geradeausfahrenden Radfahrenden, x_{19} den DTV der geradeausfahrenden Kfz und x_{20} den DTV der rechtsabbiegenden Kfz. Dementsprechend beeinflussen diese drei Größen die erwartete Anzahl der Unfälle vom Typ 23/24 am meisten. Bei Betrachtung der Beteiligtenkonstellationen des Unfalltyps 23/24 in Abbildung 3.1 wird zudem ersichtlich, dass diese Größen die für die betrachteten Unfalluntertypen charakteristischen Bewegungsrichtungen der Beteiligten widerspiegeln, wodurch das Ergebnis sinnhaft erscheint. Die bedeutsamsten Infrastrukturmerkmale hingegen sind die Roteinfärbung der Radfahrfurt (x_{17}) sowie der Ausrundungsradius der Rechtsabbiegenden (x_{10}) und die Anzahl der Linksabbiegefahrstreifen (x_4). Im Gegensatz dazu bilden die Merkmale Abbiegeverbote (x_{12}), Anzahl der Mischfahrstreifen (x_7) sowie die Radverkehrsführung im Knotenpunkt (x_{13}) die Infrastrukturmerkmale mit dem geringsten Einfluss. Demzufolge determinieren die Ausprägungen dieser Variablen die erwartete Anzahl der Unfälle vom Typ 23/24 am wenigsten.
Bei Betrachtung Unfalltyps 30/32/34 wird ersichtlich, dass die fünf bedeutsamsten Variablen abermals DTV-Kenngößen bzw. Schwerverkehrsanteile sind. Diese sind gemäß Tabelle 4.7 der DTV der rechtsabbiegenden Kfz (x_{20}), der DTV der geradeausfahrenden Kfz (x_{19}), der Schwerverkehrsanteil der Geradeausfahrenden (x_{22}), der DTV des linksabbiegenden Radverkehrs (x_{24}) sowie der Schwerverkehrsanteil der Rechtsabbiegenden (x_{23}). Diese Angaben erscheinen zunächst etwas durcheinander, aber bei Betrachtung der Beteiligtenkonstellationen des Unfalltyps 30/32/34 in Abbildung 3.2 findet sich ein möglicher Grund in der Einbeziehung unterschiedlich charakterisierter Unfalluntertypen. Des Weiteren sind die bedeutsamsten Infrastrukturmerkmale die Verkehrsregelungsart (x_2), die Signalisierung der Rechtsabbiegenden (x_9) sowie die Anzahl der Geradeausfahrstreifen (x_5). Im Gegensatz dazu stellen die Radverkehrsführung der Linksabbiegenden (x_{16}), der Abstand des ruhenden Verkehrs zum Knotenpunkt (x_{11}) sowie die Anzahl der Mischfahrstreifen (x_7) die Variablen dar, die die erwartete Anzahl der Unfälle vom Typ 30/32/34 am wenigsten beeinflussen.
Da die VI lediglich eine Aussage darüber erlaubt, welche unabhängigen Variablen die erwarteten Unfallzahlen am meisten bzw. am wenigsten beeinflussen, beleuchten die nachfolgenden Abschnitte den Effekt der einzelnen Ausprägungen der Variablen detaillierter. Dafür werden PDP's und ALE herangezogen, die sich pro untersuchten Unfalltyp beispielhaft auf die jeweils bedeutsamste DTV-Kenngröße sowie - aufgrund der untersuchten Fragestellung - auf die drei jeweiligen bedeutsamsten Infrastrukturmerkmale gemäß VI beschränken. Die übrigen PDP's sowie ALE, welche analog interpretiert werden können, befinden sich in den Abbildungen D.1 bis D.4 in Anhang D. Wie in Kapitel 5.2.4 bereits beschrieben, isolieren die ALE - im Gegensatz zu PDP's - partiell

die Effekte der übrigen Variablen, woraus einerseits eine Robustheit gegenüber korrelierten Variablen und andererseits eine zuverlässigere Aussagekraft gegenüber PDP's resultiert.
Abbildung 6.4 zeigt den Einfluss der Ausprägungen des DTV des geradeausfahrenden Radverkehrs (x_{25}) auf die erwartete Anzahl der Unfälle vom Typ 23/24. Dabei wird bei Betrachtung des PDP ersichtlich, dass die erwartete Unfallzahl bis zu einem DTV von etwa 400 Radfahrenden pro 24 Stunden leicht ansteigt, wobei das RFV bei einem DTV von etwa 80 Radfahrenden pro 24 Stunden die minimale Unfallzahl von etwa 0,15 Unfällen vorhersagt. Die erwartete Unfallzahl steigt dann bis zu einem DTV von etwa 1.150 Radfahrenden pro 24 Stunden stärker an, wobei das RFV bei einem DTV von etwa 1.150 Radfahrenden pro 24 Stunden die höchste Unfallzahl von etwa 0,9 Unfällen prognostiziert. Ein höherer DTV als die eben genannten 1.150 Radfahrenden pro 24 Stunden wiederum führt laut RFV zu einer konstanten erwarteten Unfallzahl von etwa 0,9 Unfällen und somit zu keiner weiteren Änderung derselben.

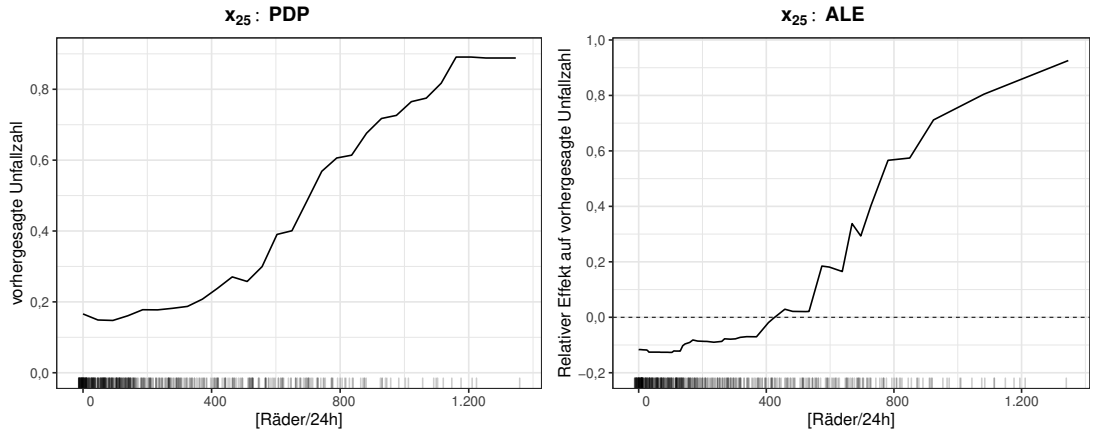


Abbildung 6.4: Unfalltyp 23/24: PDP und ALE der bedeutsamsten DTV-Größe [Quelle: eigene Darstellung]

Der ALE der Variable x_{25} weist ein ähnliches Muster auf. Allerdings stellt die Ordinate dabei anstelle des absoluten Effekts den relativen Effekt der einzelnen Ausprägungen auf die erwartete Unfallzahl dar, wobei der mittlere Effekt von x_{25} auf die erwartete Unfallzahl durch die gestrichelte Linie dargestellt ist. Dabei zeigt sich, dass bei einem DTV bis etwa 410 Radfahrende pro 24 Stunden weniger Unfälle vorhergesagt werden als im Mittel prognostiziert. Bei einem DTV von etwa 80 Radfahrenden pro 24 Stunden sagt das RFV den größten positiven Effekt auf die erwartete Unfallzahl vorher, denn diese beträgt dabei etwa 0,12 Unfälle weniger als im Mittel erwartet. Dieser DTV-Wert deckt sich mit dem DTV des PDP, der die minimale Unfallzahl prognostiziert. Bei mehr als etwa 410 Radfahrenden pro 24 Stunden erwartet das RFV mehr Unfälle als im Mittel, wobei der Anstieg des relativen Effekts zunächst bis zu einem DTV von etwa 800 Radfahrenden pro 24 Stunden stärker zunimmt, während dieser ab etwa 800 Radfahrenden

pro 24 Stunden wieder abflacht, jedoch kein konstantes Niveau erreicht. Nach dem ALE wird die maximale Unfallzahl bei dem maximalen DTV von etwa 1.350 Radfahrenden pro 24 Stunden erreicht. Der zugehörige relative Effekt beträgt dabei etwa 0,92 Unfälle mehr als im Mittel erwartet. Da sich die Verläufe des PDP und des ALE sehr ähneln, scheint bei der Untersuchung des Unfalltyps 23/24 mit dem RFV kein schwerwiegendes Problem mit korrelierten Variablen vorzuliegen. Anzumerken ist außerdem, dass die Anzahl der Beobachtungen, welche jeweils durch die Dichte der Striche oberhalb der Abszisse gekennzeichnet ist, ab einem DTV von etwa 900 Radfahrenden pro 24 Stunden abnimmt, wodurch die Vorhersagen ab diesem DTV ungenau sein können. Zudem lässt sich in beiden Abbildungen ab einem DTV von etwa 400 Radfahrenden pro 24 Stunden ein positiv degressiver Funktionsverlauf erkennen, welcher für den in Kapitel 2.1 beschriebenen *Safety in Numbers-Effekt* charakteristisch ist. Gemäß der Abbildungen D.1 und D.2 in Anhang D zeigt sich dieser ebenfalls für die zweit- und drittbedeutsamste DTV-Kenngröße des Unfalltyps 23/24 DTV der geradeausfahrenden Kfz (x_{19}) sowie DTV der rechtsabbiegenden Kfz (x_{20}).

Die PDP's bzw. ALE der drei bedeutsamsten Infrastrukturgrößen des Unfalltyps 23/24 sind Abbildung 6.5 dargestellt, wobei die Ausprägungen der Roteinfärbung der Radfahrfurt (x_{17}) aus Platzgründen mit a (nein), b (ja, über gesamten Knoten) und c (ja, teilweise) abgekürzt werden. Auf Grundlage des PDP prognostiziert das RFV für x_{17} mit 0,8 Unfällen die meisten Unfälle, wenn die Radfahrfurt einer Zufahrt teilweise über den Knotenpunkt rot eingefärbt verläuft. Im Gegensatz dazu erwartet das Verfahren etwa 0,35 Unfälle bei Roteinfärbung derselben über den gesamten Knotenpunkt sowie lediglich 0,2 Unfälle, wenn keine Roteinfärbung vorliegt. Der zugehörige ALE[2] zeigt zudem, dass das Vorhandensein einer teilweise über den Knotenpunkt rot eingefärbten Radfahrfurt den größten relativen Effekt auf die erwartete Unfallzahl aufweist, obwohl diese Ausprägung die wenigsten Beobachtungen enthält. Demzufolge ergeben sich in diesem Fall 0,6 Unfälle mehr als die mittlere durch x_{17} erwartete Unfallzahl. Während eine Roteinfärbung der Radfahrfurt über den gesamten Knoten etwa 0,1 Unfälle mehr als im Mittel erwarten lässt, verringert sich die Unfallzahl ohne Roteinfärbung der Radfahrfurt im Vergleich zum Mittel minimal. Dieses Ergebnis erscheint zunächst entgegen der Intuition, denn eine durch Roteinfärbung erhöhte Sichtbarkeit der Radfahrfurt sollte zu einer Erhöhung der Sicherheit beitragen. Der Grund für das erhöhte Unfallrisiko bei einer lediglich teilweise über den Knotenpunkt rot eingefärbten Radfahrfurt könnte jedoch darin liegen, dass der abbiegende Kfz-Verkehr die Radfahrfurt nicht ausreichend wahrnimmt, da der relevante Teil der Furt nicht eingefärbt ist. Demzufolge könnten ankommende Radfahrende missachtet werden, woraus ein erhöhtes Unfallrisiko resultiert. Im Gegensatz dazu lässt sich für die höhere erwartete Unfallzahl bei Vorliegen einer Roteinfärbung über den gesamten Knotenpunkt kein sachlogischer Grund herleiten. Es ist daher denkbar, dass das RFV bzw. die Daten die vorliegende Kausalität zwischen der

[2]Die Reihenfolge der Ausprägungen von x_{17} liegt verändert vor, da die Software *R* diese nach Ähnlichkeit der übrigen Variablen ordnet. (vgl. Kapitel 5.2.4)

Einfärbung der Radfahrfurt und dem Unfallgeschehen des Typs 23/24 nicht korrekt abbilden. Einen Grund dafür könnte der *Maßnahmenkatalog gegen Unfallhäufungen* liefern, denn dieser schlägt vor, die Sichtbarkeit des Radverkehrs durch Roteinfärbung der Radfahrfurten zu erhöhen, wenn an den entsprechenden Stellen Unfallhäufungen auftreten. (vgl. BASt (2017)) Demzufolge ist es möglich, dass die Furten aufgrund einer erhöhten Unfallzahl teilweise bzw. komplett rot über den Knotenpunkt eingefärbt wurden, sodass das Unfallgeschehen die Einfärbung bedingte und nicht umgekehrt. Das wiederum könnte einen Erklärungsansatz der höheren geschätzten Unfallzahlen bei Vorliegen einer teilweise oder komplett über den Knotenpunkt rot eingefärbten Radfahrfurt darstellen.

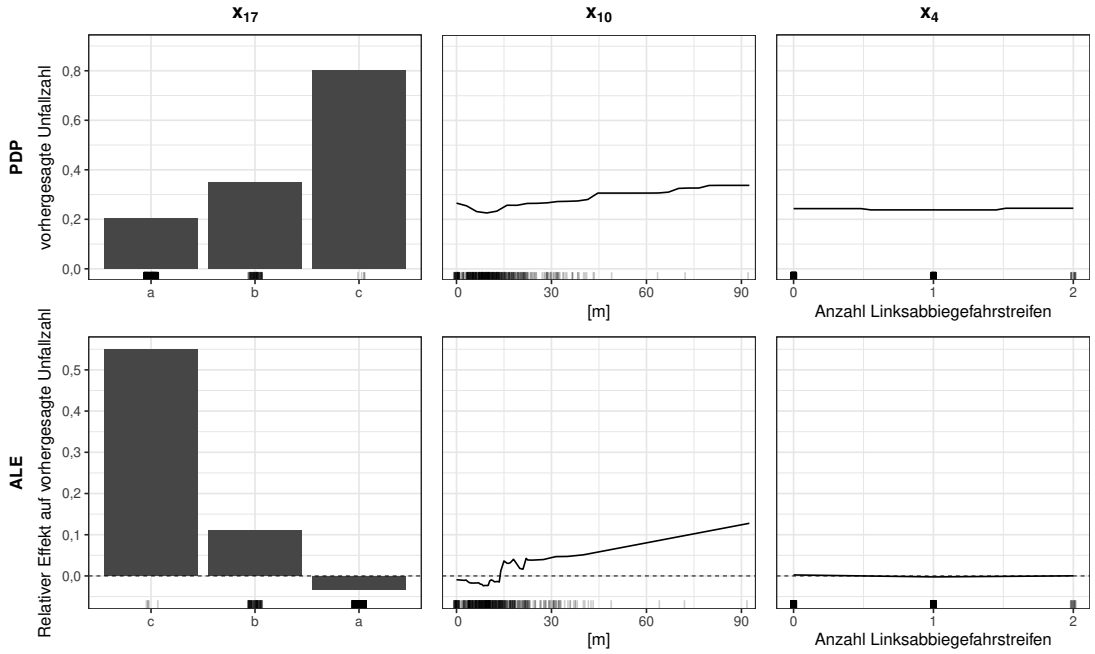


Abbildung 6.5: Unfalltyp 23/24: PDP und ALE bedeutsamer Infrastrukturgrößen [Quelle: eigene Darstellung]

Die zweitbedeutsamste Infrastrukturgröße des Unfalltyps 23/24 ist der Ausrundungsradius der Rechtsabbiegenden (x_{10}). Dabei prognostiziert sowohl der PDP als auch der ALE, dass zwischen etwa fünf und zehn Metern Ausrundungsradius die wenigsten Unfälle zu erwarten sind. Zudem zeigt der PDP ab einem Ausrundungsradius von etwa zehn Metern einen geringen Anstieg der erwarteten Unfallzahl. Da diese im gesamten Ausprägungsbereich von x_{10} lediglich gering zwischen etwa 0,22 und 0,32 Unfällen schwankt, kann der Einfluss des Ausrundungsradius für Rechtsabbiegende auf Basis des PDP als gering angesehen werden. Im Gegensatz dazu zeigt der ALE ab etwa zehn Metern einen etwas stärken Anstieg des relativen Effekts von x_{10} auf die erwartete Unfallzahl, wobei dieser einen kleinen Knick bei etwa 20 Metern aufweist. Der größte relative Effekt von etwa plus 0,13 Unfällen im Vergleich zu den im Mittel prognostizier-

ten Unfällen wird dabei bei dem maximalen Ausrundungsradius von etwa 90 Metern erreicht. Dennoch ist auch dieser Effekt vergleichsweise gering. Hinzu kommt außerdem, dass die Anzahl der Beobachtungen für Ausrundungsradien ab etwa 45 Metern abnimmt, wodurch die Schätzung der erwarteten Unfallzahlen bzw. des relativen Effekts auf die erwarteten Unfallzahlen in diesem Bereich ungenau sein kann. Generell zeigen also sowohl zu kleine als auch zu große Ausrundungsradien der Rechtsabbiegenden ein erhöhtes Unfallrisiko. Der Grund dafür kann einerseits darin liegen, dass die Rechtsabbiegenden bei einem sehr geringen Ausrundungsradius ihre Geschwindigkeit stärker reduzieren müssen, wodurch deren Konzentration vermehrt auf dem Verzögerungsvorgang liegt. Das wiederum führt möglicherweise dazu, dass aufholende Radfahrende in den Konfliktbereich einfahren ohne vom abbiegenden Kfz-Verkehr ausreichend wahrgenommen zu werden. Andererseits führen zu große Ausrundungsradien möglicherweise dazu, dass die Verkehrsteilnehmenden beim Rechtsabbiegen lediglich eine geringe Reduzierung ihres Fahrverhalten bzw. ihrer Geschwindigkeit vornehmen müssen, sodass deren Aufmerksamkeit für den Abbiegevorgang sowie die anderen Verkehrsteilnehmenden geringer ausfällt. Beides kann vermehrt zu Kollisionen führen. Weiterhin stellt die Anzahl der Linksabbiegefahrstreifen (x_4) das drittbedeutsamste Infrastrukturmerkmal dar. Dennoch weist der zugehörige PDP bzw. ALE keinen spürbaren Effekt von x_4 auf. Sichtbar wird dies bspw. anhand der horizontalen Linie im PDP. Diese besagt, dass das RFV für jede der Ausprägungen etwa 0,25 Unfälle vorhersagt. Im Falle des ALE, bei dem die horizontale Linie entlang der gestrichelten Linie verläuft, gilt zudem, dass die einzelnen Ausprägungen keine wesentlich höhere bzw. geringere Unfallzahl erwarten lassen als im Mittel vorhergesagt. Da der Unfalltyp 23/24 jedoch gemäß Abbildung 3.1 lediglich rechtsabbiegende sowie geradeaus fahrende Beteiligte umfasst, scheint es plausibel, dass die Anzahl der Linksabbiegefahrstreifen die betrachtete Unfallhäufigkeit nicht sichtbar determiniert. Da der Fokus der vorangegangenen Erläuterungen lediglich auf drei der einbezogenen Infrastrukturvariablen lag, sei an dieser Stelle auf die Abbildungen D.1 und D.2 in Anhang D verwiesen, welche die restlichen PDP's bzw. ALE darstellen. Dabei zeigen vor allem die ALE, dass die Einflüsse der Ausprägungen der übrigen Variablen auf die erwartete Anzahl der Unfälle vom Typ 23/24 vergleichsweise gering ausfallen.

Abbildung 6.6 zeigt den PDP sowie den zugehörigen ALE der Ausprägungen des DTV der rechtsabbiegenden Kfz (x_{20}), welcher die bedeutsamste DTV-Kenngröße des Unfalltyps 30/32/34 gemäß VI darstellt. Im Falle des PDP prognostiziert das RFV bei etwa 500 Kfz pro 24 Stunden die geringste Unfallzahl von etwa 0,35 Unfällen. Im weiteren Verlauf steigt diese bis zu einem DTV von etwa 4.000 Kfz pro 24 Stunden stark an und flacht anschließend bis zu einem DTV von etwa 8.000 Kfz pro 24 Stunden ab, wobei in diesem Punkt auch die maximale erwartete Unfallzahl von 1,2 Unfällen erreicht wird. Ein DTV von mehr als 8.000 Kfz pro 24 Stunden hat zudem konstant 1,2 erwartete Unfälle zur Folge. Der ALE der Variable x_{20} zeigt, dass bis zu einem DTV von etwa

1.100 Kfz pro 24 Stunden weniger Unfälle vom RFV prognostiziert werden als im Mittel vorhergesagt, wobei ein DTV zwischen null und etwa 500 Kfz pro 24 Stunden die erwartete Unfallzahl am meisten senkt. Diese beträgt dabei etwa 0,18 Unfälle weniger als im Mittel von x_{20} vorhergesagt wird. Somit deckt sich dieser DTV-Wert annähernd mit dem DTV des PDP, der die minimale Unfallzahl erwartet. Ab einem DTV von etwa 1.100 Kfz pro 24 Stunden steigt der relative Effekt auf die erwartete Unfallzahl stark an, während dieser ab einem DTV von etwa 4.000 Kfz pro 24 Stunden wieder abflacht. Auch dieser Wendepunkt deckt sich mit dem des PDP. Der größte relative Effekt wird bei dem maximalen DTV von etwa 13.500 Kfz pro 24 Stunden erzielt, wobei dieser 0,8 Unfälle mehr beträgt als im Mittel erwartet. Wie bereits bei der Variable x_{25} sind sich die Verläufe des PDP und des ALE auch bei dieser Variable ähnlich, sodass korrelierte Variablen für das Ergebnis ebenfalls unbedeutend erscheinen. Abermals nimmt die Anzahl der Beobachtungen für größere DTV-Werte ab, sodass die Schätzungen in diesem Fall ab einem DTV von etwa 5.000 Kfz pro 24 Stunden ungenau ausfallen können. Zudem zeigt sich in beiden Abbildungen ein positiv degressiver Verlauf ab etwa 500 Kfz pro 24 Stunden, wodurch sich beim Unfalltyp 30/32/34 für die Variable x_{20} ebenfalls der in Kapitel 2.1 vorgestellte *Safety in Numbers-Effekt* abzeichnet.

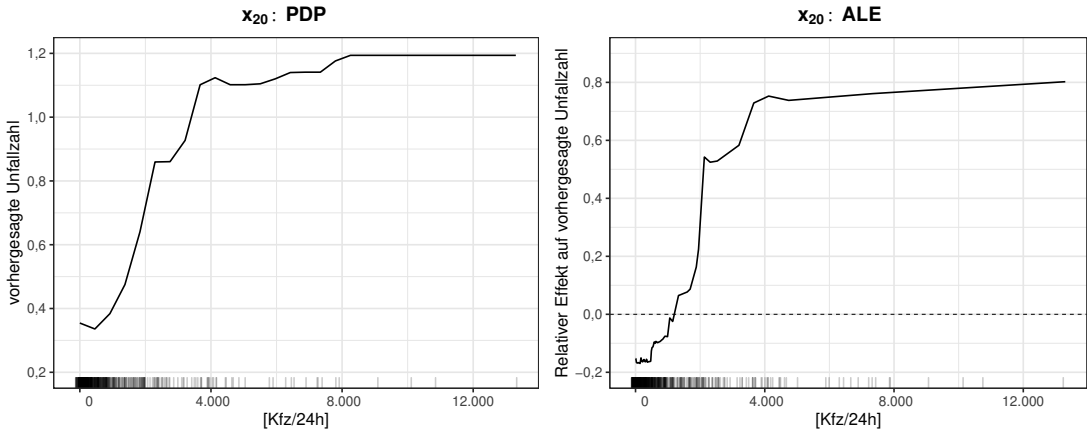


Abbildung 6.6: Unfalltyp 30/32/34: PDP und ALE der bedeutsamsten DTV-Größe [Quelle: eigene Darstellung]

Während sich dieser bei der zweitbedeutsamsten DTV-Kenngröße des entsprechenden Unfalltyps 30/32/34 DTV der geradeausfahrenden Kfz (x_{19}) gemäß Abbildung D.3 und D.4 in Anhang D nicht nachweisen lässt, wird der *Safety in Numbers-Effekt* anhand der eben genannten Abbildungen für die drittbedeutsamste DTV-Kenngröße des Unfalltyps 30/32/34 DTV des linksabbiegenden Radverkehrs (x_{24}) ebenfalls ersichtlich. Abbildung 6.7 stellt die PDP's sowie ALE der drei bedeutsamsten Infrastrukturmerkmale der Unfälle vom Typ 30/32/34 dar. Der PDP der Verkehrsregelungsart (x_2) zeigt, dass das RFV beim Verkehrszeichen Z205 (*Vorfahrt gewähren*) mit 0,6 Unfällen die höchste Unfallzahl prognostiziert. Gefolgt wird diese von einer Teil-LSA mit

etwa 0,48 erwarteten Unfällen. Im Gegensatz dazu weisen die Ausprägungen Rechts-vor-Links (R.-v.-L.), Verkehrszeichen Z206 sowie das Vorhandensein einer LSA eine ähnliche erwartete Unfallzahl von etwa 0,39 Unfällen auf. Auch bei Betrachtung des zugehörigen ALE ergibt sich für das Verkehrszeichen Z205 (*Vorfahrt gewähren*) der größte relative Effekt. Dieser beträgt 0,2 Unfälle mehr als im Mittel prognostiziert wird. Ein Grund für das erhöhte Unfallrisiko beim Unfalltyp 30/32/34 durch das Verkehrszeichen Z205 (*Vorfahrt gewähren*) liegt möglicherweise darin, dass beim Einbiegen in eine übergeordnete Straße keine Pflicht zum Anhalten besteht, wodurch kreuzende vorfahrtsberechtigte Verkehrsteilnehmende übersehen werden könnten. Weiterhin weist eine Teil-LSA einen positiven relativen Effekt von 0,05 Unfällen im Vergleich zur im Mittel erwarteten Unfallzahl auf. Im Gegensatz dazu kann die erwartete Unfallzahl durch eine Rechts-vor-Links-Regelung bzw. durch eine LSA relativ um etwa 0,07 bzw. 0,05 Unfälle gesenkt werden. Denkbare Gründe könnten dabei sein, dass die Aufmerksamkeit der Verkehrsteilnehmenden bei einer Rechts-vor-Links-Regelung erhöht ist und dass eine Rechts-vor-Links-Regelung üblicherweise Verkehrssituationen mit geringen Geschwindigkeiten bzw. Verkehrsstärken impliziert, wodurch es i.d.R zu weniger Unfällen kommt. Im Gegensatz dazu führt das Vorhandensein einer LSA zu einem verträglichen Abfließen der sich gemäß Unfalltyp 30/32/34 kreuzenden Verkehrsflüsse, wodurch das Unfallrisiko ebenfalls gesenkt wird. Das Verkehrszeichen Z206 hingegen weist keinen sichtbaren relativen Effekt auf die erwartete Unfallzahl auf, was bedeutet, dass diese Verkehrsregelungsart die Unfallzahl weder positiv noch negativ beeinflusst.

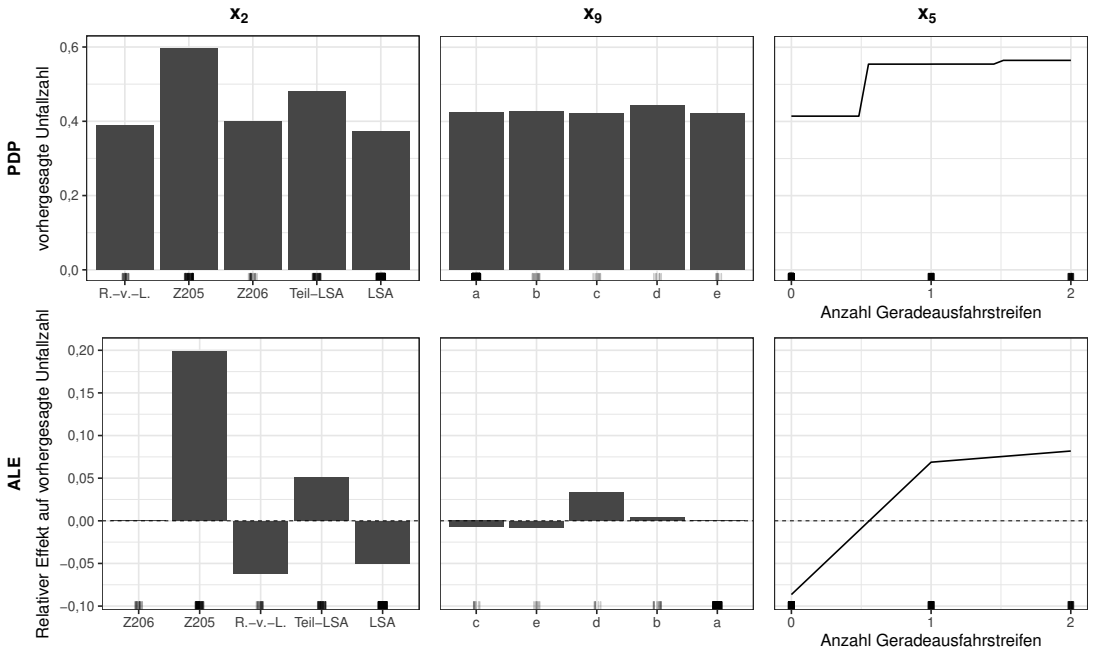


Abbildung 6.7: Unfalltyp 30/32/34: PDP und ALE bedeutsamer Infrastrukturgrößen [Quelle: eigene Darstellung]

Die zweitbedeutsamste Infrastrukturvariable des Unfalltyps 30/32/34 ist die Signalisierung der Rechstabbiegenden (x_9), deren Ausprägungen aus Platzgründen mit a bis e abgekürzt werden. Diese sind neben keiner separaten Signalisierung (a) eine separate dreifeldige Signalisierung (b), eine Dreiecksinsel ohne LSA (c), eine Dreiecksinsel mit LSA (d) sowie eine separate zweifeldige Signalisierung (e). Generell zeigt der zugehörige PDP keine allzu großen Unterschiede zwischen den erwarteten Unfallzahlen der einzelnen Ausprägungen. Demzufolge sagt das RFV bei jeder der Ausprägungen etwa 0,41 Unfälle voraus. Wird hingegen der zugehörige ALE betrachtet, so weist eine Dreiecksinsel mit LSA (d) einen vergleichsweise hohen relativen Effekt auf. Dieser beträgt etwa 0,03 Unfälle mehr als im Mittel prognostiziert werden. Im Vergleich dazu sind die relativen Effekte der restlichen Ausprägungen vernachlässigbar gering. Im Vergleich mit den relativen Effekten der zuvor beschriebenen Variable x_2 sind die selbigen von x_9 jedoch so gering, dass der Effekt dieser Variable im Gesamten als unbedeutend angesehen werden kann. Des Weiteren stellt die Anzahl der Geradeausfahrstreifen (x_5) das drittbedeutsamste Infratsrukturmerkmal der Unfälle vom Typ 30/32/34 dar. Der PDP dieser Variable sagt aus, dass ohne separate Geradeausfahrstreifen etwa 0,41 Unfälle prognostiziert werden, während die erwartete Anzahl der Unfälle bei einem bzw. zwei Geradeausfahrstreifen auf etwa 0,56 bzw. 0,57 Unfälle ansteigt. Ein ähnliches Ergebnis spiegelt sich beim zugehörigen ALE wider, denn dieser sagt einen positiven relativen Effekt auf die erwartete Unfallzahl voraus, wenn kein separater Geradeausfahrstreifen vorhanden ist. Demzufolge ist die erwartete Unfallzahl um etwa 0,08 Unfälle geringer als im Mittel prognostiziert. Im Gegensatz dazu steigert ein Geradeausfahrstreifen die erwartete Unfallzahl um etwa 0,07 Unfälle im Vergleich zur mittleren vorhergesagten Unfallzahl, während zwei Geradeausfahrstreifen diese um etwa 0,08 Unfälle im Vergleich zur mittleren vorhergesagten Unfallzahl erhöhen. Ein möglicher Grund für dieses Ergebnis könnte sein, dass das Nichtvorhandensein eines separaten Geradeausfahrstreifens - was die Existenz von Mischfahrstreifen impliziert - i.d.R. dazu führt, dass der Verkehr an der entsprechenden Zufahrt langsamer abfließt, wodurch die Geschwindigkeiten der Verkehrsteilnehmenden geringer ausfallen. Diese verringerten Geschwindigkeiten könnten wiederum zur Unfallvermeidung beitragen. Zudem könnte das Vorhandensein ein oder mehrerer separater Geradeausfahrstreifen die Knotenpunktkomplexität und damit auch das Unfallrisiko erhöhen. Die PDP's bzw. ALE der übrigen Infrastrukturmerkmale des Unfalltyps 30/32/34 befinden sich in den Abbildungen D.3 und D.4 in Anhang D. Anhand dessen lässt sich bspw. erkennen, dass eine zunehmende Anzahl der Mischfahrstreifen (x_7) die erwartete Unfallzahl ebenfalls positiv beeinflusst.
Zusammenfassend lässt sich sagen, dass sich die vorhergesagten Unfallzahlen bzw. die relativen Effekte auf dieselben bei beiden Unfalltypen in betragsmäßig kleinen Bereichen befinden, was der rechtsschiefen Verteilung der Unfallzahlen geschuldet ist. Der Effekt der betrachteten Merkmale beträgt dementsprechend in den meisten Fällen weniger als ein Unfall pro Zufahrt in fünf Jahren.

6.3 Negative Binomialregression

Die Schätzung der Negativen Binomialregression erfolgt ebenfalls getrennt für Unfälle vom Typ 23/24 mit der abhängigen Variable y_1 bzw. 30/32/34 mit der abhängigen Variable y_2 unter Verwendung der Software R, wobei im Vorfeld eine Variablenselektion mit Hilfe des LASSO-Verfahrens erfolgt. Da die negative Binomialverteilung Variablen mit Überdispersion beschreibt, erfolgt zunächst ein Test der abhängigen Variablen auf eine solche. Die dabei zugrunde gelegten Hypothesen lauten wie folgt.

$$H_0 : Var(y_k) = \mu_k \quad vs. \quad H_1 : Var(y_k) = \mu_k + \theta \cdot \mu_k^2$$

Dabei entspricht die Nullhypothese der Annahme der Equidispersion der Poissonverteilung, während die Alternativhypothese den Parameter θ enthält, der die Form der Dispersion beschreibt. Nimmt θ dabei einen Wert größer null an, liegt Überdispersion vor. Bei einem θ kleiner null hingegen liegt Unterdispersion vor, während θ gleich null eine Equidispersion beschreibt. Somit können die Hypothesen des Testproblems ebenfalls durch

$$H_0 : \theta = 0 \quad vs. \quad H_1 : \theta \neq 0$$

beschrieben werden. Um die Hypothesen zu testen, kommt eine unter H_0 asymptotisch normalverteilte t-Teststatistik zur Anwendung. (vgl. Cameron und Trivedi (1990), S. 350f.) Das Ergebnis zeigt sowohl für y_1 als auch für y_2, dass die Nullhypothese abgelehnt werden kann und sich ein θ größer null ergibt, was die Verwendung der Negativen Binomialregression für die vorliegenden Daten rechtfertigt.
Wie in Kapitel 5.3.1 unter dem Punkt *Poisson-, Gamma- und negative Binomialverteilung* beschrieben, ergibt sich die negative Binomialverteilung aus der Überlagerung von Poisson- und Gammaverteilung. Dabei beschreibt die Poissonverteilung diskrete, nicht-negative Zähldaten mit einer rechtsschiefen Verteilung, was - wie bereits mehrfach zuvor erwähnt - auf y_1 und y_2 zutrifft. Genauer gesagt beschreibt die Poissonverteilung die Abweichung der Anzahl der Unfälle an einer Zufahrt über den betrachteten Zeitraum, also die zeitliche Determinante des Unfallgeschehens. Die Gammaverteilung, die die Variabilität des Erwartungswertes abbildet, beschreibt hingegen die Abweichung der Anzahl der Unfälle zwischen den verschiedenen Zufahrten und bildet somit die räumliche Determinante des Unfallgeschehens ab. (vgl. Turner et al. (2011), S. 67)
Die Negative Binomialregression kommt in Form des in Kapitel 5.3.2 beschriebenen *Accident Prediction Models* zum Einsatz. Dabei bilden die Merkmale zum DTV des Kfz- sowie Radverkehrs (x_{18} bis x_{20}, x_{24} bis x_{26}) die Expositionsvariablen, während die restlichen Merkmale in Form von Risikovariablen in das jeweilige Modell einfließen.
Aufgrund der technischen Beschränkungen der Software R bzgl. des LASSO-Verfahrens werden die kategorialen unabhängigen Variablen in Form von Dummyvariablen einbezogen, wobei die Anzahl der Dummyvariablen jeweils der Anzahl der Ausprägungen der

entsprechenden Variable minus eins entspricht.[3] Die Ausprägung, welche dabei keine Dummyvariable zugewiesen bekommt, stellt die Referenzausprägung der betrachteten Variable dar. Der Effekt einer bestimmten Ausprägung wird dann in Bezug auf die Referenzausprägung interpretiert. Die einbezogenen Referenzausprägungen lauten dabei wie folgt.

x_1	Knotenpunktform:	Kreuzung
x_2	Verkehrsregelungsart:	Rechts-vor-Links
x_3	Grünpfeil (Z720):	nicht vorhanden
x_8	Signalisierung Linksabbiegende:	keine separate Signalisierung
x_9	Signalisierung Rechtsabbiegende:	keine separate Signalisierung
x_{11}	Abstand ruhender Verkehr zum Knotenpunkt:	mehr als 25 m
x_{12}	Abbiegeverbote:	keine
x_{13}	Radverkehrsführung im Knotenpunkt:	ohne Furt
x_{14}	Absetzung Radfahrfurt:	ohne Furt
x_{15}	Signalgeber Radverkehr:	mit MIV
x_{16}	Radverkehrsführung Linksabbiegende:	keine/Mischverkehr
x_{17}	Roteinfärbung Radfahrfurt:	nein

Weiterhin werden die Expositionsvariablen - gemäß des zu schätzenden Modells in Gleichung 5.3.8 - in logarithmierter Form in die Schätzungen einbezogen. Der Vorteil der Logarithmmierung liegt darin, dass für die entsprechenden Variablen eine symmetrischere Verteilung generiert wird. Außerdem wirkt sich die Logarithmierung zugunsten der Annahmen eines Regressionsmodells aus, da dadurch bspw. einerseits mehr Linearität zwischen den entsprechenden Expositionsvariablen und der jeweiligen abhängigen Variable erzielt und andererseits Heteroskedastizität vorgebeugt wird. (vgl. Dunn und Smyth (2018), S. 121) Um eine Logarithmierung von null zu vermeiden, werden dabei nicht die Ausprägungen selbst, sondern die mit eins addierten logarithmierten Ausprägungen einbezogen. Zudem wird die Risikovariable x_{10} in logarithmierter Form aufgenommen, da sich dadurch ebenfalls eine symmetrischere Verteilung ergibt. Da die Logarithmierung der übrigen metrisch skalierten Risikovariablen keine symmetrischere Verteilung zur Folge hat, werden diese jeweils in ihrer Ursprungsform einbezogen. Die so erlangten Verteilungen der Expositionsvariablen bzw. der Risikovariable x_{10} sind in den Abbildungen D.5 und D.6 in Anhang D dargestellt.

6.3.1 Variablenselektion

Da das Verfahren - anders als das RFV - kein zufälliges Vorgehen aufweist, ist im Vorfeld der Schätzungen eine Variablenauswahl nötig, um die Modellkomplexität zu verringern und somit Overfitting entgegenzuwirken, sodass letztendlich jeweils weniger als die 26 zur Verfügung stehenden unabhängigen Variablen einbezogen werden. Dabei

[3] Dieses Vorgehen wird sowohl bei der Schätzung des LASSO-Verfahrens als auch bei der darauffolgenden Schätzung der Negativen Binomialregression angewandt, um konsistente Ergebnisse zu erzielen.

kommt das in Kapitel 5.3.3 beschriebene LASSO-Verfahren zur Anwendung, wobei der optimale Regularisierungsparameter λ durch die Software *R* mit Hilfe von zehnfacher Kreuzvalidierung ermittelt wird. Dieses beläuft sich demnach auf 0,015 für die Unfälle vom Typ 23/24 und auf 0,046 für die Unfälle vom Typ 30/32/34. Die geschätzten Koeffizienten des LASSO-Verfahrens sind in Tabelle D.2 in Anhang D dargestellt, wobei ein Koeffizient von null bedeutet, dass die betroffenen Variablen aus der darauffolgenden Schätzung der Negativen Binomialregression ausgeschlossen werden. Dies betrifft für das Modell der Unfälle vom Typ 23/24 bspw. die Variablen x_1, x_8 sowie x_{10}. Für das Modell der Unfälle vom Typ 30/32/34 hingegen werden bspw. die Variablen x_8, x_{13} und x_{15} ausgeschlossen.
Die mittels LASSO-Verfahren für die jeweiligen Unfalltypen als relevant ermittelten Variablen werden im nächsten Schritt in die Schätzung der Negativen Binomialregression einbezogen, deren Erkenntnisse im Folgenden dargestellt sind.

6.3.2 Modellgütekriterien

Tabelle 6.2 stellt die Modellgütekriterien der Negativen Binomialregression, welche gemäß Kapitel 5.3.4 ermittelt werden, gegenüber. Dabei zeigt der LR-Test, dass die Nullhypothese in beiden Modellen bei einem zugrunde gelegten Signifikanzniveau von fünf Prozent verworfen werden kann, da der p-Wert jeweils kleiner als 0,05 ist. Somit kann davon ausgegangen werden, dass die vollständigen Modelle aussagekräftiger sind als die zugehörigen Nullmodelle. Dementsprechend weisen die unabhängigen Variablen sowohl bei dem Modell des Unfalltyps 23/24 als auch bei dem Modell des Unfalltyps 30/32/34 einen signifikanten Erklärungsanteil an der entsprechenden abhängigen Variable auf. Weiterhin kennzeichnet ein McF-R^2 von 0,28 eine gute Modellanpassung im Falle des Unfalltyps 23/24. Da McF-R^2 im Fall des Unfalltyps 30/32/34 0,14 beträgt, weist dieses Modell eine lediglich halb so gute Modellanpassung wie das Modell des Unfalltyps 23/24 auf.

Tabelle 6.2: Modellgüte nach Unfalltyp: Negative Binomialregression

	Unfalltyp 23/24	Unfalltyp 30/32/34
LR-Test	p<2,2e-16***	p=1,9e-13***
McF-R^2	0,28	0,14
MSE	0,52	1,47
RMSE	0,72	1,21

Signifikanzen des LR-Test werden wie folgt ausgewiesen:
*p<0,1; **p<0,05; ***p<0,01.

Der anhand von Kreuzvalidierung ermittelte MSE beträgt für Unfälle vom Typ 23/24 0,52 (Unfälle)^2, wodurch sich der zugehörige RMSE zu 0,72 Unfällen ergibt. Im Gegensatz dazu weist der Unfalltyp 30/32/34 einen MSE von 1,47 (Unfällen)^2 auf, weshalb

sich der RMSE entsprechend zu 1,21 Unfällen ergibt. Somit erzielt die Negative Binomialregression genauere Vorhersagen für die Testdaten des Unfalltyps 23/24 als für die Testdaten des Unfalltyps 30/32/34.

6.3.3 Variablenbedeutsamkeit

Tabelle 6.3 stellt die Ergebnisse der Negativen Binomialregression für beide geschätzten Modelle gegenüber. Neben den gemäß Gleichung 5.3.8 geschätzten Koeffizienten, deren Standardfehlern sowie Signifikanzen geben die Spalten *exp(Koeffizient)* zudem die eigentlichen Effekte der Risikovariablen auf die erwartete Unfallzahl an. Diese sagen aus, um welchen Faktor sich die erwartete Unfallzahl des entsprechenden Unfalltyps erhöht bzw. verringert, wenn sich die zugehörige metrische Risikovariable um eine Einheit erhöht. Im Gegensatz dazu gibt *exp(Koeffizient)* im Fall einer kategorialen Risikovariable die relative Veränderung der erwarteten Unfallzahl in Bezug auf deren Referenzausprägung an. Dabei spiegeln Werte größer eins eine erhöhte erwartete Unfallzahl wider, während Werte kleiner eins eine Verringerung dieser implizieren. Weiterhin stellt Tabelle 6.3 lediglich Variablen mit einem Signifikanzniveau von fünf Prozent ($p < 0,05$) dar, da dieses Signifikanzniveau üblicherweise zur Auswertung von Regressionskoeffizienten herangezogen wird. (vgl. Backhaus et al. (2008), S. 95) Die vollständige Regressionstabelle beider Unfalltypen befindet sich in Tabelle D.3 in Anhang D.

Tabelle 6.3: Ergebnisse nach Unfalltyp: Negative Binomialregression (Auszug)

abhängige Variable: Anzahl der Unfälle vom Unfalltyp ..			
	.. 23/24 (y_1)		.. 30/32/34 (y_2)	
	Koeffizient	exp(Koeffizient)	Koeffizient	exp(Koeffizient)
ln(β_0): Absolutglied				
	−12,244***	4,814e-06	−1,572***	0,208
	(1,934)		(0,522)	
ln(x_{19}+1): DTV Kfz Geradeausfahrende [Kfz/24h]				
	0,720***	-	−0,282***	-
	(0,215)		(0,047)	
ln(x_{20}+1): DTV Kfz Rechtsabbiegende [Kfz/24h]				
	0,305***	-	0,172**	-
	(0,084)		(0,070)	
ln(x_{24}+1): DTV Radverkehr Linksabbiegende [Räder/24h]				
			0,198***	-
			(0,072)	
ln(x_{25}+1): DTV Radverkehr Geradeausfahrende [Räder/24h]				
	0,420***	-		

	(0,144)			
x_1: Knotenpunktform				
Kreuzung			0	1
			(Referenz)	
Einmündung			$-1{,}223^{***}$	0,294
			(0,341)	
x_2: Verkehrsregelungsart				
R.-v.-L.			0	1
			(Referenz)	
Z205			$1{,}056^{***}$	2,875
			(0,327)	
Teil-LSA			$1{,}297^{***}$	3,658
			(0,379)	
x_3: Grünpfeil (Z720)				
nicht vorh.			0	1
			(Referenz)	
vorhanden			$0{,}772^{**}$	2,164
			(0,305)	
x_5: Anzahl Geradeausfahrstreifen				
			$1{,}047^{***}$	2,849
			(0,220)	
x_{13}: Radverkehrsführung im Knotenpunkt				
ohne Furt	0	1		
	(Referenz)			
Furt irrel.[1]	$2{,}103^{**}$	8,191		
	(0,899)			
x_{16}: Radverkehrsführung Linksabbiegende				
keine/Mischv.	0	1		
	(Referenz)			
ind.: geführt	$0{,}914^{**}$	2,494		
	(0,417)			
x_{17}: Roteinfärbung Radfahrfurt				
nein	0	1	0	1
	(Referenz)		(Referenz)	
ja, gesamt	$0{,}694^{**}$	2,002	$-0{,}963^{***}$	0,382
	(0,290)		(0,371)	
ja, teilweise	$3{,}123^{***}$	22,714		
	(0,634)			
Beobachtungen	484		484	
Dispersion $\hat{\theta}$	$1{,}731^{**}$		$0{,}509^{***}$	
	(0,879)		(0,124)	
Log Likelihood	$-190{,}369$		$-307{,}704$	

Anmerkung: Standardfehler entsprechen dem Klammerausdruck. Signifikanzen der Wald-Teststatistik werden wie folgt ausgewiesen: $^{*}p<0{,}1$; $^{**}p<0{,}05$; $^{***}p<0{,}01$.

[1]nur Radverkehr als Rechtsabbiegende

Das geschätzte Absolutglied beträgt 4,814e-06 für den Unfalltyp 23/24 und 0,208 für den Unfalltyp 30/32/34. Da das Absolutglied eines *Accident Prediction Models* jedoch keine inhaltliche Bedeutung aufweist, wird dieses im Folgenden nicht näher betrachtet. Für die erwartete Anzahl der Unfälle vom Unfalltyp 23/24 stellen der DTV der geradeausfahrenden Kfz (x_{19}), der DTV der rechtsabbiegenden Kfz (x_{20}) sowie der DTV des geradeausfahrenden Radverkehrs (x_{25}) die einzigen signifikanten Expositionsgrößen dar. Dieses Ergebnis erscheint plausibel, denn diese Bewegungsrichtungen treten typischerweise - gemäß der Beteiligtenkonstellationen in Abbildung 3.1 - bei Unfällen vom Typ 23/24 auf. Die Koeffizienten dieser Expositionsvariablen liegen dabei allesamt zwischen null und eins, wobei der DTV der geradeausfahrenden Kfz (x_{19}) mit 0,72 den größten Koeffizienten aufweist. Dieses Ergebnis ist ein Indiz für den bereits mehrfach erwähnten *Safety in Numbers-Effekt*, da Expositionskoeffizienten zwischen null und eins den für diesen Effekt typischen positiv degressiven Zusammenhang zwischen der entsprechenden Variable und der erwarteten Anzahl der Unfälle beschreiben.
Die Ausprägung *Furt irrelevant: nur Radverkehr als Rechtsabbiegende* der Variable Radverkehrsführung im Knotenpunkt (x_{13}) stellt zudem eine signifikante Risikovariable des Unfalltyps 23/24 dar. Demzufolge ist die erwartete Unfallzahl etwa 8,2 mal größer, wenn diese Furtirrelevanz anstelle gar keiner Furt im Knotenpunkt vorliegt. Ein möglicher Grund dafür könnte sein, dass der hinter dem Radverkehr geradeausfahrende Kfz-Verkehr den Abbiegewunsch und die damit einhergehende Verzögerung des Radverkehrs nicht rechtzeitig wahrnimmt, da das Abbiegen für den Kfz-Verkehr am entsprechenden Knotenpunkt nicht möglich ist. Das wiederum könnte bspw. ein erhöhtes Auffahrrisiko darstellen. Auch die indirekte Radverkehrsführung des linksabbiegenden Radverkehrs (x_{16}) stellt eine signifikante Risikovariable dar. Demzufolge ist die erwartete Anzahl der Unfälle vom Typ 23/24 bei einem indirekt geführten linksabbiegenden Radverkehr etwa 2,5 mal so groß im Vergleich mit einem im Mischverkehr geführten linksabbiegenden Radverkehr. Das Ergebnis scheint für den untersuchten Sachverhalt entgegen der Intuition, da Unfällen vom Typ 23/24 nicht durch linksabbiegende Verkehrsteilnehmende gekennzeichnet sind, sondern durch geradeausfahrende und rechtsabbiegende Beteiligte. Aufgrund dieses Sachverhalts kann an der Stelle kein eindeutiger Grund für dieses Ergebnis abgeleitet werden. Die Art der Roteinfärbung der Radfahrfurt (x_{17}) spielt als Risikovariable ebenfalls eine bedeutende Rolle bei der Beschreibung der erwarteten Unfallzahl. Demzufolge erwartet das Modell eine doppelt so hohe Unfallzahl, wenn anstelle keiner rot eingefärbten Radfahrfurt eine über den gesamten Knotenpunkt rot eingefärbte Radfahrfurt vorliegt. Im Gegensatz dazu ist die erwartete Unfallzahl bei einer teilweise rot eingefärbten Radfahrfurt sogar etwa 22,7 mal so groß im Vergleich zu Knotenpunkten ohne eine solche Einfärbung. Die Gründe für das erhöhte Unfallrisiko einer teilweise bzw. komplett rot über den Knotenpunkt eingefärbten Radfahrfurt könnten abermals darin liegen, dass einerseits bei einer teilweisen Einfärbung der relevante Teil der Furt nicht eingefärbt vorliegt, sodass der abbiegende

Kfz-Verkehr ankommende Radfahrende nicht ausreichend beachtet, und andererseits, dass das Verfahren bzw. die Daten die vorliegende Kausalität zwischen der Roteinfärbung der Furt und dem Unfallgeschehen vom Typ 23/24 nicht korrekt abbilden.
Mittels der Koeffizienten der eben genannten Einflussgrößen kann die rücktransformierte Schätzfunktion für die Anzahl der erwarteten Unfälle vom Typ 23/24 gemäß Gleichung 5.3.9 aufgestellt werden. Gleichung 6.3.1 zeigt die sich so ergebende Berechnungsvorschrift.

$$\begin{aligned} \mathrm{E}(\mathrm{y}_1) = \exp(-12,244) \cdot \mathrm{x}_{19}^{0,720} \cdot \mathrm{x}_{20}^{0,305} \cdot \mathrm{x}_{25}^{0,420} \cdot \exp(2,103 \cdot \mathrm{x}_{13}^{\text{Furt irrel.: 2}} \\ + 0,914 \cdot \mathrm{x}_{16}^{\text{ind.: gef.}} + 0,694 \cdot \mathrm{x}_{17}^{\text{gesamt}} + 3,123 \cdot \mathrm{x}_{17}^{\text{teilweise}}) \end{aligned} \tag{6.3.1}$$

Für die erwartete Anzahl der Unfälle vom Typ 30/32/34 stellen der DTV der geradeausfahrenden Kfz (x_{19}), der DTV der rechtsabbiegenden Kfz (x_{20}) sowie der DTV des linksabbiegenden Radverkehrs (x_{24}) die signifikanten Expositionsvariablen dar. Jedoch zeigt der Koeffizient des DTV der geradeausfahrenden Kfz (x_{19}) in Höhe von -0,28 einen negativen Zusammenhang mit der erwarteten Unfallzahl, was der Idee des in Kapitel 2.1 beschriebenen *Safety in Numbers-Effekt* wiederspricht. Im Gegensatz dazu weisen der DTV der rechtsabbiegenden Kfz (x_{20}) und der DTV des linksabbiegenden Radverkehrs (x_{24}) Koeffizienten zwischen null und eins auf, wenngleich diese betragsmäßig gering ausfallen. Dennoch lässt sich bei diesen Größen eine Tendenz zum *Safety in Numbers-Effekt* erkennen. Die durch die signifikanten Expositionsvariablen beschriebenen Bewegungsrichtungen scheinen für den Unfalltyp 30/32/34 plausibel, da dieser - wie in Abbildung 3.2 dargestellt - verschiedenste Bewegungsrichtungen abdeckt.
Die Knotenpunktform (x_1) determiniert als Risikovariable ebenfalls die Anzahl der erwarteten Unfälle vom Typ 30/32/34. Demzufolge ist die erwartete Unfallzahl lediglich etwa 0,3 mal so hoch, wenn es sich bei dem entsprechenden Knotenpunkt um eine Einmündung anstelle einer Kreuzung handelt. Ein möglicher Grund dafür könnte sein, dass sich die Verkehrsteilnehmenden bei einer Einmündung auf drei Zufahrten konzentrieren müssen, während sich die Zahl der zu überblickenden Zufahrten bei einer Kreuzung auf vier beläuft. Demzufolge könnte ein komplexerer Knotenpunkt ein erhöhtes Unfallrisiko darstellen. Das Verkehrszeichen Z205 (*Vorfahrt gewähren*) sowie das Vorhandensein einer Teil-LSA determinieren die erwartete Unfallzahl in Form der Risikovariable Verkehrsregelungsart (x_2) ebenfalls. Demzufolge ist die erwartete Anzahl der Unfälle etwa 2,9 bzw. 3,7 mal so hoch im Vergleich mit einer Rechts-vor-Links-Regelung. Wie schon beim RFV erwähnt, kann bspw. das erhöhte Unfallrisiko beim Verkehrszeichen Z205 (*Vorfahrt gewähren*) daher kommen, dass vor dem Einbiegen in eine übergeordnete Straße keine Pflicht zum Halten besteht, was möglicherweise zum Übersehen kreuzender vorfahrtsberechtigter Verkehrsteilnehmender führt. Neben der Verkehrsregelungsart beeinflusst das Vorhandensein eines Grünpfeils (x_3) die erwartete Unfallzahl ebenfalls in Form einer Risikovariable. Somit lassen sich etwa 2,2 mal

so viele Unfälle erwarten, wenn ein Grünpfeil vorhanden ist. Der Grund dafür liegt möglicherweise in der psychologischen Wirkung eines Grünpfeils. Dieser animiert womöglich zum Einbiegen ohne ausreichend auf etwaigen kreuzenden Verkehr zu achten, was wiederum zu einer erhöhten Unfallzahl führen kann. Des Weiteren führt ein zusätzlicher Geradeausfahrstreifen (x_5) als Risikovariable zu einer um etwa den Faktor 2,8 erhöhten erwarteten Unfallzahl. Demzufolge impliziert ein komplexerer Knotenpunkt ein erhöhtes Unfallrisiko für Unfälle vom Typ 30/32/34. Eine über den gesamten Knotenpunkt rot eingefärbte Radfahrfurt (x_{17}) stellt eine weitere Risikovariable des Unfalltyps 30/32/34 dar. Liegt eine solche Einfärbung vor, so ist die erwartete Unfallzahl lediglich etwa 0,4 mal so groß als ohne eine Roteinfärbung. Der Grund dafür könnte in der Sichtbarkeit der Radfahrfurt liegen, wodurch der Kfz-Verkehr auf das Vorhandensein des Radverkehrs hingewiesen wird. Das wiederum führt dazu, dass der Kfz-Verkehr verstärkt auf den sich nähernden Radverkehr achtet, wodurch potenzielle Konfliktsituationen frühzeitig erkannt und vermieden werden können.
Die sich analog zum Unfalltyp 23/24 ergebende rücktransformierte Schätzfunktion für die erwartete Anzahl der Unfälle vom Typ 30/32/34 ist in Gleichung 6.3.2 dargestellt.

$$\begin{aligned} \mathrm{E}(\mathrm{y}_2) = \exp(-1,572) \cdot \mathrm{x}_{19}^{-0,282} \cdot \mathrm{x}_{20}^{0,172} \cdot \mathrm{x}_{24}^{0,198} \cdot \exp(-1,223 \cdot \mathrm{x}_1^{\text{Einmündung}} \\ + 1,056 \cdot \mathrm{x}_2^{\text{Z205}} + 1,297 \cdot \mathrm{x}_2^{\text{Teil-LSA}} + 0,772 \cdot \mathrm{x}_3^{\text{vorhanden}} + 1,047 \cdot \mathrm{x}_5 \\ - 0,963 \cdot \mathrm{x}_{17}^{\text{gesamt}}) \end{aligned} \tag{6.3.2}$$

Neben den Koeffizienten schätzt die Negative Binomialregression zudem $\hat{\theta}$, anhand dessen erkennbar ist, wie stark die jeweilige Überdispersion ausfällt. Wie in Kapitel 5.3.1 unter dem Punkt *Poisson-, Gamma- und negative Binomialverteilung* beschrieben, deutet ein sehr großes $\hat{\theta}$ darauf hin, dass die negative Binomialverteilung annähernd einer Poissonverteilung folgt. Im Umkehrschluss bedeutet dies, dass kleinere $\hat{\theta}$ eine größere Überdispersion anzeigen und demzufolge die These der negativen Binomialverteilung stützen. Da die geschätzten $\hat{\theta}$ mit 1,731 bzw. 0,509 relativ kleine Werte aufweisen, kann vom Vorliegen einer Überdispersion ausgegangen werden, was abermals für die Verwendung der Negativen Binomialregression spricht.

6.3.4 Modelldiagnostik

Wie in Kapitel 5.3.6 beschrieben, besteht eine Annahme linearer Modelle in der korrekten Spezifizierung des gewählten Modells. Dazu sei gesagt, dass bei der Schätzung ökonometrischer Modelle aus Gründen der Erfassbarkeit o.Ä. generell nie alle relevanten erklärenden Variablen herangezogen werden können. Hinzu kommt, dass die Nutzung des LASSO-Verfahrens zur Variablenselektion möglicherweise zu einem Ausschluss relevanter Merkmale aus den Schätzungen führte. Demzufolge ist eine Korrelation zwischen den unabhängigen Variablen und den Residuen und damit eine Verzerrung der betroffe-

nen Koeffizienten denkbar. Abgesehen davon kann von einem wohlspezifizierten Modell ausgegangen werden, da in beiden Modellen bspw. die Zahl der zu schätzenden Koeffizienten kleiner als die Zahl der Beobachtungen ausfällt. Zudem erfolgte im Vorfeld der Schätzungen eine Überprüfung der linearen Beziehung zwischen den mittels der logarithmischen Linkfunktion transformierten abhängigen Variablen und den metrisch skalierten unabhängigen Variablen. Dabei wurde sowohl bei der Logarithmierung der Expositionsvariablen als auch bei derselben der Variable x_{10} ersichtlich, dass die Nutzung des Logarithmus - im Vergleich zu deren Einbeziehung in ursprünglicher Form - auf einen stärkeren linearen Zusammenhang hindeutet.
Um die übrigen Aspekte der Modelldiagnostik zu beleuchten, zeigt Abbildung 6.8 die Diagnostikplots beider untersuchten Unfalltypen, wobei die in Kapitel 5.3.6 eingeführten standardisierten Quantilresiduen zum Einsatz kommen.

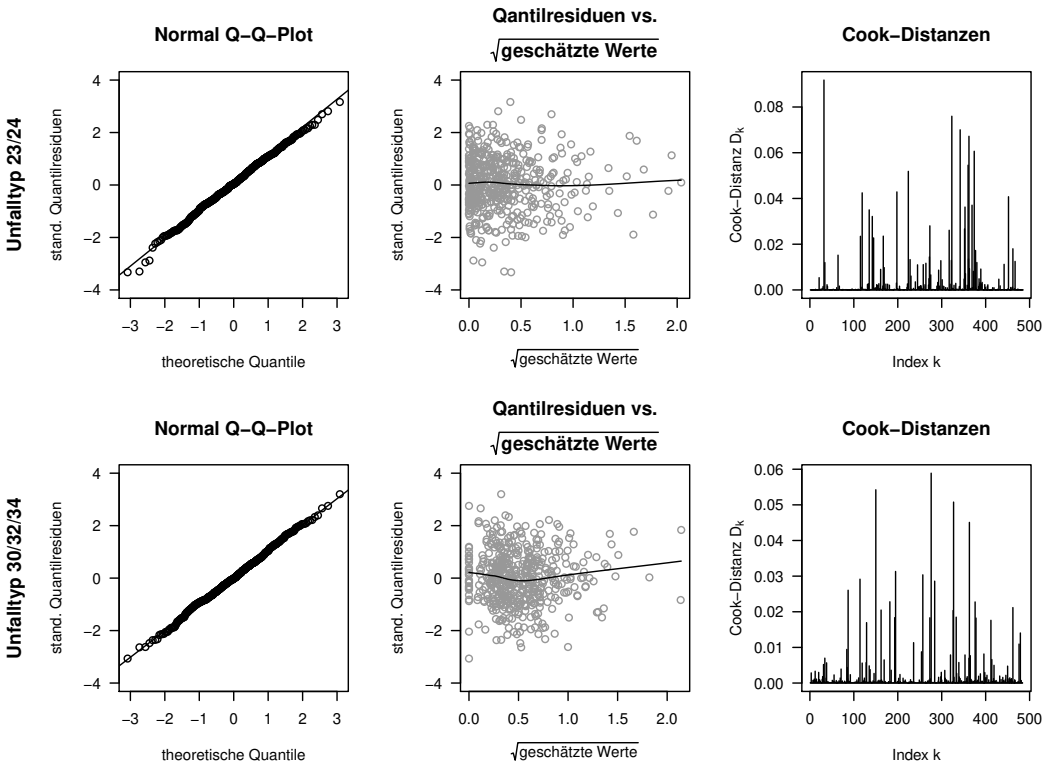


Abbildung 6.8: Modelldiagnostikplots der Negativen Binomialregression [Quelle: eigene Darstellung]

Der Normal Q-Q-Plot zeigt dabei, dass die Residuen weitestgehend auf einer Geraden liegen, wobei lediglich sehr kleine bzw. sehr große Residuen in geringem Maße von dieser abweichen. Demzufolge kann alles in Allem davon ausgegangen werden, dass die Residuen einer Normalverteilung mit einem Erwartungswert von null folgen. Weiterhin lässt sich daraus ableiten, dass die betrachteten abhängigen Variablen der gewählten

negativen Binomialverteilung folgen, was ebenfalls auf eine korrekte Spezifizierung der Modelle hindeutet.
Die Annahme der Homoskedastizität kann anhand des Plots der Residuen gegen die geschätzten Werte überprüft werden. Dabei kommt eine varianz-stabilisierende Transformation der geschätzten Werte in Form der Quadratwurzel zum Einsatz, welche zu einer gleichmäßigeren horizontalen Streuung der Residuen und damit zu einer besseren Erkennbarkeit von Trends führt. (vgl. Dunn und Smyth (2018), S. 307f.) Im Vergleich dazu befinden sich die entsprechenden Plots der Residuen gegen die geschätzten Werte ohne varianz-stabilisierende Transformation in Abbildung D.7 in Anhang D. Bei Betrachtung der hier vorliegenden Plots ist bei beiden untersuchten Unfalltypen vor allem bei größeren geschätzten Werten ein Trend erkennbar, was ein Indiz einer nicht-konstanten Varianz der Residuen darstellt. Demzufolge nimmt die Varianz in beiden Fällen mit zunehmender geschätzter Unfallzahl ab. Diese Heteroskedastizität führt unter Umständen zu ineffizienten Schätzungen sowie verzerrten Standardfehlern. Um dem Problem Abhilfe zu schaffen, könnte die Nutzung einer anderen Transformation der jeweiligen abhängigen Variable - anstelle der logarithmierten Transformation gemäß Negativer Binomialregression - oder die Nutzung einer anderen Erwartungswert-Varianz-Beziehung zur Modellierung der Überdispersion in Betracht gezogen werden. Nichtsdestotrotz kann die vorliegende Heteroskedastizität aufgrund der rechtsschiefen Verteilung der abhängigen Variablen sowie der Tatsache, dass die Varianz der Residuen lediglich für größere geschätzte Werte inhomogen ist, als nicht allzu problematisch angesehen werden.
Obwohl der Plot der Residuen gegen die geschätzten Werte keine Hinweise auf ein mögliches Autokorrelationsproblem liefert, ist es aufgrund der Natur der vorliegenden Daten möglich, dass sowohl räumliche als auch zeitliche Autokorrelation vorliegt, da die Unfallzahlen jeweils über einen Zeitraum von fünf Jahren an mehreren Zufahrten erhoben wurden. Demzufolge ist das Vorliegen verzerrter Standardfehler aufgrund dieser Tatsache ebenfalls nicht ausgeschlossen.
Des Weiteren erfolgte die Prüfung des Vorliegens perfekter Multikollinearität zwischen den unabhängigen Variablen anhand der Höhe der Korrelationen zwischen den im jeweiligen Modell einbezogenen Dummyvariablen. Um die entsprechenden Korrelationskoeffizienten nach Pearson zu bestimmen, wurden die null-eins-Kodierungen der Dummyvariablen als metrisch skaliert behandelt. Für das Vorliegen perfekter Multikollinearität besagt eine Daumenregel, dass Korrelationen, die betragsmäßig größer als 0,8 sind, ein Problem darstellen. (vgl. Backhaus et al. (2008), S. 89) Da dies jedoch in beiden Modellen nicht der Fall ist, kann von der Abwesenheit perfekter Multikollinearität ausgegangen werden.
Die Plots der Cook-Distanzen zeigen zudem, dass alle Cook-Distanzen weitaus kleiner als eins sind. Daher gibt es sowohl im Modell des Unfalltyps 23/24 als auch in dem des Unfalltyps 30/32/34 keine Beobachtungen, die einen substanziell großen Einfluss auf

die geschätzten Koeffizienten ausüben und deren Ausschluss zu erheblich veränderten Ergebnissen führen würde.

6.4 Vergleich beider Verfahren

Die folgenden Abschnitte vergleichen die Erkenntnisse des zufälligen Blackbox-RFV und der parametrischen Negativen Binomialregression hinsichtlich der untersuchten Unfalltypen. Ein wichtiger Unterschied in der Durchführung bestand dabei darin, dass beim RFV aufgrund des zufälligen Vorgehens jeweils alle 26 zur Verfügung stehenden unabhängigen Variablen einbezogen werden konnten, während im Vorfeld der Negativen Binomialregression eine Variablenselektion in Form des LASSO-Verfahrens stattfand. Zudem erforderte die Negative Binomialregression die logarithmische Transformation einiger Variablen. Neben einem Vergleich der Modellgütekriterien stellt das vorliegende Kapitel zudem die bedeutsamsten Einflussfaktoren beider Verfahren und Unfalltypen gegenüber, anhand derer letztendlich Empfehlungen für die Praxis abgeleitet werden.

6.4.1 Modellgütekriterien

Tabelle 6.4 zeigt die Modellgütekriterien beider Verfahren in Form des MSE sowie RMSE auf.

Tabelle 6.4: Modellgütevergleich beider Verfahren nach Unfalltyp

Unfalltyp ..	.. 23/24		.. 30/32/34	
	MSE	RMSE	MSE	RMSE
Random Forest	0,53	0,73	1,89	1,37
Negative Binomialregression	0,52	0,72	1,47	1,21

Dabei wird ersichtlich, dass die OOB-Beobachtungen des RFV beim Modell des Unfalltyps 23/24 annähernd genauso gute Modellvorhersagen generieren wie die Testdaten der Negativen Binomialregression. Beim Modell des Unfalltyps 30/32/34 hingegen zeigt sich eine leichte Überlegenheit der Negativen Binomialregression gegenüber des RFV. Des Weiteren liefern beide Verfahren beim Unfalltyp 23/24 bessere Modellvorhersagen als beim Unfalltyp 30/32/34, was im Fall der Negativen Binomialregression ebenfalls anhand der Größe von McF-R^2 in Tabelle 6.2 deutlich wird. Demzufolge deuten der MSE bzw. RMSE darauf hin, dass die Modelle der Negativen Binomialregression bei den vorliegenden Daten etwas genauere Vorhersagen aufweisen als die des RFV. Obwohl die Negative Binomialregression eine leichte Überlegenheit gegenüber dem RFV aufweist, sind die Fehler der Verfahren bei beiden Unfalltypen betragsmäßig gering. Der Grund dafür liegt in der rechtsschiefen Verteilung beider abhängiger Variablen, die bei den meisten Beobachtungen den Wert null annehmen. Demzufolge sagen die Verfahren

für die meisten Beobachtungen relativ geringe Unfallzahlen vorher, wodurch die mittlere (quadrierte) Abweichung zwischen vorhergesagten und realen Unfallzahlen - also der MSE bzw. RMSE - ebenfalls gering ausfällt. Um dies zu verdeutlichen, stellt Abbildung 6.9 die realen und mittels den jeweiligen Testdaten generierten Unfallzahlen der betrachteten Unfalltypen beider Verfahren grafisch gegenüber, wobei ein Punkt einer der 484 Beobachtungen bzw. Zufahrten entspricht. Würden diese entlang der jeweiligen Diagonale verlaufen, so würden die Verfahren exakte Vorhersagen der Testdaten liefern.

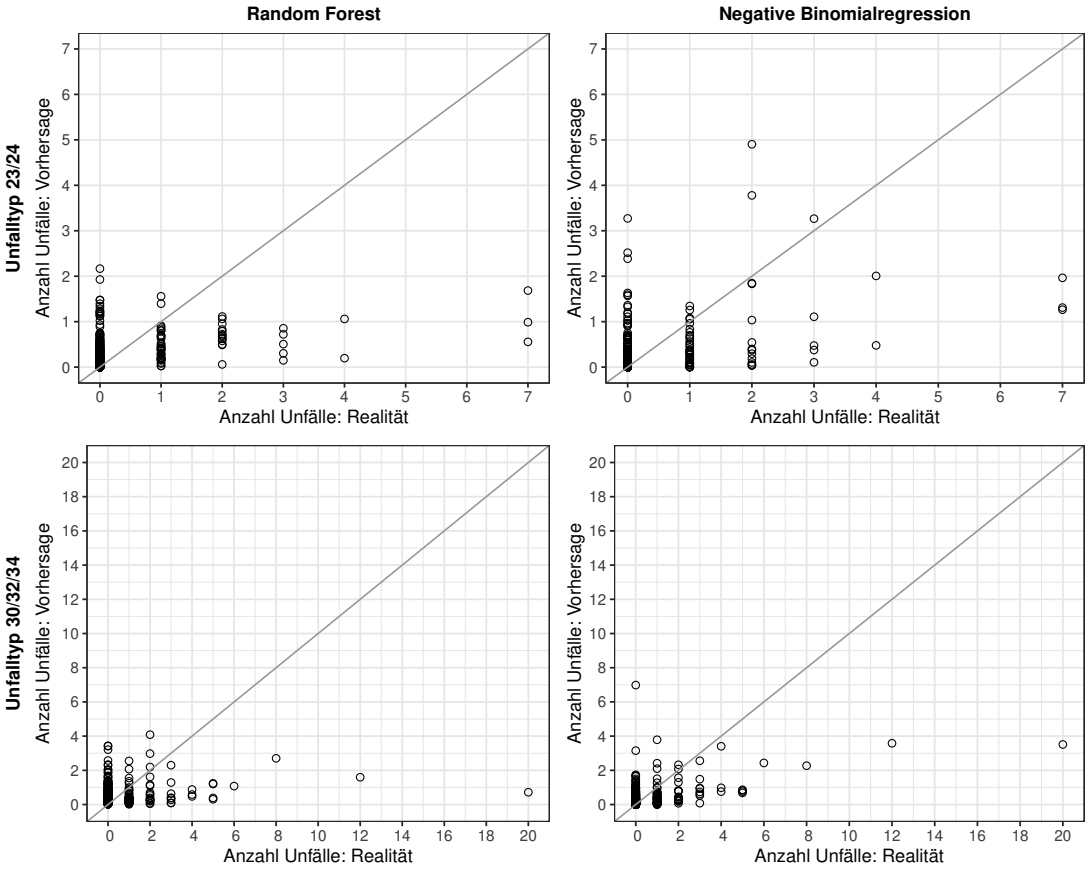


Abbildung 6.9: Gegenüberstellung realer und vorhergesagter Unfallzahl beider Verfahren nach Unfalltyp [Quelle: eigene Darstellung]

Dabei fällt bspw. beim Unfalltyp 23/24 auf, dass das RFV tendenziell geringere Vorhersagen generiert, während die Vorhersagen der Negativen Binomialregression z.T. größere Ausprägungen aufweisen. Während das RFV dabei eine maximale Unfallzahl von etwa 2,2 Unfällen generiert, beträgt das Maximum der Negativen Binomialregression etwa 4,9 Unfälle. Dennoch unterscheidet sich die mittlere (quadrierte) Abweichung zwischen realen und vorhergesagten Unfällen gemäß MSE bzw. RMSE nicht wesentlich zwischen beiden Verfahren. Ein ähnliches Muster ist auch beim Unfalltyp 30/32/34 erkennbar, denn auch hier weisen die Vorhersagen der Negativen Binomialregression z.T. etwas größere Werte auf als die des RFV. Während die maximal vorhergesagte

Unfallzahl des RFV etwa 4,1 Unfälle beträgt, so liegt diese bei der Negativen Binomialregression bei etwa sieben Unfällen. Beim Unfalltyp 30/32/34 kommen die größeren Ausprägungen der generierten Vorhersagen der Negativen Binomialregression zugute, denn dadurch erzielt das Verfahren einen geringeren MSE bzw. RMSE als das RFV.
Generell lässt sich über beide Verfahren sagen, dass diese bei beiden Unfalltypen Probleme bei der Vorhersage größerer Unfallzahlen aufweisen. Der Grund dafür könnte abermals in der rechtsschiefen Verteilung der abhängigen Variablen liegen, sodass große Ausprägungen zu wenige Beobachtungen aufweisen, um zuverlässige Vorhersagen generieren zu können. Dies wird außerdem durch die Tatsache gestützt, dass größere Ausprägungen der abhängigen Variablen nur lückenhaft auftreten. So gibt es bspw. beim Unfalltyp 30/32/34 drei Ausreißerzufahrten gemäß Abbildung 4.14, die acht, zwölf bzw. 20 Unfälle aufweisen, während für alle diskreten Ausprägungen dazwischen keine Beobachtungen vorliegen. Anhand Abbildung 6.9 wird zudem deutlich, warum die Verfahren beim Unfalltyp 30/32/34 eine schlechtere Modellgüte aufweisen als beim Unfalltyp 23/24. Der Grund dafür liegt in der maximalen Ausprägung der abhängigen Variablen beider Unfalltypen. Dabei ist diese beim Unfalltyp 30/32/34 mit 20 Unfällen weitaus höher als beim Unfalltyp 23/24 mit sieben Unfällen. Dementsprechend wirkt sich die Diskrepanz der vorhergesagten und realen Unfallzahlen beim Unfalltyp 30/32/34 stärker auf den MSE bzw. RMSE aus als beim Unfalltyp 23/24.

6.4.2 Variablenbedeutsamkeit und Handlungsempfehlungen

Erwartungsgemäß stellen sowohl bei beiden Verfahren als auch bei beiden Unfalltypen die DTV-Kenngrößen die wichtigsten Determinanten der erwarteten Unfallzahlen dar. Für die erwartete Anzahl der Unfälle vom Unfalltyp 23/24 sind dabei jeweils der DTV der geradeausfahrenden Kfz (x_{19}), der DTV der rechtsabbiegenden Kfz (x_{20}) sowie der DTV des geradeausfahrenden Radverkehrs (x_{25}) die bedeutsamsten DTV-Kenngrößen. Anhand der zugehörigen ALE (vgl. Abbildung 6.4 und Abbildung D.2 in Anhang D) bzw. Koeffizienten (vgl. Tabelle 6.3) zeigt sich zudem für all diese Variablen der *Safety in Numbers-Effekt*. Beim Modell des Unfalltyps 30/32/34 hingegen erweisen sich bei beiden Verfahren der DTV der geradeausfahrenden Kfz (x_{19}), der DTV der rechtsabbiegenden Kfz (x_{20}) sowie der DTV des linksabbiegenden Radverkehrs (x_{24}) als bedeutsamste DTV-Kenngrößen. Anhand der zugehörigen ALE (vgl. Abbildung 6.6 und Abbildung D.4 in Anhang D) bzw. Koeffizienten (vgl. Tabelle 6.3) kann der *Safety in Numbers-Effekt* in beiden Verfahren jedoch lediglich für x_{20} und x_{24} nachgewiesen werden. Zusammenfassend lässt sich also sagen, dass beide Verfahren sowohl beim Unfalltyp 23/24 als auch beim Unfalltyp 30/32/34 die gleichen Variablen als bedeutsamste DTV-Kenngrößen identifizieren. Hinzu kommt, dass beide Verfahren bei beiden Unfalltypen die gleichen Schlüsse bzgl. des Auftretens des *Safety in Numbers-Effekt* zulassen. Da der Fokus der Arbeit in der Untersuchung der Infrastruktureigenschaften auf das Unfallgeschehen liegt, stellt Tabelle 6.5 zudem die jeweils fünf bedeutsamsten Variablen

der Infrastruktur beider Verfahren gegenüber. Die entsprechenden Variablen werden dabei absteigend nach ihrer Bedeutsamkeit sortiert, wobei sich die Bedeutsamkeit im RFV nach der Höhe der zugehörigen VI gemäß Abbildung 6.3 richtet. Bei kategorialen Variablen ist dabei jeweils die bedeutsamste Ausprägung gemäß ALE (vgl. Abbildung 6.5, 6.7 sowie D.2, D.4 in Anhang D) angeführt. Im Fall der Negativen Binomialregression sind die fünf bedeutsamsten Variablen absteigend nach der betragsmäßigen Größe der zugehörigen Koeffizienten in Tabelle 6.3 sortiert.[4] Bei kategorialen Variablen ist dabei deren bedeutsamste Ausprägung sowie die zugehörige Referenzausprägung angegeben. Um die beiden Verfahren besser vergleichen zu können, werden die Effekte des RFV in Form des ALE der entsprechenden kategorialen Ausprägung im Vergleich mit dem ALE der gemäß Negativer Binomialregression zugehörigen Referenzausprägung interpretiert, sofern diese Ausprägung bei der Negativen Binomialregression ebenfalls einen relevanten Effekt aufweist. Zudem verdeutlichen die Pfeile die Wirkungsrichtung der Variablen auf die erwartete Unfallzahl. Demnach beschreibt ein Pfeil nach oben eine Erhöhung der erwarteten Unfallzahl und ein Pfeil nach unten eine Verringerung derselben. Ein horizontaler Pfeil hingegen deutet auf einen nicht erkennbaren Trend hin. Bei metrisch skalierten Variablen bezieht sich die Wirkungsrichtung zudem auf einen zunehmenden Wert der entsprechenden Variable.

Für die Modelle des Unfalltyps 23/24 zeigt sich bei beiden Verfahren, dass die teilweise Roteinfärbung der zur entsprechenden Zufahrt gehörenden Radfahrfurt (x_{17}) das bedeutsamste Infrastrukturmerkmal darstellt, wobei dieses im jeweiligen Modell zu einer Erhöhung der erwarteten Unfallzahl im Vergleich mit keiner Roteinfärbung führt. Da jedoch - wie in Kapitel 6.2.2 beschrieben - die Möglichkeit besteht, dass das Unfallgeschehen die Einfärbung bedingte und nicht umgekehrt, lässt sich die letztendliche Wirkung einer teilweisen Roteinfärbung der Radfahrfurt auf das Unfallgeschehen nicht eindeutig ableiten. Daher kann an dieser Stelle keine Handlungsempfehlung bzgl. der erwarteten Unfälle vom Typ 23/24 ausgesprochen werden. Ein weiteres in beiden Verfahren bedeutsames Merkmal des Unfalltyps 23/24 ist die indirekte Führung des linksabbiegenden Radverkehrs (x_{16}), wobei eine solche Führung die erwartete Unfallzahl beider Verfahren erhöht im Vergleich mit dessen Führung im Mischverkehr. Da dieses Ergebnis für das Entstehen der Unfälle vom Typ 23/24 sachlogisch irrelevant erscheint, kann an dieser Stelle ebenfalls keine Handlungsempfehlung hinsichtlich der Reduktion des Unfallgeschehens vom Typ 23/24 abgeleitet werden. Gemäß den Ergebnissen des RFV führt zudem ein größerer Ausrundungsradius des rechtsabbigenden Verkehrs (x_{10}) zu einer höheren erwarteten Zahl der Unfälle vom Typ 23/24, weshalb in der Praxis auf große Ausrundungsradien verzichtet werden sollte. Weiterhin ermittelt das RFV die Anzahl der Linksabbiegefahrstreifen (x_4) und die Anzahl der Rechtsabbie-

[4] Eine Sortierung anhand von *exp(Koeffizient)* ist aufgrund des Wertebereichs der e-Funktion nicht sinnvoll [e $\in (0;\infty)$]. Demnach zeigen Werte im Bereich $(0;1)$ eine Verringerung der erwarteten Unfallzahl an, während Werte im Bereich $(1;\infty)$ eine Erhöhung dieser implizieren. Aufgrund der unterschiedlichen Größe der Intervalle wird die Ordnung der Effekte nach betragsmäßiger Größe unmöglich.

gefahrstreifen (x_6) als Merkmale großer Bedeutsamkeit. Da anhand deren ALE jedoch kein Trend der Wirkung auf die erwartete Unfallzahl erkennbar ist, kann an dieser Stelle keine Empfehlung abgeleitet werden. Zudem stellt die Radverkehrsführung im Knotenpunkt (x_{13}) die drittbedeutsamste und damit die letzte der signifikanten Infrastrukturvariablen der Negativen Binomialregression dar. Anhand der aufgeführten bedeutsamen Ausprägung dieser Variable kann jedoch ebenfalls kein Handlungshinweis gegeben werden, da dadurch keine Aussage darüber möglich ist, ob das Vorhandensein einer Furt die erwartete Unfallzahl des Unfalltyps 23/24 erhöht oder senkt.

Tabelle 6.5: Vergleich bedeutsamer infrastruktureller Variablen beider Verfahren nach Unfalltyp

Random Forest		Negative Binomialregression	
Unfalltyp 23/24			
x_{17} (ja, teilweise)	↑	x_{17} (ja, teilweise / Ref.: nein)	↑
x_{10}	↑	x_{13} (Furt irrelevant[1] / Ref.: ohne Furt)	↑
x_4	↔	x_{16} (indirekt geführt / Ref.: keine/Mischverkehr)	↑
x_{16} (indirekt geführt)	↑	-	
x_6	↔	-	
Unfalltyp 30/32/34			
x_2 (Z205)	↑	x_2 (Teil-LSA / Ref.: Rechts-vor-Links)	↑
x_9 (Dreiecksinsel mit LSA)	↑	x_1 (Einmündung / Ref.: Kreuzung)	↓
x_5	↑	x_5	↑
x_3 (vorhanden)	↑	x_{17} (ja, gesamt / Ref.: nein)	↓
x_1 (Einmündung)	↓	x_3 (vorhanden / Ref.: nicht vorhanden)	↑

[1]nur Radverkehr als Rechtsabbiegende
Quelle: eigene Darstellung

Für die Modelle des Unfalltyps 30/32/34 erweist sich die Verkehrsregelungsart (x_2) bei beiden Verfahren als bedeutsamstes Infrastrukturmerkmal. Während das RFV dabei das Verkehrszeichen Z205 (*Vorfahrt gewähren*) als bedeutsamste Ausprägung anführt, stellt die Teil-LSA die bedeutsamste Ausprägung bei der Negativen Binomialregression dar. Dennoch ist die erwartete Zahl der entsprechenden Unfälle bei beiden Ausprägungen höher als die einer Rechts-vor-Links-Regelung. Demnach sollte in der Praxis die zuletzt genannte Regelung bevorzugt Anwendung finden. Die Anzahl der Geradeausfahrstreifen (x_5) ist ebenfalls ein von beiden Verfahren als bedeutsam ermitteltes Merkmal, wobei die erwartete Unfallzahl in beiden Fällen mit der Anzahl der Geradeausfahrstreifen steigt. Um die erwarteten Unfälle vom Unfalltyp 30/32/34 zu reduzieren, sollte demzufolge die Anzahl der Geradeausfahrstreifen gering gehalten werden. Auch

das Vorhandensein eines Grünpfeils (x_3) erhöht die erwartete Unfallzahl bei beiden Verfahren, weshalb der Einsatz eines Grünpfeils auf Basis der vorliegenden Ergebnisse nicht empfohlen werden kann. Während zudem ein Knotenpunkt in Form einer Einmündung (x_1) beim RFV lediglich das fünftbedeutsamste Infrastrukturmerkmal darstellt, ist diese Variable die zweitbedeutsamste der Negativen Binomialregression. Dennoch zeigt sich bei beiden Verfahren eine geringere Zahl der Unfälle vom Typ 30/32/34, wenn der Knotenpunkt eine Einmündung anstelle einer Kreuzung ist. Neben diesen Variablen bekommt die Signalisierung der Rechtsabbiegenden (x_9) in Form einer Dreiecksinsel mit LSA im RFV ebenfalls eine große Bedeutsamkeit. Demzufolge kann beim Vorliegen dieser eine erhöhte Unfallzahl erwartet werden. Aufgrund dieses Ergebnisses kann die Empfehlung ausgesprochen werden, dass der Einsatz dieser überdacht werden sollte. Im Gegensatz dazu ermittelt die Negative Binomialregression abermals die Roteinfärbung der Radfahrfurt (x_{17}) als bedeutsame Größe. Dabei wirkt sich die Einfärbung der Furt über den gesamten Knotenpunkt gegenüber gar keiner Einfärbung positiv auf die erwartete Unfallzahl aus. Demzufolge könnte in der Praxis darüber nachgedacht werden, nicht-eingefärbte Radfahrfurten über den gesamten Knotenpunkten rot einzufärben, um die erwartete Zahl der Unfälle vom Typ 30/32/34 zu verringern.
An dieser Stelle sei erneut darauf hingewiesen, dass bei der Negativen Binomialregression - im Vergleich zum RFV - nicht alle Variablen in die Schätzungen einbezogen wurden, wodurch der eben vorgestellte Vergleich mit Vorsicht zu interpretieren ist. Obwohl die Ergebnisse beider Verfahren aufgrund dieser Tatsache nicht unmittelbar vergleichbar sind, zeigt sich, dass die bedeutsamste Infrastrukturvariable des jeweils betrachteten Unfalltyps in beiden Verfahren übereinstimmt, obgleich sich die bedeutsamste Ausprägung der Variable x_2 beim Unfalltyp 30/32/34 zwischen den Verfahren unterscheidet. Zudem gibt es eine relativ große Übereinstimmung der Variablen, den die Verfahren die größten Bedeutsamkeiten beim jeweiligen Unfalltyp zuschreiben. Neben diesen Aspekten ermittelten das RFV und die Negative Binomialregression zudem bei den in jeweils beiden Verfahren übereinstimmenden Variablen die selben Wirkungsrichtungen dieser. Schlussendlich lässt sich also folgern, dass beide Verfahren sowohl beim Unfalltyp 23/24 als auch beim Unfalltyp 30/32/34 qualitativ ähnliche Ergebnisse liefern.

6.5 Vergleich mit Literaturerkenntnissen

Da die in Kapitel 2 behandelten Autoren zumeist Knotenpunkte im Ganzen anstelle von einzelnen Zufahrten betrachten und diese zudem oftmals nicht nach Unfalltyp differenzieren, kann an dieser Stelle keine abschließende Einschätzung nach Unfalltyp auf Basis von Zufahrten erfolgen. Als Alternative dazu zeigen die folgenden Abschnitte dennoch überblicksartig auf, ob sich die erzielten Ergebnisse beider untersuchter Unfalltypen im Großen und Ganzen in den Literaturerkenntnissen widerspiegeln.
Ein Vergleich der geschätzten DTV-Wirkungen mit den in der Literatur ermittelten Er-

kenntnissen bzgl. des Verkehrsaufkommens zeigt, dass die vorliegende Arbeit durchaus plausible Ergebnisse liefert. Demnach stellen diese - wie auch in der Literatur - die treibenden Größen des Unfallgeschehens zwischen Kfz- und Radverkehr dar. Eine Gegenüberstellung der entsprechenden Koeffizienten der Negativen Binomialregression (vgl. Tabelle 6.3) mit den in der Literatur ermittelten Koeffizienten gemäß Tabelle 2.1 zeigt zudem, dass sowohl die Koeffizienten des Modells des Unfalltyps 23/24 als auch die des Unfalltyps 30/32/34 plausible Werte annehmen. Weiterhin zeichnet sich der typische *Safety in Numbers-Effekt* bei fast allen ermittelten Koeffizienten ab. Eine Ausnahme im Modell des Unfalltyps 30/32/34 bildet dabei der DTV der geradeausfahrenden Kfz mit einem negativen Koeffizienten. Diese Ergebnisse finden sich ebenfalls im RFV anhand der zu den bedeutsamsten DTV-Kenngrößen gemäß VI gehörigen PDP's bzw. ALE beider Unfalltypen wieder, denn auch hier zeichnet sich der *Safety in Numbers-Effekt* - mit Ausnahme des DTV der geradeausfahrenden Kfz beim Unfalltyp 30/32/34 - ab. Das der *Safety in Numbers-Effekt* jedoch nicht immer nachgewiesen werden kann, zeigt auch die Literatur. Demnach ermitteln bspw. Turner et al. (2011) sowie Wang und Nihan (2004) ebenfalls negative Koeffizienten des Verkehrsaufkommens.
Im Gegensatz dazu erweist sich ein Vergleich der Ergebnisse der infrastrukturellen Einflussgrößen mit den Literaturerkenntnissen als schwierig, da die Autoren einerseits oftmals unterschiedliche Kenngrößen einbeziehen und andererseits, da diese selbst - wie in Kapitel 2.2 beschrieben - zumeist keinen Konsens bzgl. gemeinsam betrachteter Merkmale finden, was eine Einordnung der ermittelten Einflüsse verkompliziert. Nichtsdestotrotz versucht der folgende Abschnitt aufzuzeigen, ob und wie sich die ermittelten Einflüsse der betrachteten Infrastrukturgrößen auf das Unfallgeschehen mit den Erkenntnissen der Literatur in Kapitel 2 decken. Dabei liegt der Fokus jedoch lediglich auf jenen Variablen, die in der Literatur ebenfalls untersucht werden.
Die Roteinfärbung der Radfahrfurt stellt sowohl die bedeutsamste Kenngröße beider Verfahren des Unfallgeschehens vom Typ 23/24 als auch eine der bedeutsamsten Kenngrößen der Negativen Binomialregression beim Unfalltyp 30/32/34 dar. Demnach führt eine teilweise über den Knotenpunkt rot eingefärbte Furt anstelle von gar keiner Einfärbung zu einem höheren Unfallgeschehen des Unfalltyps 23/24, während eine Roteinfärbung der Furt über den gesamten Knotenpunkt anstelle von gar keiner Einfärbung die erwartete Anzahl der Unfälle vom Typ 30/32/34 verringert. Im Vergleich dazu ermitteln Schepers et al. (2011) einen positiven Zusammenhang zwischen einer rot eingefärbten Radverkehrsinfrastruktur und dem Unfallgeschehen. Des Weiteren prognostizieren Turner et al. (2011) - bei Vorliegen einer eingefärbten Radverkehrsinfrastruktur - im Fall eines rechtsabbiegenden Kfz und eines geradeausfahrenden Radfahrenden eine Verringerung der Unfallzahlen und im Fall von sich kreuzenden Beteiligten eine Erhöhung dieser. Da die erstgenannte Konstellation von Turner et al. (2011) dem Unfalltyp 23/24 und die zweitgenannte Konstellation dem Unfalltyp 30/32/34 zugeordnet werden kann, lässt sich ein gegensätzlicher Trend der in der Literatur und in dieser Arbeit ermittelten

Einflüsse erkennen. Neben einer rot eingefärbten Radverkehrsinfrastruktur thematisiert die Literatur zudem die Markierung einer solchen, welche in dieser Arbeit in Form der Führung des linksabbiegenden Radverkehrs bei beiden Verfahren beim Unfalltyp 23/24 von Bedeutung ist. Dabei zeigen die Ergebnisse, dass eine indirekte Führung - d.h. Markierungen, Aufstellflächen oder Signalgeber - das Unfallgeschehen vom Typ 23/24 steigert im Vergleich zu einer Führung im Mischverkehr, während sich die Literatur unschlüssig bzgl. des Einflusses einer markierten Radverkehrsinfrastruktur ist. Während gut markierte Radwege die Unfallzahlen bei Schepers et al. (2011) erhöhen, lassen Fahrbahnmarkierungen diese bei Hamann und Peek-Asa (2013) geringer ausfallen. Demnach ist die Einordnung des in der Arbeit erzielten Ergebnisses schwierig. Hinzu kommt, dass der in der Arbeit ermittelte Effekt der Führung des linksabbiegenden Radverkehrs in Hinblick auf die Entstehung der Unfälle vom Typ 23/24 sachlogisch irrelevant erscheint, was einen aussagekräftigen Vergleich mit den Literaturerkenntnissen ebenfalls erschwert. Weiterhin lässt die Anzahl der Rechtsabbiegefahrstreifen nach dem RFV beim Unfalltyp 23/24 keinen Trend bzgl. des Unfallgeschehens erkennen, wohingegen Wang und Nihan (2004) im Fall eines rechtsabbiegenden Kfz und einem in gleicher oder entgegengesetzter Richtung geradeausfahrenden Radfahrenden, was dem in der Arbeit untersuchten Unfalltyp 23/24 entspricht, einen positiven Zusammenhang zwischen der Anzahl der Rechtsabbiegefahrstreifen und der Unfallzahl ermitteln. Somit stimmt der in der Arbeit ermittelte Einfluss nicht mit den Erkenntnissen von Wang und Nihan (2004) überein. Ein weiterer Vergleich zwischen den in dieser Arbeit und den in der Literatur ermittelten Einflüssen bietet sich zudem für die Signalisierung der Rechtsabbiegenden in Form einer Dreiecksinsel mit LSA an, deren Vorhandensein die Anzahl der Unfälle vom Typ 30/32/34 gemäß RFV erhöht. Da Kim et al. (2012) einen positiven Zusammenhang zwischen dem Unfallgeschehen und der Existenz einer Verkehrsinsel ermitteln, kann in diesem Fall von einem konsistenten Ergebnis ausgegangen werden.

Da die Literatur, bei der oftmals keine nach Unfalltypen separierte Betrachtung stattfindet, bei einigen Einflussgrößen selbst verschiedene Schlüsse zieht und die Ergebnisse zudem möglicherweise vom untersuchten Standort abhängen, kann an dieser Stelle kein abschließendes Urteil darüber gefällt werden, ob die ermittelten Einflüsse - abgesehen von den eingangs beschriebenen DTV-Kenngrößen - plausibel sind.

7 Kritische Würdigung

Die vorliegende Arbeit untersuchte das Unfallgeschehen an Knotenpunktzufahrten zwischen dem Kfz- und Radverkehr in der Großstadt Dresden. Demzufolge ist es möglich, dass die analoge Untersuchung einer Kleinstadt oder eines ländlichen Gebietes dazu führt, dass andere als die in dieser Arbeit ermittelten Merkmale das Unfallgeschehen signifikant beeinflussen. Denkbar wären dabei bspw. Größen wie die Beleuchtung oder Sichtbarkeit von Knotenpunkten, da dies einerseits in ländlichen Gebieten nicht immer gegeben ist und andererseits in der Dunkelheit oftmals ein erhöhtes Risiko dafür besteht, dass Radfahrende von Kfz übersehen werden, wodurch es schließlich zu Unfällen kommen kann. Außerdem beschränkte sich die Arbeit auf konkrete Untertypen der Unfalltypen AB und EK, weshalb die erlangten Ergebnisse nicht unmittelbar auf andere Unfalltypen oder das Unfallgeschehen im Allgemeinen übertragbar sind. Weiterhin ist bspw. die Variable *Abstand ruhender Verkehr zum Knotenpunkt* einseitig verteilt, wobei fast alle Zufahrten sowie fast alle Unfälle beider betrachteter Unfalltypen die Ausprägung *mehr als 25 Meter* aufweisen. Demnach stellt sich die Frage nach der Sinnhaftigkeit der Einbeziehung dieser Variable. Da diese jedoch bei beiden Verfahren und Unfalltypen vergleichsweise unbedeutend ist, scheint diese Tatsache nicht allzu problematisch zu sein. Dennoch ist es denkbar, dass der Ausschluss dieser Variable zu veränderten Ergebnissen führt. Weiterhin bedingte der Kontingenzkoeffizient nach Pearson ein kategoriales Vorliegen aller Variablen, weshalb eine Kategorisierung der metrisch skalierten Variablen erfolgte. Diese Herabstufung metrisch skalierter auf kategoriale Variablen geht jedoch mit einem Informationsverlust der entsprechenden Variablen einher. Zudem wurden sowohl die zugehörigen Kategorien als auch die Anzahl der Kategorien der entsprechenden Variablen willkürlich festgelegt, wodurch eine alternative Festlegung der einzelnen Kategorien sowie der Anzahl der gebildeten Kategorien zu abweichenden Kontingenzkoeffizienten führen könnte. Des Weiteren wurden die Parameter für die Anzahl der zu wählenden Variablen pro Knoten im RFV gemäß den Standardeinstellungen der Software *R* gewählt. Demzufolge könnte die Zuweisung anderer Parameter zu veränderten Ergebnissen des RFV führen. Ein weiterer Aspekt, der kritisch angesehen werden kann, ist die im LASSO-Verfahren durch technische Gründe bedingte Einbeziehung von Dummyvariablen anstelle der entsprechenden kategorialen Variablen. Demnach schloss das Verfahren einzelne Ausprägungen kategorialer Variablen aus den eigentlichen Schätzungen der Negativen Binomialregression aus, wodurch die Möglichkeit verzerrter Koeffizienten besteht. Es stellt sich also die Frage, ob die Einbeziehung aller zu einer kategorialen Variable gehörenden Dummyvariablen

- sofern einzelne Ausprägungen dieser Variable gemäß dem LASSO-Verfahren einbezogen werden sollten - zu veränderten Ergebnissen der Negativen Binomialregression führt. Neben diesen Aspekten könnten die Verfahren zudem andere Ergebnisse aufweisen, wenn einerseits im RFV nicht alle zur Verfügung stehenden Variablen einbezogen werden, sondern lediglich jene, die für den betrachteten Unfalltyp von sachlogischer Relevanz erscheinen. Andererseits bietet es sich für die Negativen Binomialregression an, die Variablen - in Anlehnung an das Vorgehen einiger Autoren - aufgrund sachlogischer Überlegungen schrittweise in die Schätzungen einzubeziehen anstelle der Verwendung des LASSO-Verfahrens. Beides würde dazu führen, dass keine entgegen der Intuition erscheinenden Infrastrukturmerkmale, wie bspw. die Führung des linksabbiegenden Radverkehrs beim Unfalltyp 23/24, als bedeutsame Einflussfaktoren ermittelt werden, sondern lediglich jene, die die unfallauslösende Konfliktsituation aktiv zu beeinflussen scheinen. Weiterhin weist der untersuchte Datensatz eine relativ geringe Fallzahl der Unfälle der betrachteten Typen auf, wodurch sich eine rechtsschiefe Verteilung der entsprechenden abhängigen Variablen, bei der die meisten Zufahrten eine Anzahl von null Unfällen vom jeweiligen Typ aufweisen, abzeichnete. Obwohl sich dadurch jeweils geringe Vorhersagefehler in Form des MSE bzw. RMSE ergaben, zeigten die Plots der Gegenüberstellung der realen und vorhergesagten Unfallzahlen der Testdaten vor allem bei großen Unfallzahlen die Vorhersageschwächen beider Verfahren auf. Demnach würde eine höhere Anzahl an Unfällen möglicherweise bei beiden Verfahren zu zuverlässigeren Schätzungen und veränderten Effekten der einzelnen Merkmale führen, was wiederum in einer erhöhten Vorhersagegenauigkeit resultieren könnte. Des Weiteren könnte sowohl die vorherige Eliminierung von Ausreißerzufahrten, also jene mit einer großen Anzahl an Unfällen vom jeweiligen Typ, als auch die alleinige Einbeziehung der Zufahrten, an denen mindestens ein Unfall vom jeweiligen Typ stattfand, eine Möglichkeit zur Verbesserung der Schätzgenauigkeit darstellen. Zudem kommt die in beiden Verfahren schlechtere Modellgüte des Unfalltyps 30/32/34 möglicherweise daher, dass für die Untersuchung des besagten Unfalltyps EK zu viele dreistellige Untertypen zusammengefasst wurden. Demzufolge besteht die Möglichkeit, dass Unfälle vom Typ EK, bei denen der bevorrechtigte Verkehrsteilnehmende von rechts kommt von anderen infrastrukturellen Merkmalen beeinflusst werden als solche, bei denen der bevorrechtigte Verkehrsteilnehmende von links kommt. Die gemeinsame Betrachtung der entsprechenden Untertypen könnte dazu geführt haben, dass die eigentlichen Effekte der einzelnen Variablen verschwimmen, was wiederum die Modellanpassung an die Daten erschwert. Des Weiteren erweist sich der direkte Vergleich der Ergebnisse beider Verfahren als schwierig, da diese jeweils verschiedene Variablen betrachteten. Während das RFV bei beiden betrachteten Unfalltypen alle zur Verfügung stehenden Variablen einbezog, wurde im Vorfeld der Negativen Binomialregression eine Variablenauswahl durch das LASSO-Verfahren getroffen, wobei sich die resultierende Auswahl zudem zwischen den Modellen beider Unfalltypen unterschied.

8 Zusammenfassung und Ausblick

Das Ziel der Arbeit lag in der Untersuchung der Einflussfaktoren des Unfallgeschehens zwischen MIV und Radverkehr am Beispiel der Stadt Dresden, wobei die Datengrundlage aus 548 Radverkehrsunfällen aus den Jahren 2015 bis 2019 sowie DTV-Kenngrößen und infrastrukturellen Merkmalen von 484 Knotenpunktzufahrten bestand. Dabei kamen die ökonometrischen Verfahren RFV sowie die Negative Binomialregression in Form von *Accident Prediction Models* zur Anwendung, wobei im Vorfeld der Negativen Binomialregression eine Variablenselektion mit Hilfe des LASSO-Verfahrens stattfand. Um differenzierte Ergebnisse zu erlangen, wurden die Verfahren dabei jeweils auf zusammengefasste Untertypen der Unfalltypen AB und EK angewandt. Dabei umfassten die betrachteten Unfälle vom Typ AB Kollisionen zwischen einem rechtsabbiegenden und einem in gleicher oder entgegengesetzter Richtung geradeausfahrenden Beteiligten. Im Gegensatz dazu beinhalteten die betrachteten Unfälle vom Typ EK Kollisionen zwischen einem vorfahrtsberechtigten Verkehrsteilnehmenden und einem einbiegenden oder kreuzenden Wartepflichtigen.

Generell zeigte sich, dass die untersuchten Infrastrukturmerkmale einen weniger starken Einfluss auf die erwarteten Unfallzahlen pro Zufahrt zwischen MIV und Radverkehr ausüben als die entsprechenden DTV-Kenngrößen. Zudem konnte der in der Literatur beschriebene *Safety in Numbers-Effekt* für die meisten der untersuchten DTV-Kenngrößen nachgewiesen werden. Weiterhin ermittelten beide Verfahren bei beiden Unfalltypen die gleichen Wirkungsrichtungen derjenigen Infrastrukturmerkmale, die sowohl vom RFV als auch von der Negativen Binomialregression als bedeutsam für die Vorhersage der erwarteten Unfallzahl vom jeweiligen Typ erachtet wurden. Demnach schätzten beide Verfahren bspw. eine höhere Anzahl der Unfälle vom Typ AB, wenn die zu einer Zufahrt gehörigen Radfahrfurten teilweise bzw. komplett über den Knotenpunkt eingefärbt vorliegen im Vergleich mit keiner Einfärbung. Zudem ergab sich bei beiden Verfahren eine höhere Anzahl der Unfälle vom Typ AB, wenn der linksabbiegende Radverkehr anstelle einer Führung im Mischverkehr indirekt geführt wird. Dieses Ergebnis erwies sich jedoch als Indiz für eine Schwäche beim Vorgehen der Variableneinbeziehung beider Verfahren, da dieses Merkmal für die Entstehung der betrachteten Unfälle vom Typ AB vor sachlogischem Hintergrund irrelevant erscheint. Für die erwartete Anzahl der Unfälle vom Typ EK zeigten die Verfahren bspw., dass ein durch das Verkehrszeichen Z205 oder eine Teil-LSA geregelter Knotenpunkt mehr Unfälle aufweist als ein solcher mit einer Rechts-vor-Links-Regelung. Zudem sagten die Modelle beider Verfahren eine höhere erwartete Unfallzahl vom Typ EK vorher, wenn die Anzahl separater Geradeausfahr-

streifen der entsprechenden Zufahrt zunimmt.
In Bezug auf die betrachteten Unfalltypen zeigte die Gegenüberstellung der Modellgütekriterien beider Verfahren zudem, dass sowohl die Testdaten des RFV als auch die der Negativen Binomialregression bessere Vorhersagen für die Anzahl der Unfälle vom Unfalltyp AB als für die des Unfalltyps EK generieren. Zudem zeigten die entsprechenden Modellgüten des Unfalltyps AB kaum einen Unterschied zwischen beiden Verfahren, was bedeutet, dass beide Verfahren in qualitativ ähnlichen Modellen des genannten Unfalltyps resultierten. Demzufolge lieferte das RFV für den Unfalltyp AB annähernd so gute Vorhersagen der Testdaten wie die Negative Binomialregression. Im Gegensatz dazu wies die Negative Binomialregression bei der Vorhersage der Testdaten des Unfalltyps EK eine geringfügige Überlegenheit gegenüber dem RFV auf. Weiterhin stellte die Vorhersage größerer Unfallzahlen eine Schwierigkeit beider Verfahren dar. Auf Grundlage dessen lässt sich schließlich schlussfolgern, dass beide Verfahren bei beiden Unfalltypen qualitativ ähnliche Ergebnisse erzielten.
Eine Erweiterung der Arbeit wäre in dem Sinne denkbar, dass einerseits andere Parameter für das RFV gewählt werden und andererseits, dass die Variablenselektion im Vorfeld der Negativen Binomialregression aufgrund sachlogischer Überlegungen durchgeführt wird. Des Weiteren bietet sich die Untersuchung weiterer Unfalltypen an, um ein umfassenderen Einblick in die Wirkungen der Infrastrukturmerkmale auf das Unfallgeschehen zwischen MIV und Radverkehr zu erlangen. Zudem könnten alternative Verfahren angewandt werden. Dabei eignet sich z.B. eine logistische Regression, mit deren Hilfe es möglich ist, die abhängige Variable kategorial einzubeziehen. Demzufolge wäre es bspw. möglich, diese in die Ausprägungen *Zufahrt mit Unfall* und *Zufahrt ohne Unfall* zu unterteilen und anschließend die beeinflussenden Infrastrukturgrößen zu untersuchen.

Literaturverzeichnis

Adminaité-Fodor, D. und G. Jost (2020). *How safe is walking and cycling in europe? PIN Flash Report 38*. Brüssel: European Transport Safety Council. URL: https://etsc.eu/how-safe-is-walking-and-cycling-in-europe-pin-flash-38/ (besucht am 04. 12. 2020).

Apley, D. W. und J. Zhu (2016). *Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models*. Cornell University. URL: https://arxiv.org/abs/1612.08468v2 (besucht am 21. 10. 2020).

Backhaus, K., B. Erichson, W. Plinke und R. Weiber (2008). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 12. vollständig überarbeitete Auflage. Berlin: Springer.

BASt Bundesanstalt für Straßenwesen (2017). *Maßnahmenkatalog gegen Unfallhäufungen*. URL: https://makau.bast.de/massnahmen/113 (besucht am 10. 12. 2020).

Benninghaus, H. (2005). *Deskriptive Statistik. Eine Einführung für Sozialwissenschaftler*. 10. durchgesehene Auflage. Studienskripten zur Soziologie. Wiesbaden: Verlag für Sozialwissenschaften.

Breiman, L. (2001). „Random forests“. In: *Machine learning* 45 (1), S. 5–32.

Breiman, L. (2002). *Manual on setting up, using, and understanding random forests v3.1*. URL: https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf (besucht am 15. 09. 2020).

Breiman, L., J. Friedman, C. J. Stone und R. A. Olshen (1984). *Classification and regression trees*. Boca Raton: Chapman & Hall.

Cameron, A. C. und P. K. Trivedi (1990). „Regression-based tests for overdispersion in the Poisson model“. In: *Journal of Econometrics* 46 (3), S. 347–364.

Cleff, T. (2012). *Deskriptive Statistik und moderne Datenanalyse. Eine computergestützte Einführung mit Excel, PASW (SPSS) und STATA*. 2. überarbeitete und erweiterte Auflage. Wiesbaden: Gabler Verlag.

Dunn, P. K. und G. K. Smyth (2018). *Generalized linear models with examples in R*. Springer texts in statistics. New York: Springer. URL: https://link.springer.com/book/10.1007/978-1-4419-0118-7 (besucht am 12. 12. 2020).

Eenink, R., M. Reurings, R. Elvik, J. Cardoso, S. Wichert und C. Stefan (2008). *Accident prediction models and road safety impact assessment. Recommendations for using these tools*. Leidschendam, Niederlande: Institute for Road Safety Research.

EESC European Economic and Social Committee (2016). *European Cycling Lexicon*. Brüssel. URL: https://www.eesc.europa.eu/en/our-work/publications-other-work/publications/european-cycling-lexicon-0 (besucht am 04. 12. 2020).

FGSV Forschungsgesellschaft für Straßen- und Verkehrswesen (2006). *Richtlinien für die Anlage von Stadtstraßen. RASt 06*. Köln: FGSV Verlag.

FGSV Forschungsgesellschaft für Straßen- und Verkehrswesen (2012). *Merkblatt zur örtlichen Unfalluntersuchung in Unfallkommissionen. M Uko*. Köln: FGSV Verlag.

Friedman, J., T. Hastie und R. Tibshirani (2010). „Regularization Paths for Generalized Linear Models via Coordinate Descent“. In: *Journal of statistical software* 33 (1), S. 1–22.

Gerike, R., S. Hubrich, F. Ließke, S. Wittig und R. Wittwer (2020). *Sonderauswertung zum Forschungsprojekt "Mobilität in Städten – SrV 2018". Städtevergleich*. Hrsg. von Professur für Integrierte Verkehrsplanung und Straßenverkehrstechnik. Technische Universität Dresden.

Grömping, U. (2009). „Variable Importance Assessment in Regression: Linear Regression versus Random Forest“. In: *The American Statistician* 63 (4), S. 308–319.

Hamann, C. und C. Peek-Asa (2013). „On-road bicycle facilities and bicycle crashes in Iowa, 2007-2010“. In: *Accident Analysis and Prevention* 56, S. 103–109.

Hilbe, J. M. (2011). *Negative Binomial Regression*. 2. Auflage. Cambridge: Cambridge University Press.

ISUP Ingenieurbüro für Systemberatung und Planung GmbH, Planungsgemeinschaft Verkehr PGV-Alrutz GbR (2017). *Radverkehrskonzept der Landeshauptstadt Dresden*. URL: https://www.dresden.de/de/stadtraum/verkehr/rad-fuss/020-Das-Radverkehrskonzept.php (besucht am 04. 12. 2020).

Jacobsen, P. L. (2015). „Safety in numbers. More walkers and bicyclists, safer walking and bicycling“. In: *Injury prevention* 21 (4), S. 271–275.

Kim, M., E. Kim, J. Oh und J. Jun (2012). „Critical Factors Associated with Bicycle Accidents at 4-legged Signalized Urban Intersections in South Korea“. In: *KSCE Journal of Civil Engineering* 16 (4), S. 627–632.

Klassen, J., K. El-Basyouny und M. T. Islam (2014). „Analyzing the severity of bicycle-motor vehicle collision using spatial mixed logit models: A City of Edmonton case study“. In: *Safety Science* 62, S. 295–304.

Liaw, A. und M. Wiener (2002). „Classification and Regression by RandomForest“. In: *R News* 2 (3), S. 18–22.

Loh, W.-Y. (2011). „Classification and regression trees“. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1 (1), S. 14–23.

Maher, M. J. und I. Summersgill (1996). „A comprehensive methodology for the fitting of predictive accident models“. In: *Accident Analysis and Prevention* 28 (3), S. 281–296.

Molnar, C. (2019). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. URL: https://christophm.github.io/interpretable-ml-book/ (besucht am 12. 11. 2020).

Nobis, C. (2019). *Mobilität in Deutschland – MiD. Analysen zum Radverkehr und Fußverkehr*. Bonn, Berlin: Studie von infas, DLR, IVT und infas 360 im Auftrag des Bundesministeriums für Verkehr und digitale Infrastruktur. URL: https://www.bmvi.de/SharedDocs/DE/Anlage/G/mid-analysen-rad-fussverkehr.html (besucht am 04. 12. 2020).

Nordback, K., W. E. Marshall und B. N. Janson (2014). „Bicyclist safety performance functions for a U.S. city". In: *Accident Analysis and Prevention* 65, S. 114–122.

Ohm, D., M. Mohaupt, K. Burggraf und F. Fiedler (2016). *Verkehrsentwicklungsplan 2025plus. Ein Überblick*. Hrsg. von Landeshauptstadt Dresden. Dresden. URL: https://www.dresden.de/de/stadtraum/verkehr/verkehrsplanung/verkehrsentwicklungsplanung/010_Verkehrsentwicklungsplan_2025.php (besucht am 04. 12. 2020).

Pruscha, H. (2006). *Statistisches Methodenbuch. Verfahren, Fallstudien, Programmcodes*. Statistik und ihre Anwendungen. Berlin und Heidelberg: Springer.

Raileanu, L. E. und K. Stoffel (2004). „Theoretical comparison between the gini index and information gain criteria". In: *Annals of Mathematics and Artificial Intelligence* 41 (1), S. 77–93.

Schepers, J. P., P. A. Kroeze, W. Sweers und J. C. Wüst (2011). „Road factors and bicycle-motor vehicle crashes at unsignalized priority intersections". In: *Accident Analysis and Prevention* 43 (3), S. 853–861.

Sutton, C. D. (2005). „Classification and Regression Trees, Bagging, and Boosting". In: *Data Mining and Data Visualization*. Bd. 24. Handbook of Statistics. Elsevier, S. 303–329.

Tabachnick, B. G. und L. S. Fidell (2007). *Using multivariate statistics*. 5. Auflage. Boston, Massachusetts: Pearson/Allyn and Bacon.

Tibshirani, R. (1996). „Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society: Series B* 58 (1), S. 267–288.

Turner, S., G. Wood, T. Hughes und R. Singh (2011). „Safety performance functions for bicycle crashes in New Zealand and Australia". In: *Transportation Research Record* 2236 (1), S. 66–73.

Urban, D. (1993). *Logit-Analyse. Statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen*. Stuttgart, Jena und New York: Gustav Fischer Verlag.

Wang, Y. und N. L. Nihan (2004). „Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections". In: *Accident Analysis & Prevention* 36 (3), S. 313–321.

Zhang, Huiming und Jinzhu Jia (2017). *Elastic-net regularized high-dimensional negative binomial regression. Consistency and weak signals detection.* Cornell University. URL: `https://arxiv.org/abs/1712.03412v3` (besucht am 01. 11. 2020).

Anhang A

Grundlagen der Unfallforschung

Unfalltyp		Erläuterung
1	grün	**Fahrunfall(F)** Der Unfall wurde ausgelöst durch den Verlust der Kontrolle über das Fahrzeug (wegen nicht angepasster Geschwindigkeit oder falscher Einschätzung des Straßenverlaufs, des Straßenzustandes o.Ä.), ohne dass andere Verkehrsteilnehmer dazu beigetragen haben. Infolge unkontrollierter Fahrzeugbewegungen kann es dann aber zum Zusammenstoß mit anderen Verkehrsteilnehmern gekommen sein.
2	gelb	**Abbiege-Unfall (AB)** Der Unfall wurde ausgelöst durch einen Konflikt zwischen einem, den Vorrang Anderer zu beachtenden Abbieger und einem aus gleicher oder entgegengesetzter Richtung kommenden Verkehrsteilnehmer (auch Fußgänger!) an Kreuzungen, Einmündungen, Grundstücks- oder Parkplatzzufahrten.
3	rot	**Einbiegen/Kreuzen-Unfall (EK)** Der Unfall wurde ausgelöst durch einen Konflikt zwischen einem einbiegenden oder kreuzenden Wartepflichtigen und einem vorfahrtberechtigten Fahrzeug an Kreuzungen, Einmündungen oder Ausfahrten von Grundstücken und Parkplätzen.
4	rot / weiß	**Überschreiten-Unfall (ÜS)** Der Unfall wurde ausgelöst durch einen Konflikt zwischen einem Fahrzeug und einem Fußgänger auf der Fahrbahn, sofern dieser nicht in Längsrichtung ging und sofern das Fahrzeug nicht abgebogen ist. Dies gilt auch, wenn der Fußgänger nicht angefahren wurde.
5	hellblau	**Unfall durch ruhenden Verkehr (RV)** Der Unfall wurde ausgelöst durch einen Konflikt zwischen einem Fahrzeug des fließenden Verkehrs und einem Fahrzeug, das parkt/hält bzw. Fahrmanöver im Zusammenhang mit dem Parken/Halten durchführte.
6	orange (magenta)	**Unfall im Längsverkehr (LV)** DerUnfall wurde ausgelöst durch einen Konflikt zwischen Verkehrsteilnehmern, die sich in gleicher oder entgegengesetzter Richtung bewegten, sofern dieser Konflikt nicht einem anderen Unfalltyp entspricht.
7	schwarz	**Sonstiger Unfall (SO)** Unfall, der sich nicht den Typen 1 - 6 zuordnen lässt. Beispiele: Wenden, Rückwärtsfahren, Parker untereinander, Hindernis oder Tier auf der Fahrbahn, plötzlicher Fahrzeugschaden (Bremsversagen, Reifenschäden o.Ä.)

Abbildung A.1: Kurzbeschreibung der Unfalltypen nach M Uko [Quelle: FGSV (2012), S. 9]

Anhang B

Datengrundlage

B.1 Unfalldaten

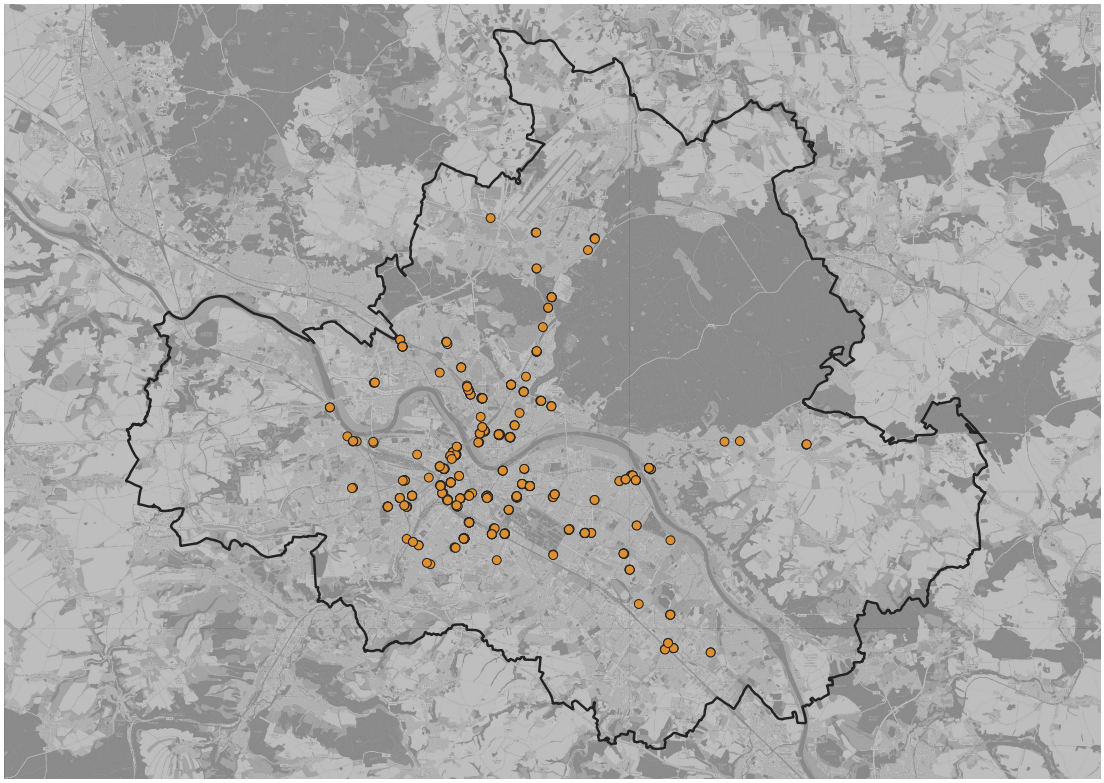


Abbildung B.1: Räumliche Verteilung der Radverkehrsunfälle in Dresden [Quelle: eigene Darstellung auf Basis von QGIS]

B.2 Infrastrukturmerkmale

Abbildung B.2 zeigt die möglichen Ausprägungen des Merkmals *Roteinfärbung der Radfahrfurt* am Beispiel eines Knotenpunktes auf. Dabei bildet die zu Zufahrt 1 (NO) gehörige Radfahrfurt die Ausprägung *nein* ab. Weiterhin sind die zu Zufahrt 2 (SO) bzw. 4 (NW) gehörigen Radfahrfurten durch die Ausprägung *ja, über gesamten Knoten* gekennzeichnet. Im Gegensatz dazu zeichnet sich die Ausprägung *ja, teilweise* bei der zu Zufahrt 3 (W) gehörigen Radfahrfurt ab, da diese nicht über die gesamte Straßenbreite rot eingefärbt vorliegt.

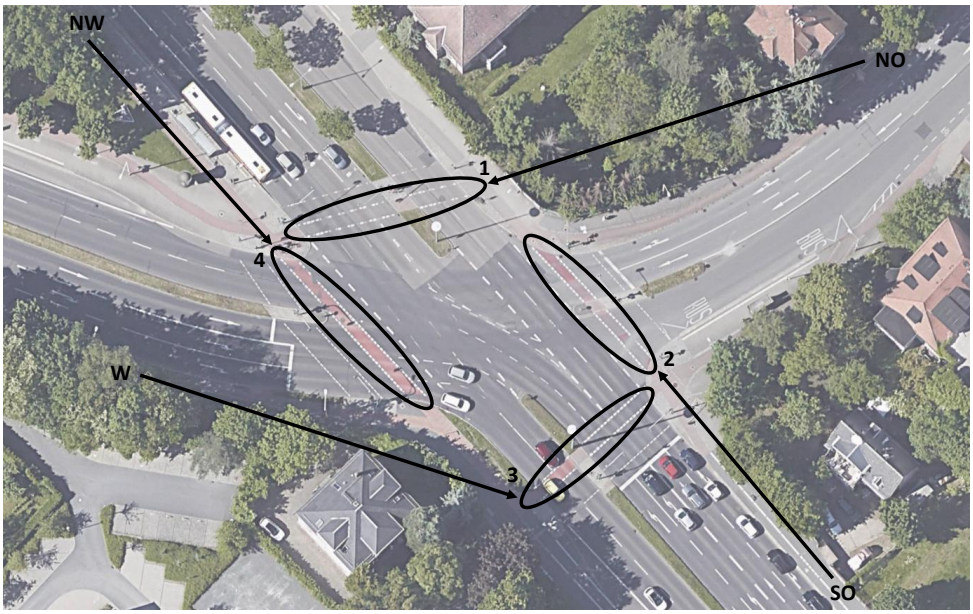


Abbildung B.2: Ausprägungen des Merkmals *Roteinfärbung der Radfahrfurt* am Beispiel eines Knotenpunktes [Quelle: eigene Darstellung auf Basis von Google Maps]

Anhang C

Methodik

Tabelle C.1 stellt den schematischen Aufbau einer Kontingenztabelle dar. Dabei stehen sich die Variable A mit v Ausprägungen (a_1 bis a_v) und die Variable B mit w Ausprägungen (b_1 bis b_w) gegenüber, wobei n_{vw} die absolute Häufigkeit des gemeinsamen Auftretens der Variablenausprägungen a_v und b_w darstellt. Weiterhin werden die Randhäufigkeiten, also die Zeilensummen $n_{v.}$ und Spaltensummen $n_{.w}$, abgetragen. Die Gesamtanzahl der Beobachtungen wird zudem durch n angegeben.

Tabelle C.1: Schematischer Aufbau einer Kontingenztabelle

	Variable B				
Variable A	b_1	b_2	...	b_w	Zeilensumme
a_1	n_{11}	n_{12}	...	n_{1w}	$n_{1.}$
a_2	n_{21}	n_{22}	...	n_{2w}	$n_{2.}$
...	...	...	...	...	...
a_v	n_{v1}	n_{v2}	...	n_{vw}	$n_{v.}$
Spaltensumme	$n_{.1}$	$n_{.2}$	...	$n_{.w}$	n

Quelle: eigene Darstellung

Des Weiteren kann n durch Aufaddieren der absoluten Häufigkeiten des Auftretens aller Ausprägungskombinationen ermittelt werden. Formal stellt sich das wie folgt dar.

$$n = \sum_{i=1}^{v} \sum_{j=1}^{w} n_{ij}$$

Anhang D

Durchführung und Ergebnisse

D.1 Korrelationsbetrachtung

Tabelle D.1: Kategorisierung der metrisch vorliegenden Variablen

Variable	Einheit	Ausprägungen
y_1		0, 1, 2, 3, 4, 7
y_2		0, 1, 2, 3, 4, 5, 6, 8, 12, 20
x_4		0, 1, 2
x_5		0, 1, 2
x_6		0, 1, 2
x_7		0, 1, 2
x_{10}	[m]	[0;5], (5;10], (10;15], (15;20], >20
x_{18}	[Kfz/24h]	[0;1.000], (1.000;2.000], (2.000;5.500], >5.500
x_{19}	[Kfz/24h]	[0;4.000], (4.000;8.000], (8.000;14.000], >14.000
x_{20}	[Kfz/24h]	[0;1.000], (1.000;2.000], (2.000;8.000], >8.000
x_{21}		[0;0,05], (0,05;0,10], (0,10;0,50], >0,50
x_{22}		[0;0,05], (0,05;0,10], (0,10;0,50], >0,50
x_{23}		[0;0,05], (0,05;0,10], (0,10;0,50], >0,50
x_{24}	[Räder/24h]	[0;50], (50;100], (100;850], >850
x_{25}	[Räder/24h]	[0;300], (300;600], (600;1.000], >1.000
x_{26}	[Räder/24h]	[0;100], (100;200], (200;1.100], >1.100

Quelle: eigene Darstellung

D.2 Random Forest

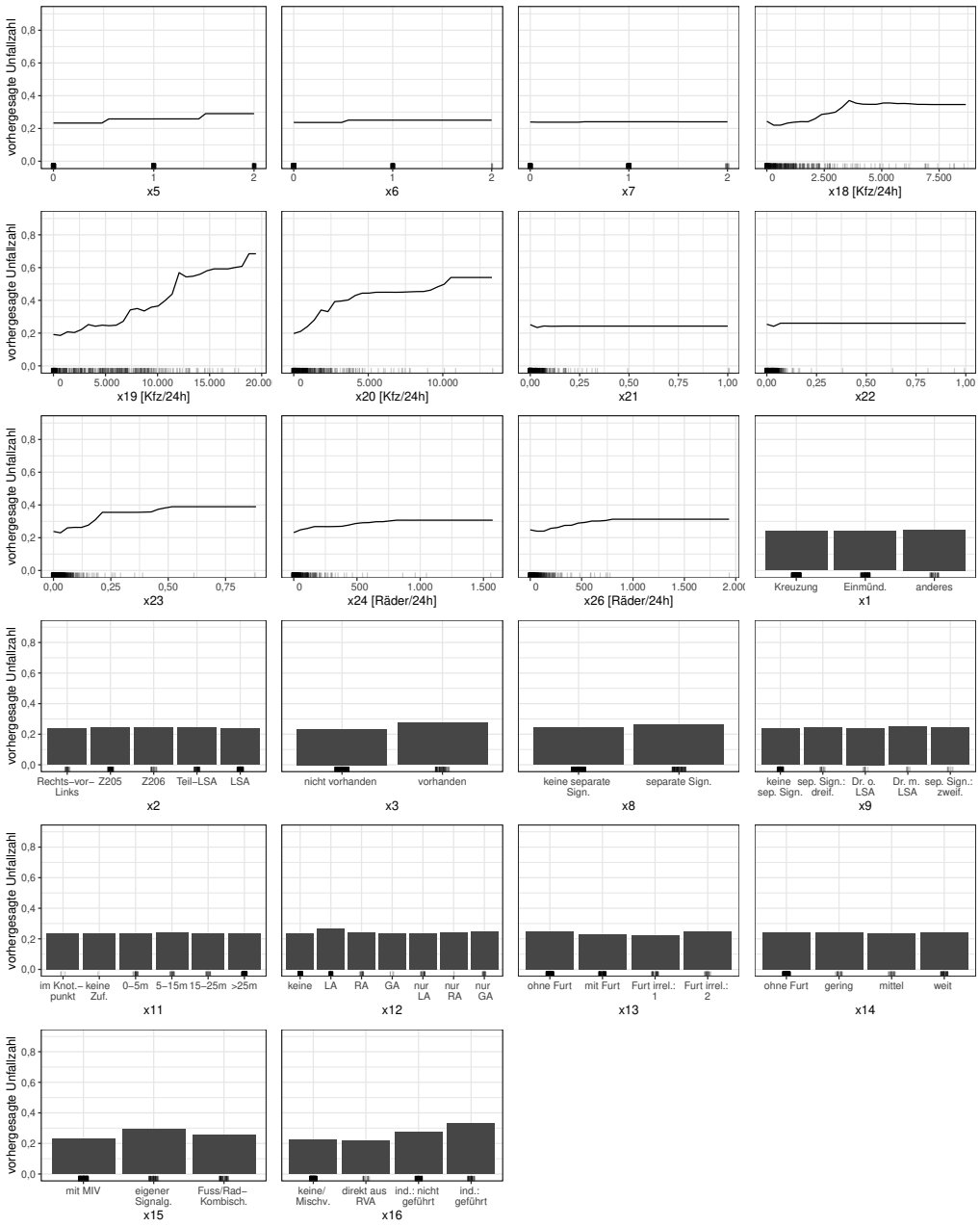


Abbildung D.1: Unfalltyp 23/24: PDP der übrigen Variablen [Quelle: eigene Darstellung]

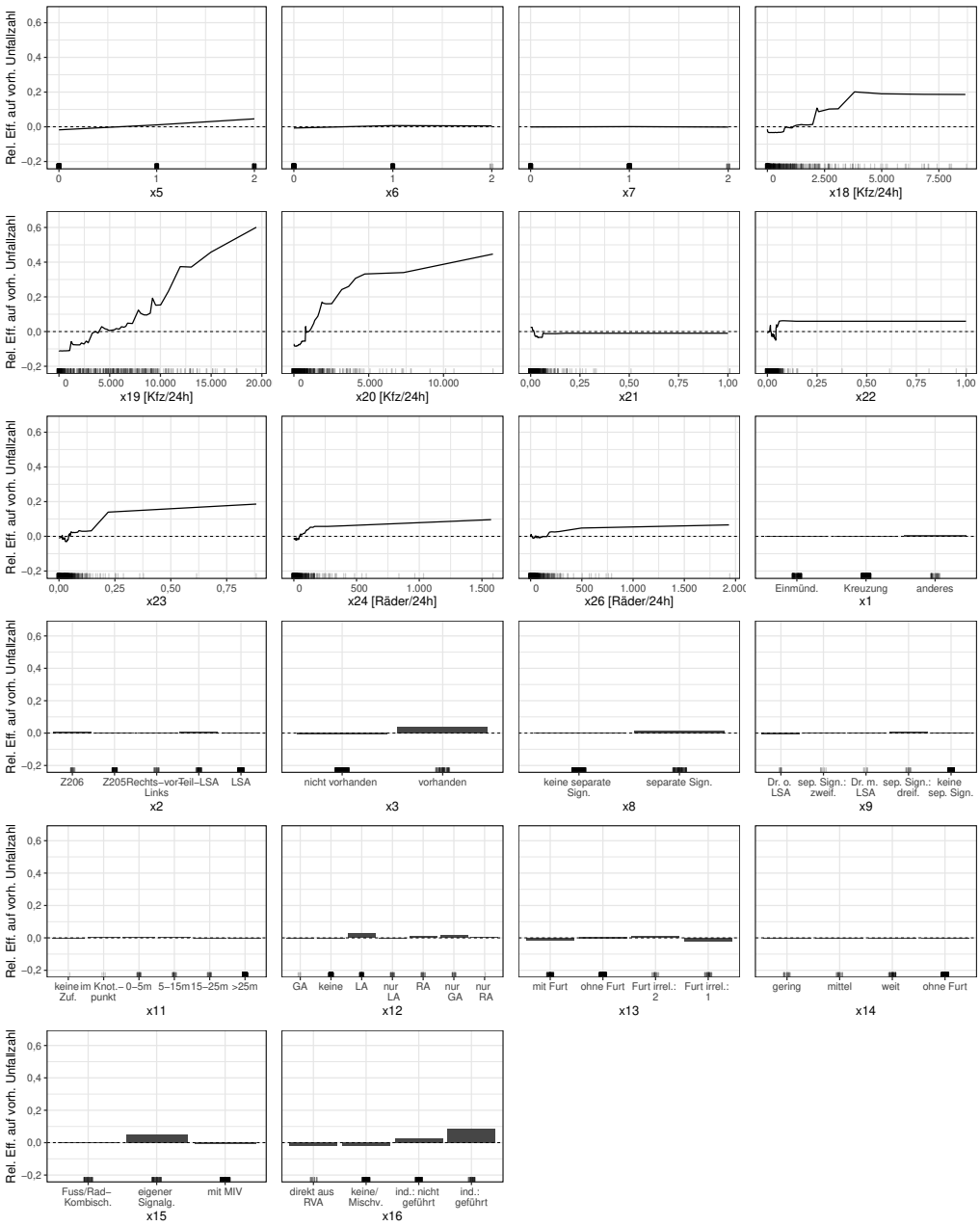


Abbildung D.2: Unfalltyp 23/24: ALE der übrigen Variablen [Quelle: eigene Darstellung]

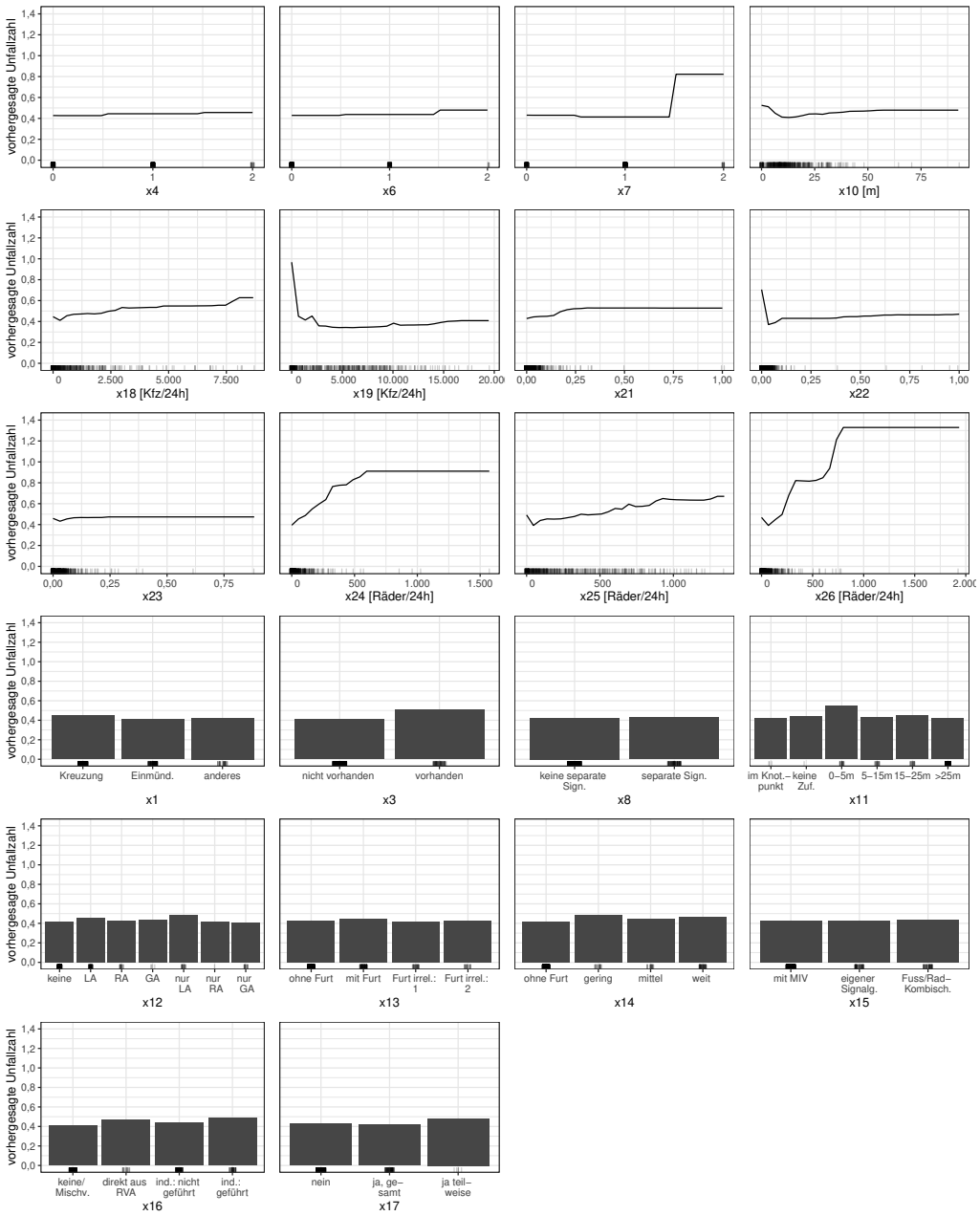


Abbildung D.3: Unfalltyp 30/32/34: PDP der übrigen Variablen [Quelle: eigene Darstellung]

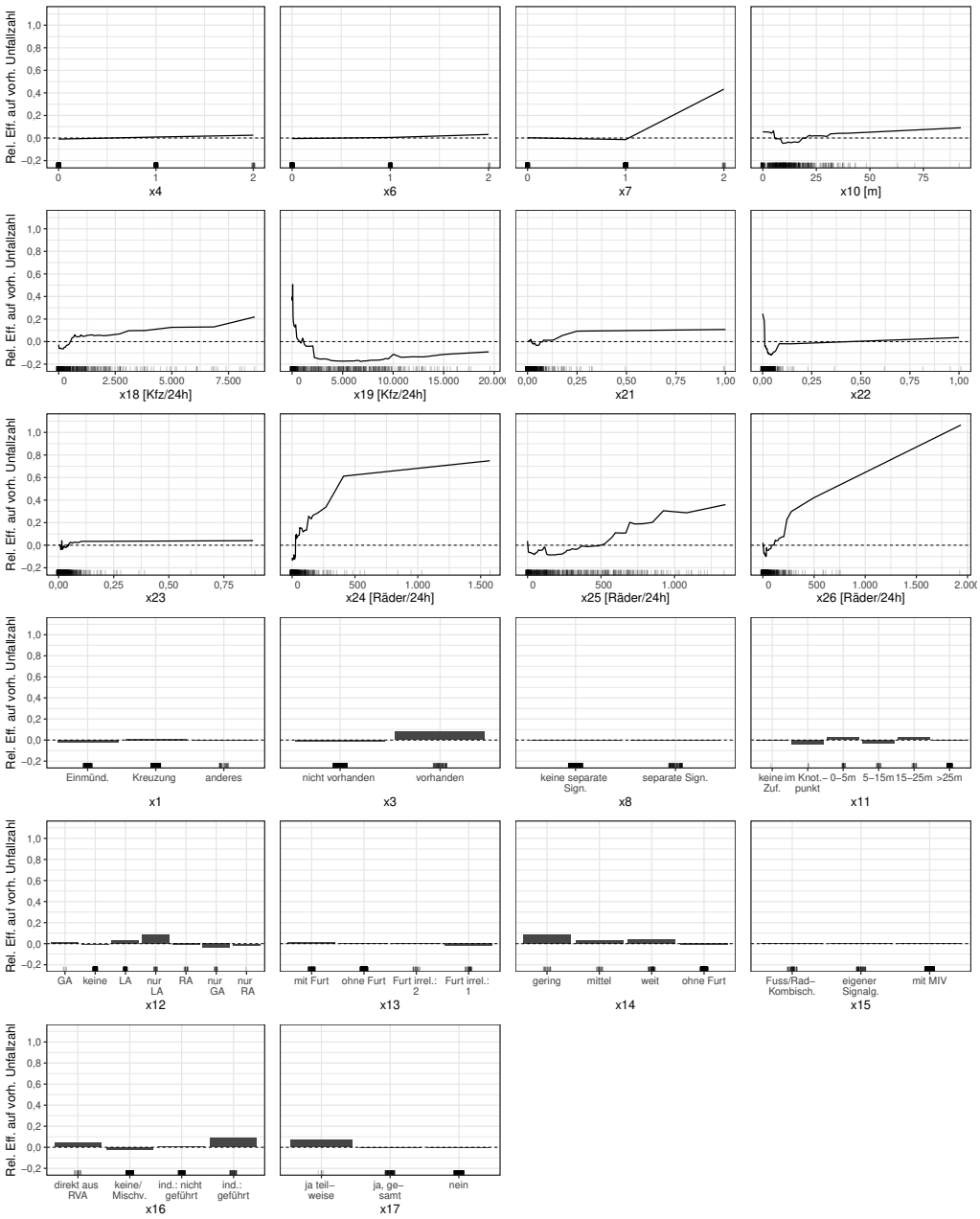


Abbildung D.4: Unfalltyp 30/32/34: ALE der übrigen Variablen [Quelle: eigene Darstellung]

D.3 Negative Binomialregression

Anmerkung zu Abbildung D.5: Die ursprünglichen Verteilungen der Expositionsvariablen lassen sich den oberen bzw. unteren Plots in Abbildung 4.21 entnehmen.

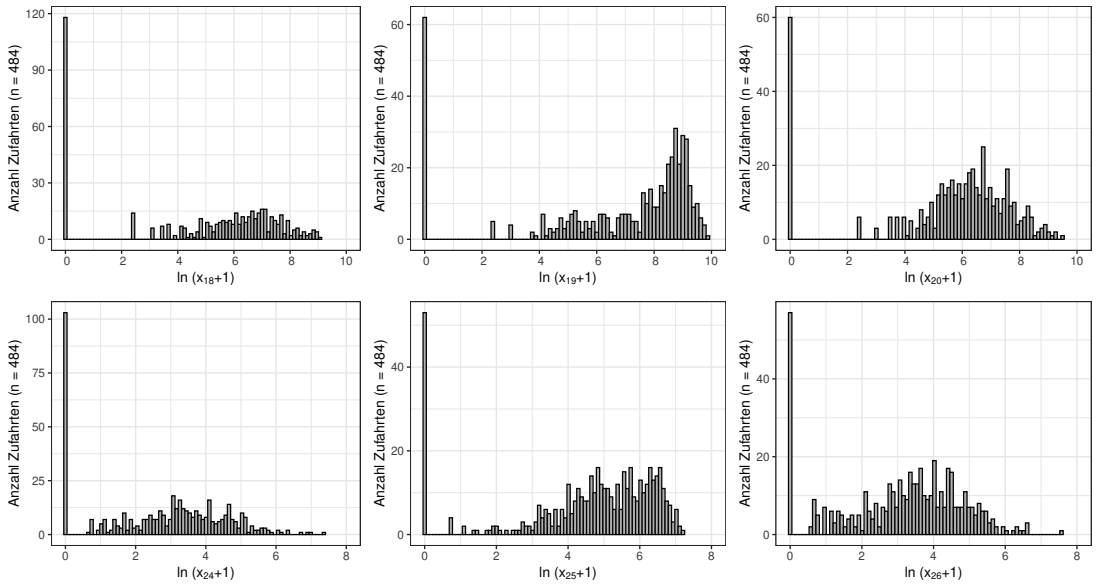


Abbildung D.5: Verteilung der Expositionsvariablen nach Logarithmierung [Quelle: eigene Darstellung]

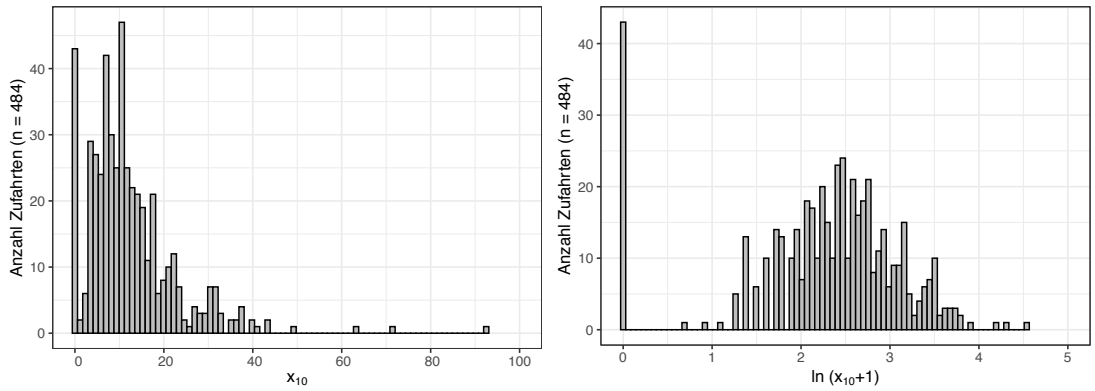


Abbildung D.6: Verteilung der Variable x_{10} vor und nach Logarithmierung [Quelle: eigene Darstellung]

Tabelle D.2: Variablenselektion mit Hilfe des LASSO-Verfahrens nach Unfalltyp

abhängige Variable: Anzahl der Unfälle vom Unfalltyp ..	.. 23/24 (y_1)	.. 30/32/34 (y_2)
$\ln(x_{18}+1)$	0,000	0,000
$\ln(x_{19}+1)$	0,294	-0,167
$\ln(x_{20}+1)$	0,187	0,159
$\ln(x_{24}+1)$	0,000	0,126
$\ln(x_{25}+1)$	0,365	0,000
$\ln(x_{26}+1)$	0,000	0,007
x_1: Kreuzung	Referenz	Referenz
x_1: Einmündung	0,000	-0,348
x_1: anderes	0,000	0,000
x_2: Rechts-vor-Links	Referenz	Referenz
x_2: Z205: Vorfahrt gewähren	0,000	0,715
x_2: Z206: Halt. Vorfahrt gewähren	0,000	0,000
x_2: Teil-LSA	0,000	0,404
x_2: LSA	0,339	0,000
x_3: nicht vorhanden	Referenz	Referenz
x_3: vorhanden	0,385	0,050
x_4	−0,234	0,000
x_5	0,191	0,334
x_6	0,040	0,000
x_7	0,000	0,000
x_8: keine separate Signalisierung	Referenz	Referenz
x_8: separate Signalisierung	0,000	0,000
x_9: keine separate Signalisierung	Referenz	Referenz
x_9: separate Signalisierung: dreifeldig	-0,561	-0,149
x_9: Dreiecksinsel ohne LSA	-0,238	-0,376
x_9: Dreiecksinsel mit LSA	0,000	0,000
x_9: separate Signalisierung: zweifeldig	-0,548	-0,099
$\ln(x_{10}+1)$	0,000	0,000
x_{11}: mehr als 25 m	Referenz	Referenz
x_{11}: Parken im Knotenpunkt	0,000	0,000
x_{11}: keine Zufahrt vorhanden	0,000	0,000
x_{11}: 0 m bis 5 m	0,000	0,000
x_{11}: mehr als 5 m bis 15 m	0,000	-0,208
x_{11}: mehr als 15 m bis 25 m	-0,124	0,000
x_{12}: keine	Referenz	Referenz
x_{12}: Linksabbiegeverbot	0,000	0,000
x_{12}: Rechtsabbiegeverbot	-0,187	0,000
x_{12}: Geradeausfahrverbot	0,000	-0,485
x_{12}: nur Linksabbiegen	0,000	0,000
x_{12}: nur Rechtsabbiegen	0,000	0,000

x_{12}: nur Geradeausfahren	0,000	-0,477
x_{13}: ohne Furt	Referenz	Referenz
x_{13}: mit Furt	-0,412	0,000
x_{13}: kein Konflikt mit Rechtsabbiegern[1]	0,000	0,000
x_{13}: nur Radverkehr als Rechtsabbieger[1]	0,466	0,000
x_{14}: ohne Furt	Referenz	Referenz
x_{14}: gering (bis 2 m)	-0,170	0,000
x_{14}: mittel (bis 4 m)	-0,081	0,000
x_{14}: weit (mehr als 4 m)	0,000	0,000
x_{15}: mit MIV	Referenz	Referenz
x_{15}: eigener Signalgeber	0,191	0,000
x_{15}: Fuß/Rad-Kombischeibe	0,048	0,000
x_{16}: keine/Mischverkehr	Referenz	Referenz
x_{16}: direkt aus Radverkehrsanlage	-0,801	0,000
x_{16}: indirekt: nicht geführt	0,293	0,000
x_{16}: indirekt: geführt	0,390	0,000
x_{17}: nein	Referenz	Referenz
x_{17}: ja, über gesamten Knoten	0,765	-0,486
x_{17}: ja, teilweise	2,370	0,000
x_{21}	-1,609	0,000
x_{22}	0,000	0,000
x_{23}	0,000	0,000

Anmerkung: Die Variablen, deren Koeffizienten nicht-null sind, bilden den Ausgangspunkt der jeweiligen Negativen Binomialregression.
[1]Furt irrelevant

Tabelle D.3: Ergebnisse nach Unfalltyp: Negative Binomialregression (vollständig)

abhängige Variable: Anzahl der Unfälle vom Unfalltyp ..	.. 23/24 (y_1)	.. 30/32/34 (y_2)
ln(K): Absolutglied		
	$-12{,}244^{***}$	$-1{,}572^{***}$
	(1,934)	(0,522)
ln(x_{19}+1): DTV Kfz Geradeausfahrende [Kfz/24h]		
	$0{,}720^{***}$	$-0{,}282^{***}$
	(0,215)	(0,047)
ln(x_{20}+1): DTV Kfz Rechtsabbieger [Kfz/24h]		
	$0{,}305^{***}$	$0{,}172^{**}$
	(0,084)	(0,070)
ln(x_{24}+1): DTV Radverkehr Linksabbieger [Räder/24h]		
		$0{,}198^{***}$
		(0,072)
ln(x_{25}+1): DTV Radverkehr Geradeausfahrende [Räder/24h]		
	$0{,}420^{***}$	
	(0,144)	
ln(x_{26}+1): DTV Radverkehr Rechtsabbieger [Räder/24h]		
		0,049
		(0,090)
x_1: Knotenpunktform		
Kreuzung		0
		(Referenz)
Einmündung		$-1{,}223^{***}$
		(0,341)
x_2: Verkehrsregelungsart		
Rechts-vor-Links	0	0
	(Referenz)	(Referenz)
Z205		$1{,}056^{***}$
		(0,327)
Teil-LSA		$1{,}297^{***}$
		(0,379)
LSA	0,474	
	(0,331)	
x_3: Grünpfeil (Z720)		
nicht vorhanden	0	0
	(Referenz)	(Referenz)
vorhanden	0,306	$0{,}772^{**}$
	(0,314)	(0,305)
x_4: Anzahl Linksabbiegefahrstreifen		

	−0,471*	
	(0,264)	
x_5: Anzahl Geradeausfahrstreifen		
	−0,055	1,047***
	(0,226)	(0,220)
x_6: Anzahl Rechtsabbiegefahrstreifen		
	−0,008	
	(0,319)	
x_9: Signalisierung Rechtsabbieger		
keine separate Signal.	0	0
	(Referenz)	(Referenz)
sep. Signal.: dreifeldig	−1,097	−1,281
	(0,719)	(0,868)
Dreiecksinsel ohne LSA	−0,621	−34,177
	(1,051)	(1,18e07)
sep. Signal.: zweifeldig	−1,659	−36,119
	(1,216)	(2,48e07)
x_{11}: Abstand ruhender Verkehr zum Knotenpunkt		
mehr als 25 m	0	0
	(Referenz)	(Referenz)
mehr als 5 m bis 15 m		−0,603
		(0,584)
mehr als 15 m bis 25 m	−32,988	
	(1,05e07)	
x_{12}: Abbiegeverbote		
keine	0	0
	(Referenz)	(Referenz)
Rechtsabbiegeverbot	−0,489	
	(0,767)	
Geradeausfahrverbot		−1,610
		(1,197)
nur Geradeausfahren		−34,927
		(8,87e06)
x_{13}: Radverkehrsführung im Knotenpunkt		
ohne Furt	0	
	(Referenz)	
mit Furt	−0,731	
	(0,490)	
Furt irrelevant: nur Radv.	2,103**	
als Rechtsabbiegende	(0,899)	
x_{14}: Absetzung Radfahrfurt		
ohne Furt	0	
	(Referenz)	
gering (bis 2 m)	−0,391	
	(0,759)	
mittel (bis 4 m)	0,040	

	(0,636)	
x_{15}: Signalgeber Radverkehr		
mit MIV	0	
	(Referenz)	
eigener Signalgeber	0,025	
	(0,327)	
Fuß/Rad-Kombischeibe	0,565	
	(0,452)	
x_{16}: Radverkehrsführung Linksabbieger		
keine/Mischverkehr	0	
	(Referenz)	
direkt aus Radv.-anlage	−1,279	
	(1,078)	
indirekt: nicht geführt	0,512*	
	(0,311)	
indirekt: geführt	0,914**	
	(0,417)	
x_{17}: Roteinfärbung Radfahrfurt		
nein	0	0
	(Referenz)	(Referenz)
ja, gesamter Knoten	0,694**	−0,963***
	(0,290)	(0,371)
ja, teilweise	3,123***	
	(0,634)	
x_{21}: Schwerverkehrsanteil Linksabbieger		
	−3,214	
	(3,554)	
Beobachtungen	484	484
Dispersion $\hat{\theta}$	1,731**	0,509***
	(0,879)	(0,124)
Log Likelihood	−190,369	−307,704

*Anmerkung: Standardfehler entsprechen dem Klammerausdruck. Signifikanzen der Wald-Teststatistik werden wie folgt ausgewiesen: *p<0,1; **p<0,05; ***p<0,01.*

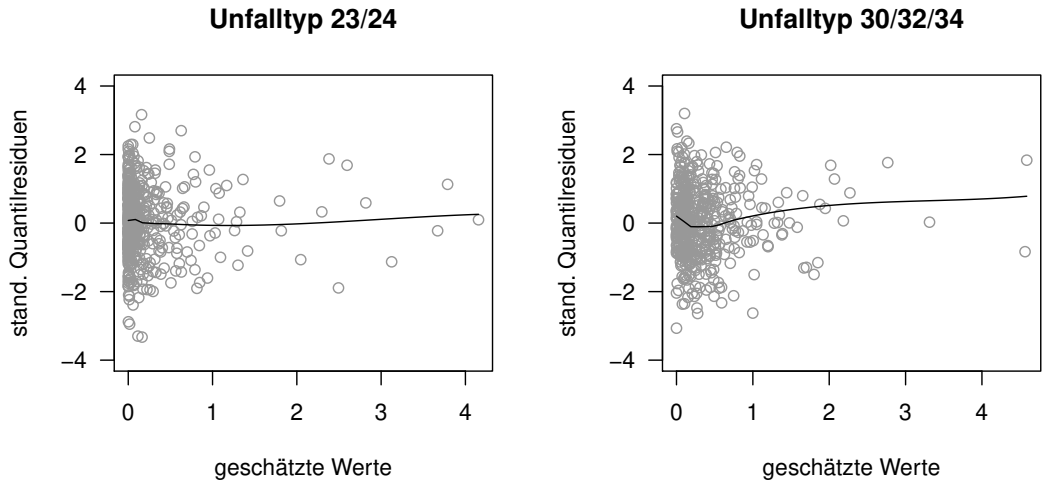


Abbildung D.7: Quantilresiduen vs. geschätzte Werte ohne varianz-stabilisierende Transformation der geschätzten Werte [Quelle: eigene Darstellung]

Selbstständigkeitserklärung

Ich erkläre hiermit, dass die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde. Die aus fremden Quellen wörtlich oder sinngemäß übernommenen Gedanken sind als solche kenntlich gemacht. Ich erkläre ferner, dass ich die vorliegende Arbeit an keiner anderen Stelle als Prüfungsarbeit eingereicht habe oder einreichen werde.

Dresden, 22. Dezember 2020



Jacqueline Martin