

Gerhard Krottmaier

**Statistische Evaluierung der Genauigkeit von BLIDS
Messsystemen für unterschiedliche Verkehrsszenarien**

DIPLOMARBEIT

**zur Erlangung des akademischen Grades eines
Diplom-Ingenieurs**

Diplomstudium Technische Mathematik



Technische Universität Graz

Betreuer:

Univ.-Prof. Dipl.-Ing. Dr. techn. Ernst Stadlober

**Institut für Statistik der Technischen Universität Graz
8010 Graz, Münzgrabenstraße 11/III**

Graz, im August 2010

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen örtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am.....

.....

(Unterschrift)

STATUARY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

.....

date

.....

(signature)

Kurzfassung

Mit Hilfe von BLIDS-Sensoren auf Basis der Bluetooth-Technologie werden Stichproben der Gesamtverkehrsfrequenz im Straßenverkehr gewonnen. Dabei haben wir zwei unterschiedliche Verkehrsszenarien untersucht und analysiert. An der Autobahn A6 (2 Fahrstreifen) im Großraum Nürnberg wurden zwischen den Messpunkten Langwasser und Alfeld die Geschwindigkeiten von Verkehrsteilnehmern (PKW und LKW) in beiden Richtungen an jeweils 5 Werktagen, Samstag und Sonntag (6.00-18.00 Uhr) erhoben. Die Geschwindigkeitsverteilung konnte durch eine Mischung von zwei oder drei Gamma-Verteilungen modelliert werden. Daraus lassen sich für jede Gruppe aussagekräftige Kenngrößen wie Durchschnittsgeschwindigkeit und Streubereich berechnen.

In zwei Kreuzungsbereichen von Graz (Wetzelsdorf und Straßgang) wurden an 42 bzw. 35 Tagen Bluetooth-Stichproben mit Gesamtverkehrsfrequenzen (Querschnittsmessungen durch Radar) verglichen. Multiple lineare Regressionsmodelle für die Gesamtverkehrsfrequenz in Abhängigkeit von Bluetoothfrequenz, Tagestyp und Tageszeit wurden erstellt und analysiert. Weiters konnten Bluetooth-Schätzungen für die Frequenzen der Routenverläufe (Längsschnittmessung) angegeben werden. Sämtliche Analysen wurden mit dem Softwarepaket R 2.10.1 (Public Domain) durchgeführt. Die Arbeit entstand in Kooperation mit der Firma c.c.com Andersen&Moser GmbH, Grambach.

Abstract

Samples of the traffic frequency are obtained by BLIDS sensors based on the Bluetooth technology. We investigated and analyzed two different traffic scenarios. On highway A6 (two lanes) in the metropolitan area of Nuremberg we collected the velocities of cars and trucks between Langwasser and Alfeld in both directions at five working days, Saturdays and Sundays (6.00 am to 6.00 pm) each. Distributions of velocity were modeled by mixtures of two or three gamma distributions. For each group meaningful statistics as average velocity and standard deviation of velocity are available.

At two crossroads in Graz (Wetzelsdorf and Straßgang) we compared Bluetooth samples with overall traffic frequencies (cross section data via radar) at 42, respectively 35 days. We represented overall traffic frequencies by multiple linear regression models based on Bluetooth frequencies, type of the day and daytime. Additionally we were able to deliver Bluetooth estimates of route frequencies (longitudinal data). Our data analysis was based on the public domain package R 2.10.1. The thesis was carried out in cooperation with c.c.com Andersen&Moser GmbH, Grambach.

Danksagung

Ich möchte mich herzlich bei Herrn Univ.-Prof. DI Dr. Ernst Stadlober für die freundliche und engagierte Unterstützung bei meiner Arbeit bedanken. Ein besonderer Dank gilt Herrn Burkhard Moser von der Firma c.c.com Anderson & Moser GmbH, der mir diese Arbeit ermöglicht hat. Sowohl Herr Stadlober als auch Herr Moser erlaubten mir ein freies und unabhängiges Arbeiten. Meinen Arbeitskollegen der Firma c.c.com verdanke ich neben einem hervorragenden Arbeitsklima auch tatkräftige Unterstützungen bei der Datenaufbereitung. Ein besonderer Dank geht an meine Mutter und an meine Freunde, die mir stets mit aufmunterten Worten zur Seite standen.

Lannach, im August 2010

Gerhard Krottmaier

Inhalt

Kapitel 1.....	1
1.1 Vorwort	1
1.1.1 Schematische Darstellung einer Bluetooth-erfassung	3
Kapitel 2.....	5
2.1 Reisezeitanalyse Einleitung:	5
2.2 Allgemeine Bemerkungen:	6
2.3 Vorgehensweise:	9
2.4 Beispiel R – Code einer Mischverteilung.....	11
2.5 Praktische Analyse.....	13
2.5.1 Vorliegende Daten.....	13
2.5.2 Bestimmung einer Verteilung.....	13
2.5.3 Mischverteilungsanalyse Werktag Route AB	17
2.5.4 Mischverteilungsanalyse Samstag Route AB.....	20
2.5.5 Mischverteilungsanalyse Sonntag Route AB.....	24
2.5.6 Mischmodelle der Route AB	29
2.5.7 Mischmodelle der Route BA	30
2.5.8 Zusammenfassung.....	32
2.5.9 Conclusio	37
2.6 Mathematische Grundlagen von Mischverteilungen	39
2.6.1 Beispiel:	39
2.6.2 Diskrete Mischung diskreter Verteilungen	40
2.6.3 Diskrete Mischung stetiger Verteilungen	41
2.6.4 Beispiele von Mischungen	42
2.6.4 Gammaverteilung.....	44
2.6.5 Chi – Quadrat – Test	45
Kapitel 3.....	46
3.1 Verkehrsfrequenzschätzung mittels Regressionsanalyse	46
3.2 Allgemeine Bemerkungen	46
3.3 Vorgehensweise	51
3.4 Beispiel R – Code	53
3.5 Praktische Analyse.....	55
3.5.1 Lineares Regressionsmodell ohne Steigung: Sensor A.....	55
3.5.2 Lineares Regressionsmodell mit Steigung: Sensor B.....	59

3.5.3 Regressionsmodell mit quadratischen Term: Sensor C.....	65
3.5.4 Lineares Regressionsmodell mit Steigung: Sensor D	69
3.5.5 Regressionsmodell mit quadratischem Term: Sensor E	73
3.5.6 Regressionsmodell mit quadratischem Term: Sensor F.....	76
3.5.7 Regressionsmodell mit quadratischem Term: Sensor G	80
3.5.8 Frequenz-, Bluetooth- und Quotenverläufe aller Messstationen	84
3.5.9 Zusammenfassung.....	87
3.5.10 Beispiel Routenauswertung	88
3.5.11 Conclusio	97
3.6 Mathematische Grundlagen von Regressionsanalysen	99
3.6.1 Einfache lineare Regression	99
3.6.2 Parameterschätzung.....	100
3.6.3 Grafische Interpretation	101
3.6.4 Multiple lineare Regression	102
3.6.5 Bestimmtheitsmaß R^2	103
3.6.6 Adjustiertes Bestimmtheitsmaß $R^2(\text{adj})$	104
3.6.7 Bemerkung: Overfitting	104
3.6.8 Angewandte statistische Tests	104
3.6.9 Residuenanalyse.....	105
3.6.10 Varianzstabilisierende Transformationen.....	108
3.6.11 Vertrauens- und Prognoseintervalle für Parameter	109
Schlussbemerkung.....	111
Literaturverzeichnis.....	112

Kapitel 1

Im Vorwort beziehen wir uns auf den *Endbericht Intelligente Infrastruktur, BLIDS 2* aus dem Jahre 2006 [9].

1.1 Vorwort

Um in der heutigen Zeit als Individuum den Anforderungen der Gesellschaft Rechnung tragen zu können, muss man mobil sein. Die Hauptsäule dieser Mobilität ist das Kfz. Sei es um zur Arbeit zu gelangen, Einkäufe zu erledigen oder nur seine Freizeit zu verbringen, alles hängt von der Erreichbarkeit eines örtlichen Zieles ab. Diese Freiheit der Mobilität bringt natürlich auch negative Seiten mit sich. Neben den umweltbedingten Auswirkungen des Verkehrs wird das derzeit existierende Straßennetz bis an seine Kapazitätsgrenzen belastet. Die Frage die sich diesbezüglich stellt ist, kann man den Verkehrsstrom so steuern, dass eine gleichmäßige Auslastung des Wegenetzes zustande kommt. Dies kann aber nur bewerkstelligt werden, wenn man über die Verkehrsfrequenz verschiedener Routen genau Bescheid weiß.

In der Vergangenheit wurden Zählungen hauptsächlich mittels Detektorschleifen, Radargeräten oder Bilderkennungssystemen durchgeführt. Diese Methoden erfordern neben zum Teil erheblichen baulichen Maßnahmen auch intensive Wartungsarbeiten. Das alles ist daher mit einem erheblichen finanziellen Aufwand verbunden.

Im Rahmen des Projekts *BLIDS - Network (Bluetooth-Based Traffic Data Collection System Network)* wird bzw. wurde ein neuartiges System entwickelt, das Bluetooth – Signale diverser elektronischer Items, wie Mobiltelefone oder Navigationsgeräte, empfängt.

Im Straßenbereich, wo eine Erfassung gewünscht wird, werden eigens dafür entwickelte Sensoren an geeignete Montageorte wie Laternen oder Hinweistafeln montiert. Diese Access Points registrieren ständig aktivierte Bluetooth – Schnittstellen in einem gewissen Umfeld. Dank der variabel einstellbaren Reichweite, ist eine relativ genaue Eingrenzung des Erfassungsbereiches möglich. Frequentiert nun ein Fahrzeug mit aktivierter Bluetooth – Schnittstelle einen Sensor, wird ein Datenpaket, das das Datum der Erfassung inklusive Uhrzeit und eine individuelle ID – Kennnummer enthält, erhoben und weitergeleitet. Wird dieselbe Identifikationsnummer an zwei Sensoren an unterschiedlichen Orten registriert, ist damit die Ermittlung einer Fahrtroute und der Reisegeschwindigkeit möglich. Somit kann man mit einem relativ geringen finanziellen und logistischen Aufwand, Stichproben des Gesamtverkehrsaufkommens erheben. In diesem Zusammenhang möchten wir auf eine aus dem Jahre 2006 datierte und vom Bundesministerium für Verkehr, Innovation und Technologie initiierte Analyse, mit dem Titel „Intelligente Infrastruktur BLIDS 2“ (Autoren: c.c.com GmbH, DI Konrad Maitisch, Institut für Statistik, TU Graz, UProf. DI Dr. Ernst Stadlober, Institut für Straßen- und Verkehrswesen, TU Graz. DI Dr. Kurt Fallast und Stadt Graz, Straßenamt, DI Dr. Winfried Höpfl) verweisen.

Ein weiteres Projekt, das mit der Erfassung von Bluetoothsignalen arbeitet, ist **PARSON (Passenger Registration System based on bluetooth and satellite Navigation)**. Ziel dieses Einsatzgebietes ist es automatisierte Analysen von Fahrgastströmen in öffentlichen Verkehrsmitteln durchzuführen, um aufgrund dieser Daten das Angebot an die Bedürfnisse der potenziellen Benutzer anzupassen. Dieses System ist seit dem Jahre 2008 für die Grazer Verkehrsbetriebe im Realbetrieb.

Ziel meiner Diplomarbeit war es nun folgende Fragen zu klären:

- Ist es aufgrund dieser erhobenen Stichproben möglich, aussagekräftige Analysen hinsichtlich Verkehrsfrequenz, Verkehrsströme und Reisezeiten zu treffen.
- Wenn ja, wie genau sind diese Aussagen?
- Welche Voraussetzungen hinsichtlich der Datenqualität und Datenumfang müssen diese Stichproben aufweisen.
- Sind kurzfristige Verkehrsprognosen hinsichtlich der Dichte und Reisezeit möglich?
- Kann man notwendige Programmpakete aus dem public domain Softwarepaket R Version 2.10.1 in das firmeneigene System implementieren?
- Welche zusätzlichen Einsatzmöglichkeiten sind mit dieser Art der Datenerhebung noch möglich?

Um all diese Fragen strukturiert zu beantworten war es notwendig diese Arbeit in 2 Kapitel zu unterteilen:

- ❖ Reisezeitanalyse mittels Mischverteilungsanalyse
- ❖ Hochrechnung bzw. Schätzung der Verkehrsfrequenz mittels Regressionsanalyse

1.1.1 Schematische Darstellung einer Bluetoothfassung

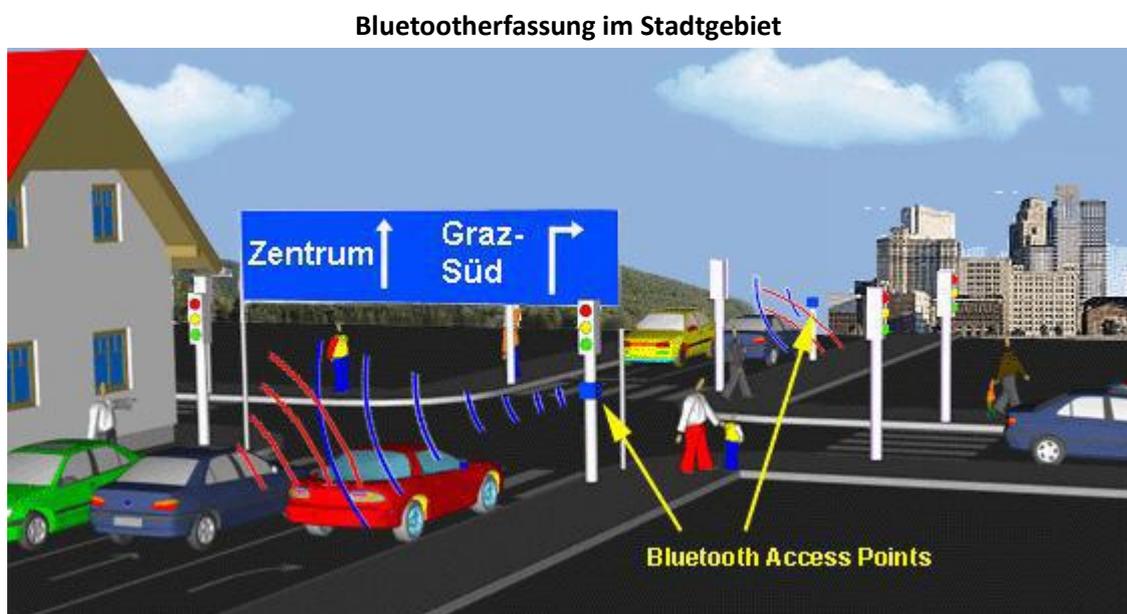
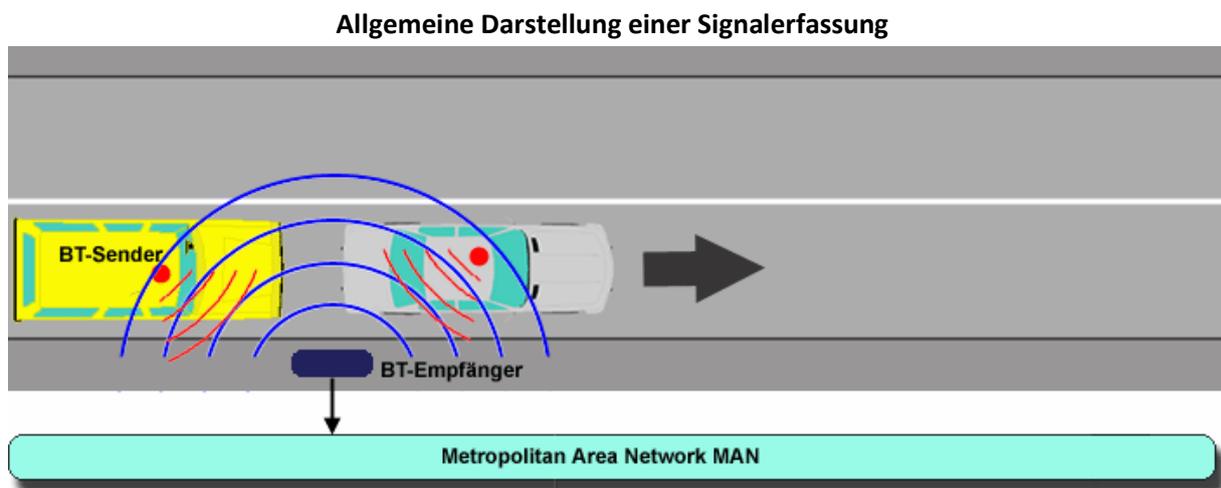
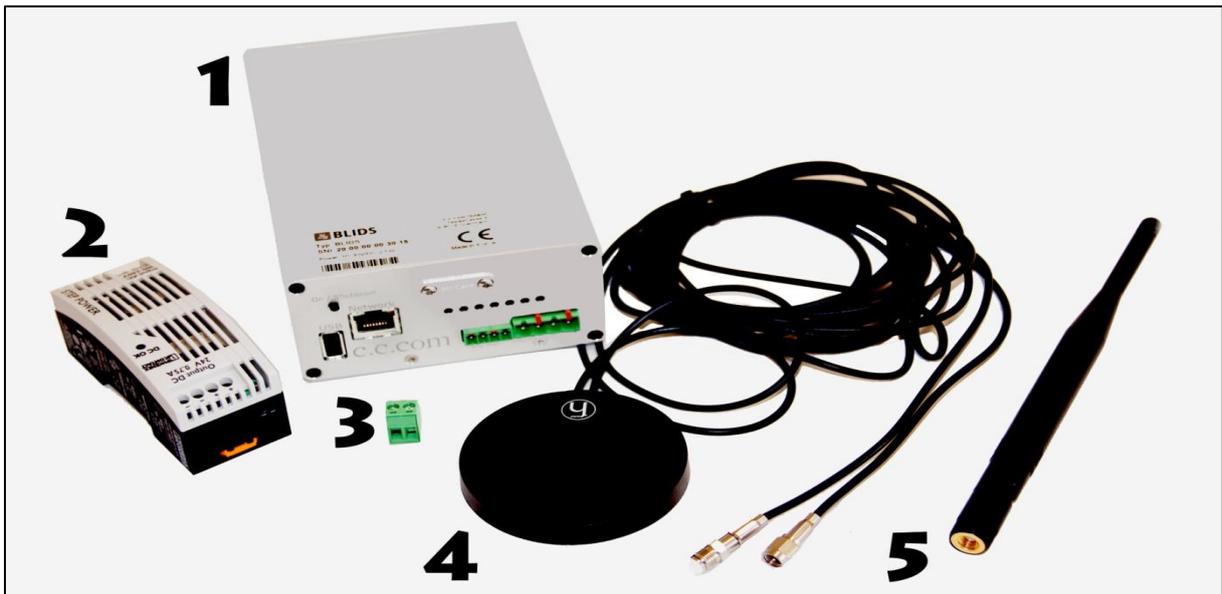


Abb. BLIDS2 Bluetooth-Based Traffic Data Collection System Demonstrator

BLIDS - Sensor



- 1 BLIDS-Sensor
- 2 Netzteil – 24 Volt Netzteil
- 3 Kabelklemme
- 4 Kombinierte GSM und GPRS-Antenne
- 5 Bluetooth-Antenne

Sensor Einstellungen :

Als Referenz dient die allgemeine Konfigurations- und Installationsanleitung der Firma c.c.com Andersen & Moser GmbH [15].

✓ **Inquiry-Length:**

Die Länge des Zyklus beeinflusst die maximale Anzahl der unterschiedlichen Bluetooth-IDs, welche pro Sequenz empfangen werden können. Die Menge der registrierten Signale hängt zusätzlich von externen Faktoren wie der Geschwindigkeit von erfassten Verkehrsteilnehmern oder der baulichen Beschaffenheit des Umfelds ab.

✓ **Inquiry-Signal-Length:**

Diese Einstellung begrenzt den Reichweitenempfang eines Sensors. Mittels einer Standard-dBi-Antenne ergeben sich folgende Empfangsleistungen:

- Class C: 50 m
- Class B: 100 m
- Class A: 200 m

Kapitel 2

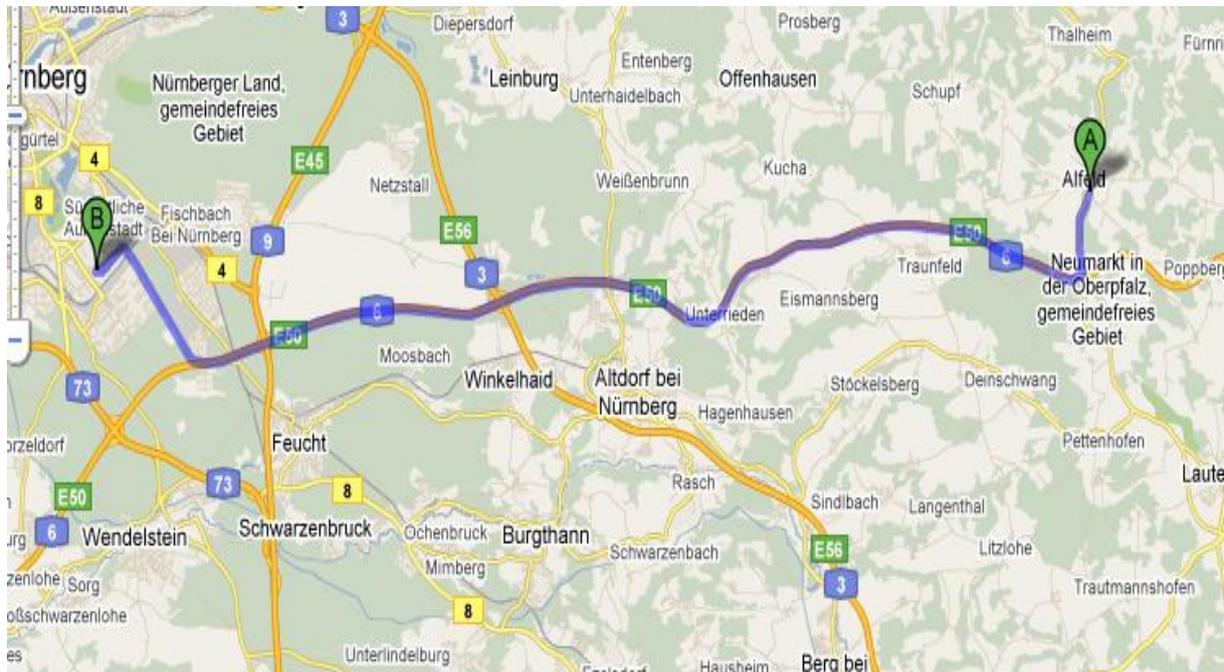
Grundlage für die folgende Analyse ist die Annahme, dass die Grundgesamtheit der erhobenen Reisezeiten aus mehreren Zeitbereichen mit Häufungen besteht. Stellt man sich nämlich den Verkehr auf einer mehrspurigen Fahrbahn vor, unter Berücksichtigung der Fahrzeugtypen (LKW und PKW), wird es zumindest 2 Geschwindigkeitsgruppen geben. Eine schnellere, die hauptsächlich den PKWs zuzuordnen ist, und eine langsamere, die überwiegend den LKWs zuzuordnen ist.

2.1 Reisezeitanalyse Einleitung:

Die Motivation einer Reisezeitanalyse ist es den Gesamtverkehr hinsichtlich seiner Reisezeit bzw. seiner Geschwindigkeit in Gruppen zu unterteilen und aufgrund der diesbezüglich ermittelten Kennzahlen Diagnosen über Abnormalitäten wie Zunahme der Verkehrsdichte oder Stauverhalten zu stellen.

Grundlage für diese Analyse sind sogenannte Pärchen. Wird die selbe eindeutige ID – Nummer an zwei verschiedenen Messpunkten registriert, kann man aus der Differenz der Erfassungszeiten die Reisezeit und somit die Reisegeschwindigkeit ermitteln. Im Normalfall kommt es vor, dass dieselbe ID – Nummer im Empfangsbereich eines Sensors mehrmals erfasst wird. Jede Registrierung wird mit einem Zeitstempel versehen. In einem solchen Fall wird diese `Reihenerfassung` zu einer einzigen Registrierung addiert, d. h. alle identischen Bluetoothadressen werden zu einer einzigen zusammengefasst. Der Zeitpunkt dieser Erfassung ist das arithmetische Mittel aller Zeitstempel derselben ID – Nummer am selben Sensor.

2.2 Allgemeine Bemerkungen:



Die für diese Analyse zugrunde liegenden Daten wurden auf der A 6 im Großraum Nürnberg, zwischen Langwasser (B) und Alfeld (A) erhoben. Die Teststrecke verfügt über mehrere Auf- bzw. Abfahrten. Dadurch ist der Pärchenbildungsindex, d. h. der Anteil der Paare an den BT – Scans, deutlich niedriger, als an jenen Strecken, die als Korridor (keine Auf- und Abfahrten) einzustufen sind.

Die Weglänge zwischen den beiden Messpunkten beträgt rund 31 km und umfasst 2 Fahrspuren pro Richtung. An Sonn- und Feiertagen sowie nachts zwischen 22.00 Uhr und 6.00 Uhr früh herrscht im Erfassungsgebiet ein LKW Nachtfahrverbot. Aufgrund unterschiedlicher Verkehrscharakteristiken an Werktagen, Samstagen und Sonntagen ist eine Unterteilung bezüglich dieser Tagestypen unbedingt erforderlich. Für diese Analyse wurden von jedem Tagestyp 5 Tage aus folgenden Zeiträumen ausgewählt:

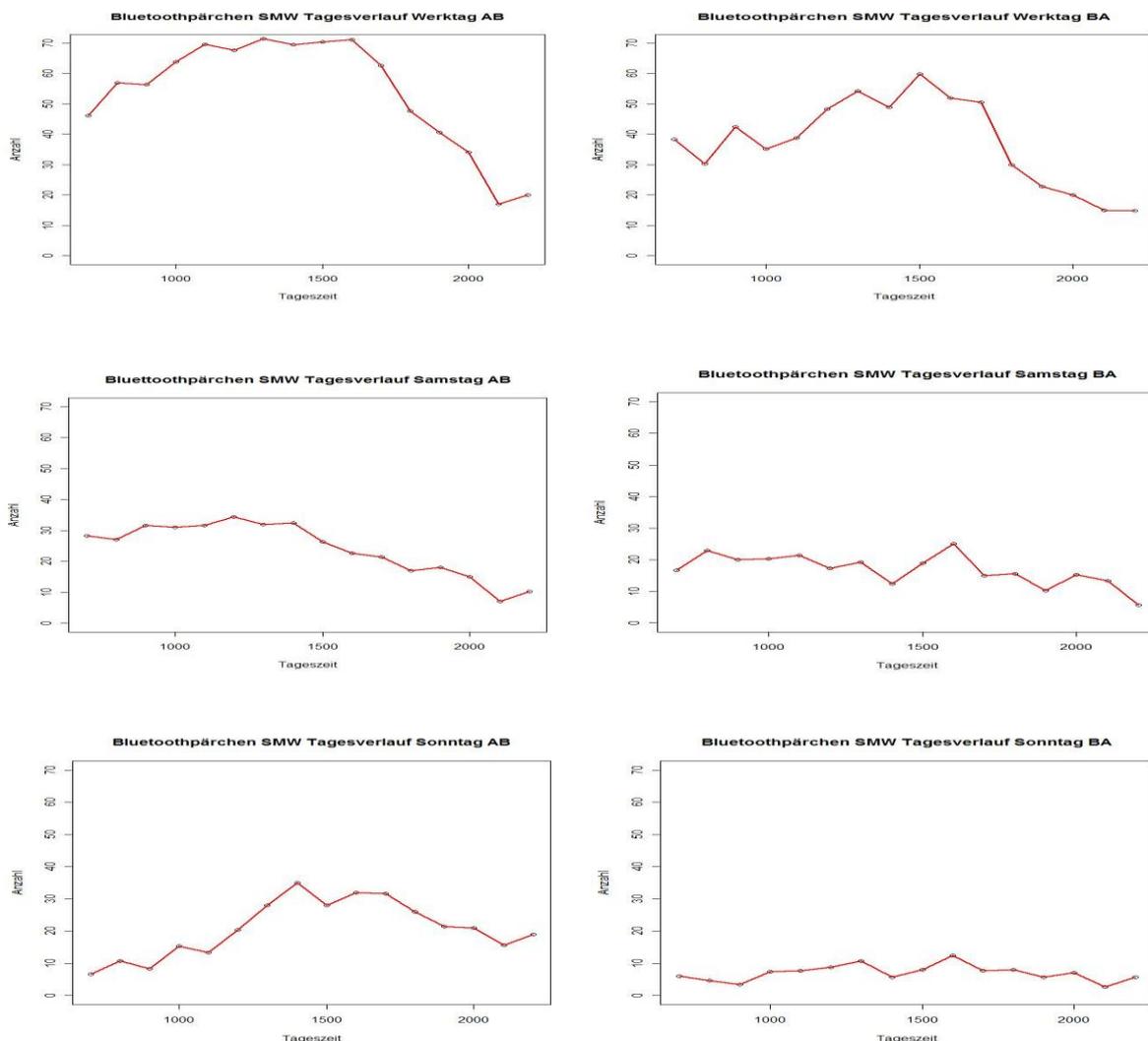
- Werktage: 25. – 29. Jänner 2010
- Samstage: 9. 16. 23. Jänner, 27. Februar, 6. März 2010
- Sonntage: 10. 17. 24. Jänner, 28. Februar, 7. März 2010

Alle Tage des gleichen Typs sind als gleichwertig zu betrachten. Darunter versteht man, dass nur jene Tage herangezogen werden, an denen keine Abweichung vom 'üblichen' Verhaltensmuster des Verkehrs zu erwarten ist (keine Feiertage, keine Umleitungen, keine Baustellen, kein Ferienbeginn, etc.). Da die Verkehrsdichte im Laufe eines Tages natürlichen

Schwankungen unterliegt, werden zusätzlich Zeitintervalle gebildet. Dazu ist es notwendig, die Stundenmittelwerte SMW, der Pärchenbildungen, in den jeweiligen Tagesverläufen zu vergleichen. Im Vordergrund steht die Überlegung jene Stunden zu einem Intervall zusammenzufassen, die sich in etwa auf dem gleichen Niveau befinden. Außerdem ist es sinnvoll, zusammenhängende Betrachtungsabschnitte festzulegen.

Intervalle: Werktag 6 – 09.59 Uhr (4h) 10 – 16.59 Uhr (7h) 17 – 22.00 Uhr (5h)
 Samstage 6 – 13.59 Uhr (8h) 14 – 16.59 Uhr (3h) 17 – 22.00 Uhr (5h)
 Sonntage 6 – 10.59 Uhr (5h) 11 – 17.59 Uhr (7h) 18 – 22.00 Uhr (4h)

Untersucht wurden die verschiedenen Zeitabschnitte sowohl auf der Route AB, als auch auf der Rückroute BA. Nachfolgend sind die für die Intervallbildung notwendigen Plots angegeben.



Allgemein ist bei Analysen von Mischverteilungen der Stichprobenumfang für die Gewährleistung einer gewissen Aussagekraft enorm wichtig. Um gültige Erkenntnisse zu gewinnen, sollen mindestens 300 Pärchen für ein Betrachtungsabschnitt registriert werden.

Ein Unterschreiten dieser Anzahl ist optisch, am sehr unruhigen Verlauf des Dichteschätzers, erkennbar und weiters an der ungewöhnlich hohen Anzahl von Geschwindigkeitsklassen. Für ein aktives Eingreifen in das Verkehrsgeschehen ist es natürlich notwendig und sinnvoll Mischverteilungsanalysen mit aktuellen Daten durchzuführen. Sollte sich die Durchdringungsrate der Bluetoothtechnologie im Verkehr weiterhin enorm steigern, ist ein solches Unterfangen durchaus möglich und eine Anpassung des Beobachtungszeitraumes, auf kürzere Zeitintervalle, zu empfehlen.

2.3 Vorgehensweise:

Mittels einer explorativen Datenanalyse, im speziellen anhand eines Histogramms mit geplottetem Dichteschätzer, ist es möglich, erstmalig zu klären, ob eine Mischverteilung vorliegt. Sind mehrere 'Pikes' erkennbar, so kann man von mehreren Geschwindigkeitsklassen ausgehen. Anhand des Verlaufes des empirischen Dichteschätzers lassen sich etwaige schiefe Verteilungen erkennen. Eine Schiefe liegt dann vor, wenn nach einem starken (schwachen) Anstieg, ein flacher (steiler) Abfall der Datenkonzentrationen erfolgt.

Der zweite Schritt beinhaltet die Aufteilung der Rohdaten, nach der Anzahl der Beobachtungen, in den verschiedenen Geschwindigkeitsklassen. Diese Aufteilung ist mit jener in einem Histogramm ident. Die Klassenbreite kann individuell durch die Angabe einer Klassenanzahl festgelegt werden. Eine solche Änderung ist aber nur dann durchzuführen, wenn es eine ersichtliche Diskrepanz zwischen der Anzahl der Pikes im Histogramm, mit Dichteschätzer und der Anzahl an Klassen mit Beobachtungsspitzen gibt. Ist ein solcher Unterschied aber nicht erkennbar, führt eine Änderung dieses Parameters zu einem instabilen Mischmodell und die Analyse wird von Fehlermeldungen begleitet. Deswegen wurden die automatisch berechneten Intervallgrenzen nur in Ausnahmefällen verändert.

Zur weiteren Analyse ist es nun unabdingbar, den vorliegenden gruppierten Datensatz als Mischverteilung zu kennzeichnen.

Im vierten Schritt werden die **Startwerte** für die Iteration und die Art der Verteilung, mit denen eine möglichst genaue Approximation an die tatsächliche Mischverteilung erzielt wird, festgelegt, wobei die Verteilungsart die Anzahl jener Parameter, die vom Benutzer angegeben werden müssen, bestimmt. Für schiefe Verteilungen sind jeweils die Iterationsstartwerte für die Geschwindigkeitsmittelwerte und Streuungen vom Benutzer anzugeben. Zusätzlich ist der Parameter 'mixconstruct' mit 'CCV' (= Gleichheit der Variationskoeffizienten zwischen den Geschwindigkeitsklassen) festzulegen [7].

Die ungefähren Mittelwerte der einzelnen Geschwindigkeitsklassen können aufgrund der vorangegangenen Gruppierung der Rohdaten erruiert werden. Die Schätzung der Standardabweichungen für die einzelnen Mittelwerte ist etwas diffiziler. Deren Evaluierung erfolgt wiederum aufgrund des gruppierten Datensatzes. Man bestimmt die ungefähren Geschwindigkeitsdifferenzen zwischen den einzelnen Mittelwerten, bzw. zwischen den Mittelwerten des höchsten und des niedrigsten Geschwindigkeitsbereiches und teilt diesen Wert durch die Anzahl der dazwischenliegenden Geschwindigkeitsklassen. Diese Berechnung basiert auf der Schätzfunktion der Standardabweichung. Zu große Abweichungen vom iterierten Idealwert werden mittels Fehlermeldungen seitens des Programmpaketes R negiert und müssen folglich korrigiert werden.

Die Eingabe eines vierten Parameters ist optional, kann also vom Benutzer entschieden werden. Dieser Gewichtungskoeffizient ordnet jeder Geschwindigkeitsklasse einen prozentuellen Anteil des Gesamtverkehrs zu. Kann aufgrund der Eingabe von Iterationsstartwerten für die Mittelwerte und den dazugehörigen Standardabweichungen keine ideale Mischverteilung gefunden werden, sollte dieser Koeffizient seitens des Benutzers angegeben werden.

Grundsätzlich werden vom verwendeten R - Programmpaket 'mixdist' die Normal-, Lognormal-, Gamma-, Weibull-, Binomial-, Nbinomial- und Poissonverteilung unterstützt [7].

Welche asymmetrische Verteilung die beste Annäherung an die vorliegende Stichprobe liefert, kann nur durch einen empirischen Vergleich der verschiedenen Analysen herausgefunden werden.

Zusammenfassend kann man sagen, dass für eine Mischverteilungsanalyse eine Startwertkonfiguration vom Benutzer gefordert wird, unter Beachtung des jeweiligen Verteilungstyps.

2.4 Beispiel R – Code einer Mischverteilung

#Installieren der Bibliothek

```
library(mixdist)
```

#Gruppierung des Datensatzes

```
woche.grouped.AB1 <- mixgroup((31000/na.omit(dur_woche_AB[,1])*3.6))
```

mixgroup : Gruppierung des Rohdatensatzes in verschiedene
Geschwindigkeitsklassen

#Definition als Mischverteilung des gruppierten Datensatzes

```
woche.mix.AB1 <- as.mixdata(woche.grouped.AB1)
```

as.mixdata: Gruppiertes Datensatz wird als Mischverteilung für weitere Analyse
definiert

#Festlegung der Startwerte für die Mischungsanteile pi, Verteilungsparameter mu und se,
und der Verteilung selbst

```
fit.woche.AB1 <- mix(woche.mix.AB1,mixparam(pi=c(0.05, 0.75, 0.2), mu=c(30,100, 130),  
sigma=c(5, 15, 15)), "gamma", mixconstr(consigma="CCV"))
```

mix: Durchführung einer Mischverteilungsanalyse, Kennzahlen der einzelnen
Geschwindigkeitsgruppen werden iteriert

pi: Anteil der Geschwindigkeitsgruppen am Gesamtverkehr, Eingabe der
Startwerte optional

mu: geschätzte Iterationsanfangswerte der einzelnen
Klassendurchschnittsgeschwindigkeiten, Eingabe erforderlich

sigma: geschätzte Standardabweichungen der
Klassendurchschnittsgeschwindigkeiten, Eingabe erforderlich

gamma: Verteilungsangabe der Rohdaten, Eingabe erforderlich

mixconstr: Festlegen von übereinstimmenden Variationskoeffizienten
Eingabe erforderlich

#Ausgabe der iterierten Klassenparameter und Ergebnis des Chi – Quadrat - Tests

```
summary(fit.woche.AB1)
```

Parameters:

pi **mu** **sigma**

1 0.03271 21.59 2.998

2 0.68393 88.13 12.237

3 0.28336 129.62 17.997

Standard Errors:

```

pi.se mu.se sigma.se
1 0.004979 0.6256 0.1392
2 0.017431 0.6032 NA
3 0.017134 1.3971 NA

```

Analysis of Variance Table

```

Df Chisq Pr(>Chisq)
Residuals 4 163.07 < 2.2e-16 ***

```

pi: Anteil am Gesamtverkehr der jeweiligen Gruppe
mu: Durchschnittsgeschwindigkeiten
sigma: Standardabweichung
Standard Errors: Standardfehler für geschätzte Parameter
2.2e-16 *:** p-Wert des Chi-Quadrat-Tests
NA: NULL (kein Wert)

Aufgrund der übereinstimmenden Variationskoeffizienten werden durch die Mischverteilungsanalyse für 3 Geschwindigkeitsklassen insgesamt 7 „freie Parameter“

geschätzt. Durch $\frac{\mu_1}{\sigma_1} = \frac{\mu_2}{\sigma_2} = \frac{\mu_3}{\sigma_3}$ wird außerdem nur ein Fehler für die

Standardabweichung angegeben, da dieser für alle 3 Standardabweichungen gilt. Die fehlenden Werte werden mit ' NA ' aufgefüllt.

#Plotten der Mischverteilung und des Dichteschätzers

```

plot(fit.woche.AB1, main="Mischmodell Wochentage Route AB 6 bis 9.59 Uhr",
xlab="Geschwindigkeiten in km/h")
lines(density((31000/na.omit(dur_woche_AB[,1])*3.6)), col="grey", lwd=2)

```

2.5 Praktische Analyse

2.5.1 Vorliegende Daten

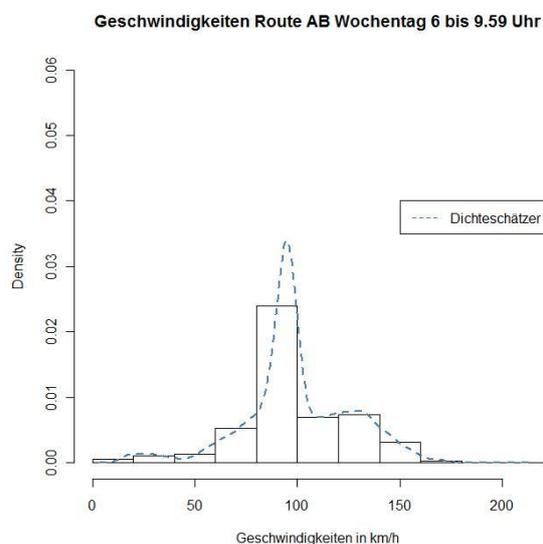
- ID – Nummer
- Zeitpunkt der erstmaligen Erfassung
- Sensor der erstmaligen Erfassung
- Zeitpunkt der letzten Erfassung
- Sensor der letzten Erfassung

Sollte dieselbe ID – Nummer am selben Sensor mehrmals erfasst werden, so wird als Zeitpunkt der Erfassung das arithmetische Mittel herangezogen. Dasselbe gilt für die letztmalige Registrierung. Übersteigt die Reisezeit einen Wert von 2 Stunden, so wird die betreffende Bluetooth – Adresse gelöscht.

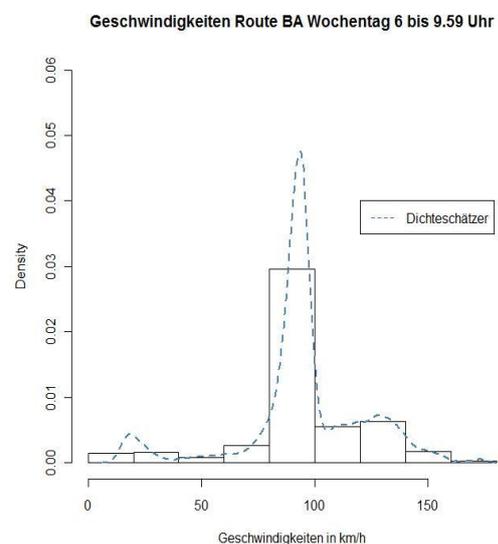
2.5.2 Bestimmung einer Verteilung

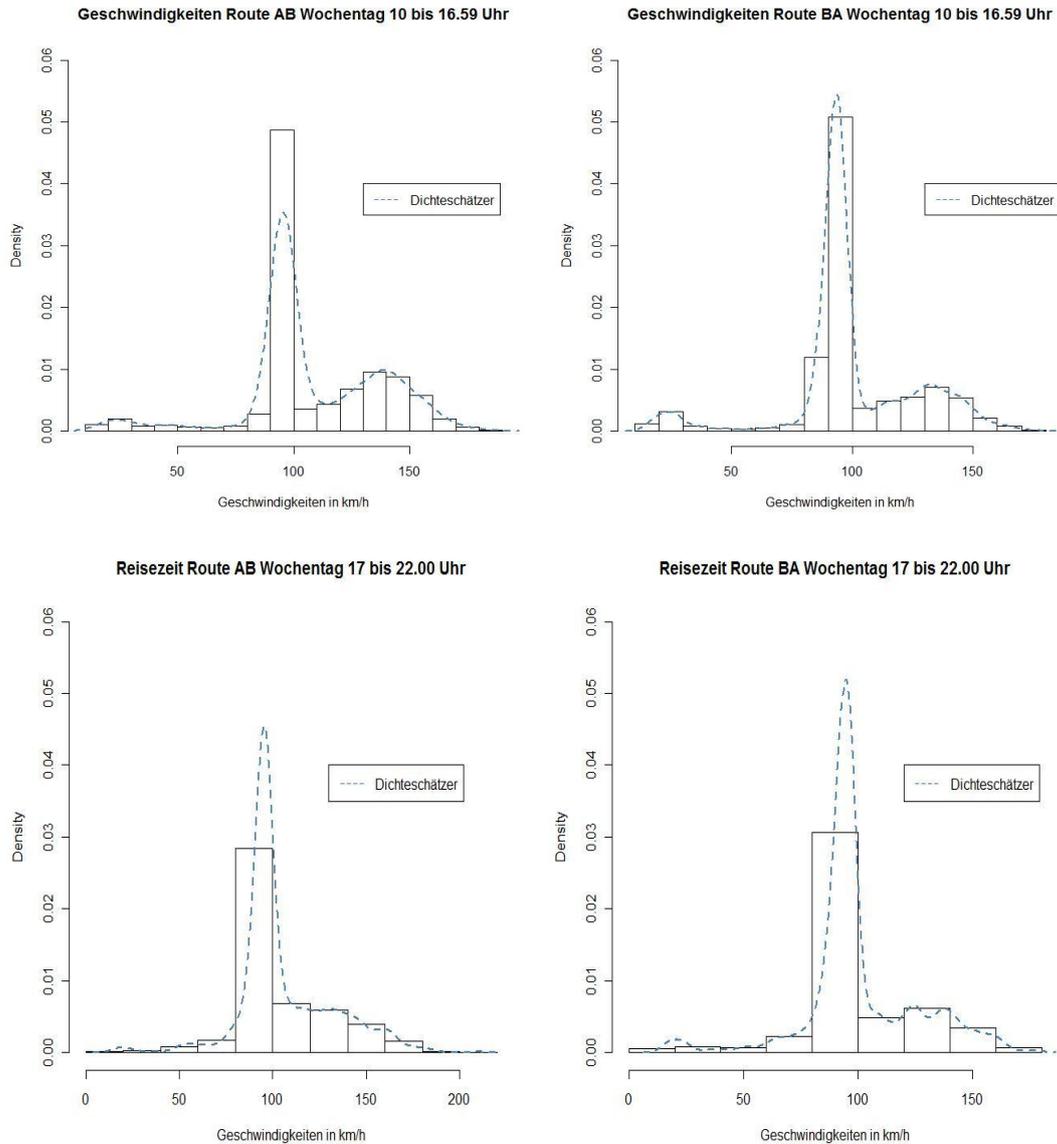
Von fundamentaler Wichtigkeit bei der praktischen Mischverteilungsanalyse ist die Festlegung einer Verteilung, anhand derer die Iteration durchgeführt wird. Mittels Histogrammen und geplotteten empirischen Dichteschätzern lassen sich anhand der Verläufe der Dichteschätzer, Datenkonzentrationen in bestimmten Geschwindigkeitsklassen und damit auch Asymmetrien erkennen.

Werktag Route AB



Werktag Route BA

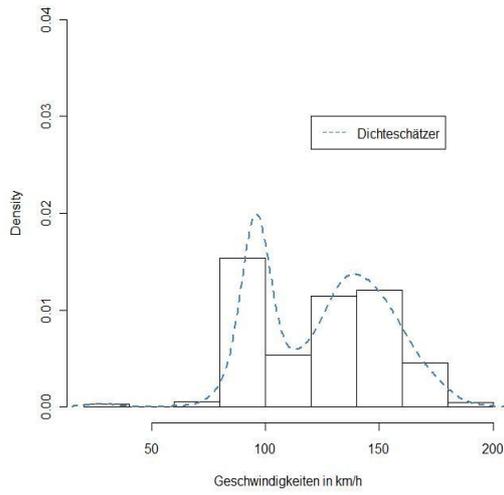




Beide Routen verdeutlichen im höheren Geschwindigkeitsbereich zwei deutlich erkennbare 'Pikes'. Im niedrigsten Geschwindigkeitsbereich sind mehr oder minder Ausreißer zu erkennen, die lediglich in den jeweils ersten und zweiten Zeitintervallen beider Routen schwach ausgeprägte Erhebungen bilden. Der zweite 'Gipfel' ist von einem sehr steilen Anstieg und einem nicht minder intensiven Abstieg geprägt und wird hauptsächlich von einer Geschwindigkeitsklasse gebildet. Im höchsten Geschwindigkeitsbereich ist jeweils ein 'Pike' zu erkennen, deren Höhe um ein Vielfaches geringer ausfällt. Charakteristisch an diesen letzten Erhebungen sind die langgezogenen moderaten Abstiege.

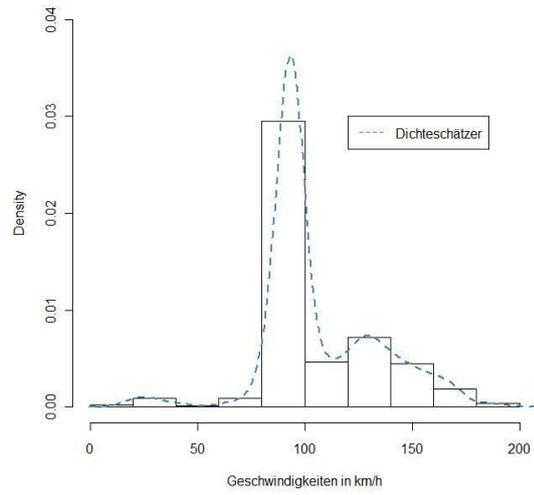
Samstage Route AB

Geschwindigkeiten Route AB Samstag 6 bis 13.59 Uhr

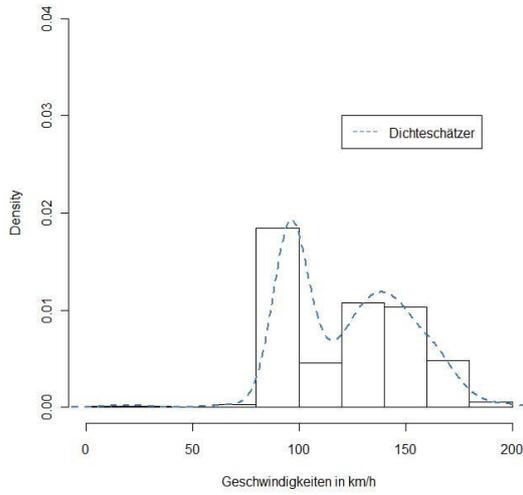


Samstage Route BA

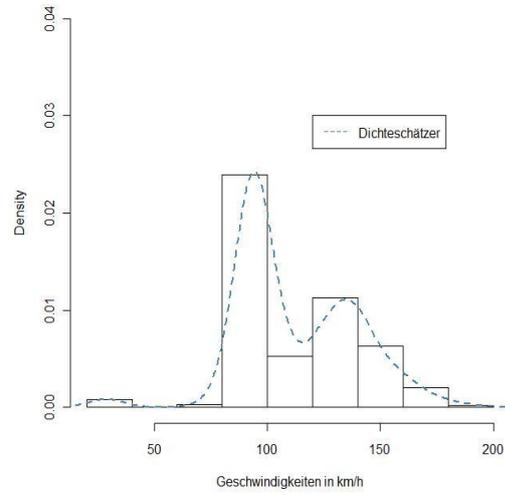
Geschwindigkeiten Route BA Samstag 6 bis 13.59 Uhr



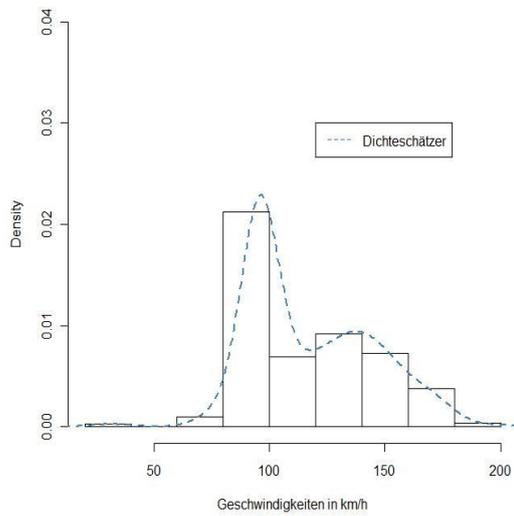
Geschwindigkeiten Route AB Samstag 14 bis 16.59 Uhr



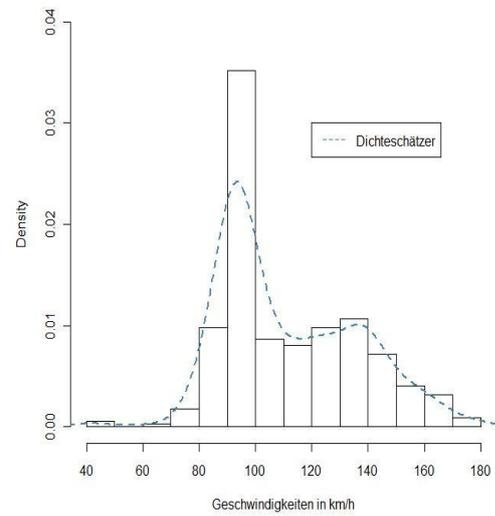
Geschwindigkeiten Route BA Samstag 14 bis 16.59 Uhr



Geschwindigkeiten Route AB Samstag 17 bis 22.00 Uhr



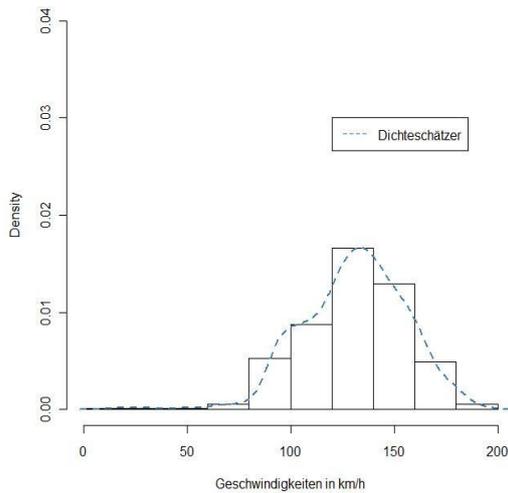
Geschwindigkeiten Route BA Samstag 17 bis 22.00 Uhr



An Samstagen kann man wiederum anhand der Verläufe der empirischen Dichteschätzer stets 2 ausgeprägte Erhebungen deutlich erkennen. Im Gegensatz zu den Werktagen hat sich deren Gewichtung jedoch etwas in den höheren Geschwindigkeitsbereich verschoben. Somit sind die Anstiege der ersten 'Pikes' etwas moderater, deren Breite fällt dadurch aber größer aus. Besonders im höchsten Geschwindigkeitsbereich sind die Höhen und Breiten deutlich ausgeprägter. Auffallend ist jedoch, dass im untersten Abschnitt weniger Ausreißer erkennbar sind. Insgesamt ist jedoch festzustellen, dass auch an Samstagen, aufgrund der vorliegenden Asymmetrie, eine Rechtsschiefe der Rohdaten vorliegt.

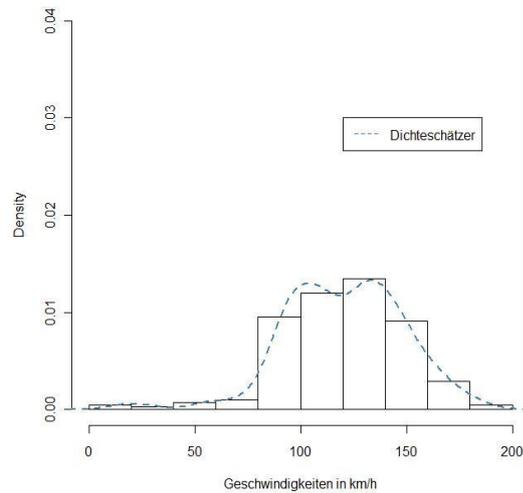
Sonntage Route AB

Geschwindigkeiten Route AB Sonntag 6 bis 10.59 Uhr

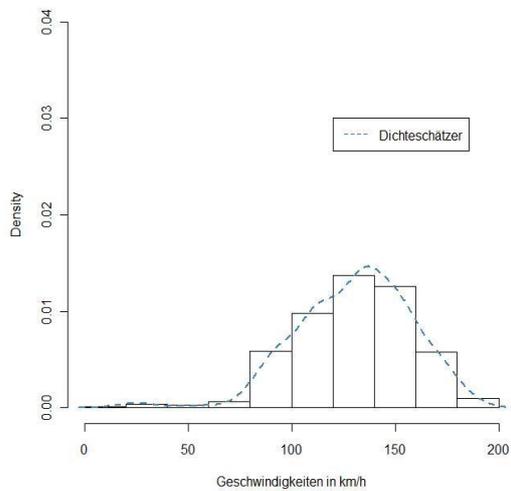


Sonntage Route BA

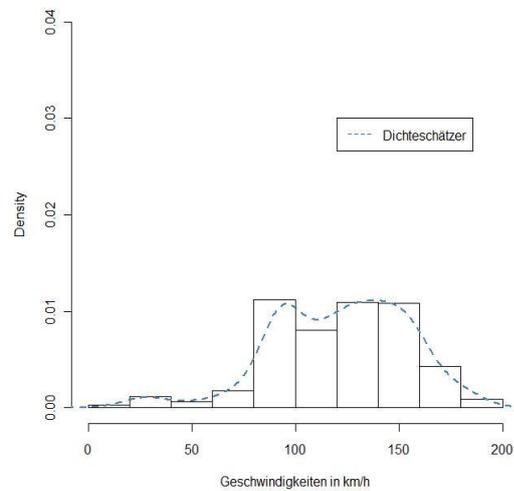
Geschwindigkeiten Route BA Sonntag 6 bis 10.59 Uhr

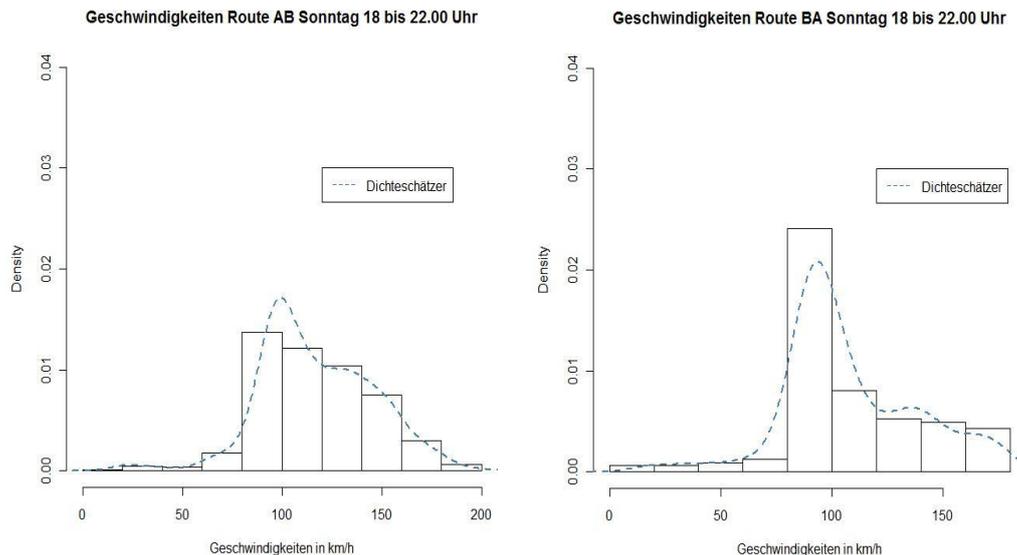


Geschwindigkeiten Route AB Sonntag 11 bis 17.59 Uhr



Geschwindigkeiten Route BA Sonntag 11 bis 17.59 Uhr





Der Tagestyp Sonntag zeichnet sich durch starke Überlappungen der Geschwindigkeitsgruppen aus. Anhand der empirischen Dichteschätzer lassen sich jedoch zumindest zwei Datenkonzentrationen im höheren Geschwindigkeitsbereich ausmachen. Auffallend ist jedoch, dass an Sonntagen ,auf der Route BA, im letzten Zeitintervall (18 – 22.00 Uhr), weniger ausgeprägte Überschneidungen existieren. Im niedrigsten Geschwindigkeitssegment sind wiederum Ausreißer zu erkennen. Zusammenfassend ist festzustellen, dass aufgrund der vorliegenden Informationen, Linkslastigkeiten (Rechtsschiefe) des vorliegenden Datenmaterials bestehen.

Aufgrund der **Rechtsschiefe** des Datenmaterials, kommen zur Mischverteilungsanalyse nur jene Verteilungen in Frage, die diese Charakteristik besitzen. Deshalb beschränkt sich die Verteilungsannahme auf die Lognormal-, Gamma- und Weibullverteilung. Welche Verteilungsklasse die beste Approximation an den jeweiligen empirischen Dichteschätzer mit sich bringt, konnte nur durch Vergleich der jeweiligen Analysen ermittelt werden.

Für die vorliegende Reisezeitanalyse wurde die **Gammaverteilung** gewählt, da sie im direkten Vergleich zur Lognormal- und Weibullverteilung bessere Anpassungstendenzen aufweist.

2.5.3 Mischverteilungsanalyse Werktag Route AB

Zeitintervall 6 bis 9.59 Uhr

```
> woche.grouped.AB1 <- mixgroup((31000/na.omit(dur_woche_AB[,1])*3.6), breaks=15);
woche.grouped.AB1      #Woche AB 1. Zeitintervall
```

	X	count
1	20	13
2	30	16
3	40	13
4	50	7
5	60	26
6	70	52
7	80	82
8	90	122
9	100	493
10	110	89
11	120	90
12	130	100
13	140	90
14	150	55
15	160	26
16	170	7
17	180	2
18	190	0
19	200	0
20	Inf	1

Die 2. Spalte beinhaltet die äquidistanten Geschwindigkeitsbereiche, wobei die Einteilung gemäß der Anfertigung eines Histogramms erfolgt. Aus der 3. Spalte kann man die jeweilige Anzahl jener Verkehrsteilnehmer herauslesen, die einer Geschwindigkeitsklasse zuzuordnen sind. Deutlich sind 3 Gruppen mit den höchsten Frequenzen, um 30, 100 und 130 km/h, erkennbar.

```
> fit.woche.AB1 <- mix(woche.mix.AB1,mixparam(pi=c(0.05, 0.65, 0.3),mu=c(30,100, 130),
+ sigma=c(3, 15, 10)), "gamma",mixconstr(consigma="CCV"))
```

```
> summary(fit.woche.AB1)
```

Parameters:

	pi	mu	sigma
1	0.03271	21.59	2.998
2	0.68393	88.13	12.237
3	0.28336	129.62	17.997

Standard Errors:

	pi.se	mu.se	sigma.se
1	0.004979	0.6256	0.1392
2	0.017431	0.6032	NA
3	0.017134	1.3971	NA

Analysis of Variance Table

	Df	Chisq	Pr(>Chisq)
Residuals	4	163.07	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Die Analyse dieses Mischmodells ergibt mit 22, 88 und 130 km/h, drei Geschwindigkeitsmittelwerte, mit den dazugehörigen Standardabweichungen von 3, 12 und 18 km/h. Rund drei Prozent der Verkehrsteilnehmer sind der ersten Gruppe, 68 Prozent der zweiten und 28 Prozent der dritten Verteilung, zuzuordnen. Die Standardfehler, dieser iterativ ermittelten Kennzahlen, betragen für die beiden ersten Stichprobenmittelwerte jeweils 0.6 km/h, und 1.4 km/h für den 3. Geschwindigkeitsmittelwert.

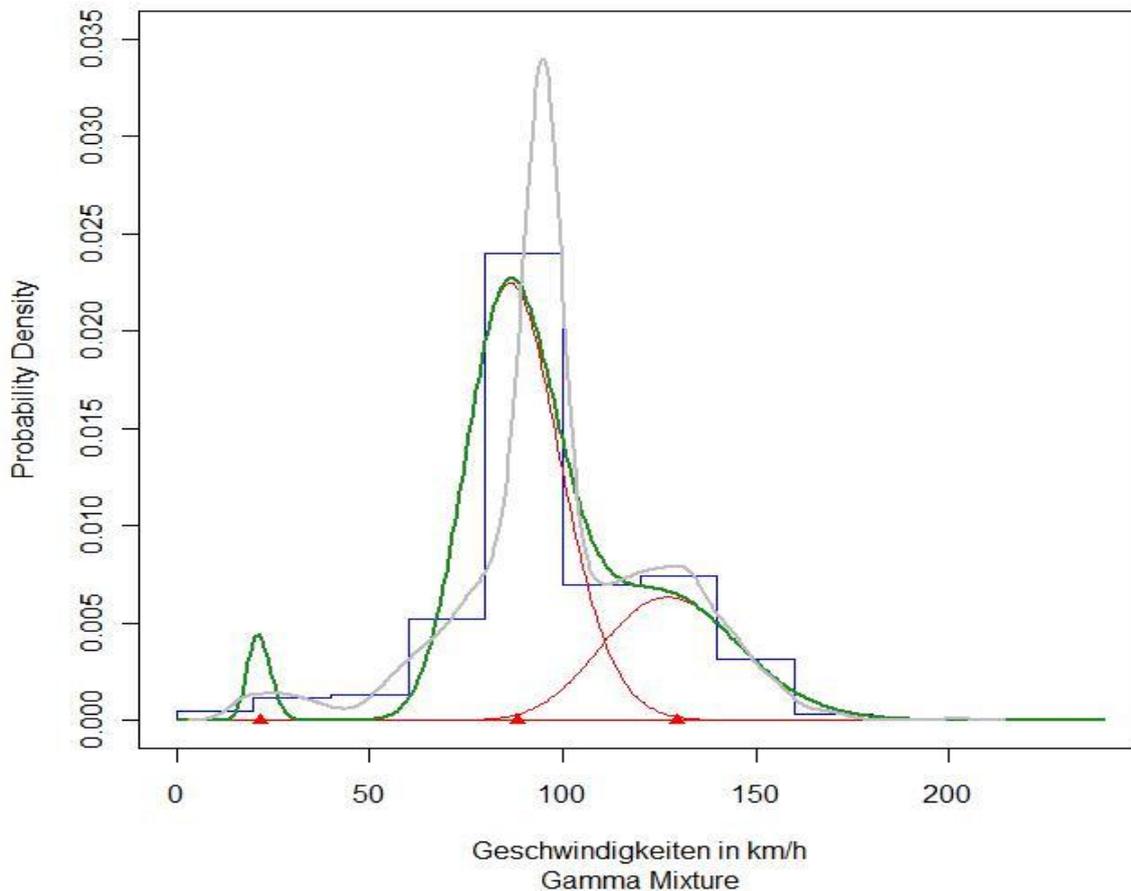
Sie repräsentieren die durchschnittlichen Abweichungen von den geschätzten Kennzahlen und sind somit ein Maß für die Genauigkeit der Schätzung. Dieser Fehler wird durch den Stichprobenumfang massiv beeinflusst, d. h. je größer der Stichprobenumfang ist, desto kleiner wird der Standardfehler.

Im Gegensatz dazu bewertet die Standardabweichung die tatsächliche Geschwindigkeitsstreuung je Klasse und ist vom Stichprobenumfang unabhängig. Eine hohe Streuung einer Geschwindigkeitsklasse lässt auf eine hohe Verkehrsfrequenz schließen und umgekehrt. Dies ist im dritten Geschwindigkeitsbereich der Fall. Der daraus ermittelte VC – Koeffizient ist mit 13.9 relativ hoch, und vermittelt eine deutliche Variabilität der Reisezeit.

Die Güte der Anpassung dieses iterierten Modells an die empirische Dichteverteilung ist aufgrund des p – Wertes des Chi – Quadrat – Tests ¹ von nahezu Null als schlecht zu bewerten.

¹ Chi – Quadrat – Test (siehe 2.6.5)

Mischmodell Wochentage Route AB 6 bis 9.59 Uhr



Anhand dieser Grafik kann eine akzeptable Annäherung der iterativ berechneten Geschwindigkeitsverteilungen an den Dichteschätzer, in den hohen Geschwindigkeitsbereichen, festgestellt werden. Der Plot dieser Mischverteilungsanalyse weist auf die Existenz dreier Geschwindigkeitsklassen hin, die sich anhand ihrer Mittelwerte und Gewichtungen signifikant voneinander unterscheiden. Da während der gesamten Fahrzeit eine ständige Vermischung der Geschwindigkeiten auf den unterschiedlichen Fahrspuren auftritt, können die errechneten Gruppen nicht eindeutig einer Fahrspur zugewiesen werden.

2.5.4 Mischverteilungsanalyse Samstag Route AB

Zeitintervall 6 bis 13.59 Uhr

```
> sam.grouped.AB1 <- mixgroup((31000/na.omit(dur_sam_AB[,1])*3.6)); sam.grouped.AB1
#Sam AB 1. Zeitintervall
```

```
  X count
1 40    7
2 60    0
```

```

3 80 15
4 100 411
5 120 143
6 140 305
7 160 322
8 180 121
9 Inf 13

```

Anhand der Klasseneinteilung können 3 Geschwindigkeitsmittelwerte ermittelt werden. Der direkte Vergleich zwischen der errechneten Klasseneinteilung und dem Histogramm zeigt diesbezüglich jedoch eine Diskrepanz hinsichtlich jener Klassen mit den höchsten Frequenzen. Die nachfolgende Analyse wird demzufolge mit 3 *Klassen* durchgeführt.

```
> fit.sam.AB1 <- mix(sam.mix.AB1,mixparam(pi=c(0.05, 0.5, 0.45),mu=c(30,100, 150),
sigma=c(5, 15, 15)), "gamma", mixconstr(consigma="CCV"))
```

```
> summary(fit.sam.AB1)
```

Parameters:

```

      pi  mu sigma
1 0.005235 30.01 2.937
2 0.410560 94.10 9.211
3 0.584205 144.90 14.184

```

Standard Errors:

```

      pi.se  mu.se sigma.se
1 0.002011 10.5933  1.039
2 0.014195  0.5204   NA
3 0.014220  0.6077   NA

```

Analysis of Variance Table

```

      Df Chisq Pr(>Chisq)
Residuals 2 33.496 5.325e-08 ***

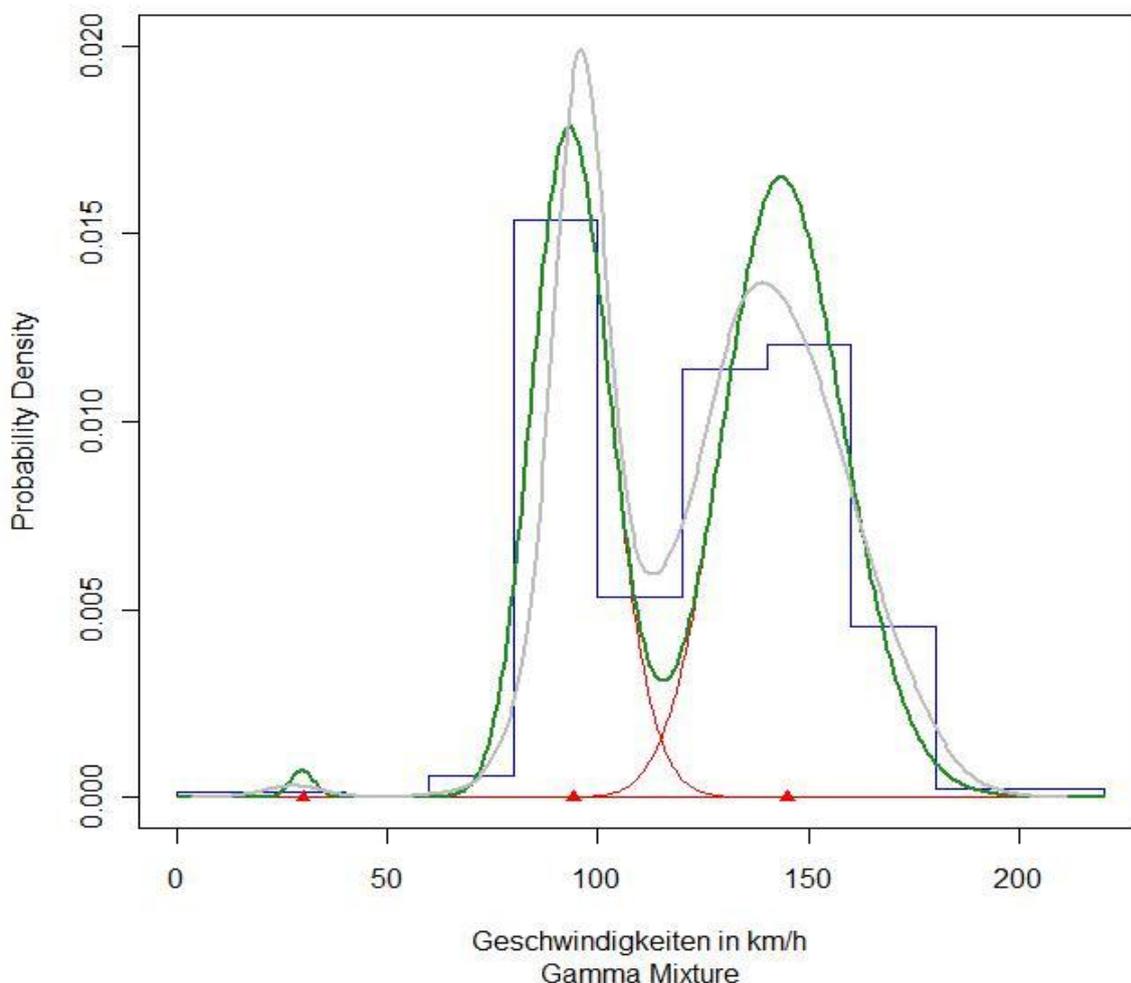
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Die iterierten Geschwindigkeitsmittelwerte betragen 30, 94 bzw. 145 km/h, die dazugehörigen Streuungen 3, 9 und 14 km/h. Aus der Frequenzgewichtung lässt sich die Verteilung 0,5, 41 bzw. 58,5 Prozentpunkte erkennen. Der Standardfehler von rund 0,2

Prozent bei der Gewichtung und 11 km/h bei der Streuung des ersten Geschwindigkeitsbereiches, negiert aber die Aussagekraft der dazugehörigen Kennzahlen, da der Stichprobenumfang für diesen Bereich zu klein ist. Aufgrund dessen ist eine nochmalige Analyse mit 2 *Geschwindigkeitsklassen* zu empfehlen. Die Standardfehler der beiden höheren Geschwindigkeitsbereiche, belegen aber für deren Parameteriterationen, ausreichend große Stichprobengrößen. Mit 9.8 ist der errechnete VC – Koeffizient deutlich niedriger als an Werktagen desselben Zeitintervalls. Somit besteht für die einzelnen Geschwindigkeitsverteilungen eine niedrigere Variabilität. Anhand des Anpassungstests liegt keine gute Anpassung des iterierten Modells an den Dichteschätzer vor.

Mischmodell Samstags Route AB 6 bis 13.59 Uhr



Anhand dieses Plots ist die anteilmäßig harmonische Aufteilung in 2 Geschwindigkeitsbereiche erkennbar, wobei aber eine leichte Zunahme beider durchschnittlichen Geschwindigkeiten zu verzeichnen ist. Der direkte Vergleich mit dem selben Zeitintervall an Werktagen veranschaulicht an Samstagen eine deutliche Zunahme von Verkehrsteilnehmern mit einer kürzeren Reisezeit. Anhand der vorliegenden Rohdaten kann eine Spezifizierung bzw. Trennung zwischen LKW und PKW nicht vorgenommen

werden. Eine mögliche Erklärung dafür wäre allerdings eine Abnahme des Schwerverkehrs an Samstagen.

Samstag Route AB, Zeitintervall 6 bis 13.59 Uhr, Geschwindigkeit > 50 km/h

```
>fit.sam.AB1.1<- mix(sam.mix.AB1.1,mixparam(mu=c(100, 140),
sigma=c(10,15)),"norm",mixconstr(consigma="CCV"))
> summary(fit.sam.AB1.1)
```

Parameters:

	pi	mu	sigma
1	0.4059	96.55	9.621
2	0.5941	144.68	14.418

Standard Errors:

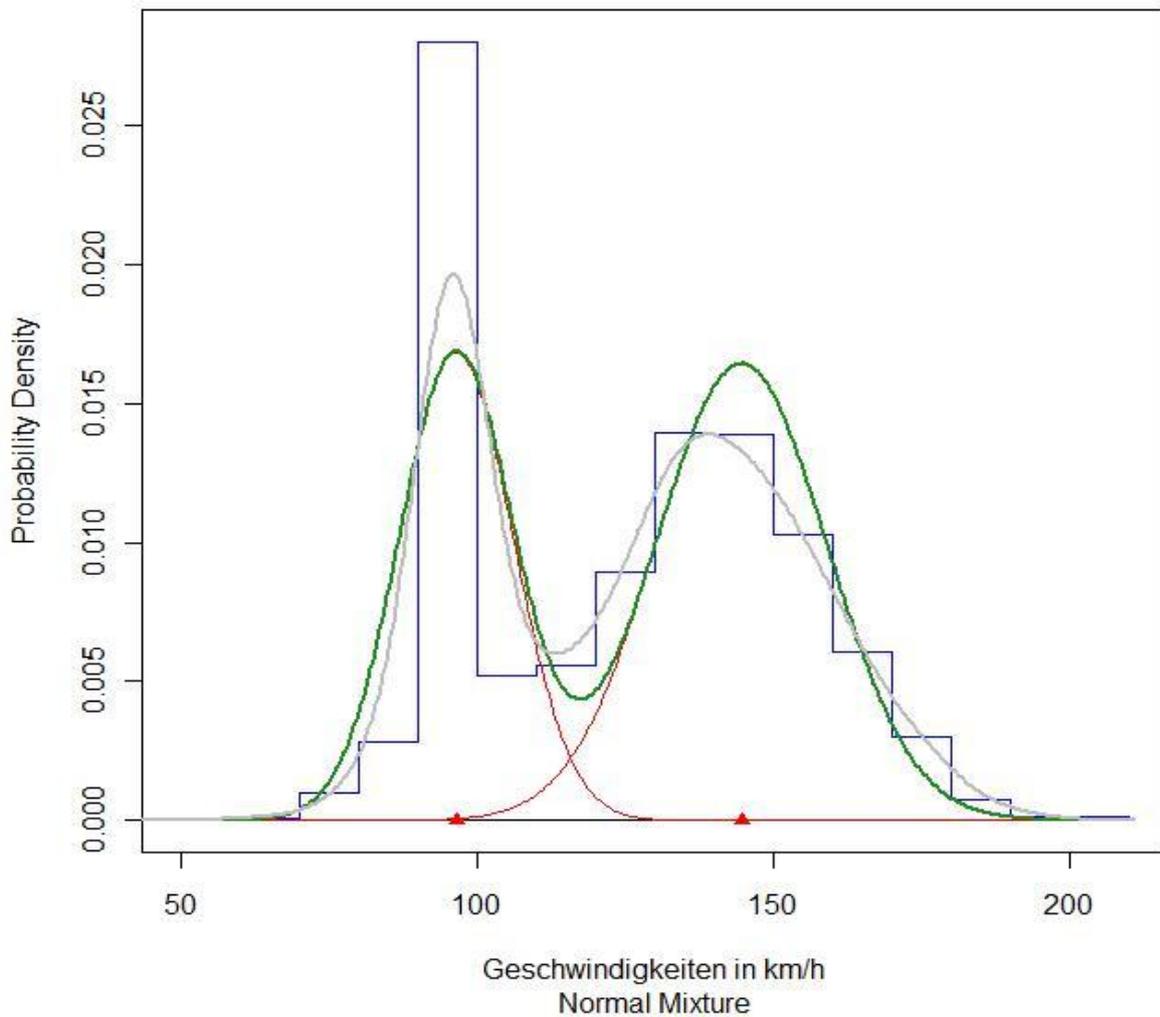
	pi.se	mu.se	sigma.se
1	0.01438	0.5133	0.2240
2	0.01438	0.5672	NA

Analysis of Variance Table

	Df	Chisq	Pr(>Chisq)
Residuals	9	263.52	< 2.2e-16 ***

Die Mischanalyse mit 2 *Geschwindigkeitsklassen* (*Geschwindigkeiten > 50 km/h*) ergibt für die beiden Mischverteilungen annähernd gleiche Kennzahlen für die Durchschnittsgeschwindigkeiten, Streuungen und den Frequenzverteilungen, wie die Analyse mit 3 Geschwindigkeitsklassen. Die Standardfehler belegen eine ausreichend große Stichprobe (siehe S. 35, Einschränkung der Reisezeit).

Mischmodell Samstage Route AB 6 bis 13.59 Uhr



2.5.5 Mischverteilungsanalyse Sonntag Route AB

Zeitintervall 6 bis 10.59 Uhr

```
> son.grouped.AB1 <- mixgroup((31000/na.omit(dur_son_AB[,1])*3.6), breaks=15);
son.grouped.AB1      #Son AB 1. Zeitintervall
```

```
X count
1 20  1
2 30  0
3 40  1
4 50  1
5 60  0
```

```

6 70 2
7 80 2
8 90 2
9 100 35
10 110 34
11 120 28
12 130 56
13 140 62
14 150 49
15 160 43
16 170 25
17 180 10
18 190 3
19 Inf 1

```

Auch mittels einer variablen Gruppierungsanpassung kann die Existenz eines dritten Geschwindigkeitssegmentes nicht eindeutig geklärt werden. Ob eine Analyse mit 2 oder 3 Geschwindigkeitsgruppen eine harmonische Anpassung an den Dichteschätzer der Rohdaten bringt, kann nur durch Durchführung verschiedener Mischverteilungsanalysen geklärt werden. Der direkte Vergleich beider Analysen zeigte eindeutige Präferenzen für 3 *Geschwindigkeitsgruppen*.

```

> fit.son.AB1 <- mix(son.mix.AB1,mixparam(pi=c(0.05, 0.25, 0.7), mu=c(50, 100, 140),
+ sigma=c(5,10, 15)), "gamma", mixconstr(consigma="CCV"))

```

```

> summary(fit.son.AB1)

```

Parameters:

```

      pi  mu sigma
1 0.008449 31.58 4.166
2 0.235678 103.12 13.606
3 0.755873 140.51 18.539

```

Standard Errors:

```

      pi.se mu.se sigma.se
1 0.004915 2.669 0.4705
2 0.044304 2.666  NA
3 0.044448 1.831  NA

```

Analysis of Variance Table

```

      Df Chisq Pr(>Chisq)

```

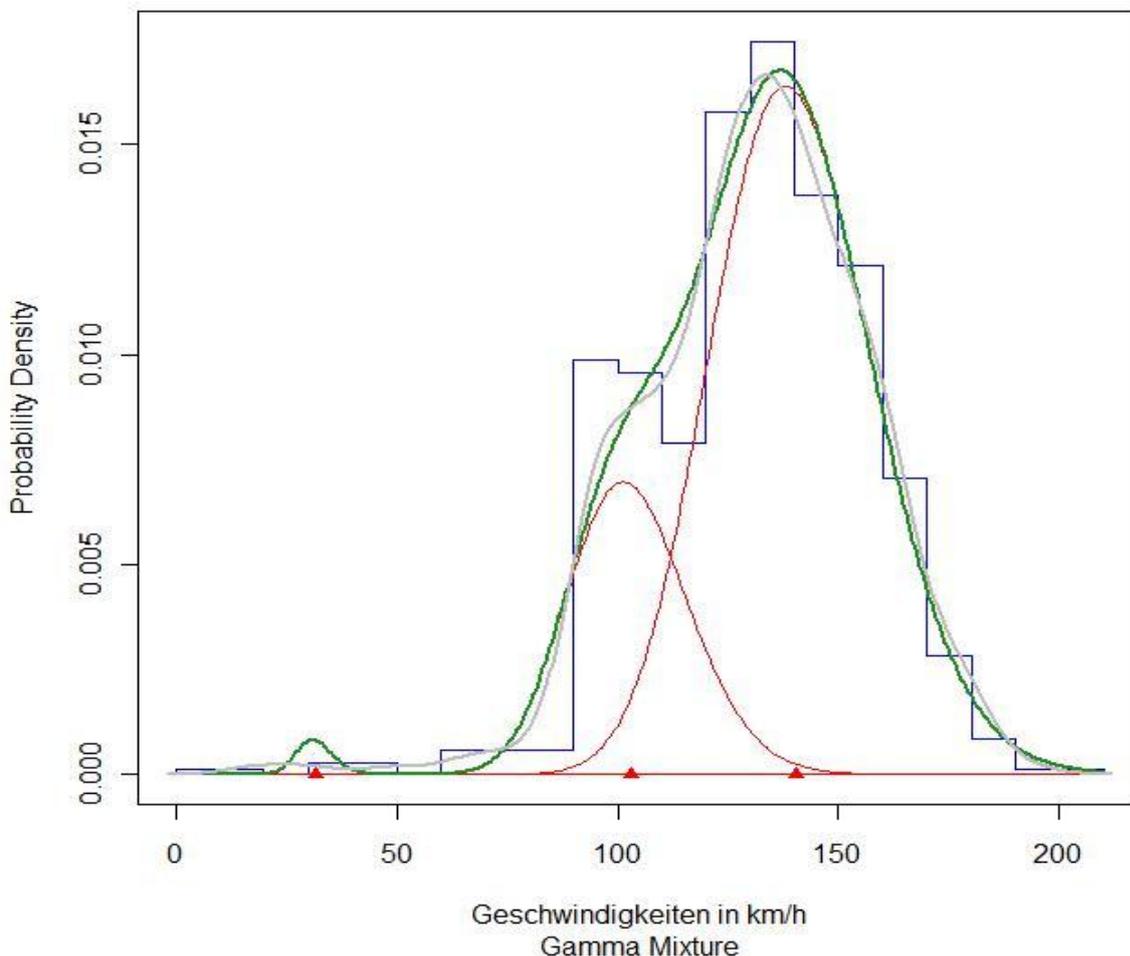
Residuals 12 43.748 1.686e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Nicht ganz ein Prozent der gesamten Verkehrsteilnehmer bewältigen die Teststrecke mit Einlegung einer Pause. 24 Prozent weisen bei einer Standardabweichung von 14 km/h eine mittlere Geschwindigkeit von 103 km/h, 76 Prozent weisen bei einer Streuung von 14 km/h eine Durchschnittsgeschwindigkeit von 141 km/h auf. Der Standardfehler von 0.49 Prozentpunkten bei einem Frequenzanteil von 0.84 Prozent der ersten Geschwindigkeitsklasse weist auf eine zu *geringe Verkehrsfrequenz* in diesem Segment hin. Bis zu einer Geschwindigkeit von 50 km/h sind lediglich 3 Beobachtungen zu registrieren. Eine nochmalige Analyse mit der Nichtbeachtung dieser Verkehrsteilnehmer ist zu empfehlen. Mit einem Variationskoeffizienten von 13.2 ist von einer hohen Variabilität der Durchschnittsgeschwindigkeiten der einzelnen Klassen auszugehen.

Der vorliegende Chi – Quadrat – Test verneint eine gute Übereinstimmung des Modells mit dem empirischen Dichteschätzer.

Mischmodell Sonntags Route AB 6 bis 10.59 Uhr



Obwohl im niedrigsten Geschwindigkeitsbereich nicht ganz ein Prozent der Verkehrsteilnehmer enthalten sind, wird erst durch Beachtung eben dieser Klasse eine harmonische Anpassung der iterierten Klassenverteilungen an den bestehenden Dichteschätzer der Rohdaten erreicht. Im höheren Geschwindigkeitssegment ist eine starke Überlappung beider Gruppierungen erkennbar, wobei ein Frequenzüberhang der letzten Klasse ersichtlich ist. Insgesamt ist ein Anstieg der jeweiligen Durchschnittsgeschwindigkeiten festzustellen. Erklärbar ist dieses Szenario mit dem bestehenden LKW – Fahrverbot an Sonntagen.

Sonntag Route AB, Zeitintervall 6 bis 10.59 Uhr, Geschwindigkeit > 50 km/h

```
> fit.son.AB1.1<- mix(son.mix.AB1.1,mixparam(mu=c(100, 140),
sigma=c(10,15)),"norm",mixconstr(consigma="CCV"))
```

```
> summary(fit.son.AB1.1)
```

Parameters:

	pi	mu	sigma
1	0.2373	103.2	13.24
2	0.7627	140.5	18.03

Standard Errors:

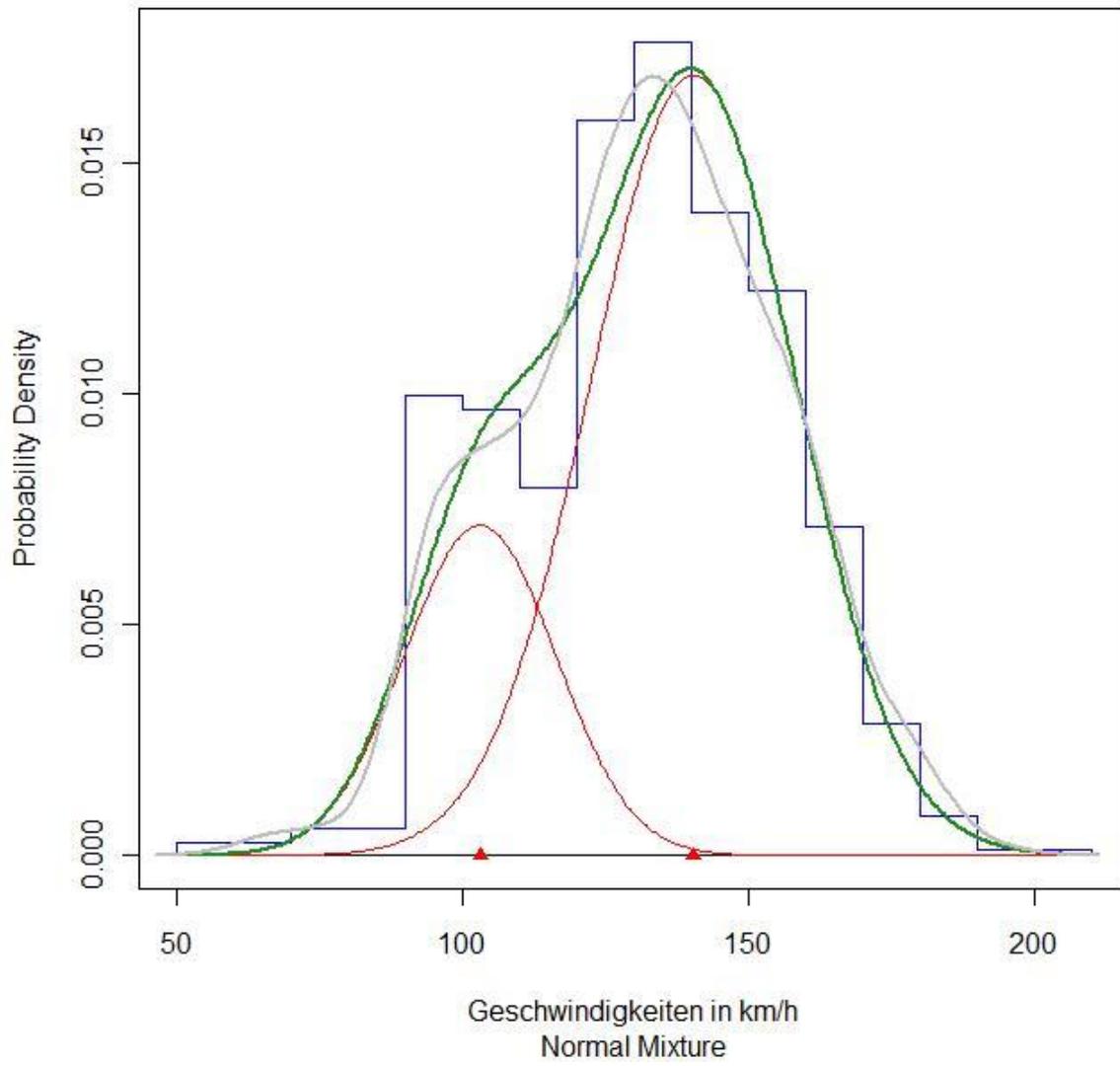
	pi.se	mu.se	sigma.se
1	0.0465	2.875	1.043
2	0.0465	1.831	NA

Analysis of Variance Table

	Df	Chisq	Pr(>Chisq)
Residuals	9	24.954	0.003023 **

Die Mischanalyse mit 2 *Geschwindigkeitsklassen (Geschwindigkeiten > 50 km/h)* ergibt für die Frequenzverteilungen, Durchschnittsgeschwindigkeiten und den Standardabweichungen annähernd gleiche Werte, wie das zweite und dritte Geschwindigkeitssegment der Mischanalyse ohne Geschwindigkeitsbeschränkung. Aufgrund der Standardfehler kann von einer genügend großen Stichprobe ausgegangen werden (siehe S. 35, Einschränkung der Reisezeit).

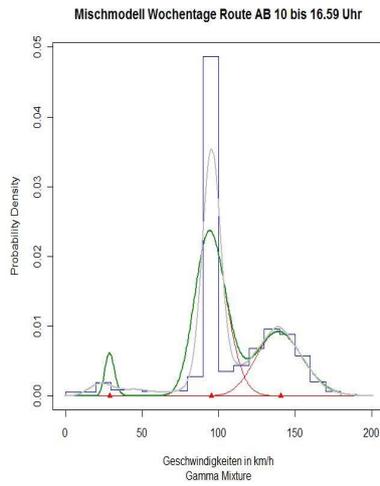
Mischmodell Sonntage Route AB 6 bis 10.59 Uhr



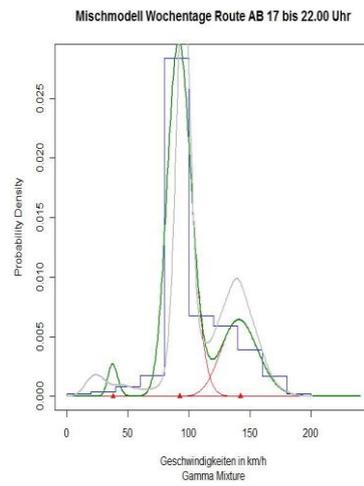
2.5.6 Mischmodelle der Route AB

Werktage

2. Zeitintervall

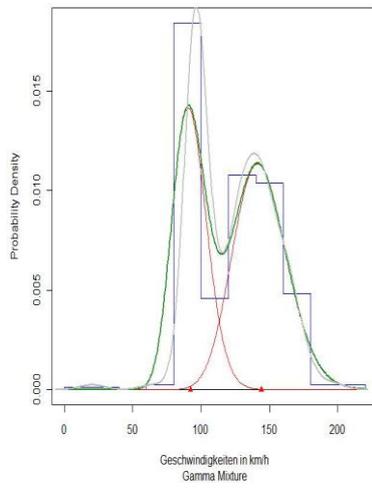


3. Zeitintervall

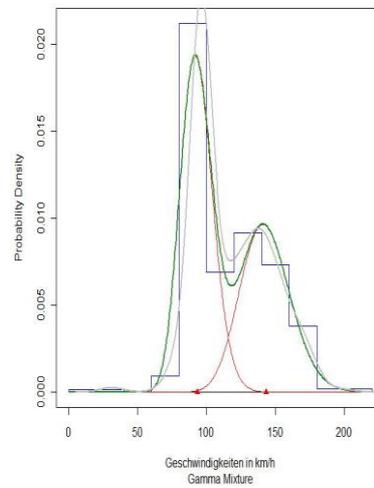


Samstage

Mischmodell Samstags Route AB 14 bis 16.59 Uhr

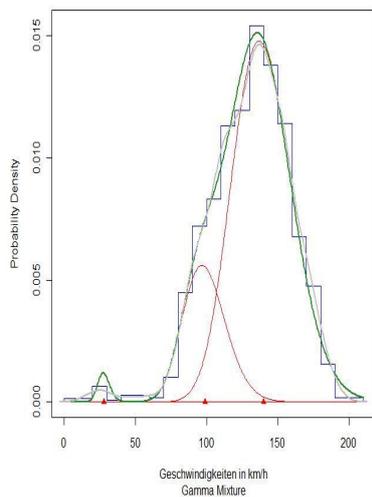


Mischmodell Samstags Route AB 17 bis 22.00 Uhr

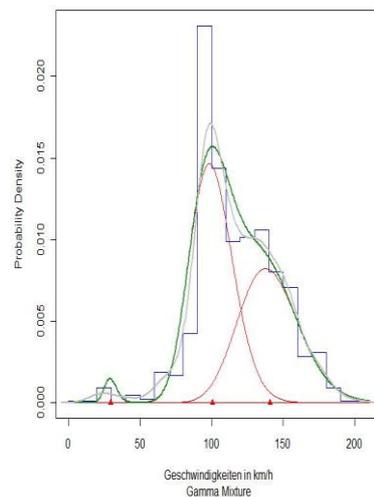


Sonntage

Mischmodell Sonntags Route AB 11 bis 17.59 Uhr



Mischmodell Sonntags Route AB 18 bis 22.00 Uhr

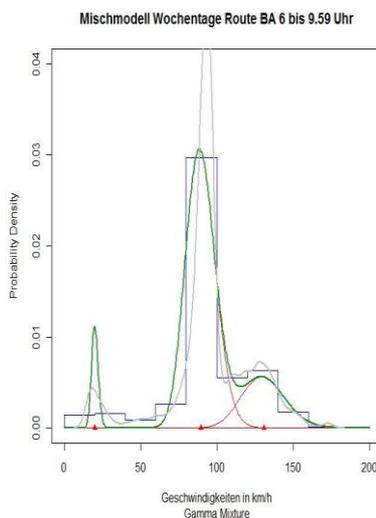


Aufgrund der vorliegenden Mischverteilungsanalysen, ist eine deutliche Frequenzverschiebung, des Verkehrs, in den höheren Geschwindigkeitsbereich, in Abhängigkeit vom jeweiligen Tagestyp, zu verzeichnen. Während an Werktagen, im mittleren Geschwindigkeitssegment, die höchste Verkehrsdichte registriert wird, ist an Sonntagen ein Trend zu mehr Teilnehmern mit höheren Geschwindigkeiten zu erkennen. In Anbetracht des an Sonntagen existierenden LKW – Fahrverbots ist ein direkter Zusammenhang mit der Verkehrszusammensetzung nicht von der Hand zu weisen.

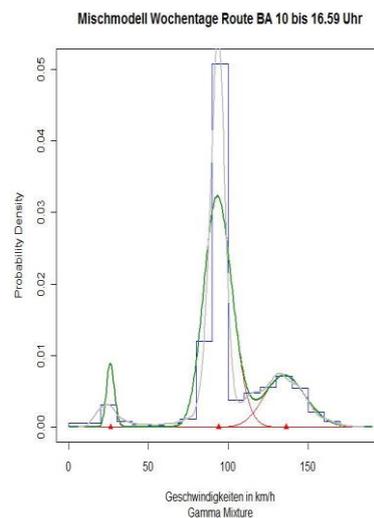
Besonders an Sonntagen liegt eine starke Überlappung der einzelnen Geschwindigkeitsklassen vor. Aus den vorliegenden Plots der Route AB ist kein Stauverhalten bzw. überproportionale Verkehrsdichte zu diagnostizieren, da die dafür typischen Merkmale fehlen. Läge nämlich ein solches Verkehrsszenario vor, wäre eine Differenzierung in verschiedene Geschwindigkeitsgruppen, nicht oder nur sehr schwer möglich. Ein weiteres Indiz für ein normales Verkehrsgeschehen sind die relativ großen Differenzen zwischen den einzelnen Durchschnittsgeschwindigkeiten und der sehr konzentrierten Verteilung um diese Geschwindigkeiten selbst.

2.5.7 Mischmodelle der Route BA

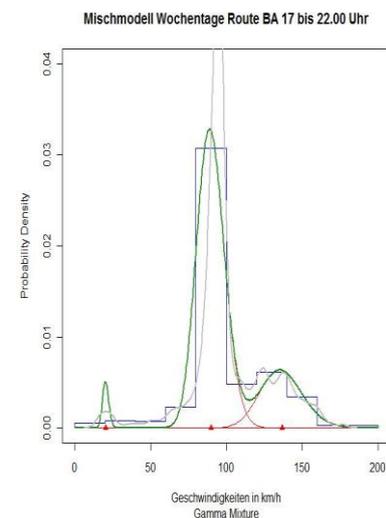
Werktage 1. Zeitintervall



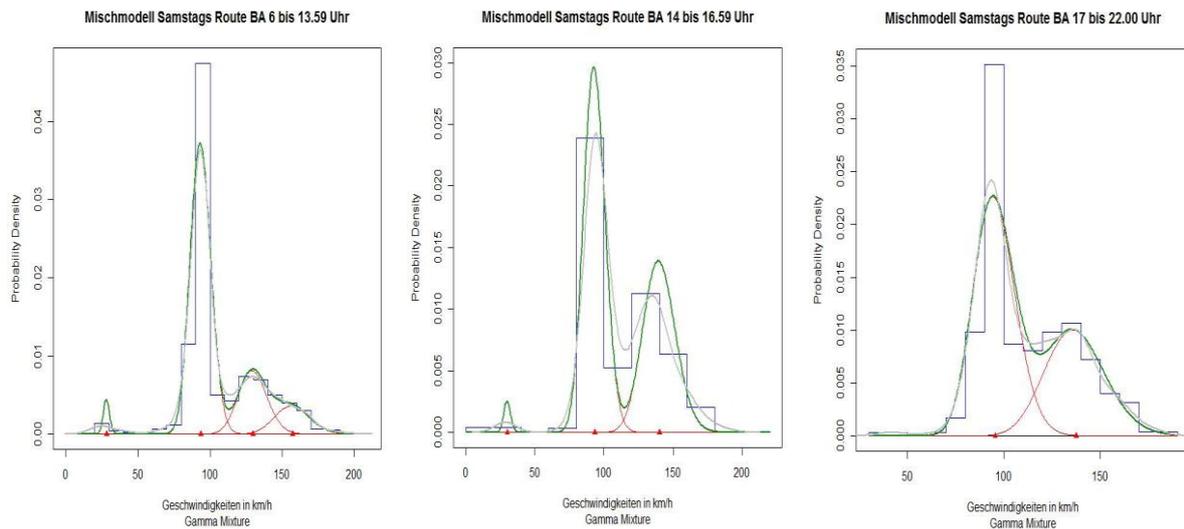
2. Zeitintervall



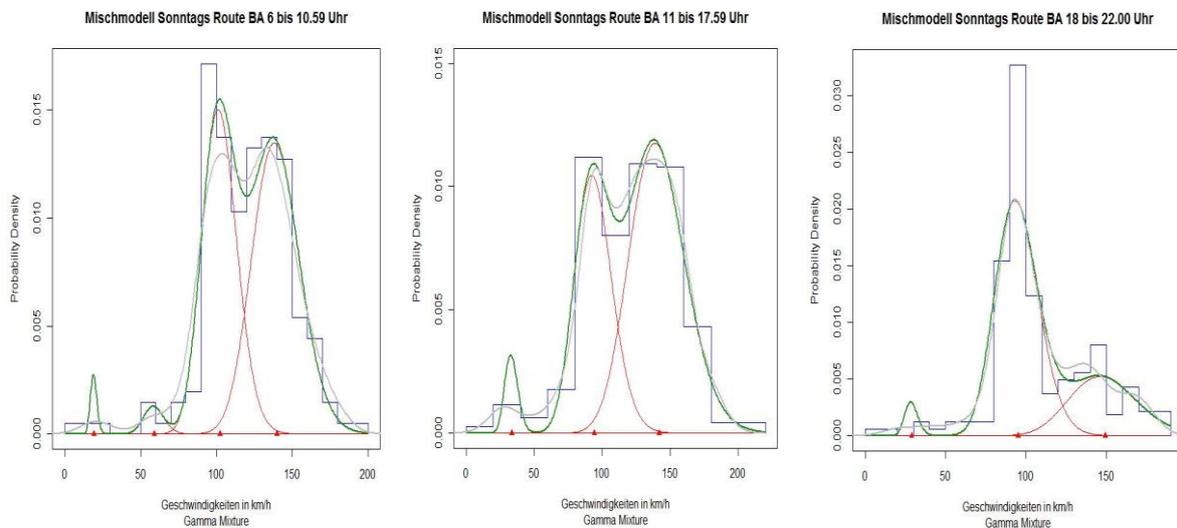
3. Zeitintervall



Samstage



Sonntage



Aufgrund einer zu geringen Stichprobengröße, kann für das 1. und 3. Zeitintervall an Sonntagen, keine aussagekräftige Prognose gestellt werden. Auffallend ist außerdem noch die hohe Anzahl der Geschwindigkeitsgruppen im ersten Zeitsegment an Samstagen. Nur mit 4 Verteilungen ist eine harmonische Anpassung, an den Dichteschätzer, der Rohdaten möglich. Anhand der Standardfehler ist die Analyse jedoch eindeutig. Die restlichen Plots weisen grundsätzlich die gleichen Merkmale wie die Route AB auf.

2.5.8 Zusammenfassung**1. Zeitintervall**

	Wochentage		Samstage		Sonntage	
Routen	AB	BA	AB	BA	AB	BA
# Verteilungen	3	3	3	4	3	4
VC - Koeffizient	13.9	10.8	9.8	7.7	13.2	11.3
Durchschnittsgeschw. in km/h	22, 88, 130	20, 90, 131	30, 94, 145	28, 93, 130, 157	32, 103, 141	19, 59, 103, 140
Standardabweichung In km/h	3, 12, 18	2, 10, 14	3, 9, 14	2, 7, 10, 12	4, 14, 19	2, 7, 12, 16
Stichprobengröße	1284	754	1337	921	355	204

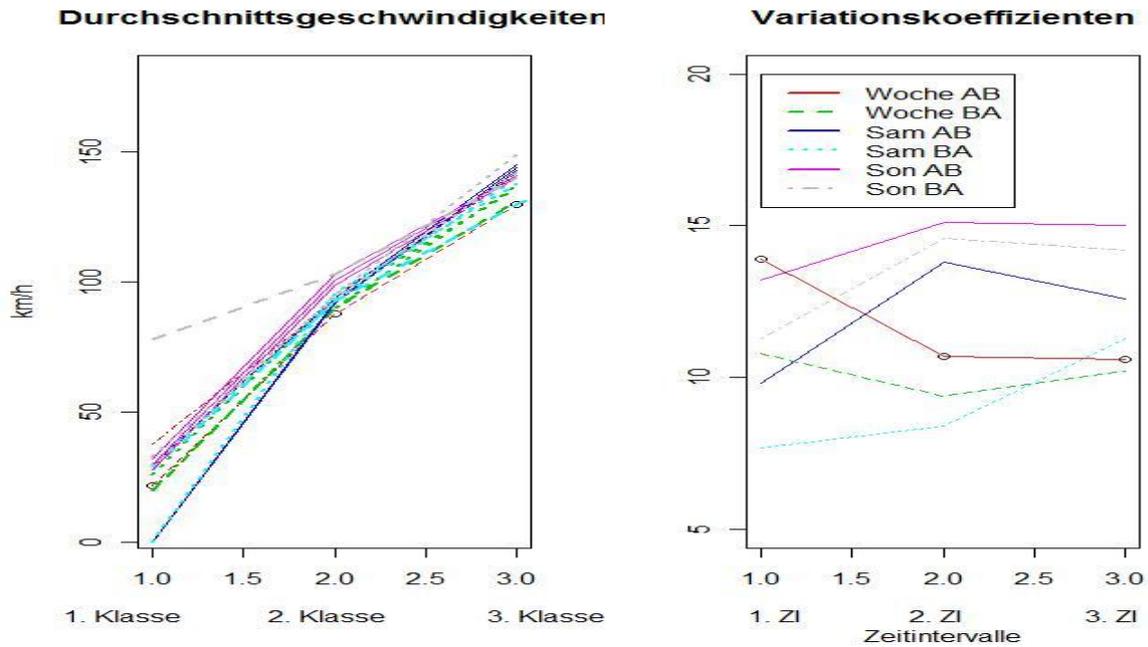
2. Zeitintervall

	Wochentage		Samstage		Sonntage	
Routen	AB	BA	AB	BA	AB	BA
# Verteilungen	3	3	2	3	3	3
VC - Koeffizient	10.7	9.4	13.8	8.4	15.1	14.6
Durchschnittsgeschw. in km/h	29, 95, 141	26, 94, 136	92, 144	30, 93, 140	28, 99, 140	33, 94, 142
Standardabweichung In km/h	3, 10, 15	2, 9, 13	13, 20	3, 8, 12	4, 15, 21	5, 14, 21
Stichprobengröße	2548	1889	372	316	1096	393

3. Zeitintervall

	Wochentage		Samstage		Sonntage	
Routen	AB	BA	AB	BA	AB	BA
# Verteilungen	3	3	2	2	3	3
VC - Koeffizient	10.6	10.2	12.6	11.3	15	14.2
Durchschnittsgeschw. in km/h	38, 93, 142	20, 90, 137	93, 143	96, 138	30, 101, 141	29, 95, 149
Standardabweichung In km/h	4, 10, 15	2, 9, 14	12, 18	11, 16	4, 15, 21	4, 13, 21
Stichprobengröße	841	717	370	347	424	162

Anhand der vorliegenden Analysen kann man grundsätzlich von 3 Geschwindigkeitsklassen, je Zeitintervall und Route, ausgehen. Eine eindeutige Zuordnung, der einzelnen Verteilungen, an eine jeweilige Fahrspur, kann jedoch, mangels geeigneter Daten, nicht getroffen werden. Es können lediglich Vermutungen diesbezüglich angestellt werden. Da der Schwerverkehr vorwiegend auf der ersten Fahrspur durchgeführt wird, kann eine tendenzielle Zuordnung der 2. Geschwindigkeitsklasse an diese Fahrspur getroffen werden. Inwieweit sich die verschiedenen Fahrzeugtypen je Fahrspur vermischen, kann ohne diesbezügliche Unterscheidung, jedoch nicht beantwortet werden. Wiederum kann nur vermutet werden, dass auf der 2. Fahrspur jene Fahrzeuge verkehren, die eine höhere Durchschnittsgeschwindigkeit aufweisen. Insgesamt kann man jedoch sagen, dass die Intensität des Schwerverkehrs massiven Einfluss auf die Gewichtung des Gesamtverkehrs, als auch auf die Höhe der einzelnen Durchschnittsgeschwindigkeiten ausübt. Auch ist zu bemerken, dass sich die Durchschnittsgeschwindigkeiten beider Routen desselben Zeitintervalls stark ähneln. Obwohl auf der Route AB eine deutlich höhere Verkehrsdichte als in der Gegenrichtung herrscht, beeinflusst dieser Faktor, die Höhe der iterierten Kennzahlen, kaum.



Grundsätzlich ist zu bemerken, dass die jeweilige Verkehrsdichte keinen Einfluss auf den Wert des VC – Koeffizienten hat. Obwohl die durchschnittliche Frequenz im Tagesverlauf stark ansteigt, spiegeln die Variationskoeffizienten diesen Trend nicht wieder.

Außerdem ist an Werktagen auf beiden Routen im Tagesverlauf ein Absinken der Variabilität der Durchschnittsgeschwindigkeiten erkennbar, während an Samstagen und Sonntagen der umgekehrte Trend ersichtlich ist. Besonders an Sonntagen werden signifikant höhere Variabilitäten der Geschwindigkeiten verzeichnet. In Anbetracht des sonntäglichen LKW – Fahrverbots, kann ein unmittelbarer Zusammenhang nicht ausgeschlossen werden.

Einschränkung der Reisezeit

Bei den meisten durchgeführten Mischanalysen wurden 3 *Geschwindigkeitsbereiche* festgestellt, wobei das niedrigste Segment jene Verkehrsteilnehmer enthält, die entlang der untersuchten Strecke eine Pause einlegen. Da deren Anteil am Gesamtverkehrsaufkommen äußerst gering ist, stellt sich natürlich die Frage, inwieweit die Reisezeitanalyse, durch diese Verkehrsteilnehmer beeinflusst wird. Folgende Zeitintervalle, mit den jeweils höchsten Bluetooth-Verkehrsfrequenzen, wurden daher explizit auf diese Frage hin nochmals untersucht, indem die Daten mit Geschwindigkeiten unter 50 km/h, nicht berücksichtigt wurden.

Route AB

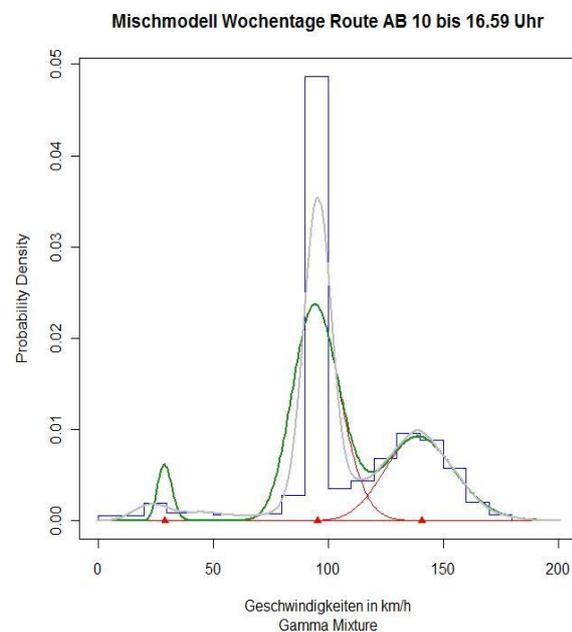
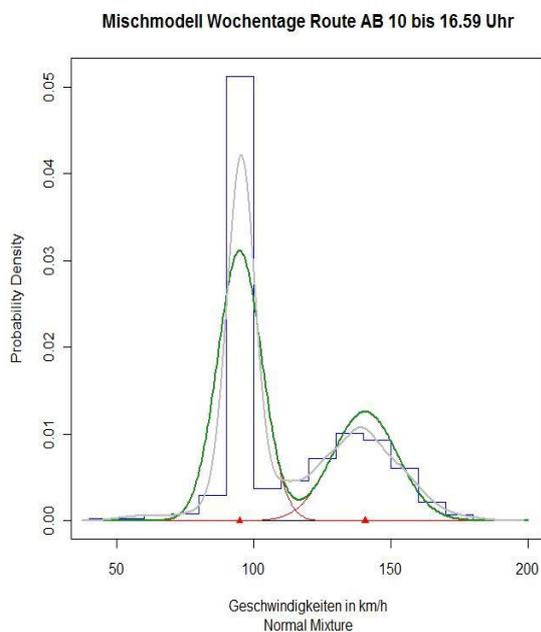
- Wochentag Zeitintervall 10 – 16.59 Uhr
- Samstag Zeitintervall 6 – 13.59 Uhr
- Sonntage Zeitintervall 11 – 17.59 Uhr

Die Analyse des reduzierten Datensatzes wurde mit einer symmetrischen **Normalverteilung** durchgeführt, da nur mittels dieser Verteilung, eine adäquate Anpassungsgüte an den empirischen Dichteschätzer erreicht wurde. Die Lognormal-, Gamma-, Binomial- oder Nbinomialverteilung lieferten entweder ähnlichen Ergebnisse, oder wurden vom Programm R durch Fehlermeldungen ausgeschlossen.

Geschwindigkeiten > 50 km/h

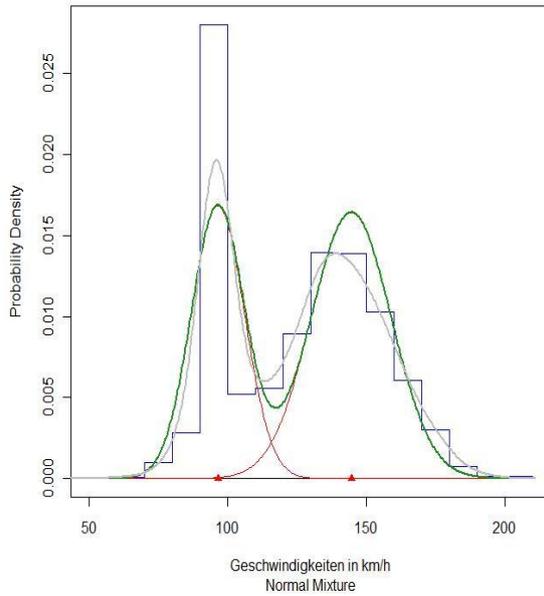
Alle Geschwindigkeiten

Wochentag

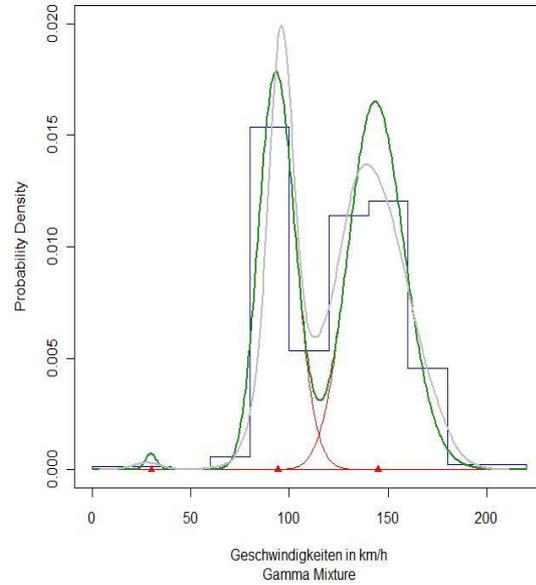


Samstag (siehe S. 23)

Mischmodell Samstage Route AB 6 bis 13.59 Uhr

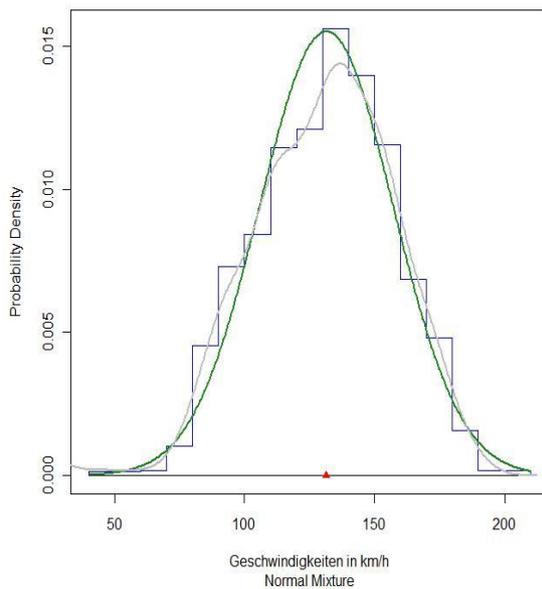


Mischmodell Samstags Route AB 6 bis 13.59 Uhr

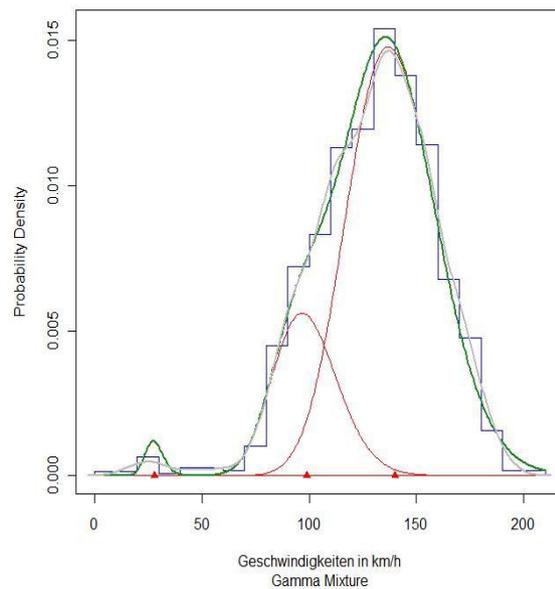


Sonntag (siehe S.27)

Mischmodell Sonntage Route AB 11 bis 17.59 Uhr



Mischmodell Sonntags Route AB 11 bis 17.59 Uhr



	Wochentage		Samstage		Sonntage	
	Geschw. >50 km/h	Alle Geschw.	Geschw. >50km/h	Alle Geschw.	Geschw. >50km/h	Alle Geschw.
# Verteilungen	2	3	2	3	1	3
VC - Koeffizient	8.4	10.7	10	9.8	19.6	15.1
Durchschnittsgeschw. in km/h	95, 141	29, 95, 141	97, 145	30, 94, 145	132	28, 99, 140
Standardabweichung In km/h	8, 12	4, 15, 21	10, 14	3, 9, 14	26	4, 15, 21
Gewichtung in %	63, 37	5, 60, 35	41, 59	0.5, 41, 58.5	100	1, 21, 78
Stichprobengröße	2421	2548	1330	1337	1082	1096

Die Einschränkung der Geschwindigkeit zeigt an den Werktagen und Samstagen nur geringfügige Änderungen, bezüglich der Durchschnittsgeschwindigkeiten. Während die Standardabweichungen an den Samstagen annähernd gleich bleiben, ist an den Werktagen eine deutlich niedrigere Streuung erkennbar. Bedingt durch die Reduktion von 3 auf eine Geschwindigkeitsklasse, an den Sonntagen, sind deutliche Unterschiede zwischen den iterierten Kennzahlen, des jeweiligen Ansatzes, zu sehen.

Unverändert blieben hingegen die Ergebnisse der Chi – Quadrat – Tests, wonach an allen 3 verglichenen Zeitintervallen keine gute Anpassung des iterierten Dichteschätzers an den empirischen Dichteschätzer erreicht werden konnte. Zusammenfassend ist zu sagen, dass die Eingrenzung der Geschwindigkeiten auf mindestens 50 km/h die Anpassungsgüte der iterierten Modelle nicht wesentlich verbessert hat.

2.5.9 Conclusio

Mittels einer Mischverteilungsanalyse ist eine Trennung in die verschiedenen Geschwindigkeitsklassen möglich. Anhand der iterierten Kennzahlen können unmittelbare Vergleiche zwischen den einzelnen Zeitintervallen, je Tagestyp, getroffen werden. Es ist aber anzumerken, dass in keinem betrachteten Zeitintervall, eine gute Anpassungsgüte erzielt werden konnte. Die Elimination der niedrigsten Geschwindigkeitsklassen, brachte keine positive Änderung. Es ist aber anzumerken, dass die Einschränkung der Reisezeit, die Modellierung mit einer symmetrischen Normalverteilung ermöglichte.

Inwieweit die ermittelten Variationskoeffizienten den tatsächlichen Verkehr widerspiegeln, kann aufgrund der vorliegenden Daten, nicht erruiert werden. Eine diesbezügliche Kombination zwischen Bluetoothdaten und Zählraten wäre von Vorteil, um

die Verkehrszusammensetzung in die bestehende Analyse einfließen zu lassen. In Anbetracht der Tatsache, dass ab einer Stichprobengröße von ca. 300 Pärchen, aussagekräftige Diagnosen gestellt werden können, wäre eine dynamische Durchführung solcher Analysen durchaus möglich. Durch Erkennung diverser Verkehrsanomalien, wie hohe Verkehrsdichten oder Staus, ist in Zukunft ein unmittelbares Eingreifen in das aktuelle Verkehrsgeschehen durchaus denkbar. Um dies durchführen zu können, müssten jedoch Anpassungen der Sensoranzahl an die jeweilige Fahrtstrecke vorgenommen werden.

Inwieweit sich die errechneten Kennzahlen bei Verkehrsabnormalitäten ändern, kann aufgrund des vorliegenden Datenmaterials nicht eruiert werden, da keines der Merkmale, die das „Stauverhalten“ abbilden, registriert werden konnte. Es ist aber anzunehmen, dass sich die Parameter der einzelnen Geschwindigkeitsgruppen, bei stark steigendem Verkehr auf einem niedrigeren Niveau, annähern. Das würde eine starke Überlappung der einzelnen Verteilungen implizieren. Inwieweit sich die Variabilität der Durchschnittsgeschwindigkeiten ändert, kann jedoch nicht prognostiziert werden.

2.6 Mathematische Grundlagen von Mischverteilungen

In diesem Kapitel wird grundlegend Wichtiges zum Thema Mischverteilungen wiederholt. Konkret geht es um die Definition einer Mischverteilung inklusive Beispielen, die Gammaverteilung und den Chi – Quadrat – Test. Diesbezügliche Referenzen sind Samuel Kotz [2], Thorn Thaler [3], Ursa Pantle [4], Uni – Oldenburg [5] und P.D.M. Macdonald [7] und [8].

Bemerkung:

Grundsätzlich versucht man ein einfaches Verteilungsmodell zu verwenden. Häufig tritt aber der Fall ein, dass ein einfaches Modell zu grob oder ungenau ist. Ein Grund dafür kann sein, dass die Häufigkeiten eines Merkmals multimodal sind. Solche heterogenen Populationen setzen sich demnach aus mehreren Gruppen (Subpopulationen) zusammen.²

2.6.1 Beispiel:

Man nehme die erwachsene Bevölkerung von Graz und erhebt von jedem einzelnen, dessen Körpergröße. Welche Verteilung wird dieses Merkmal am besten beschreiben? Stellt man die erhobenen Daten beispielsweise mit einem Histogramm dar, wird man 2 Erhebungen feststellen. Grund dafür ist, dass Männer eine höhere Durchschnittsgröße aufweisen als Frauen.

Solche Beobachtungen werden in der Regel mit der Mischung zweier Dichtefunktionen von Normalverteilungen beschrieben:

$$f(x) = pf_1(x) + (1-p)f_2(x)$$

mit $f_1 \sim N(\mu_f, \sigma_f)$ und $f_2 \sim N(\mu_m, \sigma_m)$

p ... Frauenanteil an der Gesamtbevölkerung

(μ_f, σ_f) ... Durchschnittsgröße und Standardabweichung der Frauen

(μ_m, σ_m) ... Durchschnittsgröße und Standardabweichung der Männer

Die Dichtefunktion der Größe setzt sich also aus einer Mischung der Dichtefunktion der Frauen und jene der Männer, zusammen³.

² Literaturverzeichnis [3]

³ Literaturverzeichnis [1]

2.6.2 Diskrete Mischung diskreter Verteilungen

Definition 1: Seien $P_1(x), \dots, P_n(x)$ Wahrscheinlichkeitsfunktionen. Dann definiert sich die Mischung von diskreten Verteilungen als

$$P(x) = r_1 \cdot P_1(x) + r_2 P_2(x) + \dots + r_n P_n(x)$$

$$0 \leq r_i \leq 1 \quad \forall i$$

$$\text{mit } \sum_{i=1}^n r_i = 1.$$

Satz 1: Die so definierte Mischverteilung ist selbst wiederum eine diskrete Verteilung, wobei für das k-te Moment gilt:

$$\mu'_k = r_1 \mu'_k(1) + r_2 \mu'_k(2) + \dots + r_n \mu'_k(n) = \sum_{i=1}^n r_i \mu'_k(i)$$

$\mu'_k(i)$ bezeichnet das k – te Moment der i – ten Mischkomponente.⁴

Beispiel einer diskreten Mischverteilung: Werfen von Münzen

Betrachten wir 2 Münzen, für die folgendes gilt:

$$P(\{\text{„Kopf“}\}) = 0.1 \quad \text{Münze 1}$$

$$P(\{\text{„Kopf“}\}) = 0.8 \quad \text{Münze 2}$$

Nun wählen wir eine dieser Münzen zufällig aus und werfen diese viermal. Sei X die Anzahl der geworfenen „Köpfe“. Münze 1 wird mit Wahrscheinlichkeit r ausgewählt, d. h.

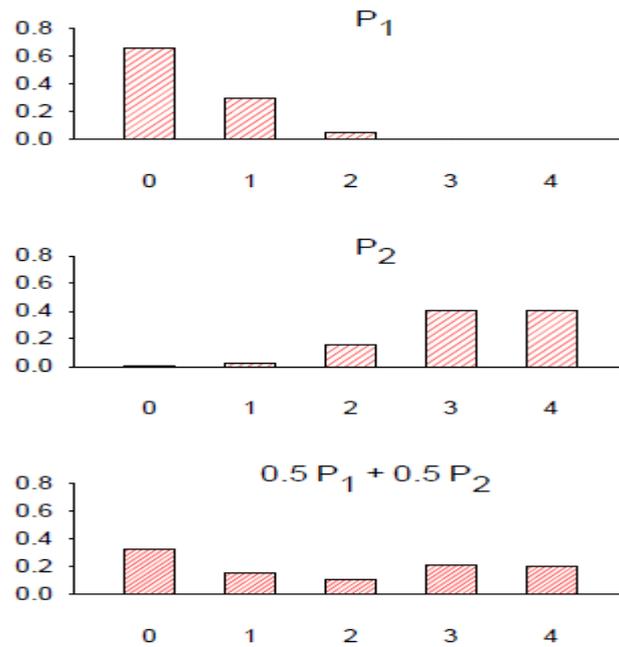
$$P(\{\text{wählen Münze 1}\}) = r.$$

Dann gilt für die Wahrscheinlichkeit von X

$$\begin{aligned} P(\{X = x\}) &= P(\{X = x, \text{Münze 1 ausgewählt}\} \cup \{X = x, \text{Münze 2 ausgewählt}\}) \\ &= P(\{X = x \mid \text{Münze 1}\})P(\{\text{Münze 1}\}) + P(\{X = x \mid \text{Münze 2}\})P(\{\text{Münze 2}\}) \\ &= \binom{4}{x} 0.1^x (1-0.1)^{4-x} r + \binom{4}{x} 0.8^x (1-0.8)^{4-x} (1-r). \end{aligned} \quad ^5$$

⁴ Literaturverzeichnis [3]

⁵ Literaturverzeichnis [1]


 Diskrete Mischung zweier Binomialverteilungen für $r = 0.5$

2.6.3 Diskrete Mischung stetiger Verteilungen

Definition 2: Seien $f_1(x), \dots, f_n(x)$ Dichtefunktionen. Dann definiert sich die Mischung stetiger Verteilungen als

$$f(x) = r_1 \cdot f_1(x) + r_2 f_2(x) + \dots + r_n f_n(x)$$

$$0 \leq r_i \leq 1 \quad \forall i$$

$$\text{mit } \sum_{i=1}^n r_i = 1.$$

Satz 1 gilt auch für stetige Verteilungen und deren Momente.⁶

Beispiel einer stetigen Mischverteilung: Sei X eine Zufallsvariable, die eine Länge von Tieren wiedergibt. Y sei eine Zufallsvariable, die das Geschlecht der Tiere repräsentiert, aber nicht beobachtet werden kann. Sei

$f_1(x)$ die Dichtefunktion der Länge für männliche Tiere

$f_2(x)$ die Dichtefunktion der Länge für weibliche Tiere

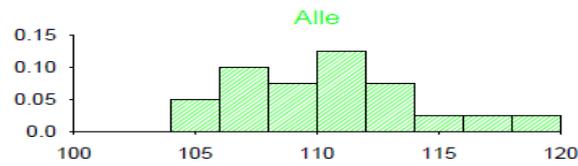
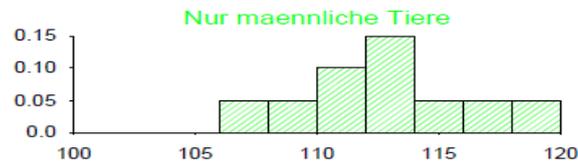
$r = P(\{\text{Männliches Tier}\})$

$1 - r = P(\{\text{Weibliches Tier}\})$.

⁶ Literaturverzeichnis [3], [2]

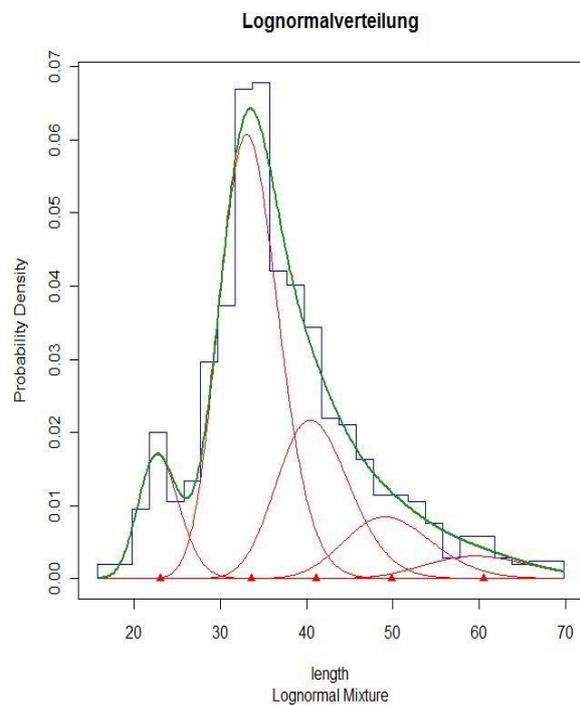
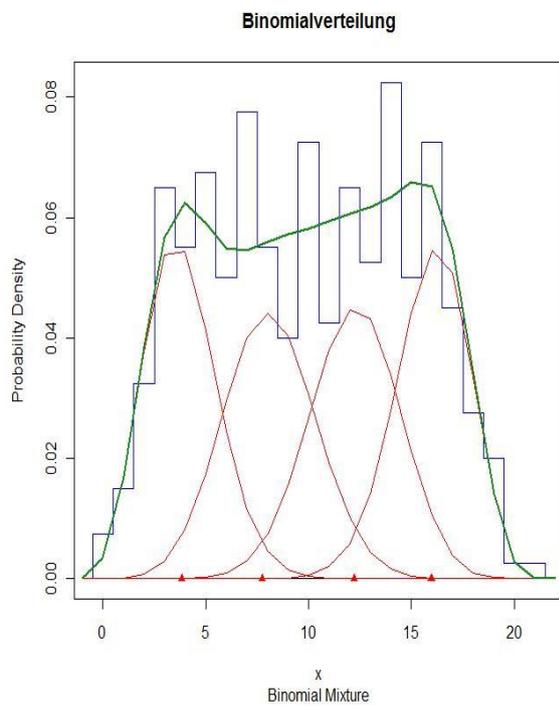
Die Dichtefunktion der Mischverteilung berechnet sich dann wie folgt:

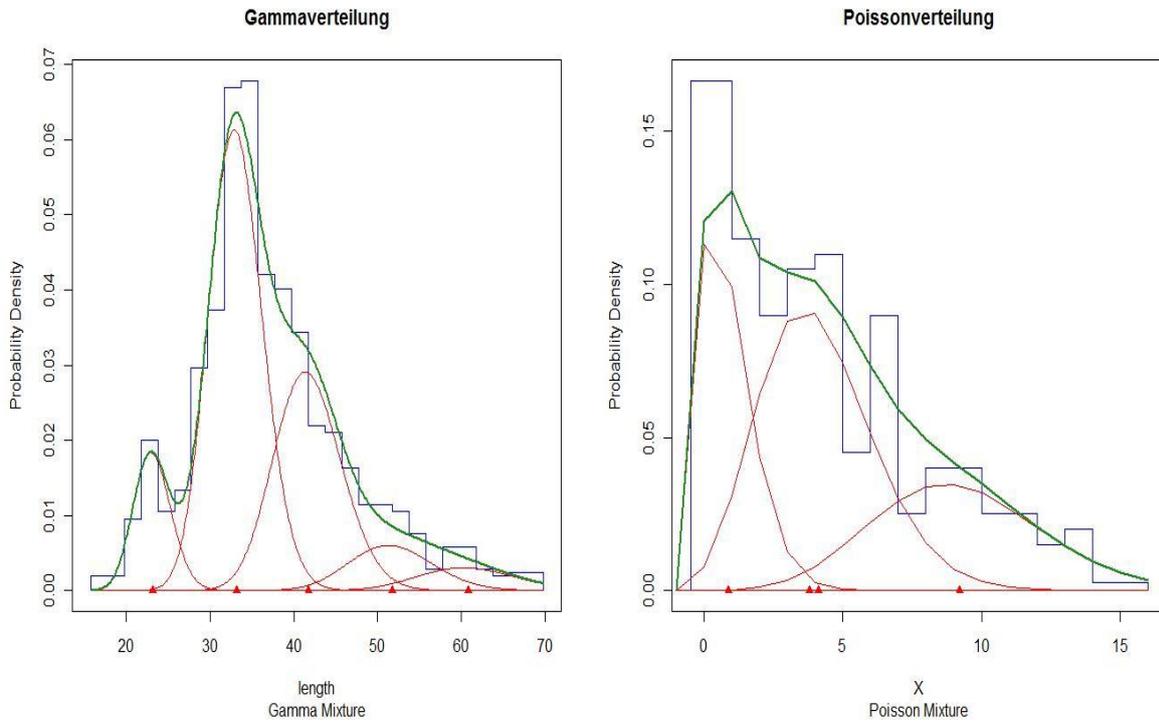
$$f(x) = P(\{\text{Männliches Tier}\})f_1(x) + P(\{\text{Weibliches Tier}\})f_2(x) \\ = rf_1(x) + (1 - r)f_2(x).$$



Kieflerlängen von weiblichen und männlichen Tieren

2.6.4 Beispiele von Mischungen





Parameterfestlegung :

R-Code

```

mix(mixdat, mixpar, dist = "norm", constr = list(conpi = "NONE", conmu = "NONE",
consigma = "NONE", fixpi = NULL, fixmu = NULL, fixsigma = NULL, cov = NULL, size = NULL),
emsteps = 1, usecondit = FALSE, exptol = 5e-06, print.level = 0, ...)
    
```

mixdat	Data frame, der den gruppierten Datensatz enthält, Eingabe erforderlich
mixpar	Data frame, der die Iterationsstartwerte enthält, Eingabe erforderlich
dist	eine der folgenden Verteilungen, „norm“, „lnorm“, „gamma“, „weibull“, „binom“, „nbinom“ oder „pois“, Eingabe erforderlich

Welche zusätzlichen Startparameter vom Benutzer bekannt gegeben werden müssen, hängt von der gewählten Verteilung ab. In diesem Zusammenhang verweise ich auf das Package `mixdist`⁷ und auf den User's Guide⁸, die detaillierte Beschreibungen enthalten.

⁷ Literaturverzeichnis [8]

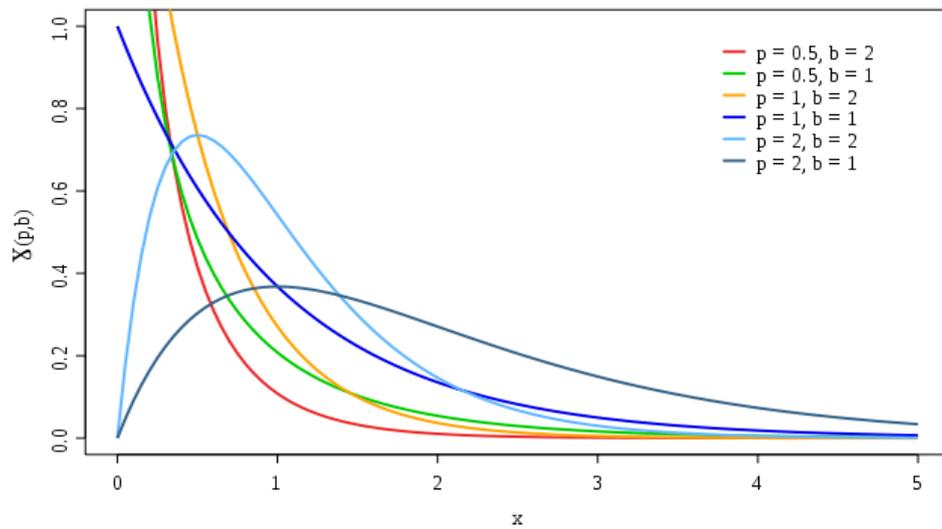
⁸ Literaturverzeichnis [7]

2.6.4 Gammaverteilung

Definition 2: Eine Zufallsvariable Y sei gammaverteilt, mit den Parametern $b > 0$ und $p > 0$, wenn die Dichte von Y gegeben ist, durch

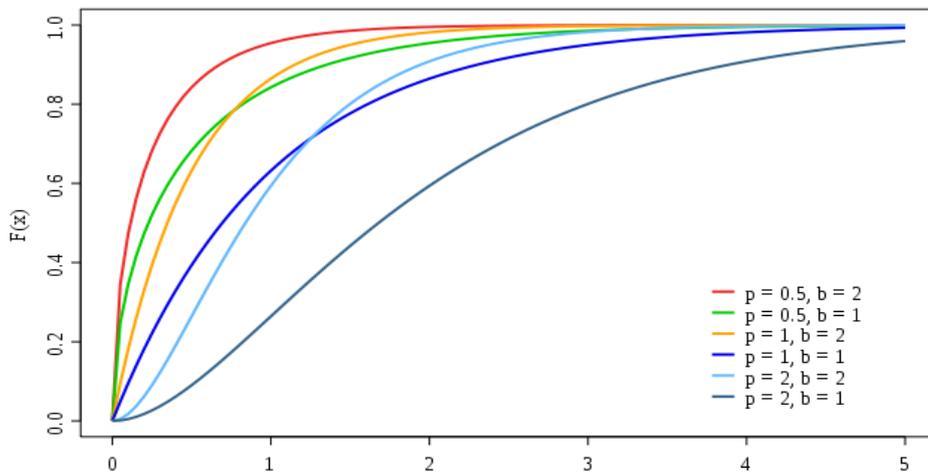
$$f_Y(y) = \begin{cases} \frac{b^p}{\Gamma(p)} e^{-by} y^{p-1} & \text{falls } y > 0 \\ 0 & \text{falls } y \leq 0 \end{cases}$$

Schreibweise: $Y \sim \Gamma(b, p)$



Dichtefunktionen einer Gammaverteilung unterschiedlicher Parameter

Die Gammaverteilung hat damit folgende Verteilungsfunktion $F_Y(y) = \int_0^y f_y(t) dt$ [4].⁹



Verteilungsfunktionen einer Gammaverteilung unterschiedlicher Parameter

⁹ Literaturverzeichnis [4]

2.6.5 Chi – Quadrat – Test

Sei F_0 die unbekannte Verteilung der Population, dann werden mittels eines Chi – Quadrat – Tests, die folgenden Hypothesen getestet:

$$H_0: F = F_0 \quad \text{gegen} \quad H_1: F \neq F_0 \quad (\text{zweiseitig})$$

Teststatistik:

$$Y^2 = \sum_{i=1}^k \frac{(n \hat{p}_i - nh_i)^2}{nh_i} \sim \chi_{k-p-1}^2$$

k ... Klassenanzahl

p ... Anzahl der freien Parameter

nh_i ... beobachtete Häufigkeit in Klasse i

\hat{np}_i ... geschätzte Häufigkeit in Klasse i

Der Chi – Quadrat – Test versteht sich hier als Anpassungstest, d. h. er testet anhand der Teststatistik, ob die Differenz zwischen der Anzahl der geschätzten und der Anzahl der beobachteten Häufigkeiten, Chi – Quadrat verteilt ist.⁹

Kapitel 3

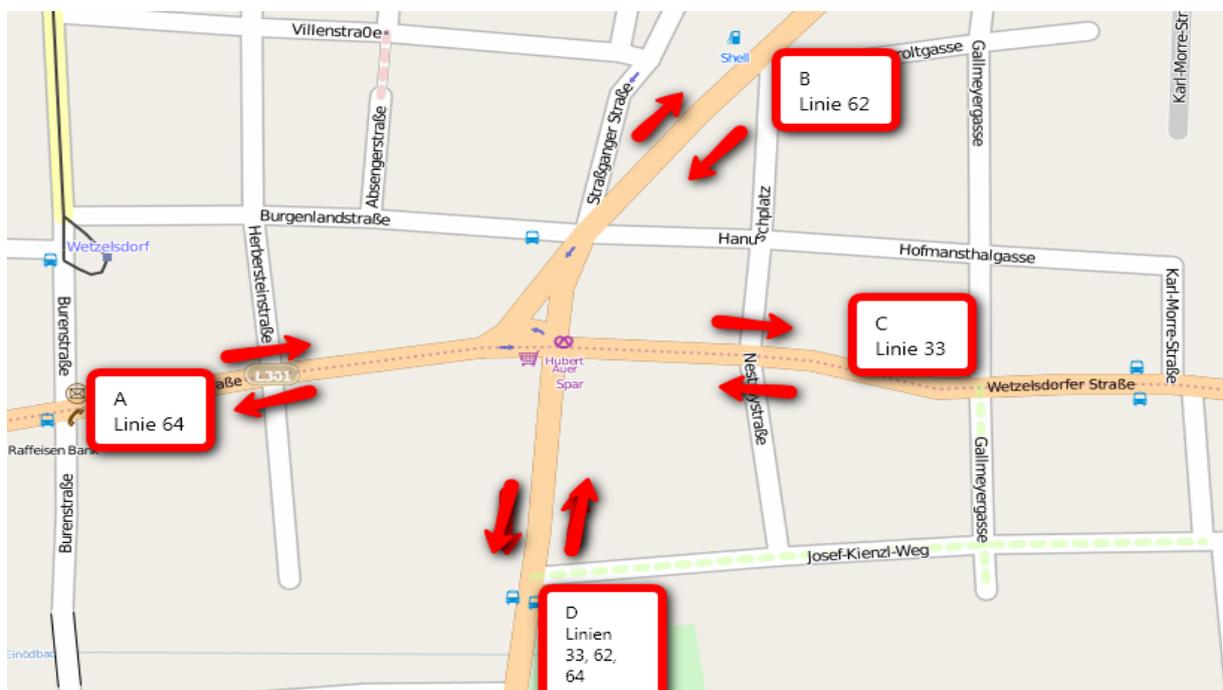
3.1 Verkehrsfrequenzschätzung mittels Regressionsanalyse

In diesem Kapitel wird die Frage geklärt, ob es möglich ist, aufgrund von Verkehrsstichproben, die mittels BLIDS-Sensoren erhoben wurden, realitätsnahe Schätzungen der Gesamtverkehrsfrequenz vorzunehmen, was somit eine Verkehrserhebung mittels Radargeräten, ersetzen könnte. Die gesamte Verkehrsfrequenz wurde mittels Radarwürfeln registriert. Dazu wurden die Daten zweier Kreuzungen im Stadtbereich von Graz (Wetzelsdorf und Straßgang) analysiert.

Die einfachste und zugleich auch ungenaueste Methode ist es, nur aufgrund von Bluetoothquoten¹⁰ auf den Gesamtverkehr rückzuschließen. Die Gefahr dieser einfachen Methode besteht darin, dass die Charakteristiken eines jeden Messortes hinsichtlich der Verkehrsschwankungen im Tagesablauf, Sensorempfindlichkeit oder baulichen Gegebenheiten nicht berücksichtigt werden.

3.2 Allgemeine Bemerkungen

Kreuzung 1: Wetzelsdorf



Pfeile sind Richtungen des öffentlichen Verkehrs.

¹⁰ Bluetoothquote = BLIDS-Frequenz/Radarerfassungen

Standorte

MAC IP-Adresse	Standort	Bezeichnung	Radar 1	Radar 2
20 133	Wetzelsdorferstr. 153	1 A	IN (Ost)	OUT (West)
39 152 (seit 2009-09-07 12:12) (früher 141)	Reinighausstr. (Shell-Tankstelle)	2 B	OUT (Süd)	IN (Nord)
30 143 (seit 2009-09-09 10:15) (früher 133)	Wetzelsdorferstr. 123	3 C	OUT (West)	IN (Ost)
40 154	Strassgangerstr. 114	4 D	OUT (Süd)	IN (Nord)

Radarkonfigurationen

Standort	Radar Nr	Winkelfaktor	Sens	PKW	LKW	sonst.
4	R1	1.320	11	100-380	509, 821, Bus: 840,731	Motorrad: 200
	R2	1.388	11	200-430	517, 492, 940, 1300, 650	Pritsche: 419, 513
3	R1	1.388	10			

Codierung

Messpunkt	A	B	C	D
A		West/Nord EA A In/B In	West/Ost EA A In/C In	West/Süd EA A In/D Out
B	Nord/West EA B Out/A Out		Nord/Ost EA B Out/C In	Nord/Süd EA B Out/D Out
C	Ost/West EA C Out/A Out	Ost/Nord EA C Out/B In		Ost/Süd EA C Out/D Out
D	Süd/West EA D In/A Out	Süd/Nord EA D In/B In	Süd/Ost EA D In/C In	

Notation: EA Ein/Aus...Kreuzungseinfahrt/Kreuzungsausfahrt

Kreuzung 2: Straßgang



Pfeile sind Richtungen des öffentlichen Verkehrs.

Standorte

MAC	IP-Adresse	Standort	Bezeichnung	Radar 1	Radar 2
31	144	Strassgangerstr. (Knusperhaeuschen)	5 E	OUT (Süd)	IN (Nord)
28	141	Kaertnerstr. 381	6 F	OUT (Süd)	IN (Nord)
36	149 (seit 2009-09-08) (frueher 136)	BIT	7 G	OUT (West)	IN (Nord)

Radarkonfigurationen

Standort	Radar Nr	Winkelfaktor	Sens	PKW	LKW	sonst.
4	R1	1.320	11	100-380	509, 821, Bus: 840,731	Motorrad: 200
	R2	1.388	11	200-430	517, 492, 940, 1300, 650	Pritsche: 419, 513
3	R1	1.388	10			

Codierung

Messpunkt	E	F	G
E		Nord/Süd EA E Out/F Out	West/Ost EA E In/G In
F	Süd/Nord EA F In/E In		Süd/Ost EA F In/G In
G	Ost/West EA G Out/E Out	Ost/Süd EA G Out/F Out	

Notation: EA Ein/Aus...Kreuzungseinfahrt/Kreuzungsausfahrt

Beide Versuchskreuzungen liegen in einem dicht bebauten Stadtgebiet. Im Erfassungsbereich aller Sensoren befinden sich daher Wohnungen, Arbeitsplätze, Geschäfte und Freizeitbereiche. Dadurch entstand das Problem des sogenannten „Datenmülls“, d. h. es wurden Bluetoothsignale registriert, die nicht dem Verkehr zuzuordnen waren. Durch die geringen Entfernungen der BLIDS-Sensoren zueinander, entstanden zudem noch Überschneidungen ihrer Erfassungsbereiche. Das führte wiederum zu Erfassungen desselben Bluetoothsignals an mehr als 2 Sensoren. Es entstanden „künstliche“ Routen, die im Realverkehr nicht möglich waren.

Mehrfacherfassungen desselben Signals konnten durch den, in Punkt 2.1 beschriebenen Algorithmus, entfernt werden. Routen, deren Dauer eine bestimmte Zeitschranke überschritten, wurden ebenfalls eliminiert. Dadurch konnte die Mehrzahl von registrierten Fußgängern ebenfalls entfernt werden. Ein größeres Problem stellten jene Verkehrsteilnehmer dar, die sich längere Zeit im Kreuzungsbereich aufhielten, weil sie dort arbeiteten, wohnten oder einkauften und ´unmögliche´ Routen produzierten.

Dieses Problem wurde damit gelöst, dass alle Sensoren, an denen dasselbe ID-Signal empfangen wurde, zu einer Gruppe zusammengefasst wurden. Bestand diese Gruppe aus mehr als 2 Messpunkten, wurde als Anfangspunkt einer Route, jener Messpunkt gewählt, an dem das ID-Signal zuerst empfangen wurde, und als Endpunkt, jener Sensor, an dem dasselbe ID-Signal zuletzt registriert wurde. Überschritt die Reisedauer, d. h. Differenz zwischen Zeitpunkt der ersten und letzten Erfassung, eine volle Stunde, so wurde diese Route als nicht verkehrsrelevant eingestuft und somit nicht berücksichtigt.

Für die Durchführung einer aussagekräftigen Regressionsanalyse, ist es erforderlich, vollständige Datensätze zu verwenden. Bedingt durch fehlende Datenregistrierungen unterscheiden sich daher die Beobachtungszeiträume beider Versuchsanlagen. Zusätzlich war, die für eine Analyse sinnvolle Tageszeit auf 6 – 19 Uhr einzuschränken, da nur zu diesen Stunden genügend Routen gebildet werden konnten.

Kreuzung 1:

Beobachtungszeitraum Sonntag 10. Jänner 2010 bis Samstag 20. Februar 2010 (42 Tage)

Kreuzung 2:

Beobachtungszeitraum Sonntag 7. Februar 2010 bis Samstag 13. März 2010 (35 Tage)

Weiters war die Richtungsdefinition der verbliebenen Bluetoothpärchen an jene der Radardaten anzugleichen. Die eingesetzten Radargeräte unterschieden nämlich nur zwischen in die Kreuzung ein- bzw. aus der Kreuzung ausfahrende Fahrzeuge. Um eine Informationsgleichwertigkeit der Daten sicherzustellen, wurde eine 'Anonymisierung' der Pärchen durchgeführt. D. h. alle Bluetooth Routen die denselben Ausgangspunkt aufweisen, wurden addiert. Die in Punkt 2.2 getroffene Einteilung in die verschiedenen Tagestypen ist auch in diesem Kapitel gültig.

3.3 Vorgehensweise

Ziel einer multiplen linearen Regressionsanalyse ist es, festzustellen, ob es eine lineare Beziehung zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen gibt. In unserem Fall ist die mittels Radar erhobene Verkehrsfrequenz die abhängige Variable, während die BLIDS – Frequenzen die einzige unabhängige Variable darstellen. Da an allen 3 Tagestypen eine unterschiedliche Verkehrsfrequenz herrscht, ist zusätzlich eine Faktorisierung, der zugrundeliegenden Daten, hinsichtlich des Tagestyps der Erhebung, notwendig.

Im ersten Schritt ist eine explorative Datenanalyse durchzuführen. Anhand der verschiedenen Quotenverläufe je Tagestyp ist nun zu entscheiden, ob die vorgenommene Faktorisierung ausreicht, oder ob zusätzlich eine Aufteilung der Tageszeit in Stunden vorgenommen werden soll. Weiters soll eruiert werden, ob die Bluetoothfrequenz linear oder quadratisch in das Modell eingeht. Obwohl es dadurch zwei Variable $x_1 = x$ und $x_2 = x^2$ im Modell gibt, handelt es sich nach wie vor um ein lineares Modell der Form

$y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$, mit $\varepsilon \sim N(0, \sigma^2)$ bei dem die Parameter a_0 , a_1 und a_2 auf Grund der

Daten zu schätzen sind und das konkrete Modell $\hat{y} = b_0 + b_1x_1 + b_2x_2$ liefern. Folgende Fälle sind diesbezüglich zu unterscheiden:

(i) Linearer Quotenverlauf ohne Steigung:

#Faktorisierung nach Tagestyp
#Errechnung eines linearen Regressionsmodells
je Tagestyp

(ii) Linearer Quotenverlauf mit Steigung:

#Faktorisierung nach Tagestyp und Stunden
#Errechnung eines linearen Regressionsmodells
je Tagestyp und Stunde

(iii) Quadratischer Quotenverlauf:

#Faktorisierung nach Tagestyp und Stunden
#Errechnung eines Regressionsmodells mit
quadratischem Term je Tagestyp und Stunde

Bestimmte Merkmale eines errechneten Regressionsmodells geben Auskunft über die Aussagekraft und Qualität der Schätzung. Ein fertiges Modell spezifiziert eine Schätzung als Linearkombination von verschiedenen Parametern. Mittels eines F – Tests wird die Nullhypothese getestet, ob die Parameter b_i gleich Null sind. Eine Ablehnung dieser

Hypothese sagt aus, dass zumindest einer dieser Parameter signifikant von Null verschieden ist, und somit eine Linearkombination bzw. Schätzung, der abhängigen Variablen, überhaupt sinnvoll ist. Ein weiteres Qualitätsmerkmal, sind die beiden Bestimmtheitsmaße R^2 und R_{adj}^2 . Sie geben an, welcher Anteil der Gesamtvarianz durch das Modell erklärt wird. Damit ein Overfitting (Überanpassung des Modells: Berücksichtigung zu vieler unabhängiger Variablen) ausgeschlossen werden kann, sollten sowohl der Multiple R-squared, als auch der Adjusted R-squared, in etwa gleich groß sein. Je größer beide Werte sind, desto größer ist der Anteil der Varianz der durch dieses Modell erklärt wird.

Mittels der verschiedenen errechneten Modelle kann nun die tatsächliche Verkehrsfrequenz y_i anhand der vorhandenen BLIDS – Frequenzen durch \hat{y}_i geschätzt werden. Schätzungen \hat{y}_i können aber von den tatsächlichen Frequenzen y_i abweichen. Die Residuen $r_i = y_i - \hat{y}_i$ müssen bestimmte Anforderungen erfüllen, um die Adäquatheit eines Modells zu bestätigen. Liegt beispielsweise eine Inhomogenität der Varianzen oder keine Normalverteilung des Zufallsvektors ε vor, dann verletzt das Modell die Kriterien, welche eine zuverlässige statistische Interpretation des Modells erlauben (siehe Abschnitt 3.6.8). Eine Stabilisierung der Varianzen kann mittels einer Box – Cox – Transformation der zugrunde liegenden Daten erreicht werden (siehe Abschnitt 3.6.9). Eine erneute Analyse mit diesen transformierten Daten ist dann notwendig. Sollte eine dieser erwähnten Bedingungen nicht erfüllt sein, ist die Aussagekraft des Regressionsmodells sehr eingeschränkt.

3.4 Beispiel R - Code

Lineares Modell ohne Steigung:

```
lm.einfahrt.a <- lm(rfrequa ~ bfrequa: tag + tag)
```

lm: Erstellung eines Regressionsmodells
 rfrequa: abhängige Variable Radarfrequenz
 bfrequa: unabhängige Variable BLIDS - Frequenz
 tag: Faktorisierung nach Tagestyp

Lineares Modell mit Steigung:

```
lm.reg.rein <- lm(rfrequb.r ~ bfrequb.r: tag.r + tag.r)
```

lm: Erstellung eines Regressionsmodells
 rfrequb: abhängige Variable Radarfrequenz
 bfrequb: unabhängige Variable BLIDS - Frequenz
 tag.r: Faktorisierung nach Tagestyp und Stunden

Modell mit quadratischem Term:

```
lm.regression.stunden.rein <- lm(rfrequer.r ~ bfrequer.r: tag.r + bfrequer.r ^2: tag.r + tag.r)
```

lm: Erstellung eines Regressionsmodells
 rfrequer: abhängige Variable Radarfrequenz
 bfrequer: unabhängige Variable BLIDS - Frequenz
 tag.r: Faktorisierung nach Tagestyp und Stunden

Merkmale eines Modells:

```
lm(formula = rfrequa ~ bfrequa:tag + tag)
```

Residuals:

Min	1Q	Median	3Q	Max
-145.788	-27.923	-3.619	25.297	162.538

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	79.20837	7.07198	11.200	< 2e-16 ***
tag2	12.54448	16.78358	0.747	0.455
tag3	-74.63128	14.92588	-5.000	7.6e-07 ***
beina:tag1	3.27039	0.08302	39.391	< 2e-16 ***
beina:tag2	3.01891	0.23250	12.985	< 2e-16 ***
beina:tag3	4.70456	0.24542	19.169	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 43.73 on 582 degrees of freedom

Multiple R-squared: 0.816, Adjusted R-squared: 0.8144

F-statistic: 516.1 on 5 and 582 DF, p-value: < 2.2e-16

(Intercept) 79.20837: Wert des Intercepts

2e-16 ***: p-Wert des Faktors, Referenzwert < 0.05, Berücksichtigung des Faktors

tag2: Tagestyp, gegebenenfalls Stunde

79.20837: Parameterschätzungen

Residual standard error: Standardfehler s_e der Regression

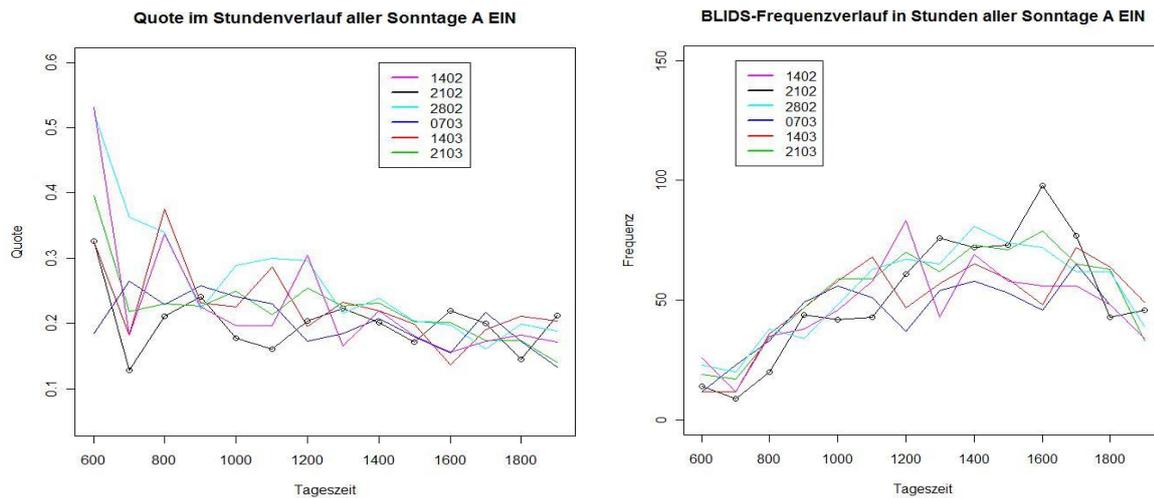
degrees of freedom: Freiheitsgrade

Multiple R-squared: Bestimmtheitsmaß

Adjusted R-squared: Korrigiertes Bestimmtheitsmaß

p-value: p – Wert des Modells, Signifikanz ab Referenzwert < 0.05

0.455: nicht signifikant

Sonntage

Sensor A zeigt unabhängig vom jeweiligen Tagestyp einen ruhigen horizontalen Quotenverlauf. Nur Sonntagmorgen bis ca. 10.00 Uhr ist eine stärkere Quotenfluktuation zu registrieren. Da die BLIDS – Frequenzen an den Wochenenden bzw. die Radarfrequenzen an den Wochentagen konträre Verläufe aufweisen, kann von einer Unabhängigkeit zwischen Quotenhöhe und Frequenz ausgegangen werden. Verkehrsspitzen bzw. -flauten, sind anhand der Quotenverläufe, nicht zu erkennen. Aufgrund dessen ist keine Differenzierung in Tagesstunden, bei der Errechnung eines Regressionsmodells, notwendig und somit eine Faktorisierung nach Tagestyp ausreichend.

Regressionsmodell Sensor A

```
> lm.einfahrt.a <- lm(rfrequa ~ bfrequa:tag + tag); summary(lm.einfahrt.a)
```

Call:

```
lm(formula = rfrequa ~ bfrequa:tag + tag)
```

Residuals:

Min	1Q	Median	3Q	Max
-145.788	-27.923	-3.619	25.297	162.538

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	79.20837	7.07198	11.200	< 2e-16 ***
tag2	12.54448	16.78358	0.747	0.455
tag3	-74.63128	14.92588	-5.000	7.6e-07 ***
beina:tag1	3.27039	0.08302	39.391	< 2e-16 ***
beina:tag2	3.01891	0.23250	12.985	< 2e-16 ***
beina:tag3	4.70456	0.24542	19.169	< 2e-16 ***

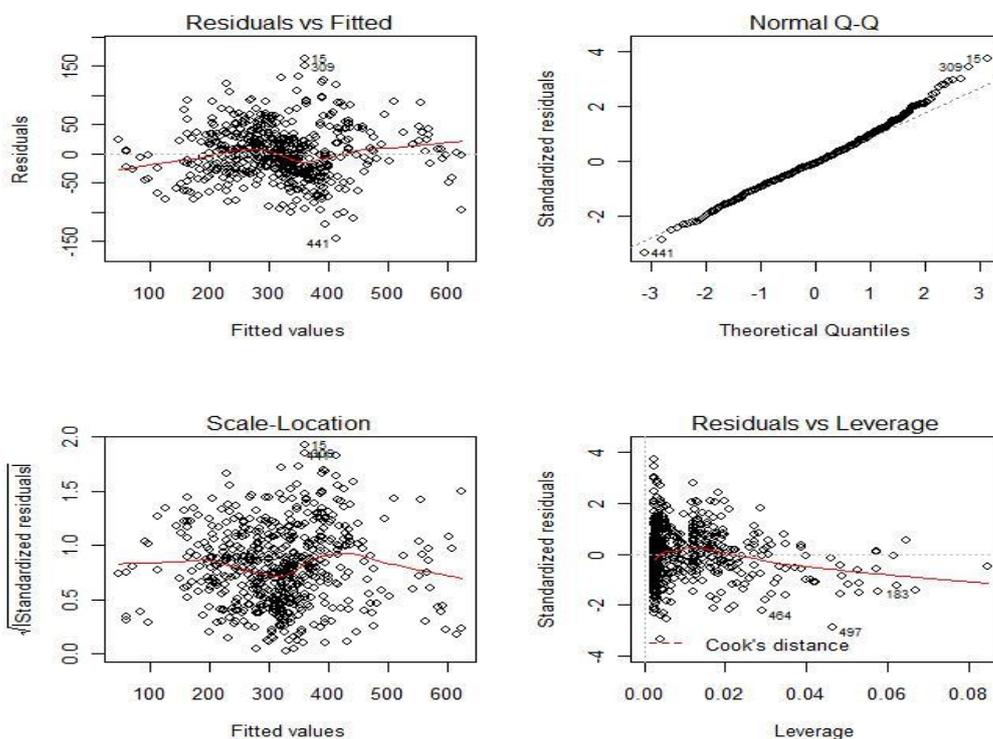
 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 43.73 on 582 degrees of freedom

Multiple R-squared: 0.816, Adjusted R-squared: 0.8144

F-statistic: 516.1 on 5 and 582 DF, p-value: < 2.2e-16

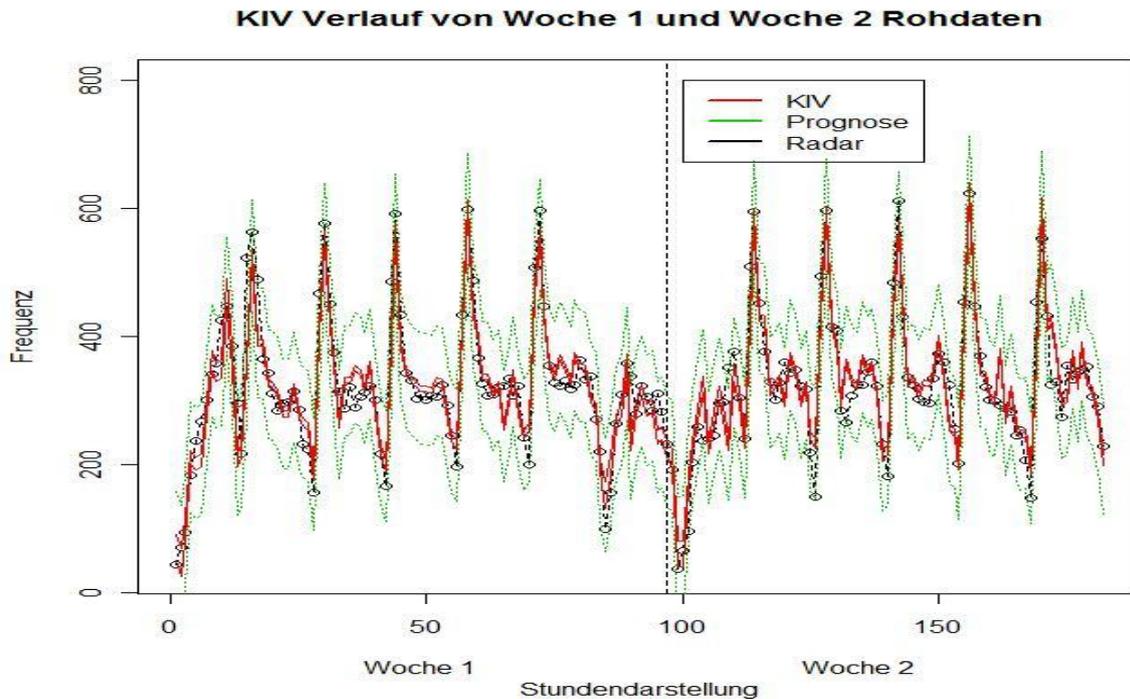
Das lineare Regressionsmodell für die Messstation A ist mit einem p-Wert nahe bei Null hochsignifikant. Beide R Werte sind mit 82 Prozentpunkten annähernd gleich hoch und beweisen eine sehr gute Schätzgenauigkeit des errechneten Modells. Ein Overfitting des Modells ist aufgrund der Wertegleichheit auszuschließen. Die Standardabweichung beträgt 43.73. Dieses errechnete Modell garantiert somit eine hohe Schätzgenauigkeit. Durch die Nichtsignifikanz von tag2 gilt der Intercept von 79.21 sowohl an Werktagen als auch an Samstagen.



Der Residuenplot zeigt eine gleichmäßige Anordnung der Punkte, in einem breiten horizontalen Band. Datenkonzentrationen und Muster sind demnach nicht zu erkennen. Durch diese Anordnung, ist kein linearer Zusammenhang zur Verkehrsfrequenz gegeben. Somit kann eine Inhomogenität der Varianzen ausgeschlossen werden.

Anhand des Q-Q-Plots ist eine sehr gute Anpassung der standardisierten Residuen, an die Quantile einer Normalverteilung, ersichtlich. Etwas stärkere Abweichungen sind nur an den

Randbereichen ersichtlich. Der Cook-Distanz-Plot schließt die Beeinflussung der Parameterschätzung, durch einzelne Datenrealisierungen, grundlegend aus.

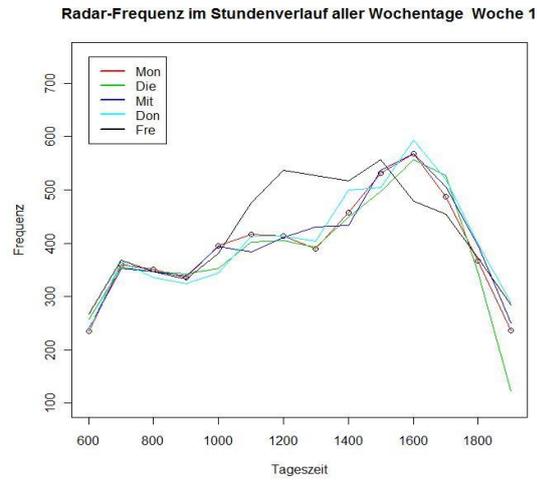
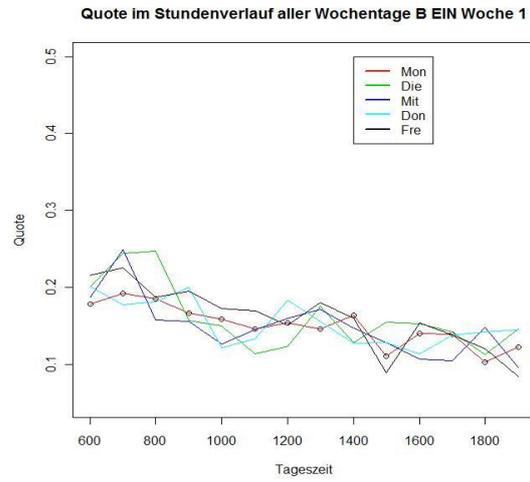


Eine zweiwöchige Gegenüberstellung der Schätzdaten mit den tatsächlichen Verkehrsfrequenzen zeigt eine sehr gute Treffergenauigkeit des errechneten Modells. Größere Abweichungen des Modells werden durch die Enge des 95 prozentigen Konfidenzintervalls¹¹ für den Mittelwert negiert. Insgesamt verdeutlicht dieser Plot eine sehr gute Anpassung des errechneten Konfidenzintervalls, und spiegelt den tatsächlichen Frequenzverlauf gut wieder. Die beobachtete Gesamtfrequenz liegt immer im 95 prozentigen Prognoseintervall für den Einzelwert. Die Schätzgenauigkeit des errechneten Regressionsmodells wird dadurch bestätigt.

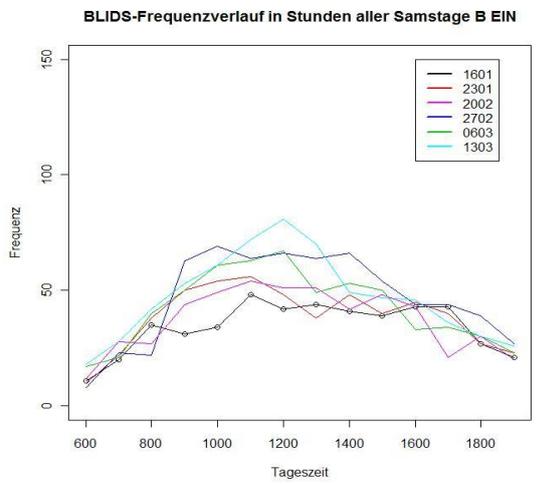
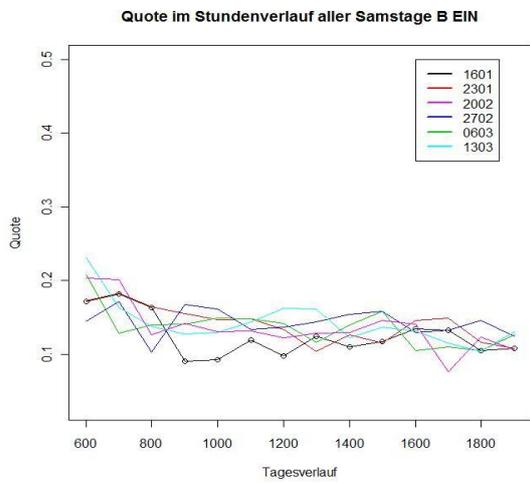
¹¹ Konfidenzintervall KIV oder Vertrauensbereich siehe 3.6.10

3.5.2 Lineares Regressionsmodell mit Steigung: Sensor B

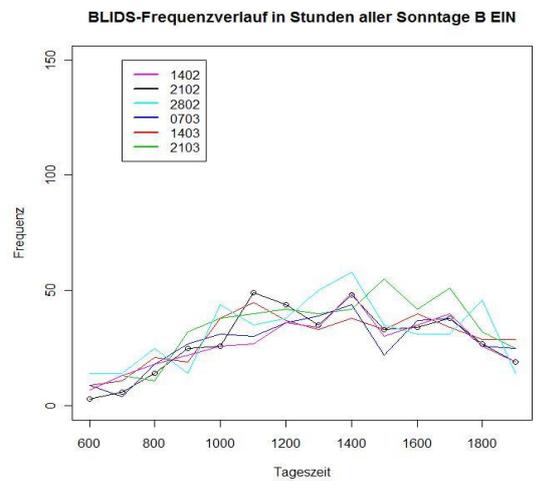
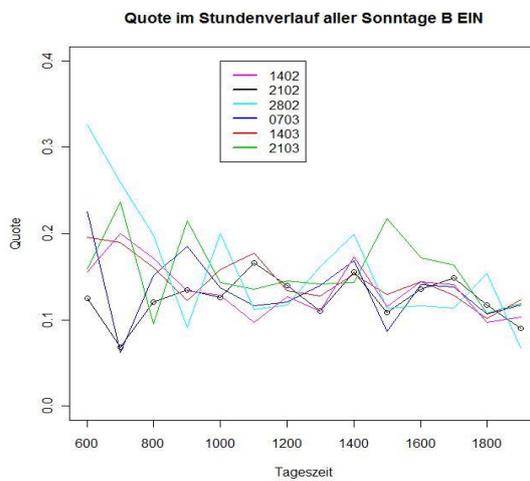
Wochentage



Samstage



Sonntage



Messstation B zeigt an Wochentagen einen linearen Quotenverlauf mit einer negativen Steigung. Vor allem Morgens geht eine niedrigere Verkehrsfrequenz mit einer höheren Bluetoothquote überein. Verkehrsspitzen hingegen, werden durch den Quotenverlauf nicht angezeigt. An Samstagen sind keine Auffälligkeiten erkennbar. Bedingt durch eine extrem niedrige Verkehrsdichte, fluktuiert die errechnete Quote Sonntags bis ca. 8.00 Uhr und beruhigt sich ab einer BLIDS – Frequenz von ca. 40 Signalen pro Stunde. Ansonsten sind keine Auffälligkeiten zu erkennen.

Regressionsmodell Sensor B für Samstage und Sonntage

```
> lm.regression.b <- lm(rfrequb ~ bfrequb:tag + tag); summary(lm.regression.b)
```

Call:

```
lm(formula = reinb ~ beinb:tag + tag)
```

Residuals:

Min	1Q	Median	3Q	Max
-159.162	-47.921	-1.588	44.020	201.618

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	158.5091	12.8443	12.341	< 2e-16 ***
tag2	-115.6483	23.5200	-4.917	1.14e-06 ***
tag3	-112.6912	22.3833	-5.035	6.39e-07 ***
beinb:tag1	3.9174	0.2033	19.270	< 2e-16 ***
beinb:tag2	6.4257	0.4478	14.349	< 2e-16 ***
beinb:tag3	5.7773	0.5657	10.213	< 2e-16 ***

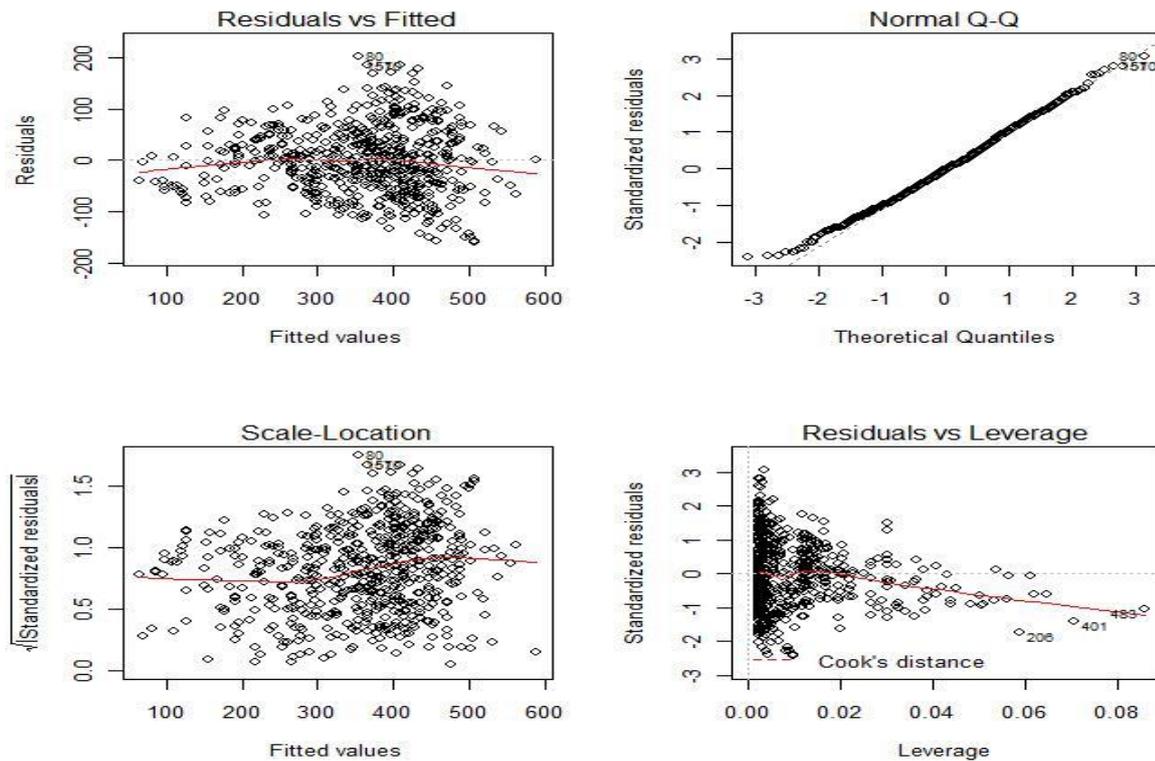
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 65.89 on 582 degrees of freedom

Multiple R-squared: 0.6855, Adjusted R-squared: 0.6827

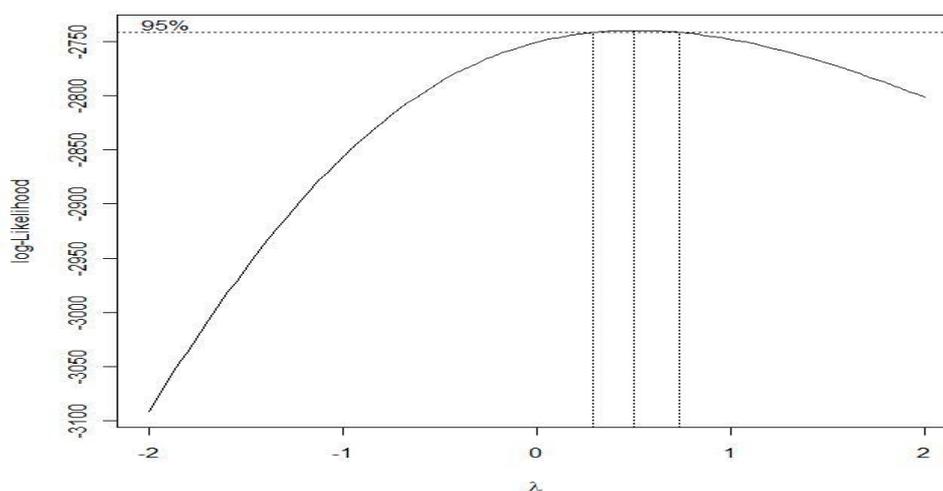
F-statistic: 253.7 on 5 and 582 DF, p-value: < 2.2e-16

Dieses Modell zeichnet sich durch einen akzeptablen Bestimmtheitsgrad aus. Beide R – Werte treffen sich mit rund 68 Prozentpunkten, auf einem relativ niedrigen Niveau. Ein Overfitting kann demnach ausgeschlossen werden. Die Standardabweichung $s_e = 68$ ist um die Hälfte größer als bei Messpunkt A, bedingt durch die schlechte Anpassung (0.68 gegen 0.81). Durch dieses Modell wird ein weitaus geringerer Anteil der Variabilität der Verkehrsfrequenz erklärt und somit besteht ein geringerer Zusammenhang zwischen der Radar- und der BLIDS-Frequenz.



Die Residuenanalyse veranschaulicht eine asymmetrische Fehlerverteilung. Besonders im höheren Frequenzbereich, ist eine Zunahme der Fehlerhäufigkeit zu registrieren. Dadurch nimmt die Punktwolke im ersten Plot die Form eines Trichters an und bezeugt dadurch eine Frequenzabhängigkeit. Diese Residuenanordnung belegt eine Inhomogenität der Varianzen. Um eine Stabilisierung der Varianzen zu erreichen, ist deshalb eine Box-Cox-Transformation der zugrunde liegenden Daten, für die Wochentage, durchzuführen. Aufgrund der Quotenverläufe, an den Wochentagen, ist zudem eine Stundenfaktorisierung durchzuführen.

Regressionsmodell Sensor B mit Stundenfaktorisierung und Wurzel - Transformation für Wochentage



Anhand dieses Plots ist zu erkennen, dass mit einem Wert $\lambda = \frac{1}{2}$ (Wurzeltransformation) eine Stabilisierung der Varianzen erreicht wird.

```
> lm.reg.rein <- lm(sqrt(rfrequb.r) ~ bfrequb.r:tag.r + tag.r-1); summary(lm.reg.rein)
```

Call:

```
lm(formula = sqrt(rfrequb.r) ~ bfrequb.r:tag.r + tag.r - 1)
```

Residuals:

```
    Min     1Q  Median     3Q     Max
-2.83255 -0.44142  0.01302  0.49895  2.69905
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
tag.r1         13.52384   0.82510  16.391 < 2e-16 ***
tag.r2         16.61919   0.83822  19.827 < 2e-16 ***
tag.r3         17.09078   1.02109  16.738 < 2e-16 ***
tag.r4         15.83733   1.24325  12.739 < 2e-16 ***
tag.r5         16.45437   1.03981  15.824 < 2e-16 ***
tag.r6         18.29862   0.92560  19.769 < 2e-16 ***
tag.r7         16.12941   0.91403  17.647 < 2e-16 ***
tag.r8         15.85166   0.74840  21.181 < 2e-16 ***
tag.r9         17.93321   0.80125  22.381 < 2e-16 ***
tag.r10        20.10704   0.83839  23.983 < 2e-16 ***
tag.r11        19.78528   0.92687  21.346 < 2e-16 ***
tag.r12        18.51434   1.27887  14.477 < 2e-16 ***
tag.r13        18.42523   1.04679  17.602 < 2e-16 ***
tag.r14        14.18782   0.71991  19.708 < 2e-16 ***
beinb.r:tag.r1  0.04778   0.01905   2.508 0.01254 *
beinb.r:tag.r2  0.02711   0.01209   2.242 0.02555 *
beinb.r:tag.r3  0.01862   0.01548   1.203 0.22981
beinb.r:tag.r4  0.04385   0.01963   2.233 0.02610 *
beinb.r:tag.r5  0.04957   0.01679   2.953 0.00334 **
beinb.r:tag.r6  0.03272   0.01368   2.393 0.01720 *
beinb.r:tag.r7  0.06720   0.01289   5.212 3.05e-07 ***
beinb.r:tag.r8  0.06930   0.01052   6.586 1.47e-10 ***
beinb.r:tag.r9  0.05554   0.01118   4.967 1.02e-06 ***
beinb.r:tag.r10 0.03741   0.01186   3.155 0.00173 **
beinb.r:tag.r11 0.04555   0.01296   3.516 0.00049 ***
beinb.r:tag.r12 0.05677   0.02017   2.815 0.00513 **
beinb.r:tag.r13 0.02337   0.02188   1.068 0.28599
beinb.r:tag.r14 0.06509   0.02202   2.956 0.00330 **
```

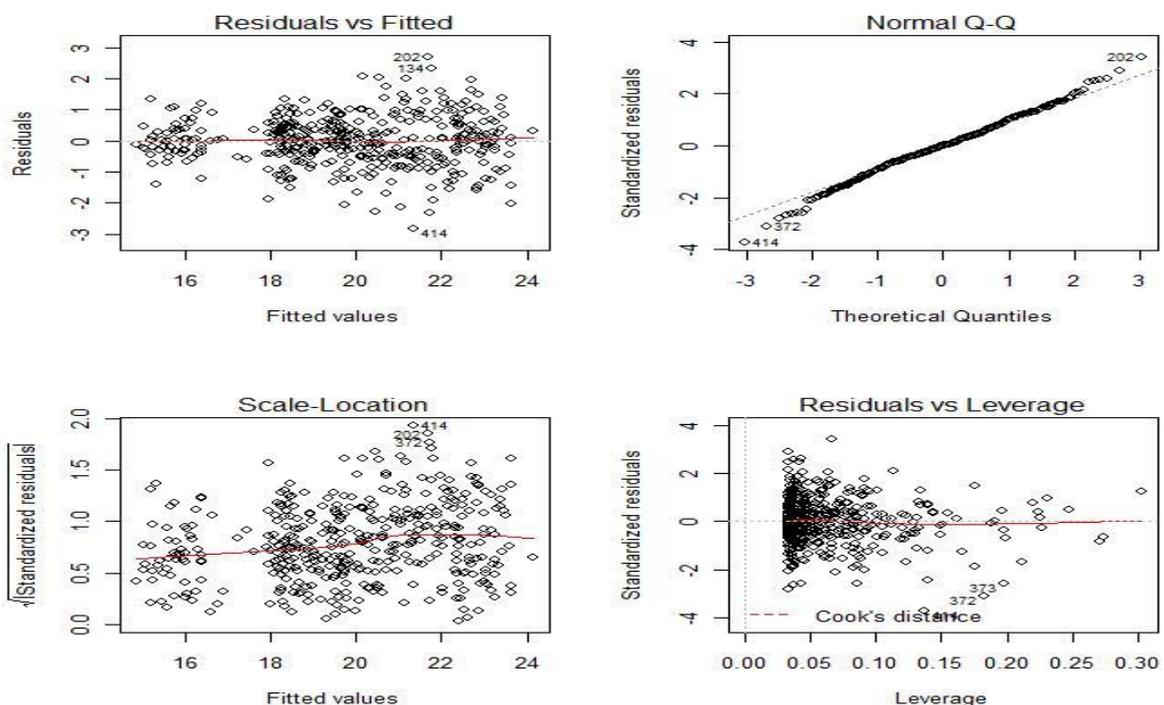
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.817 on 388 degrees of freedom

Multiple R-squared: 0.9984, Adjusted R-squared: 0.9983

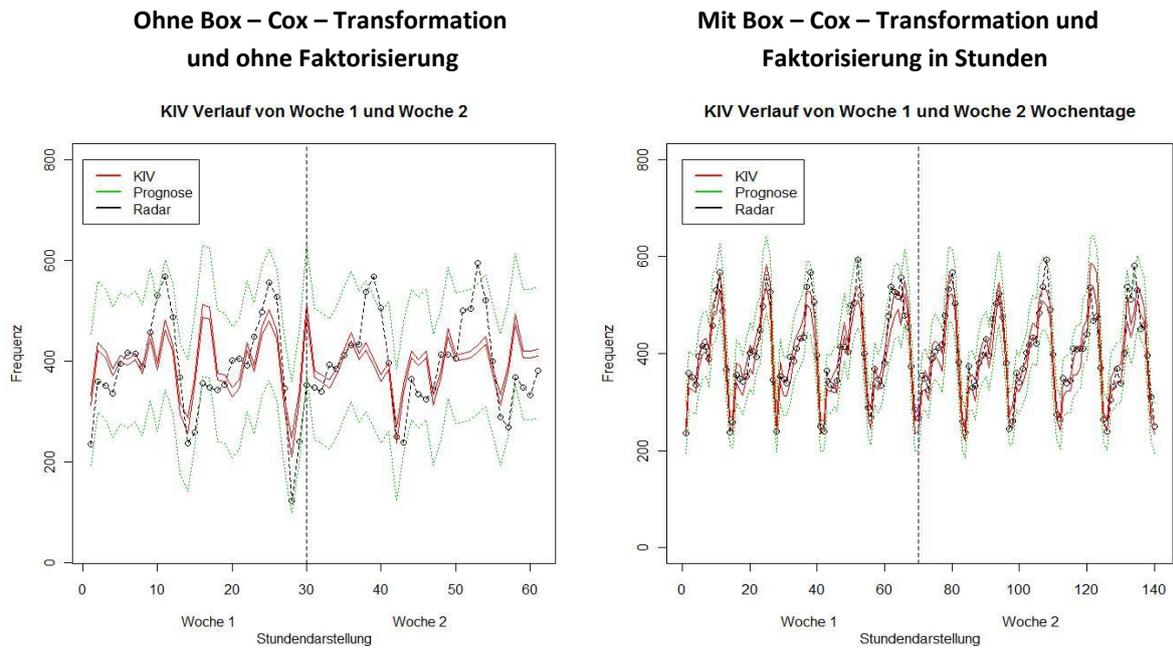
F-statistic: 8887 on 28 and 388 DF, p-value: < 2.2e-16

Dieses Modell ist wiederum hochsignifikant. Beide R – Werte sind nach Entfernung aller redundanten (nicht signifikanten) Parameter, mit 99.8 Prozentpunkten, weitaus höher, als im ursprünglichen Modell und belegen eine bei weitem bessere Approximation an die Realdaten. Ein Overfitting kann, aufgrund der Wertegleichheit, ausgeschlossen werden. Zudem wird durch die Signifikanz fast aller errechneten Stundenparameter, eine hohe Flexibilität gegenüber Verkehrsschwankungen garantiert. Durch die Faktorisierung in Stunden, wird eine deutlich bessere Schätzgenauigkeit erreicht.



Anhand der Residuenanalyse, ist eine deutliche Stabilisierung der Varianzen ersichtlich. Datenkonzentrationen, bzw. Muster sind kaum mehr vorhanden. Nur im untersten Frequenzbereich ist eine leichte Verdichtung zu erkennen. Anhand des Q – Q – Plots wird eine sehr gute Anpassung der standardisierten Residuen an die Quantile einer theoretische Normalverteilung verdeutlicht. Etwas stärkere Abweichungen sind nur an den Randbereichen zu erkennen.

Zudem wird eine Beeinflussung der Schätzparameter durch einzelne Datenrealisierungen, aufgrund des Cook-Distanz-Plots, verneint. Überschreitungen von Distanzlinien sind demnach nicht gegeben.

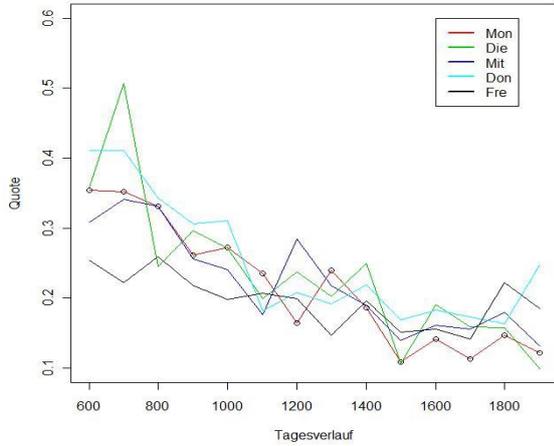


Anhand des linken Plots kann man, besonders an Wochentagen, größere Abweichungen der Schätzungen vom tatsächlichen Verkehrsaufkommen erkennen, während an Samstagen und Sonntagen, eine gute Homogenisierung erreicht wird. Diese Ungenauigkeit wird durch die Breite des Prognoseintervalls, und deren Verlauf, unterstrichen. Mittels der Verfeinerung des Modells konnte für die Wochentage eine sehr gute Approximation der Schätzdaten an die tatsächliche Verkehrsfrequenz erreicht werden. Eine deutliche Verkleinerung, sowohl des Konfidenzintervalls als auch des Prognoseintervalls, unterstreicht dieses Ergebnis.

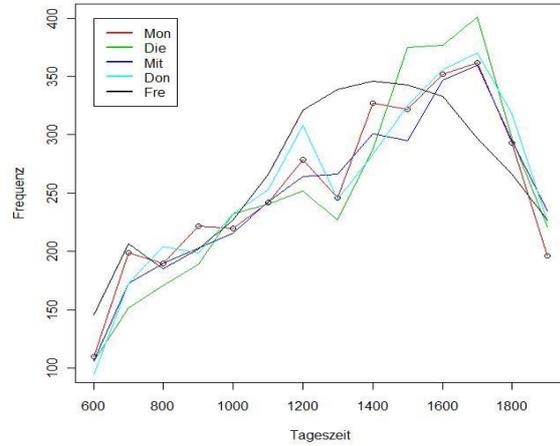
3.5.3 Regressionsmodell mit quadratischen Term: Sensor C

Wochentage

Quote im Stundenverlauf aller Wochentage C EIN Woche 1

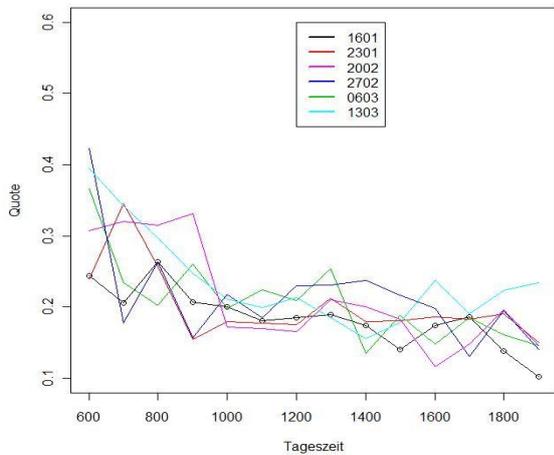


Radar-Frequenz im Stundenverlauf aller Wochentage Woche 1

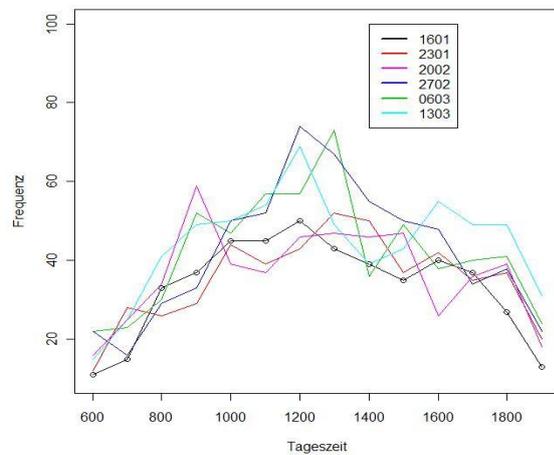


Samstage

Quote im Stundenverlauf aller Samstage C EIN

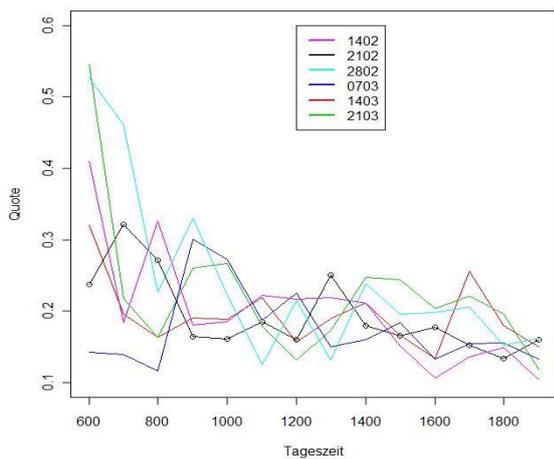


BLIDS-Frequenzverlauf in Stunden aller Samstage C EIN

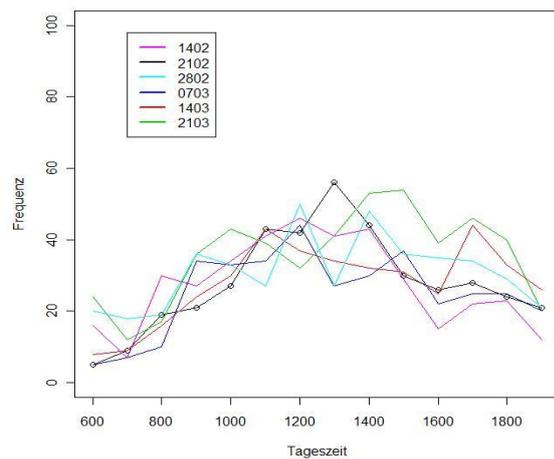


Sonntage

Quote im Stundenverlauf aller Sonntage C EIN

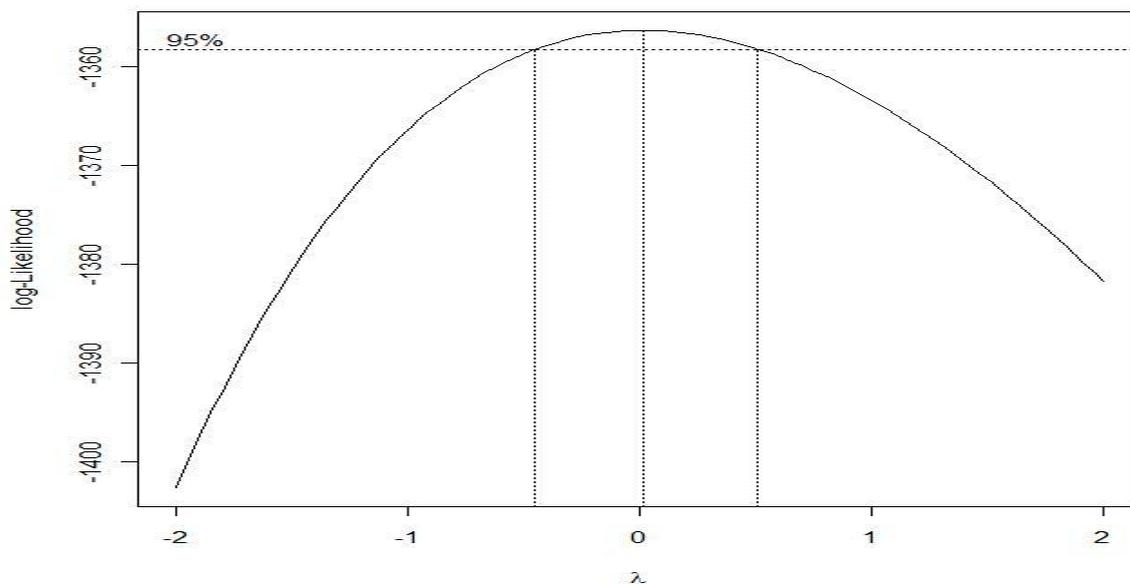


BLIDS-Frequenzverlauf in Stunden aller Sonntage C EIN



An den Wochentagen ist ein starkes quadratisches Absinken der Bluetoothquote im Tagesverlauf zu beobachten. Zudem besteht morgens eine stärkere Quotenfluktuation. Die tatsächliche Verkehrsfrequenz verläuft hingegen konträr. Ein Zusammenhang beider Faktoren kann daher nicht ausgeschlossen werden. Samstags und Sonntags beruhigen sich die Quotenverläufe, nach anfänglichen starken Schwankungen, nur marginal. Aufgrund dessen wird die folgende Regressionsanalyse mit Daten von 9 – 18 Uhr durchgeführt. Zudem erfolgt eine Glättung im größeren 2-Stunden Intervall.

**Log - Transformiertes geglättetes Modell (2 Stunden Intervall) von 9 – 18 Uhr für
Samstage und Sonntage**



```
> regression.c.glatt.log <- lm(log(rfreqc.glatt) ~ bfreqc.glatt:tag.glatt + tag.glatt);
summary(regression.c.glatt.log)
```

Call:

```
lm(formula = log(rfreqc.glatt) ~ bfreqc.glatt:tag.glatt + tag.glatt)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.49754	-0.10636	0.01283	0.12360	0.32868

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.184254	0.114205	45.394	< 2e-16 ***
tag.glatt2	-0.163743	0.212556	-0.770	0.441982
tag.glatt3	-0.464758	0.185633	-2.504	0.013076 *

```

beinc.glatt:tag.glatt1 0.008508 0.002041 4.170 4.51e-05 ***
beinc.glatt:tag.glatt2 0.009740 0.003887 2.506 0.012987 *
beinc.glatt:tag.glatt3 0.013760 0.004116 3.343 0.000986 ***

```

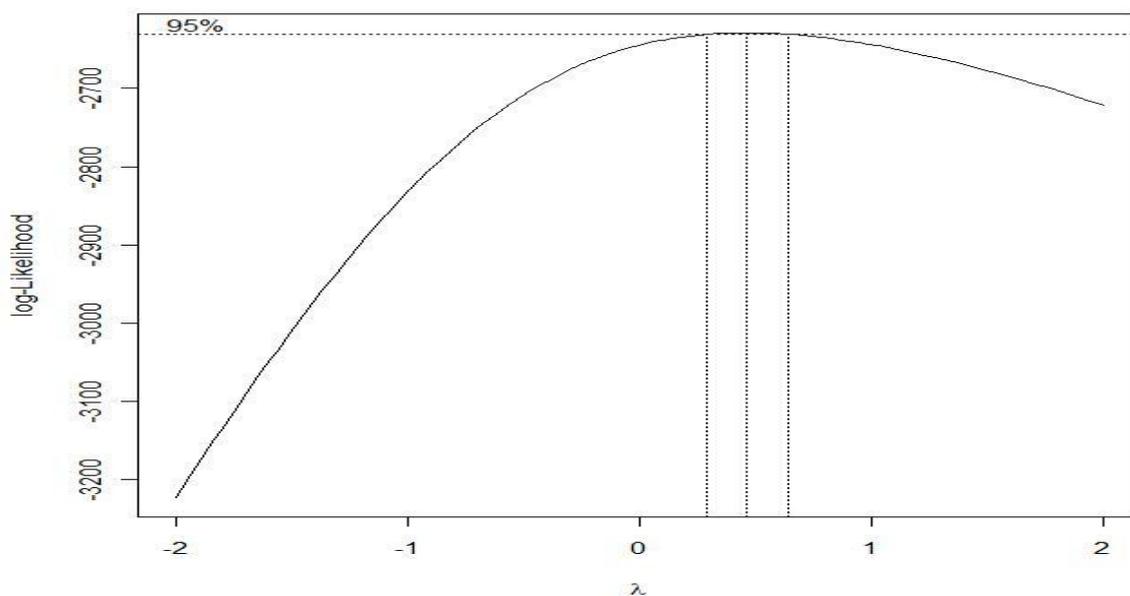
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1714 on 204 degrees of freedom

Multiple R-squared: 0.5262, Adjusted R-squared: 0.5146

F-statistic: 45.31 on 5 and 204 DF, p-value: < 2.2e-16

Wurzel - Transformiertes, geglättetes Modell mit quadratischem Term (2 Stunden Intervall) von 9 – 18 Uhr für Wochentage



```

> lm.reg.sqrt.rein <- lm(sqrt(rfreqc.r) ~ bfreqc.r:tag.r + bfreqc.r^2:tag.r + tag.r-1);
summary(lm.reg.sqrt.rein)

```

Call:

```
lm(formula = sqrt(rfreqc.r) ~ bfreqc.r:tag.r + beinc.r^2:tag.r + tag.r-1)
```

Residuals:

```

    Min     1Q  Median     3Q     Max
-2.00046 -0.45418  0.04434  0.49770  2.65963

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
tag.r1         9.475227   0.917872  10.323 < 2e-16 ***
tag.r2        11.309625   0.742067  15.241 < 2e-16 ***
tag.r3        12.375300   1.093070  11.322 < 2e-16 ***

```

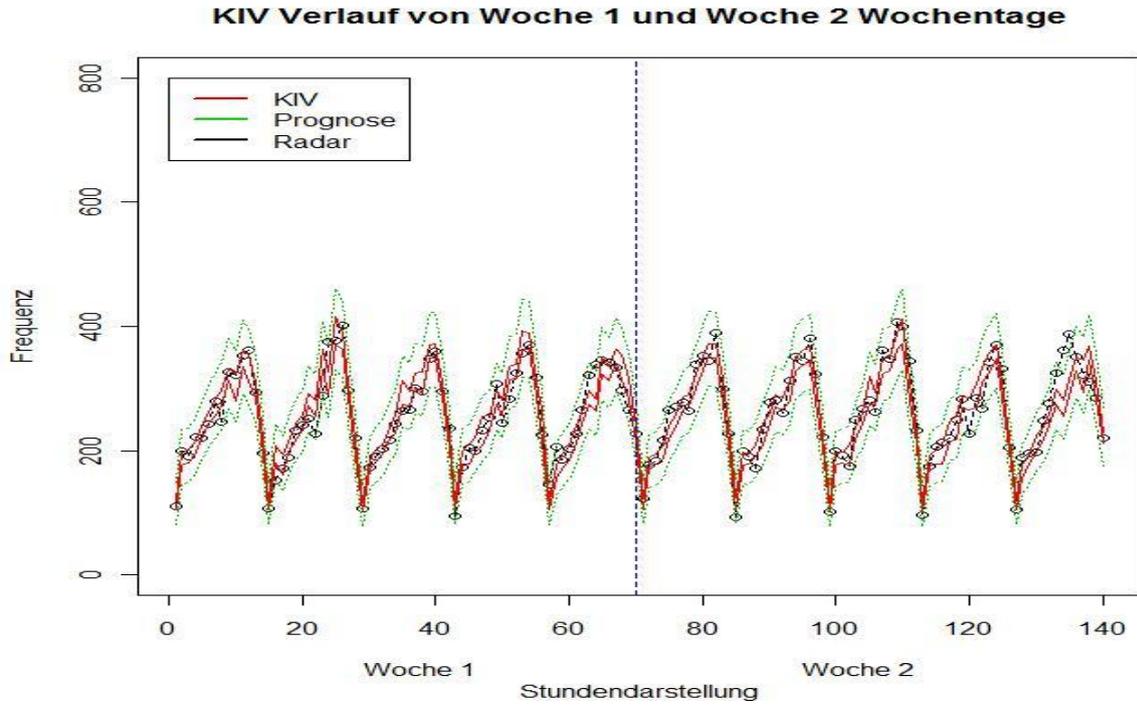
tag.r4	12.796768	1.003135	12.757	< 2e-16	***
tag.r5	14.771551	1.146277	12.887	< 2e-16	***
tag.r6	14.409101	0.941301	15.308	< 2e-16	***
tag.r7	15.388773	0.772602	19.918	< 2e-16	***
tag.r8	14.526164	1.035082	14.034	< 2e-16	***
tag.r9	14.221783	0.974491	14.594	< 2e-16	***
tag.r10	16.049434	0.702352	22.851	< 2e-16	***
tag.r11	15.999300	0.904686	17.685	< 2e-16	***
tag.r12	16.140530	0.898952	17.955	< 2e-16	***
tag.r13	16.984900	0.941630	18.038	< 2e-16	***
tag.r14	13.511057	0.627938	21.517	< 2e-16	***
beinc.r:tag.r1	0.028237	0.025339	1.114	0.265818	
beinc.r:tag.r2	0.034081	0.012187	2.796	0.005424	**
beinc.r:tag.r3	0.022215	0.019246	1.154	0.249095	
beinc.r:tag.r4	0.027198	0.018418	1.477	0.140567	
beinc.r:tag.r5	0.005927	0.020866	0.284	0.776509	
beinc.r:tag.r6	0.026473	0.016985	1.559	0.119903	
beinc.r:tag.r7	0.023224	0.013278	1.749	0.081066	.
beinc.r:tag.r8	0.040011	0.019340	2.069	0.039223	*
beinc.r:tag.r9	0.061005	0.015983	3.817	0.000157	***
beinc.r:tag.r10	0.037235	0.011625	3.203	0.001472	**
beinc.r:tag.r11	0.052940	0.015752	3.361	0.000854	***
beinc.r:tag.r12	0.050814	0.015274	3.327	0.000962	***
beinc.r:tag.r13	0.005466	0.019037	0.287	0.774157	
beinc.r:tag.r14	0.035403	0.017976	1.969	0.049616	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7889 on 388 degrees of freedom

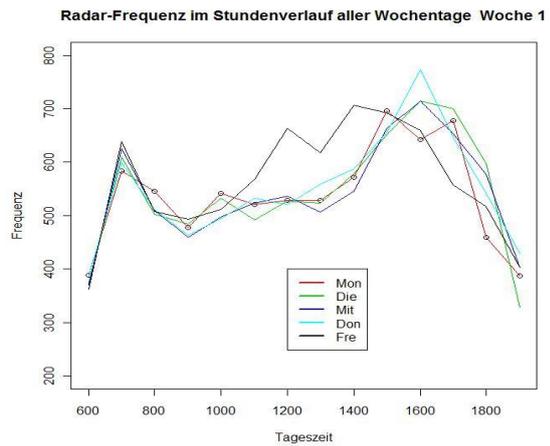
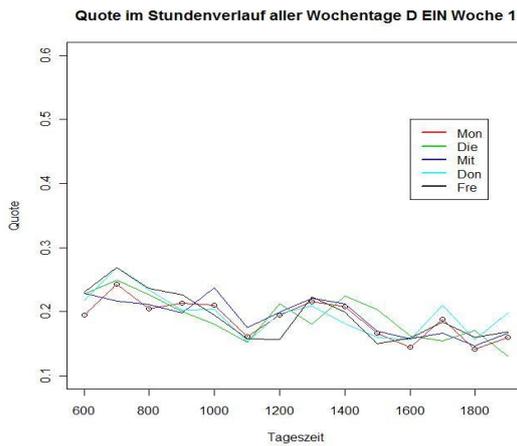
Multiple R-squared: 0.9978, Adjusted R-squared: 0.9976

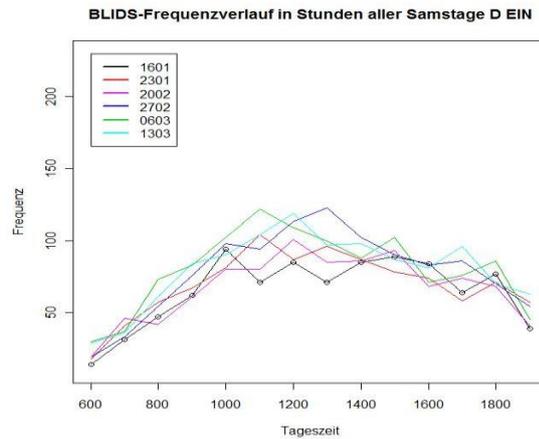
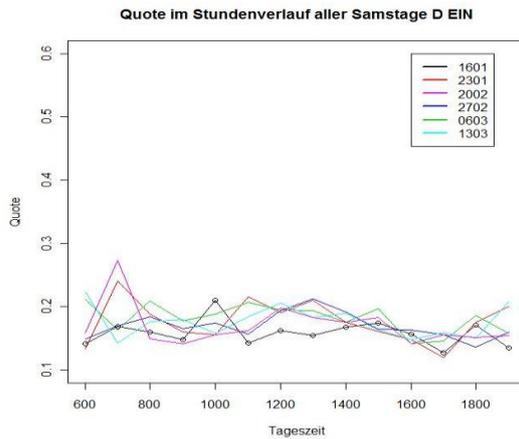
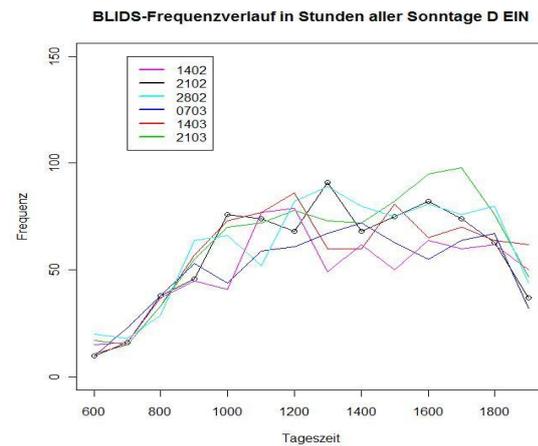
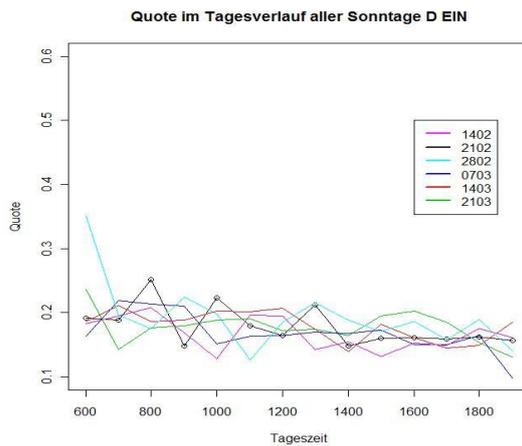
F-statistic: 6156 on 28 and 388 DF, p-value: < 2.2e-16



3.5.4 Lineares Regressionsmodell mit Steigung: Sensor D

Wochentage



Samstage**Sonntage****Regressionsmodell für Samstage und Sonntage**

```
> lm.einfahrt.d <- lm(rfrequd ~ bfrequd:tag + tag); summary(lm.einfahrt.d)
```

Call:

```
lm(formula = rfrequd ~ bfrequd:tag + tag)
```

Residuals:

Min	1Q	Median	3Q	Max
-182.40	-51.96	-9.25	49.60	188.53

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	218.8824	16.2393	13.479	< 2e-16 ***
tag2	-152.0562	28.7958	-5.280	1.82e-07 ***
tag3	-192.8900	26.5988	-7.252	1.32e-12 ***
beind:tag1	3.0913	0.1493	20.704	< 2e-16 ***
beind:tag2	4.9135	0.3062	16.049	< 2e-16 ***
beind:tag3	5.3347	0.3437	15.521	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 71.15 on 582 degrees of freedom

Multiple R-squared: 0.7427, Adjusted R-squared: 0.7405

F-statistic: 335.9 on 5 and 582 DF, p-value: < 2.2e-16

Regressionsmodell mit Stundenfaktorisierung Wochentage

```
> lm.reg <- lm(rfrequd ~ bfrequd:tag.stunden + tag.stunden-1); summary(lm.reg)
```

Call:

```
lm(formula = reind ~ beind:tag.stunden + tag.stunden - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-134.7311	-25.0063	0.6829	24.5213	113.8554

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
tag.stunden1	147.0771	42.0878	3.495	0.000529 ***
tag.stunden2	116.2098	36.7753	3.160	0.001700 **
tag.stunden3	374.6984	63.5760	5.894	8.15e-09 ***
tag.stunden4	300.3589	80.9454	3.711	0.000237 ***
tag.stunden5	354.7058	62.4353	5.681	2.61e-08 ***
tag.stunden6	428.3850	58.1325	7.369	1.03e-12 ***
tag.stunden7	334.0051	60.1454	5.553	5.18e-08 ***
tag.stunden8	288.0767	56.0320	5.141	4.32e-07 ***
tag.stunden9	208.7898	85.1653	2.452	0.014659 *
tag.stunden10	447.8463	61.8553	7.240	2.39e-12 ***
tag.stunden11	234.1335	67.7260	3.457	0.000606 ***
tag.stunden12	371.7056	56.1444	6.621	1.18e-10 ***
tag.stunden13	334.2225	57.4821	5.814	1.26e-08 ***
tag.stunden14	242.3328	50.1942	4.828	1.98e-06 ***
beind:tag.stunden1	2.8088	0.5160	5.443	9.25e-08 ***
beind:tag.stunden2	3.2241	0.2566	12.563	< 2e-16 ***
beind:tag.stunden3	1.1901	0.5388	2.209	0.027759 *
beind:tag.stunden4	1.7541	0.7663	2.289	0.022609 *
beind:tag.stunden5	1.5338	0.5801	2.644	0.008523 **
beind:tag.stunden6	0.9964	0.5465	1.823	0.069044 .
beind:tag.stunden7	2.0783	0.5476	3.795	0.000171 ***
beind:tag.stunden8	2.4364	0.5000	4.873	1.60e-06 ***

```

beind:tag.stunden9  3.4484  0.7342  4.697 3.66e-06 ***
beind:tag.stunden10 1.8259  0.5246  3.481 0.000556 ***
beind:tag.stunden11 3.9838  0.5933  6.715 6.64e-11 ***
beind:tag.stunden12 2.5558  0.4955  5.158 3.98e-07 ***
beind:tag.stunden13 2.3495  0.6158  3.815 0.000158 ***
beind:tag.stunden14 2.3444  0.8060  2.909 0.003835 **

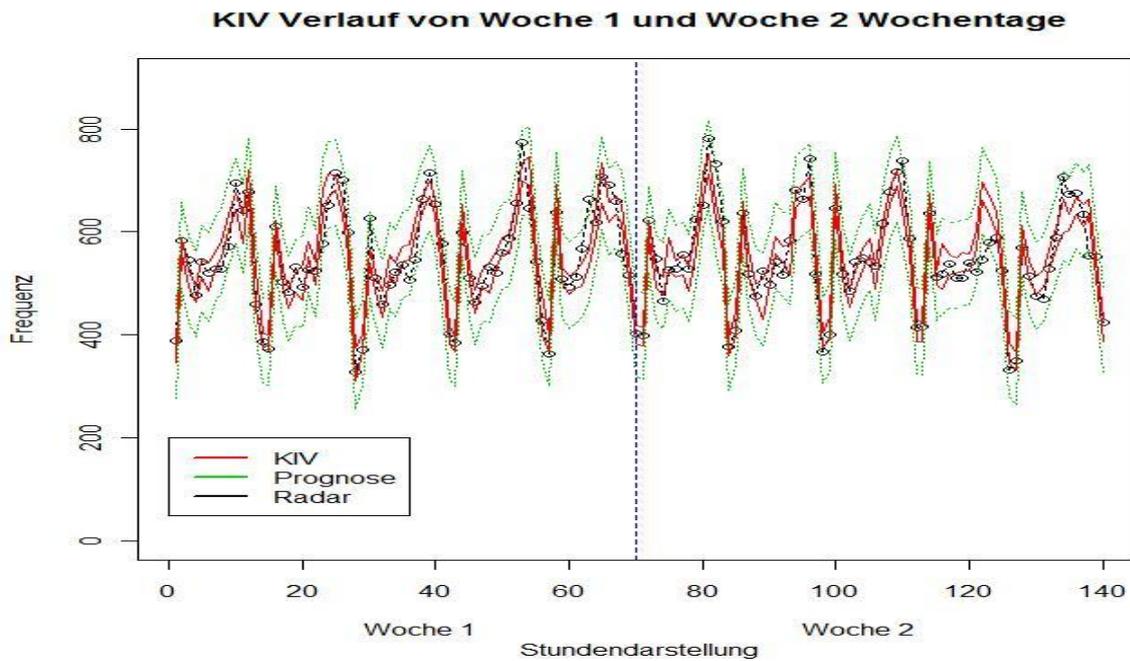
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.45 on 392 degrees of freedom

Multiple R-squared: 0.9948, Adjusted R-squared: 0.9945

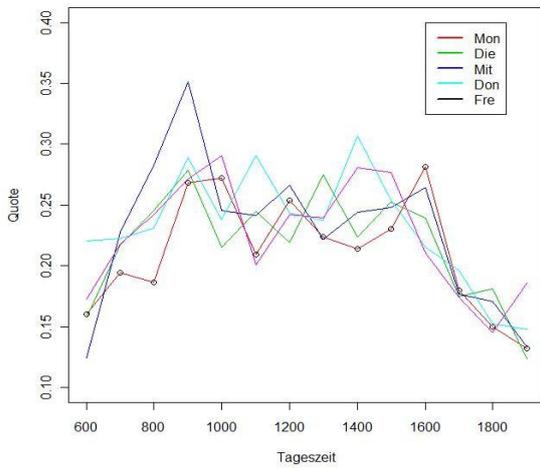
F-statistic: 2698 on 28 and 392 DF, p-value: < 2.2e-16



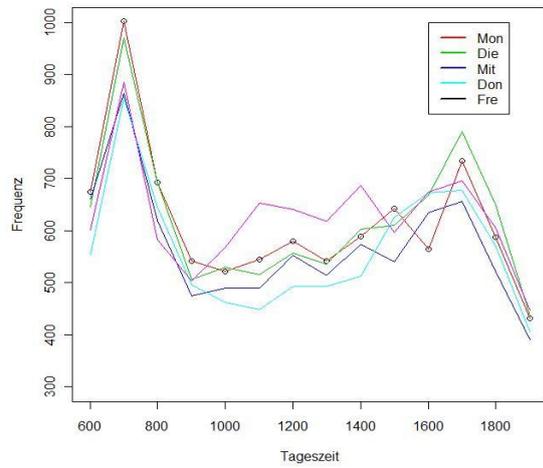
3.5.5 Regressionsmodell mit quadratischem Term: Sensor E

Wochentage

Quote im Stundenverlauf aller Wochentage E EIN Woche 1

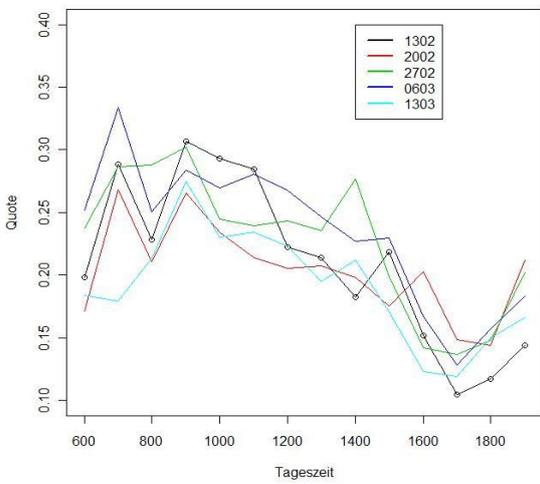


Radar-Frequenz im Stundenverlauf aller Wochentage Woche 1

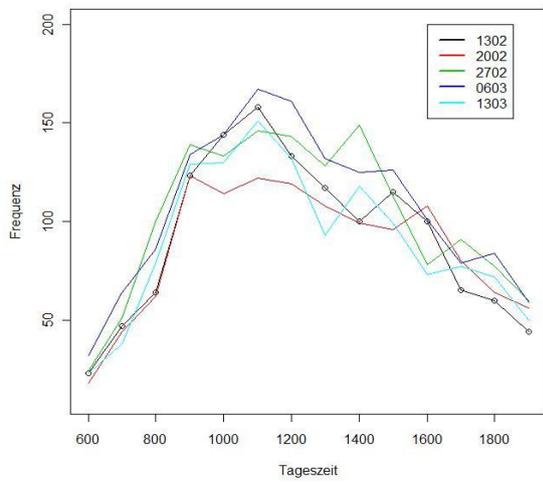


Samstage

Quote im Stundenverlauf aller Samstage E EIN

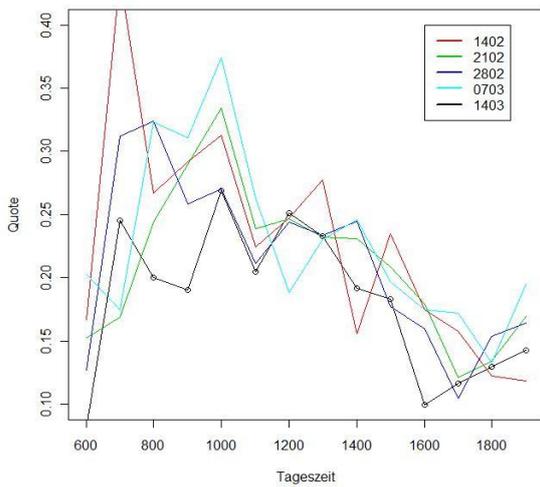


BLIDS-Frequenzverlauf in Stunden aller Samstage E EIN

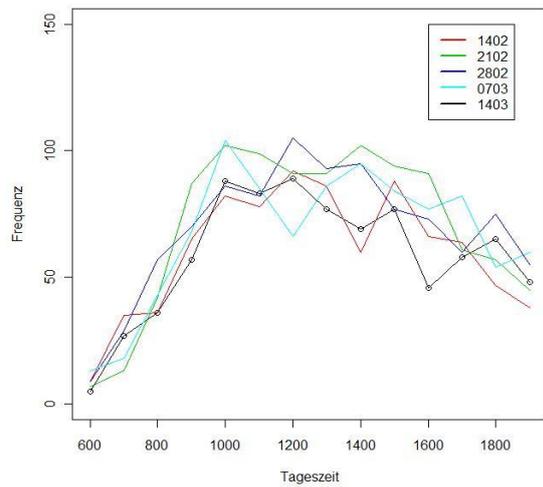


Sonntage

Quote im Stundenverlauf aller Sonntage E EIN



BLIDS-Frequenzverlauf in Stunden aller Sonntage E EIN



Regressionsmodell Samstage und Sonntage

```
> lm.einfahrt.e <- lm(rfreque ~ bfreque:tag + tag); summary(lm.einfahrt.e)
```

Call:

```
lm(formula = rfreque ~ bfreque:tag + tag)
```

Residuals:

Min	1Q	Median	3Q	Max
-219.53	-70.97	-17.18	57.87	270.09

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	343.5839	19.9776	17.198	< 2e-16 ***
tag2	-182.5565	37.3516	-4.888	1.39e-06 ***
tag3	-253.1242	36.8513	-6.869	2.00e-11 ***
beine:tag1	2.0222	0.1429	14.148	< 2e-16 ***
beine:tag2	3.0992	0.3064	10.115	< 2e-16 ***
beine:tag3	3.5408	0.4407	8.035	7.20e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 97.94 on 484 degrees of freedom

Multiple R-squared: 0.6679, Adjusted R-squared: 0.6645

F-statistic: 194.7 on 5 and 484 DF, p-value: < 2.2e-16

Regressionsmodell mit quadratischem Term: Stundenfaktorisierung Wochentage

```
> lm.regression.stunden.rein <- lm(rfreque.r ~ bfreque.r:tag.r + bfreque.r^2:tag.r + tag.r-1);
summary(lm.regression.stunden.rein) #Regressionsmodell Wochentage
```

Call:

```
lm(formula = rfreque.r ~ bfreque.r:tag.r + bfreque.r^2:tag.r + tag.r - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-162.2855	-31.1325	0.5654	32.0135	116.1010

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
--	----------	------------	---------	----------

tag.r1	551.5459	77.1398	7.150	5.89e-12	***
tag.r2	210.3728	64.7273	3.250	0.001276	**
tag.r3	202.8839	74.9464	2.707	0.007152	**
tag.r4	415.9011	77.5381	5.364	1.56e-07	***
tag.r5	450.5209	79.3915	5.675	3.11e-08	***
tag.r6	334.6580	58.3901	5.731	2.30e-08	***
tag.r7	347.1090	72.7981	4.768	2.83e-06	***
tag.r8	320.9115	73.7216	4.353	1.81e-05	***
tag.r9	436.8342	63.3370	6.897	2.83e-11	***
tag.r10	541.9457	62.8698	8.620	3.10e-16	***
tag.r11	483.6501	67.1938	7.198	4.36e-12	***
tag.r12	322.7492	69.1870	4.665	4.54e-06	***
tag.r13	431.8911	59.6057	7.246	3.22e-12	***
tag.r14	347.2186	54.5302	6.367	6.65e-10	***
beine.r:tag.r1	0.7609	0.7321	1.039	0.299394	
beine.r:tag.r2	3.6649	0.3422	10.709	< 2e-16	***
beine.r:tag.r3	2.9215	0.4704	6.210	1.63e-09	***
beine.r:tag.r4	0.8294	0.5331	1.556	0.120780	
beine.r:tag.r5	0.7118	0.5591	1.273	0.203911	
beine.r:tag.r6	1.6987	0.4245	4.001	7.83e-05	***
beine.r:tag.r7	1.7538	0.5214	3.364	0.000862	***
beine.r:tag.r8	1.5850	0.5088	3.115	0.002006	**
beine.r:tag.r9	1.1318	0.4277	2.646	0.008537	**
beine.r:tag.r10	0.4111	0.3927	1.047	0.295961	
beine.r:tag.r11	1.1726	0.4361	2.689	0.007545	**
beine.r:tag.r12	3.1603	0.5235	6.037	4.33e-09	***
beine.r:tag.r13	1.9081	0.6364	2.999	0.002925	**
beine.r:tag.r14	1.6228	0.8995	1.804	0.072162	.

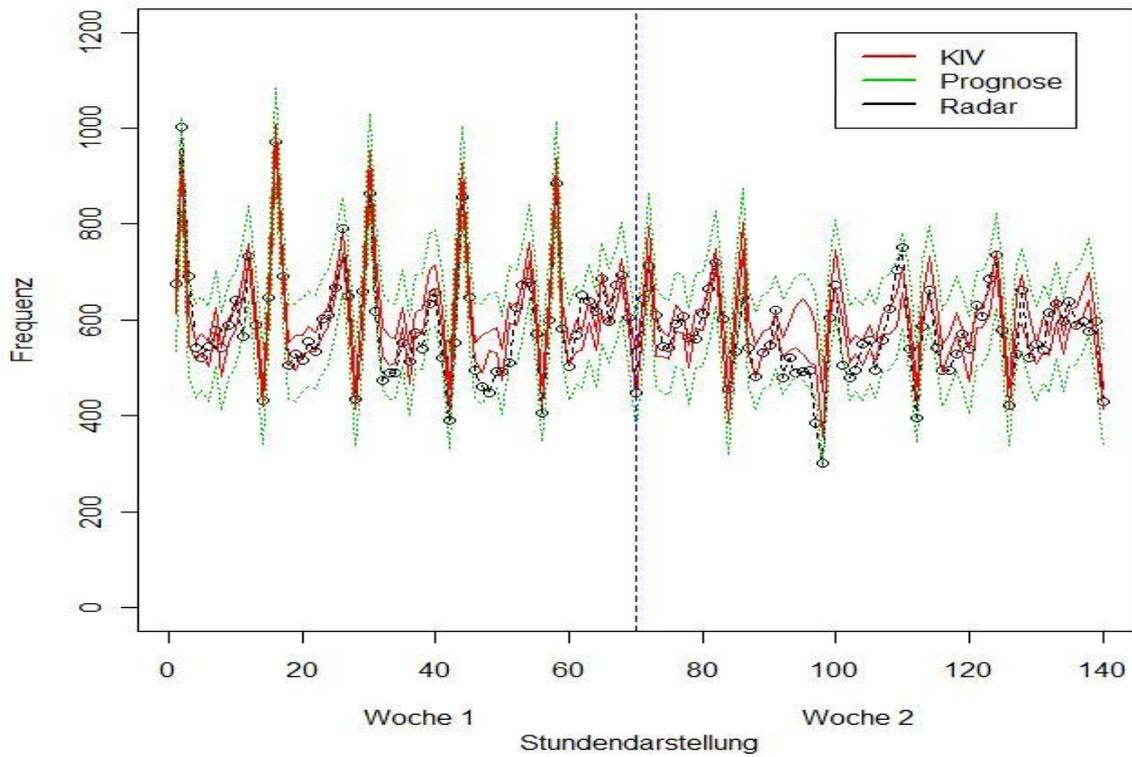
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 48.87 on 321 degrees of freedom

Multiple R-squared: 0.9944, Adjusted R-squared: 0.994

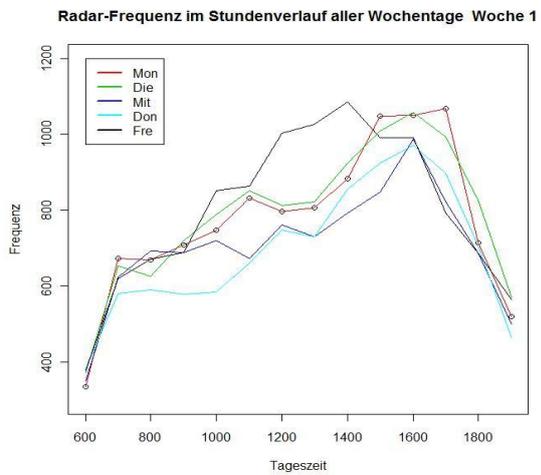
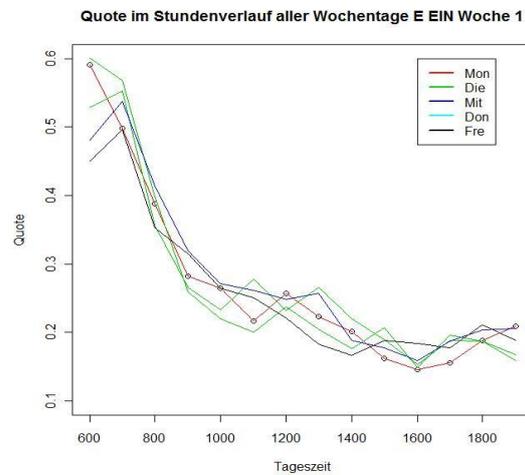
F-statistic: 2053 on 28 and 321 DF, p-value: < 2.2e-16

KIV Verlauf von Woche 1 und Woche 2 Wochentage



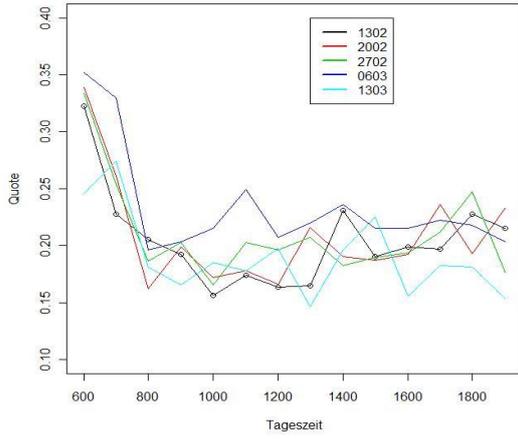
3.5.6 Regressionsmodell mit quadratischem Term: Sensor F

Wochentage

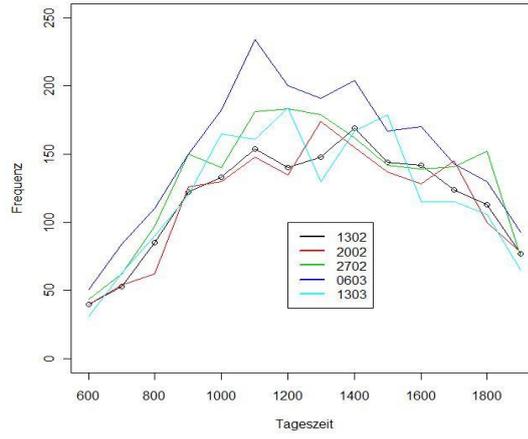


Samstage

Quote im Stundenverlauf aller Samstage F EIN

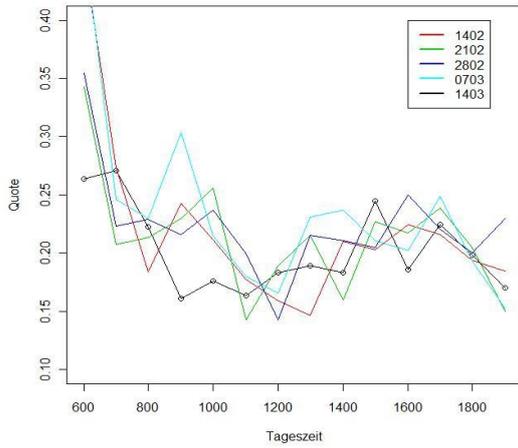


BLIDS-Frequenzverlauf in Stunden aller Samstage F EIN

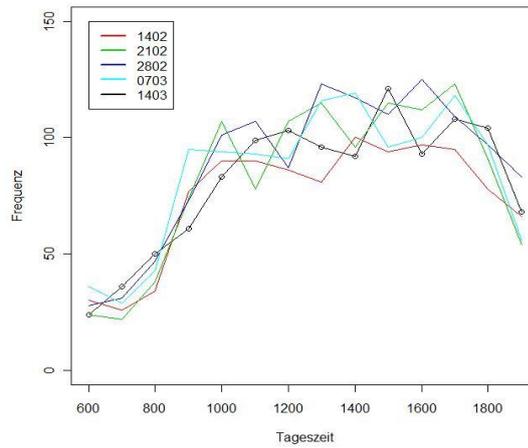


Sonntage

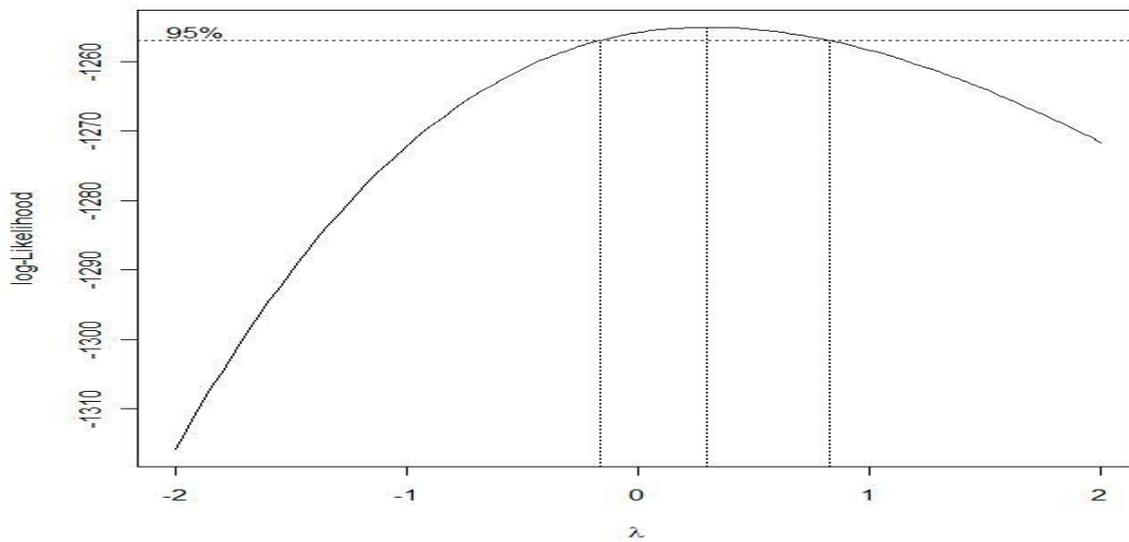
Quote im Stundenverlauf aller Sonntage F EIN



BLIDS-Frequenzverlauf in Stunden aller Sonntage F EIN



Log - Transformiertes, geglättetes Modell von 9 – 18 Uhr für Samstage und Sonntage



```
> regression.glatt.f.trans <- lm(log(rfreqf.glatt) ~ bfreqf.glatt:tag.glatt + tag.glatt);
summary(regression.glatt.f.trans)
```

Call:

```
lm(formula = log(rfreqf.glatt) ~ bfreqf.glatt:tag.glatt + tag.glatt)
```

Residuals:

```
   Min     1Q  Median     3Q    Max
-0.45877 -0.07513 -0.00919  0.10042  0.22857
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.5296464  0.0902258  72.370 < 2e-16 ***
tag.glatt2     -0.6062875  0.1859221  -3.261 0.001343 **
tag.glatt3     -0.9497895  0.2392132  -3.970 0.000106 ***
beinf.glatt:tag.glatt1 0.0013272  0.0005089   2.608 0.009920 **
beinf.glatt:tag.glatt2 0.0047192  0.0010631   4.439 1.63e-05 ***
beinf.glatt:tag.glatt3 0.0060772  0.0022299   2.725 0.007099 **
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1287 on 169 degrees of freedom

Multiple R-squared: 0.7326, Adjusted R-squared: 0.7247

F-statistic: 92.6 on 5 and 169 DF, p-value: < 2.2e-16

Log - Transformiertes, Regressionsmodell mit quadratischem Term für Wochentage

```
> lm.regression.stunden.log <- lm(log(rfreqf.rein) ~ bfreqf.rein:tag.stunden.rein +
bfreqf.rein^2:tag.stunden.rein+tag.stunden.rein-1) > summary(lm.regression.stunden.log)
```

Call:

```
lm(formula = log(rfreqf.rein) ~ bfreqf.rein:tag.stunden.rein +
  bfreqf.rein^2:tag.stunden.rein + tag.stunden.rein - 1)
```

Residuals:

```
   Min     1Q  Median     3Q    Max
-0.2921541 -0.0478961 -0.0009905  0.0509139  0.2148794
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
tag.stunden.rein1    5.5738414  0.1524156  36.570 < 2e-16 ***
tag.stunden.rein2    5.7153106  0.1265504  45.162 < 2e-16 ***
```

tag.stunden.rein3	6.1154000	0.1400439	43.668	< 2e-16	***
tag.stunden.rein4	6.1135080	0.1621014	37.714	< 2e-16	***
tag.stunden.rein5	6.4438568	0.1743068	36.968	< 2e-16	***
tag.stunden.rein6	5.9231521	0.1527844	38.768	< 2e-16	***
tag.stunden.rein7	6.0672017	0.1787165	33.949	< 2e-16	***
tag.stunden.rein8	6.4405946	0.1699926	37.888	< 2e-16	***
tag.stunden.rein9	6.4874945	0.1908021	34.001	< 2e-16	***
tag.stunden.rein10	6.4635965	0.1375310	46.997	< 2e-16	***
tag.stunden.rein11	6.8121291	0.1574916	43.254	< 2e-16	***
tag.stunden.rein12	6.2931806	0.1116487	56.366	< 2e-16	***
tag.stunden.rein13	6.6201804	0.1242996	53.260	< 2e-16	***
tag.stunden.rein14	5.9945992	0.1092363	54.877	< 2e-16	***
beinf.rein:tag.stunden.rein1	0.0016737	0.0007733	2.164	0.031174	*
beinf.rein:tag.stunden.rein2	0.0021472	0.0003782	5.677	3.08e-08	***
beinf.rein:tag.stunden.rein3	0.0017169	0.0006016	2.854	0.004604	**
beinf.rein:tag.stunden.rein4	0.0024852	0.0008349	2.977	0.003138	**
beinf.rein:tag.stunden.rein5	0.0012603	0.0009325	1.352	0.177480	
beinf.rein:tag.stunden.rein6	0.0042043	0.0008096	5.193	3.69e-07	***
beinf.rein:tag.stunden.rein7	0.0036552	0.0009399	3.889	0.000122	***
beinf.rein:tag.stunden.rein8	0.0016405	0.0009145	1.794	0.073766	.
beinf.rein:tag.stunden.rein9	0.0018999	0.0010586	1.795	0.073633	.
beinf.rein:tag.stunden.rein10	0.0023737	0.0007819	3.036	0.002596	**
beinf.rein:tag.stunden.rein11	0.0008260	0.0009392	0.879	0.379794	
beinf.rein:tag.stunden.rein12	0.0034294	0.0006608	5.190	3.75e-07	***
beinf.rein:tag.stunden.rein13	0.0002846	0.0008952	0.318	0.750727	
beinf.rein:tag.stunden.rein14	0.0028893	0.0011345	2.547	0.011343	*

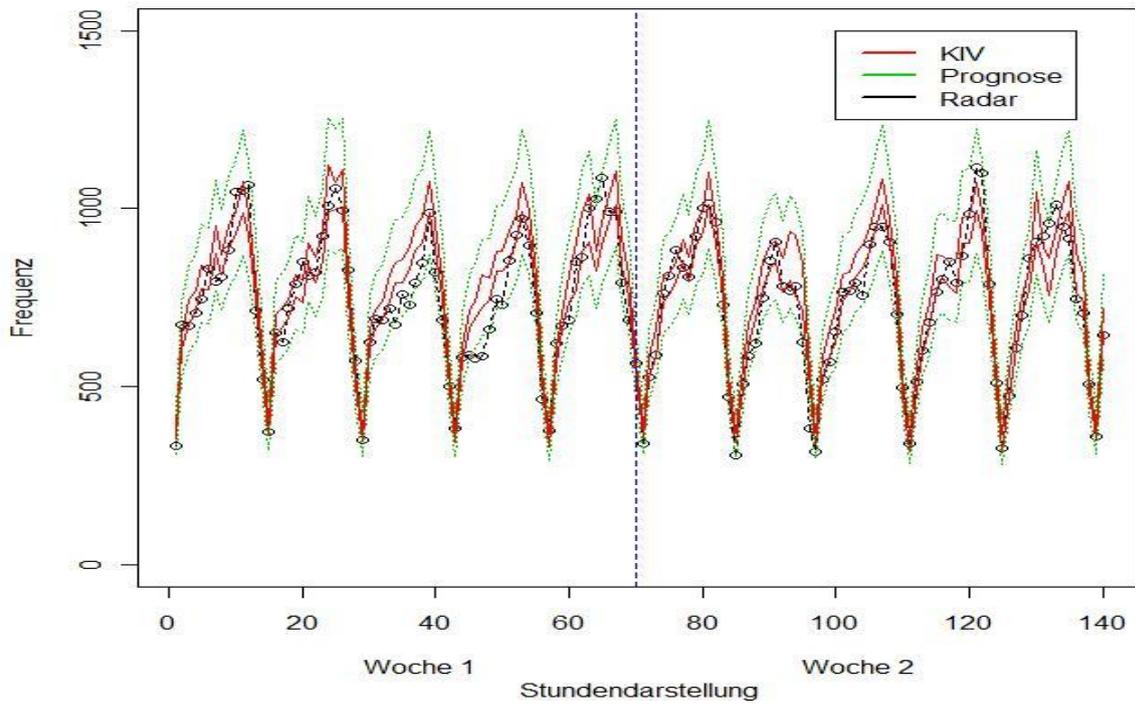
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08327 on 319 degrees of freedom

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9998

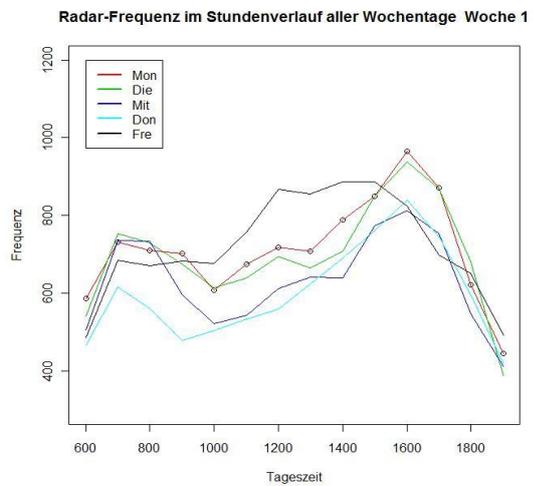
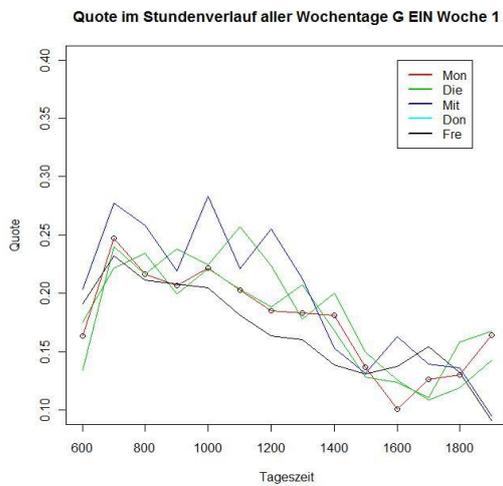
F-statistic: 7.86e+04 on 28 and 319 DF, p-value: < 2.2e-16

KIV Verlauf von Woche 1 und Woche 2 Wochentage



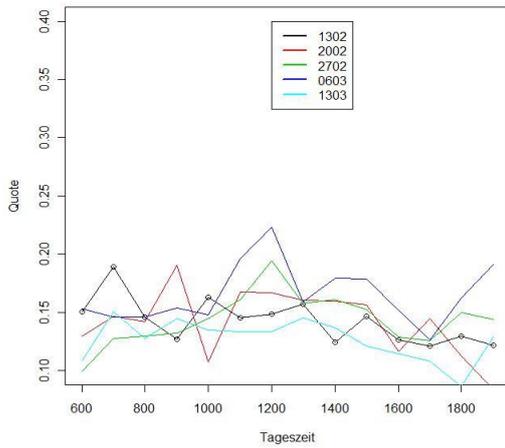
3.5.7 Regressionsmodell mit quadratischem Term: Sensor G

Wochentage

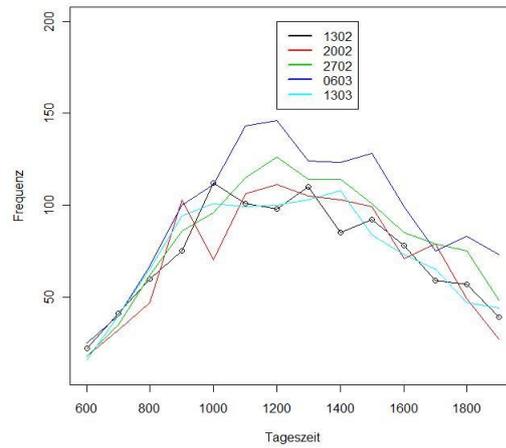


Samstage

Quote im Stundenverlauf aller Samstage G EIN

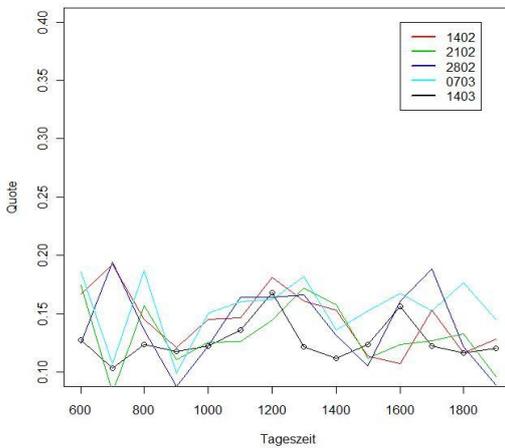


BLIDS-Frequenzverlauf in Stunden aller Samstage G EIN

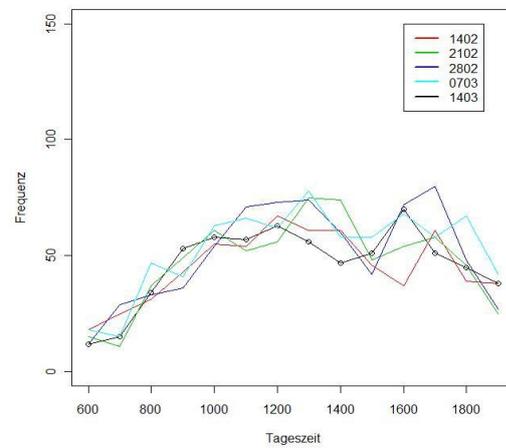


Sonntage

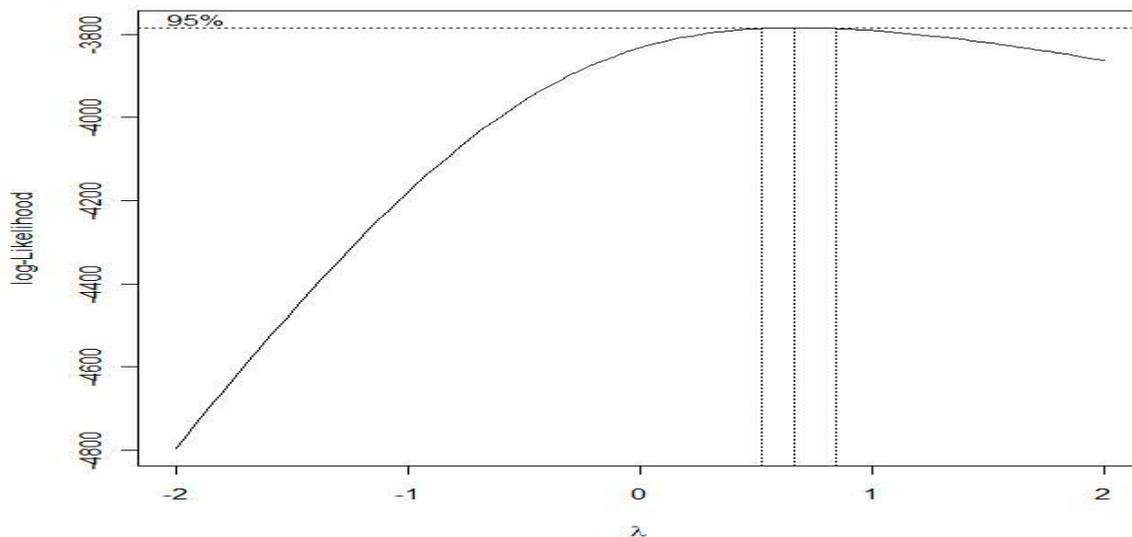
Quote im Stundenverlauf aller Sonntage E EIN



BLIDS-Frequenzverlauf in Stunden aller Sonntage G EIN



Wurzel - Transformiertes Regressionsmodell für Samstage und Sonntage



```
> lm.einfahrt.g <- lm(sqrt(rfrequg)~bfrequg:tag+tag); summary(lm.einfahrt.g)
```

Call:

```
lm(formula = sqrt(rfrequg) ~ bfrequg:tag + tag)
```

Residuals:

```
   Min   1Q   Median   3Q   Max
-6.5769 -1.5294 -0.2394  1.5848  5.8281
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 21.267615  0.426358  49.88 <2e-16 ***
tag2        -8.150363  0.806455 -10.11 <2e-16 ***
tag3       -10.997371  0.851253 -12.92 <2e-16 ***
being:tag1  0.040915  0.003409  12.00 <2e-16 ***
being:tag2  0.123406  0.007971  15.48 <2e-16 ***
being:tag3  0.166760  0.014236  11.71 <2e-16 ***
---
```

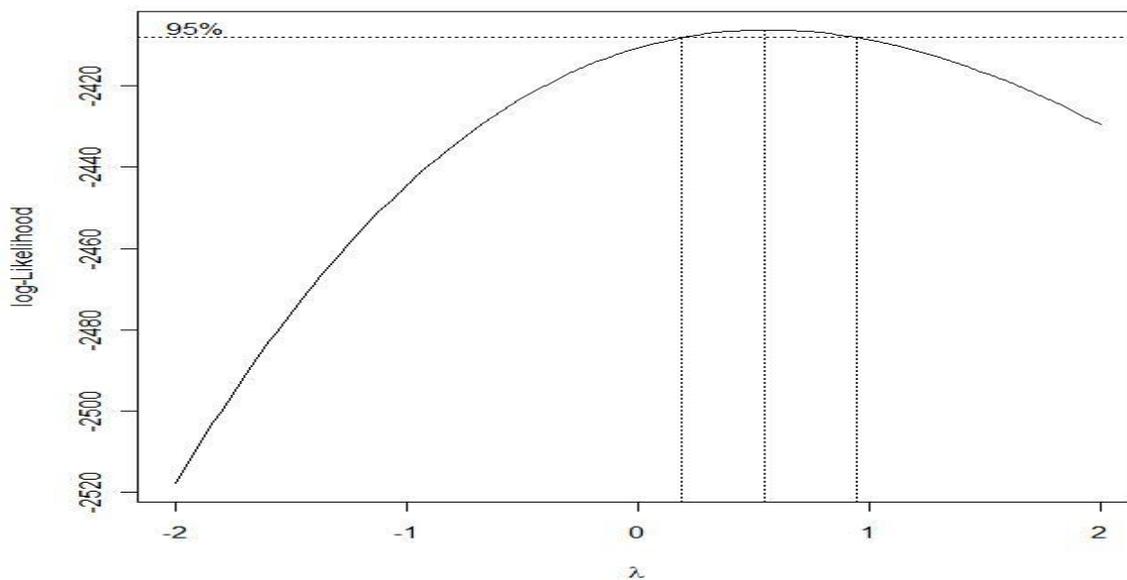
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.134 on 484 degrees of freedom

Multiple R-squared: 0.737, Adjusted R-squared: 0.7343

F-statistic: 271.3 on 5 and 484 DF, p-value: < 2.2e-16

Wurzel - Transformiertes Regressionsmodell mit quadratischem Term für Wochentage mit Stundenfaktorisierung



```
> lm.regression.trans <- lm(sqrt(rfrequg.r) ~ bfrequg.r:tag.r + bfrequg.r^2:tag.r + tag.r-1)
> summary(lm.regression.trans)
```

Call:

```
lm(formula = sqrt(rfrequg.r) ~ bfrequg.r:tag.r + bfrequg.r^2:tag.r + tag.r - 1)
```

Residuals:

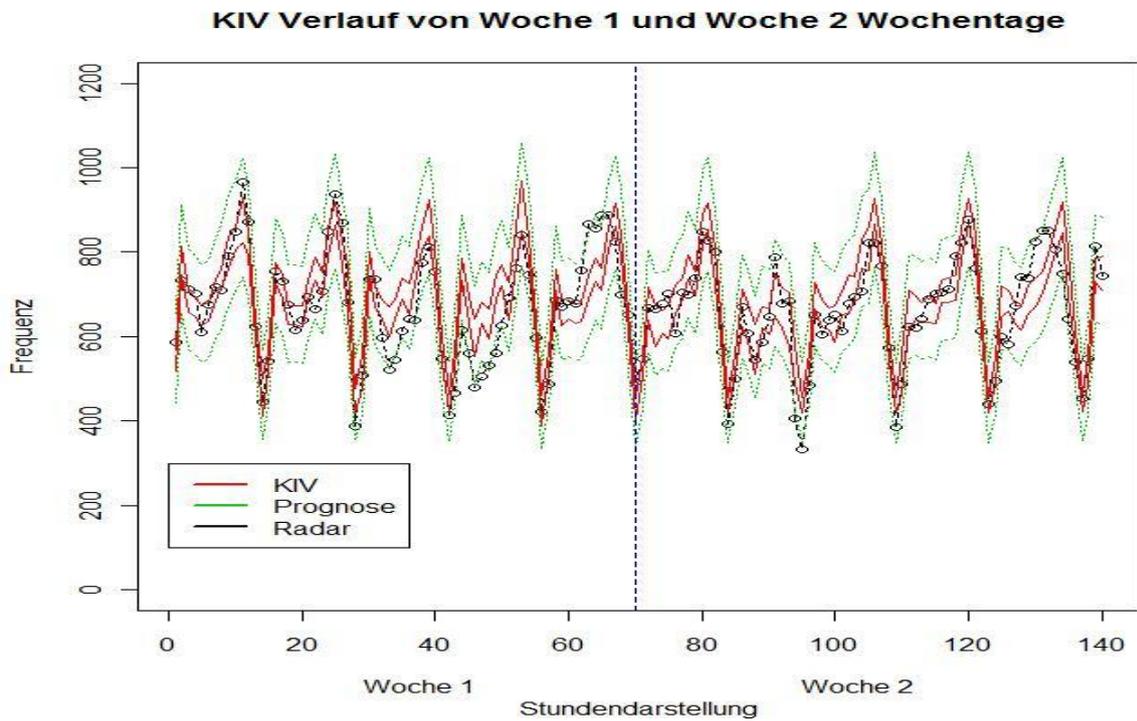
```
    Min     1Q  Median     3Q     Max
-3.139628 -0.588879 -0.003668  0.658347  3.611915
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
tag.r1         21.21614   1.66234  12.763 < 2e-16 ***
tag.r2         21.11848   1.66106  12.714 < 2e-16 ***
tag.r3         19.98455   2.01439   9.921 < 2e-16 ***
tag.r4         20.64167   1.62964  12.666 < 2e-16 ***
tag.r5         23.38933   1.75324  13.341 < 2e-16 ***
tag.r6         21.44855   1.87044  11.467 < 2e-16 ***
tag.r7         25.23533   2.38268  10.591 < 2e-16 ***
tag.r8         19.80987   2.27518   8.707 < 2e-16 ***
tag.r9         20.02426   2.05683   9.736 < 2e-16 ***
tag.r10        26.42191   2.03415  12.989 < 2e-16 ***
tag.r11        28.24973   2.42155  11.666 < 2e-16 ***
tag.r12        25.50291   1.21318  21.022 < 2e-16 ***
tag.r13        19.46531   1.35301  14.387 < 2e-16 ***
tag.r14        19.72920   1.45558  13.554 < 2e-16 ***
being.r:tag.r1  0.02182   0.01861   1.172 0.242108
being.r:tag.r2  0.03758   0.01022   3.677 0.000277 ***
being.r:tag.r3  0.03988   0.01243   3.207 0.001476 **
being.r:tag.r4  0.03604   0.01153   3.125 0.001940 **
being.r:tag.r5  0.01554   0.01254   1.240 0.216056
being.r:tag.r6  0.03079   0.01343   2.292 0.022530 *
being.r:tag.r7  0.01072   0.01770   0.606 0.545223
being.r:tag.r8  0.05629   0.01736   3.242 0.001313 **
being.r:tag.r9  0.05757   0.01591   3.619 0.000344 ***
being.r:tag.r10 0.02092   0.01694   1.235 0.217685
being.r:tag.r11 0.01379   0.02077   0.664 0.507214
being.r:tag.r12 0.02841   0.01151   2.468 0.014115 *
being.r:tag.r13 0.06693   0.01674   3.997 7.97e-05 ***
being.r:tag.r14 0.02187   0.02622   0.834 0.404823
```

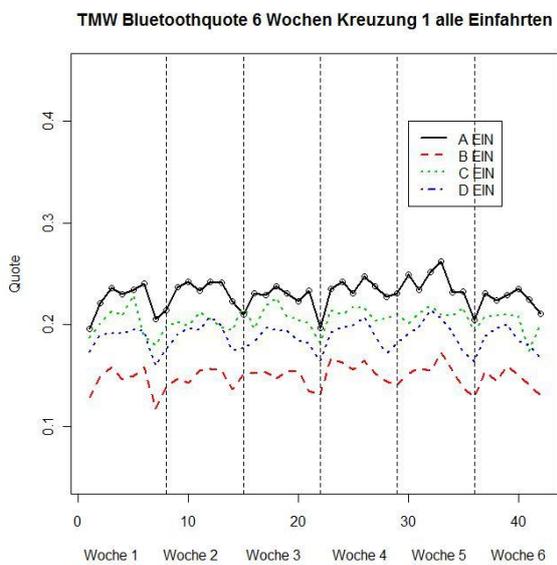
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.125 on 318 degrees of freedom
 Multiple R-squared: 0.9983, Adjusted R-squared: 0.9982
 F-statistic: 6784 on 28 and 318 DF, p-value: < 2.2e-16

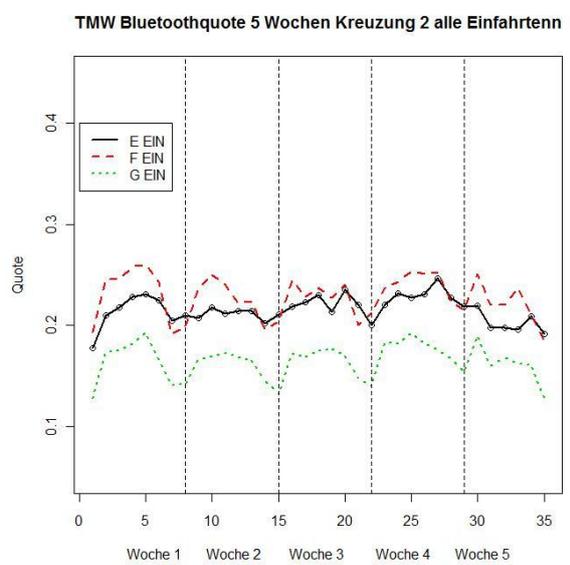


3.5.8 Frequenz-, Bluetooth- und Quotenverläufe aller Messstationen

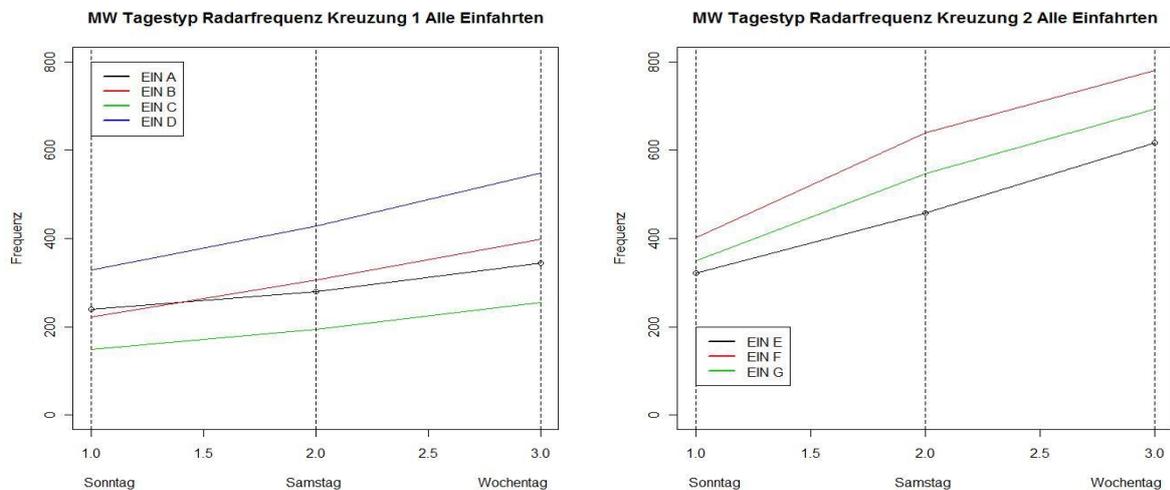
Kreuzung 1



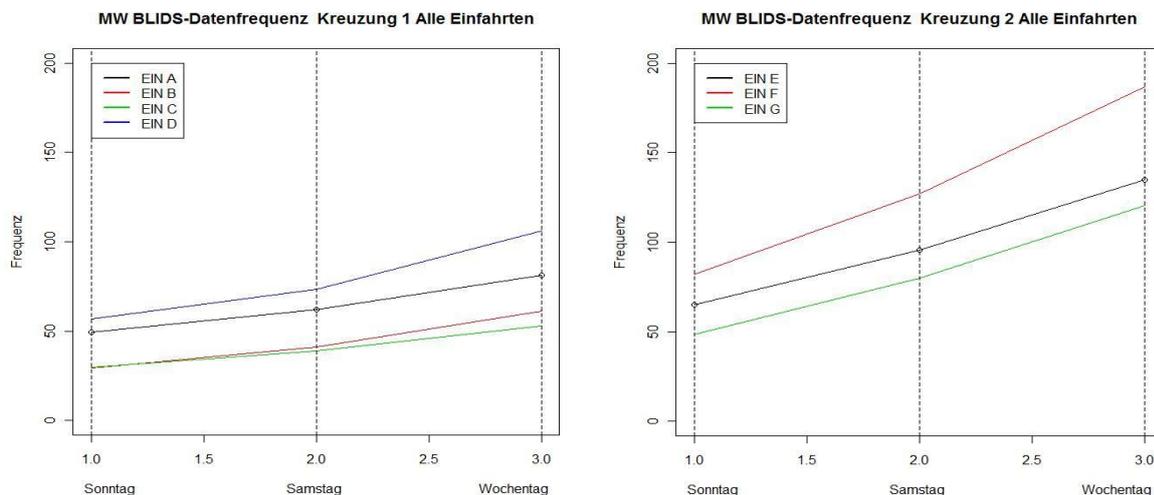
Kreuzung 2



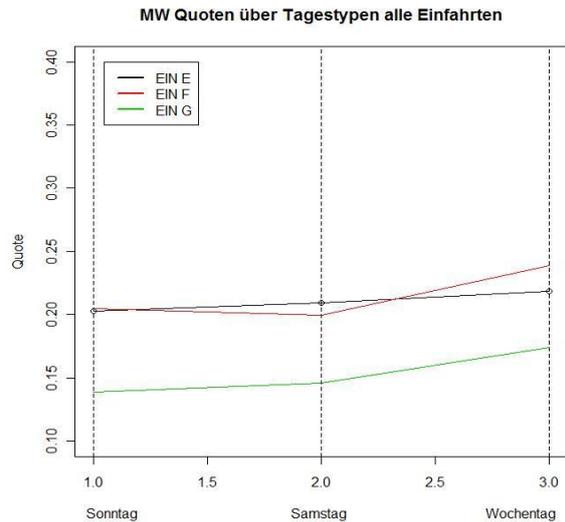
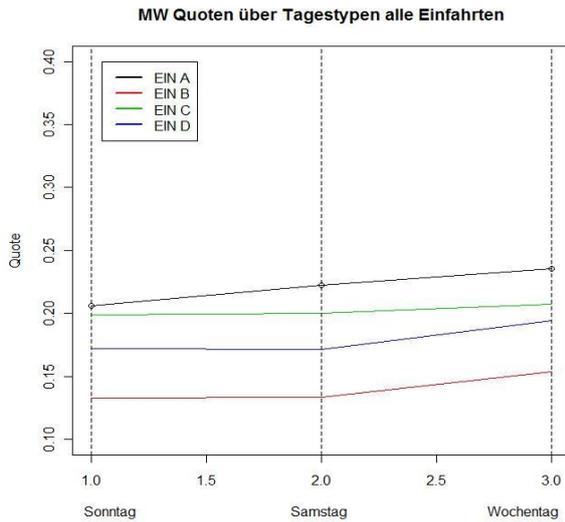
Anhand dieser Plots kann man sehr gut die unterschiedlichen Niveaus der Bluetoothquoten jeder einzelnen Messstation erkennen. Sensor C zeigt, im Gegensatz zu alle anderen Standorten, einen etwas gleichmäßigeren Verlauf. Grundsätzlich werden an den Wochenenden etwas geringere Quotenhöhen registriert, als an den Wochentagen. Die Systematik der Quotenverläufe aller Sensoren, ist in etwa gleich. Ausreißer sind nicht zu erkennen.



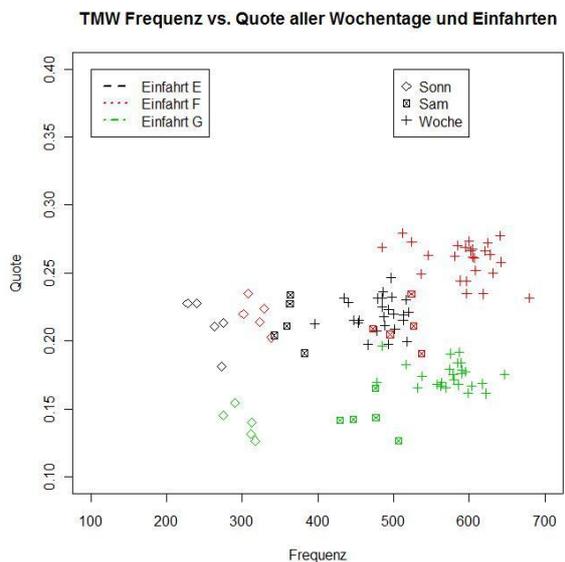
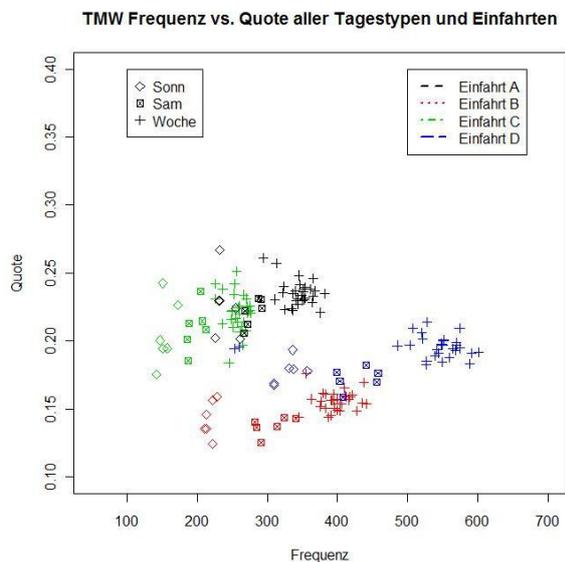
An allen 7 Messstationen werden unterschiedliche Radarfrequenzen registriert, wobei die Spitzenwerte an Wochentagen erzielt werden. Grundsätzlich ist festzustellen, dass an der Kreuzung 2 eine deutlich höhere Verkehrsdichte gegeben ist.



Im Großen und Ganzen, ist bei den Radar- und BLIDS-Frequenzen, dieselbe Verkehrscharakteristik zu erkennen, wonach an den Wochentagen ein bedeutend höheres Verkehrsaufkommen zu registrieren ist. Auffallend ist jedoch, dass die Reihenfolge der Frequenzen, zwischen Radar- und BLIDS-Sensoren, nicht völlig ident ist. Besonders die Messstationen C und D sind in dieser Hinsicht herauszuheben. Im Verhältnis zu den tatsächlichen Verkehrsfrequenzen, werden an diesen beiden BLIDS-Geräten weniger Bluetoothsignale registriert.



Besonders die Messstationen C und E zeigen, unabhängig vom jeweiligen Tagestyp, konstante Bluetoothquoten. Alle anderen Sensoren registrieren vor allem an den Wochentagen mehr Bluetoothsignale als an den anderen Tagestypen. Hervorzuheben ist in dieser Hinsicht Station F, wo eine deutlicher Niveausprung erkennbar ist.



Markante Quotenausreißer sind in Abhängigkeit zur Frequenz nicht zu erkennen. Ein Zusammenhang zwischen der Quotenhöhe und der tatsächlichen Verkehrsdichte kann nicht festgestellt werden. Im Großen und Ganzen ist im zweiten Kreuzungsbereich eine etwas größere horizontale Punktstreuung festzustellen.

3.5.9 Zusammenfassung**95 % - Konfidenzintervalle¹² nach Tagestyp und Messstation in Prozent**

Tagestyp	Wochentag	Samstage	Sonntage
Einfahrt			
A	23 – 24	21 – 23	21 – 24
B	18 – 19	13 – 14	13 – 15
C	21 – 23	20 – 22	19 – 22
D	19 – 20	16 – 18	17 – 18
E	21 – 22	20 – 23	20 – 23
F	25 – 27	20 – 22	20 – 24
G	17 – 18	13 – 15	13 – 15

Kennzahlen

Messpunkt	Tagestyp	Bestimmtheitsmaß	Faktorisierung	Box–Cox-Transformation	Stichprobenumfang BLIDS-Signale
A	Alle	0.816	Tage		43522
B	Samstage, Sonntage	0.686	Tage		31636
	Wochentage	0.998 ¹³	Stunden	$\lambda = \frac{1}{2}$ Wurzel	
C	Samstag, Sonntage	0.526	Tage	$\lambda = 0$ Log	27931
	Wochentage	0.998 ¹³	Stunden	$\lambda = \frac{1}{2}$ Wurzel	
D	Samstage, Sonntage	0.743	Tage		55577
	Wochentage	0.995	Stunden		

$$^{12} KIV_{12} = \bar{x} \pm \frac{s}{\sqrt{n}} t_{n-1,0.975}$$
¹³ Modell ohne redundante Parameter

E	Samstage, Sonntage	0.668	Tage		58460
	Wochentage	0.994 ¹⁴	Stunden		
F	Samstage, Sonntage	0.733	Tage	$\lambda = 0$ Log	79973
	Wochentage	0.99 ¹⁴	Stunden	$\lambda = 0$ Log	
G	Samstage, Sonntage	0.737	Tage	$\lambda = \frac{1}{2}$ Wurzel	51162
	Wochentage	0.998 ⁶	Stunden		

Zusammenfassend ist zu sagen, dass einzig in Messpunkt A, an allen 3 Tagestypen, keinerlei Transformation oder Glättung, der zugrundeliegenden Daten durchgeführt wurde. Mit rund 82 Prozentpunkten, weist dieses errechnete Modell, eine für alle Tage, gültige Schätzgenauigkeit, auf. An allen anderen Messstationen bedarf es hingegen einiger zusätzlicher Maßnahmen, um aussagekräftige Analysen, zu berechnen. Besonders die diversen Quoteneinheiten, an Wochentagen, erforderten kürzere Faktorisierungen, um deren Verläufe abzubilden. Bedingt durch diese Erfordernisse ist ein direkter Vergleich der Modellgüten nur anhand der Bestimmtheitsmaße recht oberflächlich.

Allein durch eine Stundenfaktorisierung wird das Bestimmtheitsmaß derart erhöht, dass ein direkter Vergleich zu einem 'herkömmlichen' Modell ausscheidet.

Messpunkt F weist, im Rahmen einer Faktorisierung nach Stunden, mit 99.9 Prozentpunkten an Wochentagen, das höchste Bestimmtheitsmaß auf. An Sensor C ist an Wochentagen der Bestimmtheitsgrad nur 53 Prozent.

Durch Berücksichtigung der Quotencharakteristiken und gegebenenfalls Datentransformationen, kann die Anpassungsgenauigkeit der Schätzungen, an die Realdaten markant verbessert werden. Für alle Messstationen können durchaus gute, wenn nicht sogar sehr gute Regressionsmodelle errechnet werden. Qualitative Ausreißer im negativen Sinn sind keine vorhanden. Alle von den verschiedenen Sensoren gelieferten Daten eignen sich zur Modellierung.

3.5.10 Beispiel Routenauswertung

Mittels der Regressionsmodelle werden für jeden Messpunkt, Schätzungen der Querschnittsfrequenz, errechnet. Für eine Routenauswertung ist dieser Wert prozentuell auf die einzelnen Routen aufzuteilen.

¹⁴ Modell ohne redundante Parameter

Berechnung:

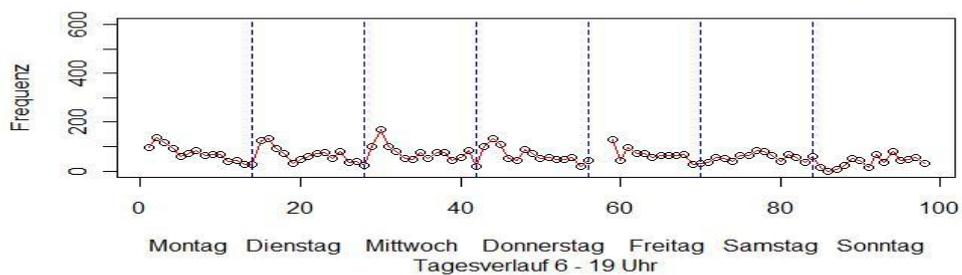
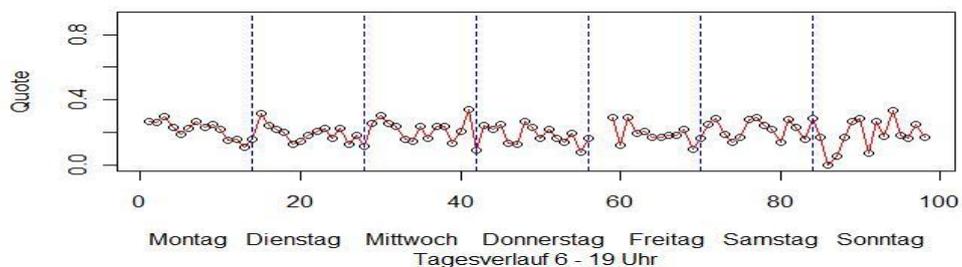
z. B. Messpunkt A

- geschätzte Querschnittsfrequenz 600 Fahrzeuge pro Stunde
- Route AB: 25 Pärchen¹⁵
Route AC: 40 Pärchen
Route AD: 60 Pärchen

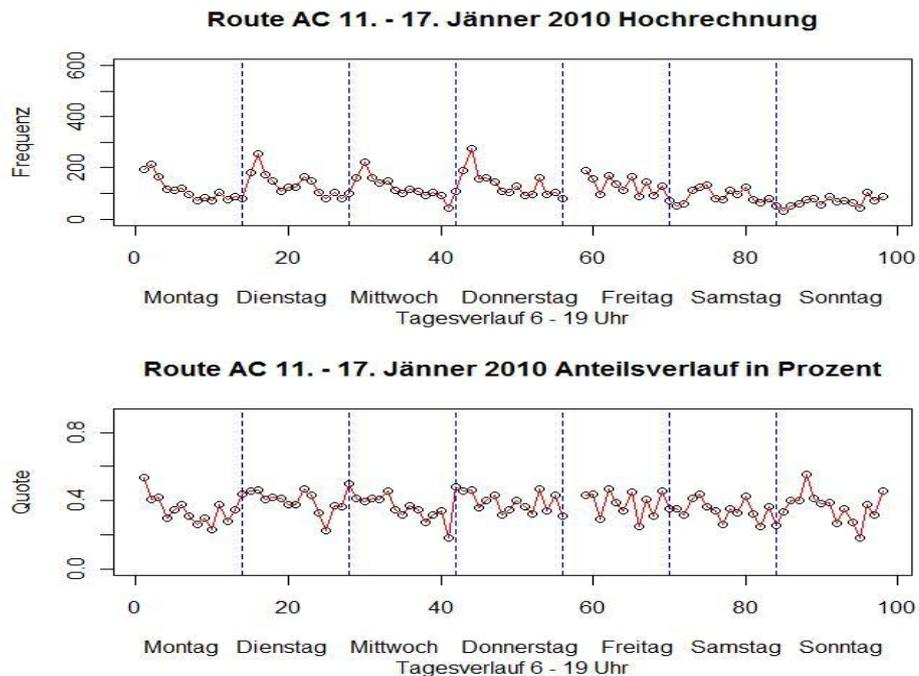
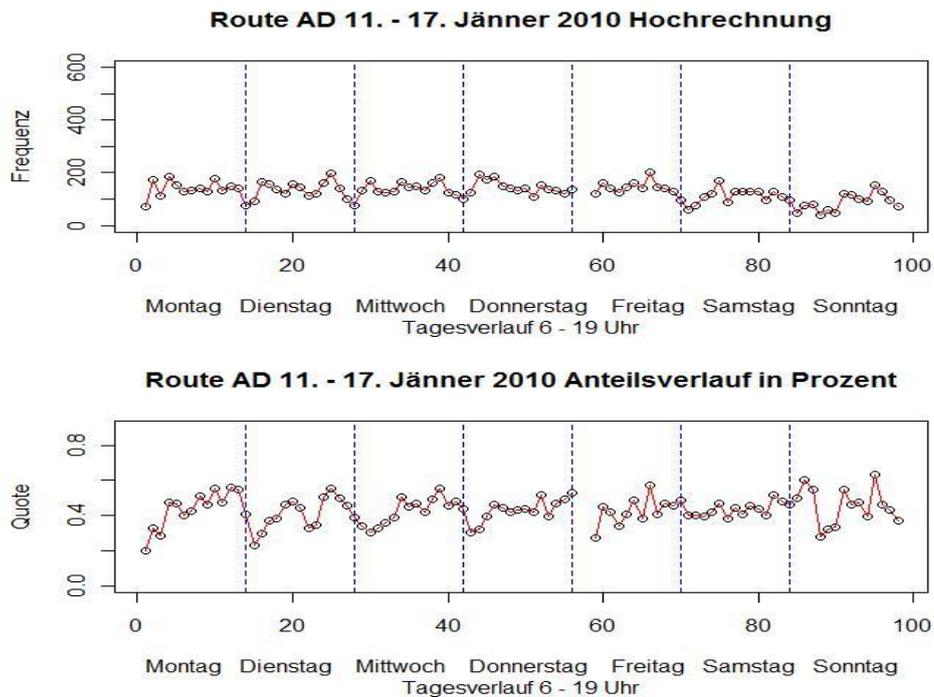
prozentuelle Aufteilung:

- Route AB: $\frac{25}{25 + 40 + 60} * 100 = 20 \%$
- Route AC: 32 %
- Route AD: 48 %

D. h. 120 (20 % von 600) Fahrzeuge benutzen die Route AB, 192 die Route AC und 288 die Route AD.

Routenauswertungen Messpunkt A**Route AB****Route AB 11. - 17. Jänner 2010 Hochrechnung****Route AB 11. - 17. Jänner 2010 Anteilsverlauf in Prozent**

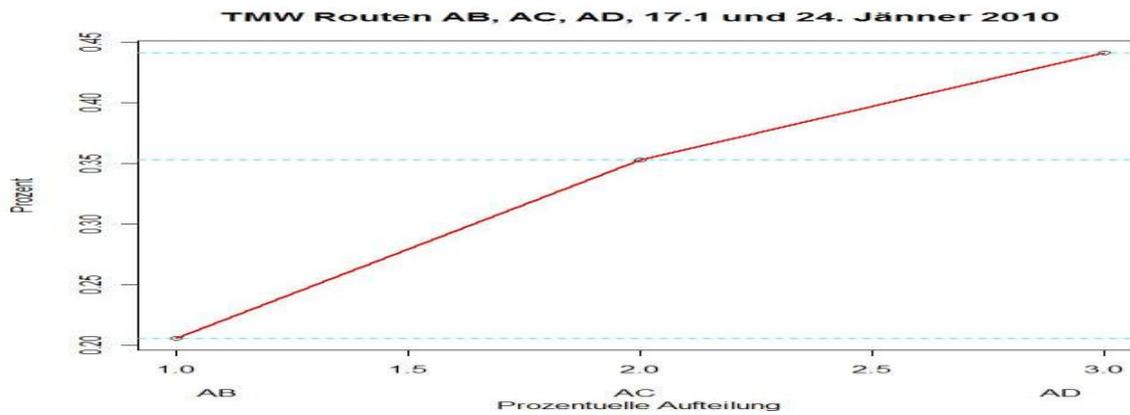
¹⁵ Erfassung derselben ID-Nummer an 2 verschiedenen Messpunkten

Route AC**Route AD**

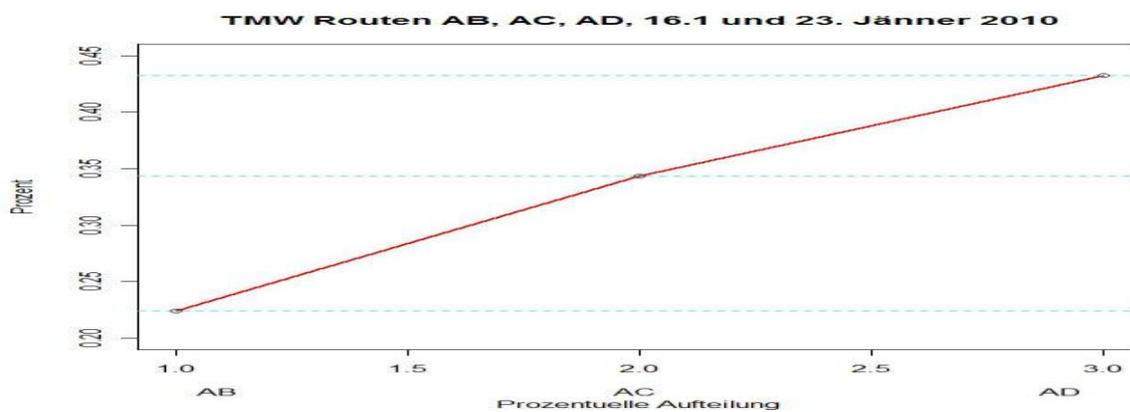
Anhand der Frequenzverläufe, sind Verkehrsspitzen bzw. -flauten, zu erkennen. Sowohl auf der Route AB, als auch auf AC, wird um 7.00 Uhr Früh die größte Verkehrsdichte geschätzt. Nachmittags hingegen, ist ein solcher Effekt nicht vorhanden. Die größten Frequenzschwankungen existieren auf der Route AC, während auf AD, Schwankungen deutlich moderater ausfallen. Anteilsmäßig ergibt die prozentuelle Auslastung, dass auf den Routen AB und AC, nach einer morgendlichen Spitze, ein leichtes Abflachen des Verkehrs

einsetzt. Route AD hingegen, zeigt ein komplett konträres Bild, wonach Auslastungsspitzen erst nachmittags einsetzen. Daraus kann man schließen, dass die Routen AB und AC eher für die Fahrt zur Arbeit genutzt werden, während die Richtung AD keine zeitliche Nutzungsprioritäten zu haben scheint.

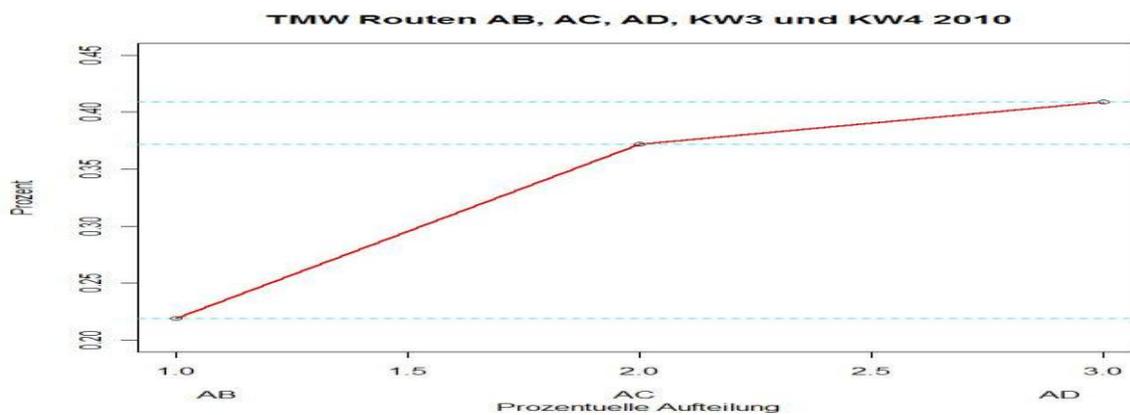
Prozentuelle Aufteilung der Routen AB, AC, AD Sonntags



Prozentuelle Aufteilung der Routen AB, AC, AD Samstags



Prozentuelle Aufteilung der Routen AB, AC, AD Werktags

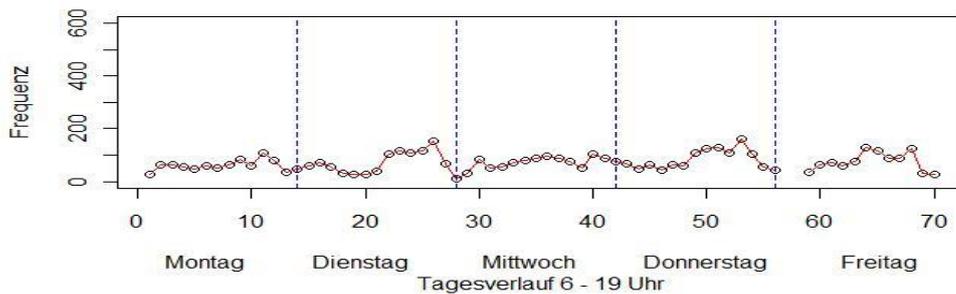


Für den Messpunkt A ergibt die Verkehrsstromanalyse, dass die Route AD prozentuell am häufigsten genutzt wird. AB hingegen, wird am wenigsten frequentiert. Eine Richtungstendenz ist eindeutig feststellbar. Grundsätzlich ist an allen 3 Tagestypen derselbe Trend, hinsichtlich der prozentuellen Routenauslastung, gegeben. Route AB wird am wenigsten frequentiert, AD hingegen, zeigt den bei weiten höchsten Nutzungsgrad.

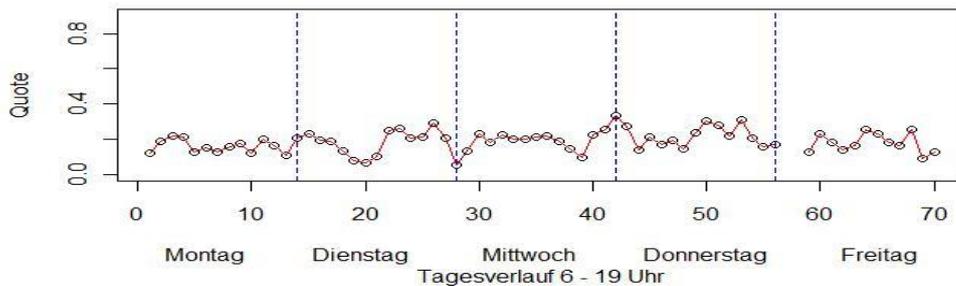
Routenauswertungen Messpunkt B

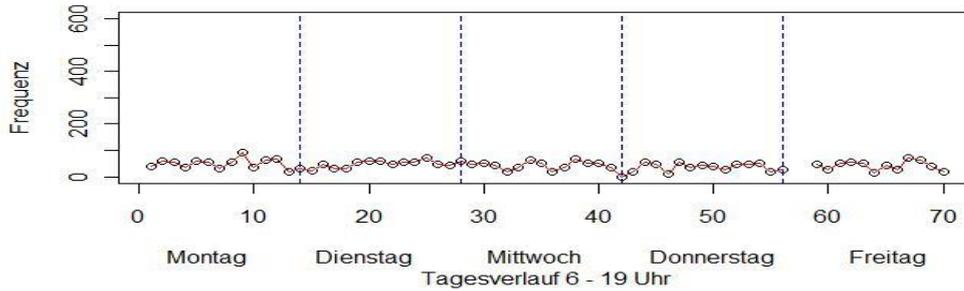
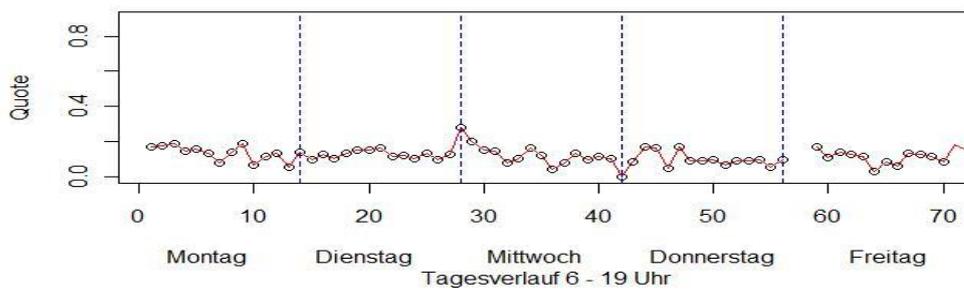
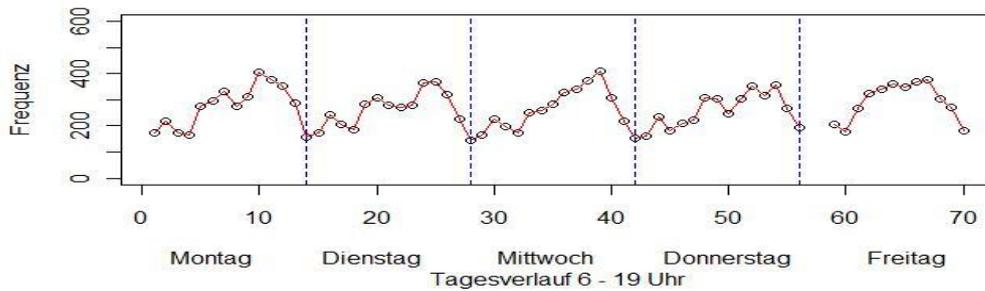
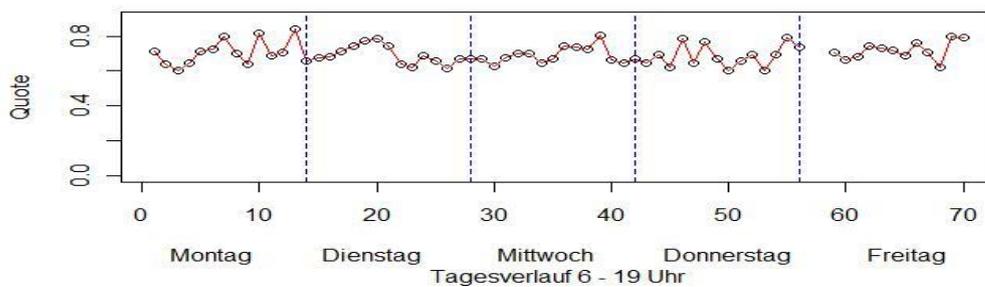
Route BA

Route BA 11. - 15. Jänner 2010 Frequenzverlauf



Route BA 11. - 15. Jänner 2010 Prozentverlauf

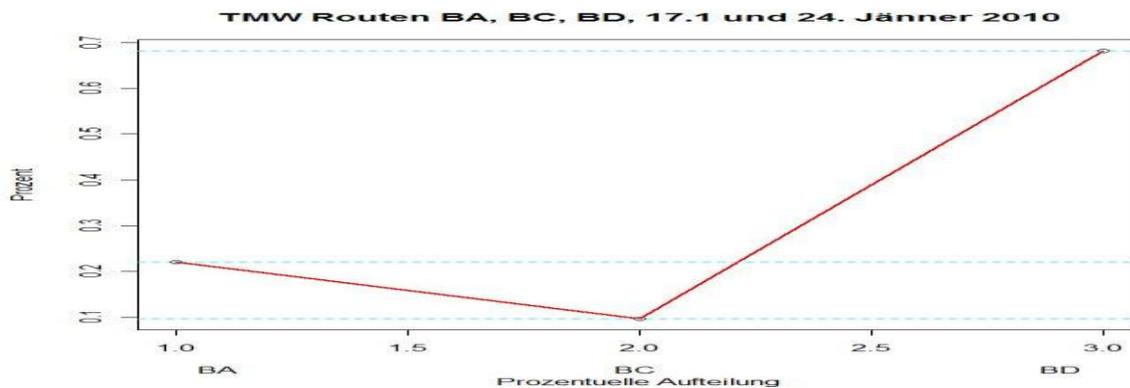


Route BC**Route BC 11. - 15. Jänner 2010 Frequenzverlauf****Route BC 11. - 15. Jänner 2010 Prozentverlauf****Route BD****Route BD 11. - 15. Jänner 2010 Frequenzverlauf****Route BD 11. - 15. Jänner 2010 Prozentverlauf**

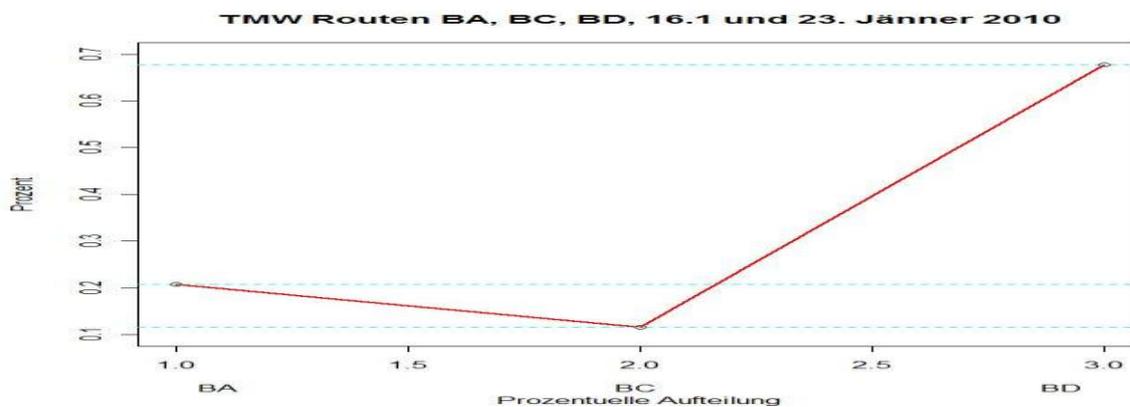
Im Messpunkt B sind große Frequenzunterschiede, hinsichtlich der Routen erkennbar. Route BD wird mit Abstand am häufigsten befahren, während Route BC nur eine sehr schwache Verkehrsauslastung aufweist. Verkehrsspitzen sind aufgrund dessen, auf dieser

Route kaum ersichtlich. Dieser große Auslastungsunterschied wiederholt sich im prozentuellen Verlauf. Schwankungen treten aber, in einem weit moderateren Umfang auf. Diese schematische Darstellung lässt eine deutliche Richtungsbevorzugung des Verkehrs erkennen.

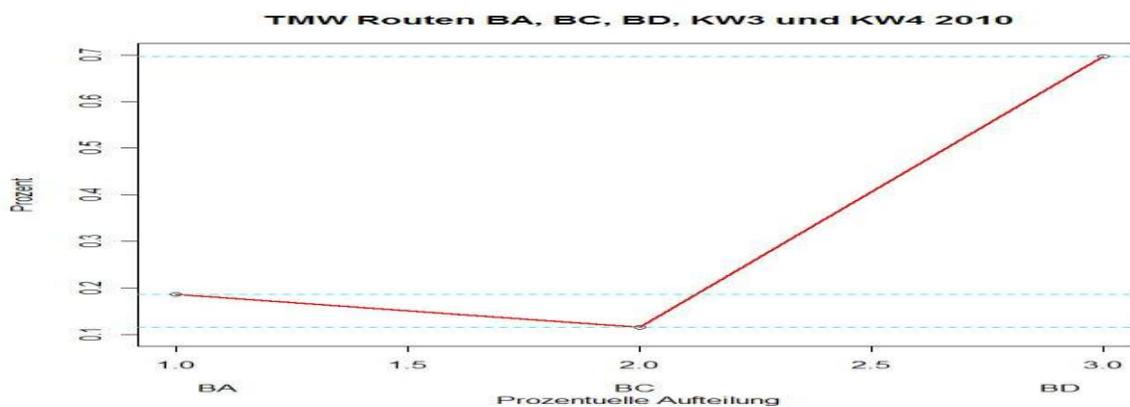
Prozentuelle Aufteilung der Routen BA, BC, BD Sonntags



Prozentuelle Aufteilung der Routen BA, BC, BD Samstags



Prozentuelle Aufteilung der Routen BA, BC, BD Werktags



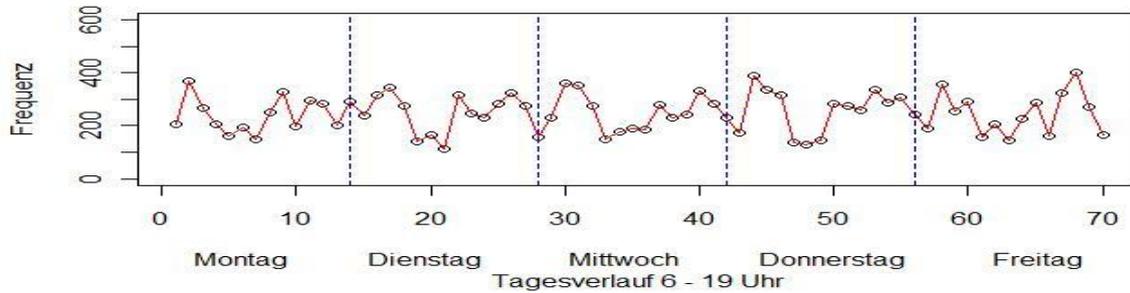
Grundsätzlich wiederholt sich die Routenauslastung, bezogen auf die jeweiligen Tagestypen. Schwankungen treten nur minimal, im einstelligen Prozentbereich, auf. Ein

deutlicher Nutzungsüberhang ist der Route BD (von Nord nach Süd) zuzuordnen. Demnach wird diese Strecke überwiegend von stadtauswärtsfahrenden Verkehrsteilnehmern benützt.

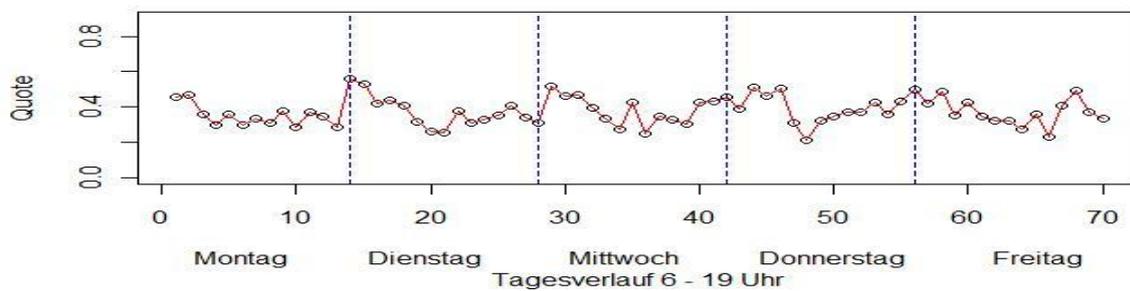
Routenauswertungen Messpunkt G

Route GE

Route GE 8. - 12. Februar 2010 Frequenzverlauf

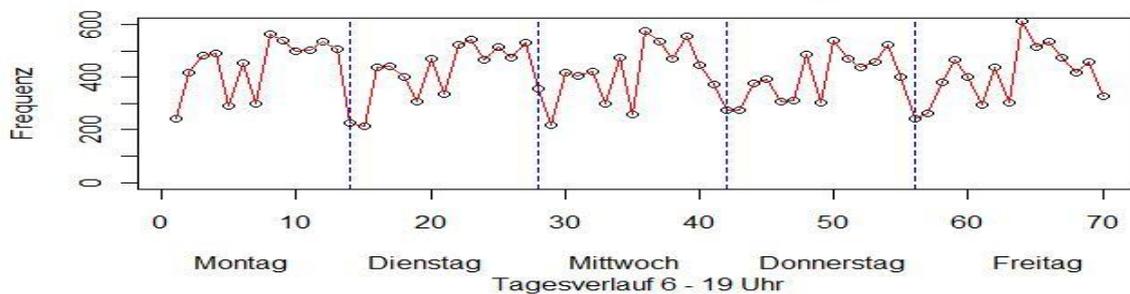


Route GE 8. - 12. Februar 2010 Prozentverlauf

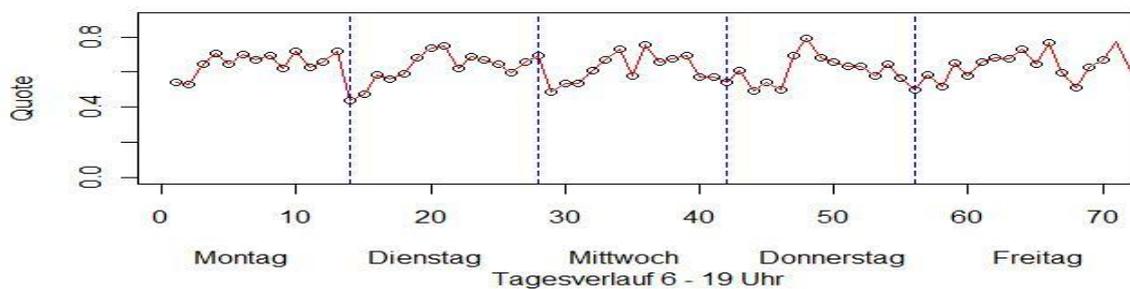


Route GF

Route GF 8. - 12. Februar 2010 Frequenzverlauf

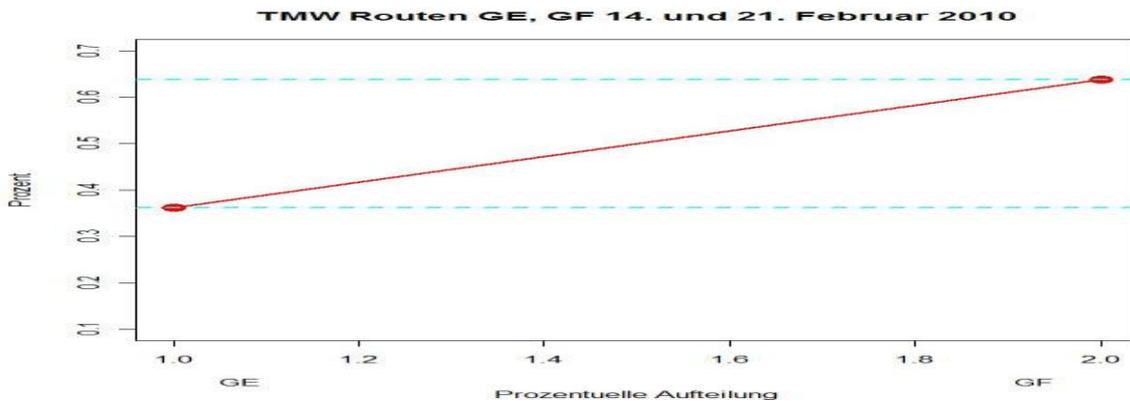


Route GF 8. - 12. Februar 2010 Prozentverlauf

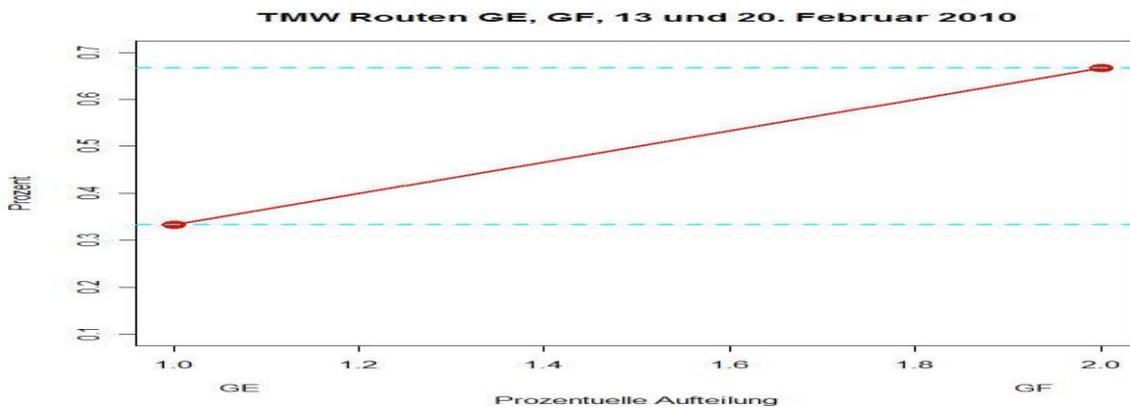


Der Frequenzvergleich beider Routen, ergibt eine bei weitem höhere Auslastung der Route GF. Während auf der Route GE vor allem morgens Verkehrsspitzen erreicht werden, zeigt die Route GF ein konträres Verhalten, wonach nachmittags höhere Frequenzdichten auftreten. Beide Prozentverläufe spiegeln dieses Verhalten wieder.

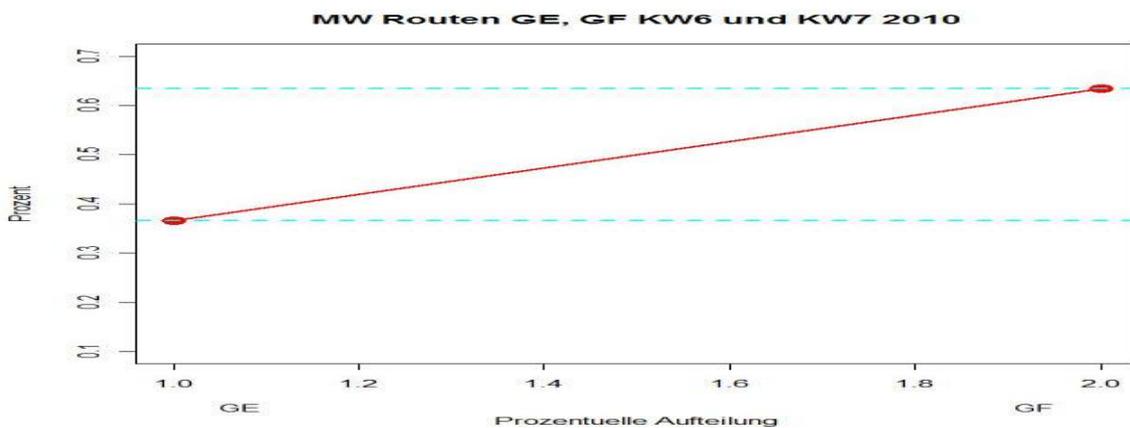
Prozentuelle Aufteilung der Routen GE, GF Sonntags



Prozentuelle Aufteilung der Routen GE, GF Samstags



Prozentuelle Aufteilung der Routen GE, GF Werktags



Der prozentuelle Vergleich der Frequenzaufteilung ergibt eine deutliche Konzentration auf der Route GF. An allen 3 Tagestypen sind nur marginale Aufteilungsunterschiede zu erkennen. Demnach ist dieser Route, unabhängig vom Tagestyp, eine Nutzungspriorität zuzuordnen.

Zusammenfassung:

Zusammenfassend ist zu sagen, dass von allen analysierten Routen, die Strecke GF, die von Westen nach Süden verläuft, die stärkste Nutzungsfrequenz besitzt. Die größte Verkehrsdichte besteht zwischen 13.00 und 17.00 Uhr, stadtauswärts. Dieselbe Frequenzdichte ist morgens hingegen nicht zu registrieren. Demnach besteht eine Nutzungspriorität dieser Route für jene Personen, die die Stadt nachmittags verlassen möchten.

Auf den Routen AB, AC und AD sind Verkehrsspitzen erkennbar, deren Schwankungsbreite, aufgrund der gesamten Verkehrsauslastung, deutlich geringer ausfällt. Anhand der Frequenzverläufe sind auch Stoßzeiten ersichtlich.

Die prozentuellen Verläufe spiegeln Streckenbevorzugungen wieder. Schlußendlich ist noch hinzuzufügen, dass bezüglich der unterschiedlichen Routen, die anteilmäßige Verkehrsauslastung an den verschiedenen Tagestypen grundsätzlich gleich bleibt. Änderungen fallen nur sehr geringfügig aus. Zusätzlich ist darauf hinzuweisen, dass der werktägliche Berufsverkehr die prozentuelle Aufteilung aller betrachteten Routen, nur marginal beeinflusst.

3.5.11 Conclusio

Für die Güte eines Modells ist nicht die Anzahl der registrierten Bluetoothsignale und somit die Quote ausschlaggebend, sondern deren Verlauf, und Fluktuation. Auf Verläufe mit linearen oder quadratischen Steigungen, wird mittels engerer Faktorisierung und eventueller Errechnung eines Regressionsmodells mit quadratischem Term, reagiert. Varianzinhomogenität, wird mittels Box-Cox-Transformation der Responsevariablen y begegnet. Quotenschwankungen werden mittels Glättungen 'begradigt'. Mit all diesen Maßnahmen erzielt man markante Modellverbesserungen. Mittels geeigneter Wahl der Messstandorte und starkes Engagement bezüglich der Filterung der Daten, wird vor Beginn der eigentlichen Regressionsanalyse, ein guter Grundstein gelegt.

Besonders wichtig ist die Erwähnung, dass die errechneten Regressionsmodelle nur solange Gültigkeit besitzen, solange sich der Mix aus den unterschiedlichsten Einflussgrößen, für die Registrierung der Bluetoothsignale, nicht markant ändert. Sollte sich einer dieser Faktoren,

wie der Standort des Sensors, Gerätekalibrierung, bauliche Gegebenheiten oder die Verkehrscharakteristik ändern, ist eine Neuberechnung eines Regressionsmodells erforderlich.

Aussagekräftige Modelle erfordern eine zumindest 4-wöchige Datenerhebung. Eingehende Analysen haben zudem ergeben, dass ab einer BLIDS-Frequenz von 40 erfassten Signalen, eine Quotenberuhigung eintritt. Somit ist eine Umlegung der Schätzfrequenz auf die einzelnen Routen, bei einer geringeren ID-Erfassung, nicht sinnvoll.

Anhand der Routenauswertung erkennt man den großen Vorteil dieses Systems, gegenüber anderen Messsystemen, die nur Querschnittsmessungen erlauben. Man kann den Gesamtverkehr in Routen aufteilen und jeder Gruppe eine Richtung zuordnen. Aufgrund dieser Informationen ist es besser möglich, Verkehrskonzepte von stark frequentierten Einzugsgebieten zu erstellen.

3.6 Mathematische Grundlagen von Regressionsanalysen

In diesem Kapitel werden die mathematischen Grundlagen zur Regressionsanalyse wiederholt bzw. näher erläutert. Die Regressionsanalyse und das dazugehörige Modell, Parameterschätzungen, Qualitätsmerkmale, Hypothesentests, Residuen und Varianzstabilisierende Transformationen werden konkretisiert. Als Referenzen hierzu dienen Ernst Stadlober [6], David Kleinbaum [10], Thomas Kälin [12], Andreas Handl [13].

Ziel einer einfachen linearen Regression ist es eine Beziehung zwischen einer abhängigen Variablen, auch Response oder Zielgröße genannt, und einer unabhängigen Variablen, der sogenannten erklärenden Variablen oder Prädiktor, in Form eines linearen Modells $Y = \beta_0 + \beta_1 x + \varepsilon$ zu quantifizieren. Dieses Modell präsentiert eine Gerade, wobei β_0 den y-Achsenabschnitt (intercept) und β_1 die Steigung der Geraden darstellt.

In unserem Fall wird durch die folgende Analyse die Beziehung zwischen der abhängigen Variablen Radarfrequenz und der einzigen unabhängigen Variablen BLIDS-Frequenz beschrieben.

3.6.1 Einfache lineare Regression

Gegeben seien zwei Vektoren $x = (x_1, \dots, x_n)$ und $Y = (Y_1, \dots, Y_n)$, wobei x_i bekannte Datenrealisierungen, und Y_i eine Zufallsvariable (ZV) ist. Des Weiteren wird angenommen, dass Y ein von x abhängiger Zufallsvektor sei. Man nimmt also an, dass die Zufallsvariable Y_i an einer festen Stelle x_i eine von x_i linear abhängige Erwartung habe.

Das einfache lineare Regressionsmodell kann wie folgt beschrieben werden:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, \dots, n$$

bzw.

$$E(Y_i) = \mu_i = \beta_0 + \beta_1 x_i \quad i = 1, \dots, n$$

Die Zufallsvariablen $\varepsilon_1, \dots, \varepsilon_n$ sind nicht beobachtbar und werden durch die sogenannten Residuen r_1, \dots, r_n geschätzt (siehe 3.6.2).

Als Voraussetzung gilt:

ε_i unabhängig identisch $N(0, \sigma^2)$ -verteilt und Y_i unabhängig $N(\mu_i, \sigma^2)$ -verteilt.

Die Eigenschaft von nicht signifikant unterschiedlichen Varianzen der Residuen, bezeichnet man als Varianzhomogenität. Y_1, \dots, Y_n sind zwar unabhängig, aber nicht identisch verteilt, wegen $E(Y_i) = \mu_i$.¹⁶

3.6.2 Parameterschätzung

Für ein einfaches lineares Regressionsmodell sind folglich die Parameter β_0 , β_1 und σ^2 zu schätzen.

Die unbekanntenen Regressionskoeffizienten β_0 und β_1 werden durch die Kleinste-Quadrate-Methode bzw. Least Square Methode geschätzt.

Aus der Minimierung der Fehlerquadratsumme

$$SSE(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 x_i))^2 \rightarrow \min \quad \text{wobei } \beta_0 + \beta_1 x_i = \mu_i$$

resultieren die LS-Schätzer $b_0 = \bar{Y} - b_1 \bar{x}$ und

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Die geschätzten Erwartungswerte des Modells lauten somit:

$$\hat{Y}_i = \hat{E}(Y_i) = \hat{\mu}_i = b_0 + b_1 x_i = (\bar{Y} - b_1 \bar{x}) + b_1 x_i = \bar{Y} + b_1 (x_i - \bar{x}).$$

Die Methode der kleinsten Quadrate ermöglicht jedoch keine Schätzung des Varianzparameters σ^2 . Bedingt durch die Annahme der unabhängigen $N(\mu_i, \sigma^2)$ Zufallsvariablen Y_i können jedoch durch die Maximum-Likelihood-Methode, Schätzer für β_0 ,

¹⁶ Literaturverzeichnis [10]

β_1 und σ^2 hergeleitet werden. Die ML-Schätzer für b_0 und b_1 entsprechen den obigen LS-Schätzern, für die Varianz ergibt sich

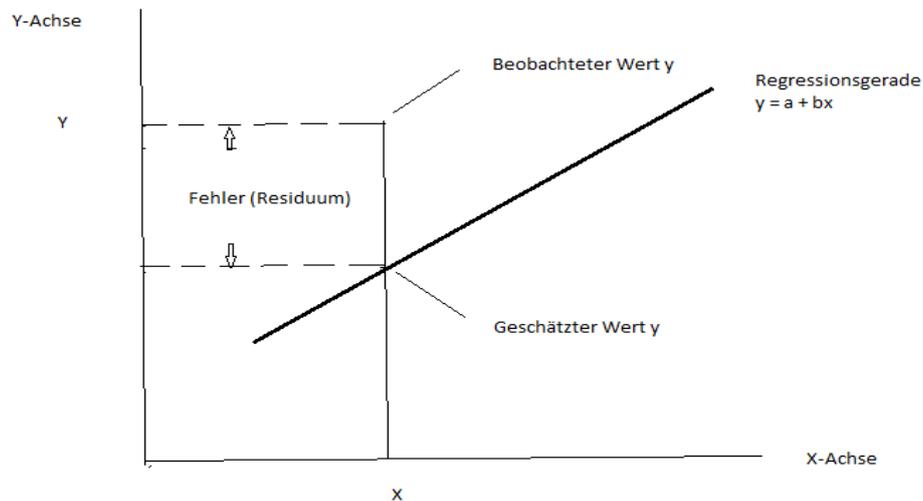
$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\mu}_i)^2 = \frac{1}{n} SSE(b_0, b_1) = \frac{\text{minimale Fehlerquadratsumme}}{n}.$$

Dieser Schätzer für σ^2 ist jedoch nicht erwartungstreu. Da sowohl b_0 als auch b_1 geschätzt wurden, ist der Schätzer für die Varianz entsprechend zu skalieren. Der erwartungstreue Schätzer für σ^2 lautet somit:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\mu}_i)^2 = \frac{SSE(b_0, b_1)}{n-2}$$

$R_i = Y_i - \hat{\mu}_i$ ist das i -te Residuum¹⁷.

3.6.3 Grafische Interpretation



Das Residuum ist jener Fehler, der die Entfernung zwischen dem beobachteten Wert Y , und dem erwarteten Wert (geschätzten Wert) auf der Regressionsgeraden $\hat{y} = a + bx$ beschreibt. Der sogenannte Intercept ist a und b ist die Steigung der Geraden. Mittels der Regressionsanalyse sollen die Parameter der Regressionsgeraden so berechnet werden, dass die Fehlerquadratsummen minimiert werden.

¹⁷ Literaturverzeichnis [10]

3.6.4 Multiple linear Regression

Sei n die Anzahl der Beobachtungen $y = (y_1, \dots, y_n)^t$ der Responsevektor, x_1, \dots, x_n die Prädiktoren, $\beta = (\beta_0, \beta_1, \dots, \beta_k)^t$ der (unbekannte) Parametervektor und $p := k + 1$ die Anzahl der Parameter im Modell. Seien weiters $\varepsilon = (\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n)^t$ unbekannte, nicht beobachtbare Fehlerterme. Dann beschreibt ε_i die Abweichung der i -ten Beobachtung y_i von der wahren Regressionsoberfläche.

Unter Annahme einer Normalverteilung kann das multiple Regressionsmodell mit

$$Y_i = \mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad \text{mit} \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

und in Matrizennotation als

$$Y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0, \sigma^2 I_n)$$

bzw.

$$E(Y) = \mu = X\beta \quad \text{mit} \quad Y \sim N(\mu, \sigma^2 I_n)$$

wobei $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}$... Vektor der Beobachtungen Y_i ,

$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}$... $(n \times p)$ Designmatrix (Matrix der unabhängigen Variablen)

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \quad \dots \text{Parametervektor}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix} \quad \dots \text{Fehlervektor}$$

beschrieben werden.

Die Regressionskoeffizienten $\beta_0, \beta_1, \dots, \beta_k$ werden mit der Least – Squares – Methode geschätzt. Somit erhält man den LS – Schätzer für β als Lösung der Normalgleichungen

$$b = (X^t X)^{-1} X^t Y$$

Als Prognosevektor bzw. Vektor der Vorhersagewerte wird der folgende Vektor definiert:

$$\hat{\mu} = Xb = X(X^t X)^{-1} X^t y = Hy \quad \text{wobei} \quad H = X(X^t X)^{-1} X^t$$

als Hatmatrix bezeichnet wird.¹⁸

3.6.5 Bestimmtheitsmaß R^2

Das Bestimmtheitsmaß R^2 gibt den Anteil der Variabilität einer abhängigen Variablen an, der durch das Regressionsmodell erklärt wird, d. h. R^2 ist ein Maß für die Anpassung der Regressionsgeraden an die zugrundeliegenden Daten.

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{SSR(b)}{SST} = 1 - \frac{SSE(b)}{SST}, \quad 0 \leq R^2 \leq 1$$

$R^2 = 1 \quad \Leftrightarrow \quad b_i = y_i, \quad i = 1, \dots, n.$ Das entspricht dem Idealfall, da das Modell eine perfekte Anpassung an die Realdaten ist.

¹⁸ Literaturverzeichnis [10]

$R^2 = 0 \Leftrightarrow \hat{\mu}_i = \bar{y}, \quad i = 1, \dots, n.$ Das entspricht genau dem Gegenteil und bedeutet, dass es keine Beziehung, sprich Abhängigkeit zwischen der abhängigen und der unabhängigen Variablen gibt.¹⁹

3.6.6 Adjustiertes Bestimmtheitsmaß $R^2(\text{adj})$

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n-1}{n-p} = 1 - \frac{SSE(b) / (n-p)}{SST / (n-1)}$$

wobei p die Anzahl der geschätzten Parameter ist und n – p bzw. n – 1 die dazugehörigen Freiheitsgrade darstellen.¹⁹

3.6.7 Bemerkung: Overfitting

Als overfitting (deutsch: Überanpassung) wird in der Regressionsanalyse ein Modell charakterisiert, das zu viele erklärende Variable enthält. Durch die zusätzliche Aufnahme von Regressoren kann das Bestimmtheitsmaß R^2 nicht sinken. Dadurch können irrelevante Regressoren zur Erklärung der Varianz beitragen und somit die Anpassung des Modells an die Daten künstlich erhöhen. Große Unterschiede zwischen R^2 und dem korrigierten Bestimmtheitsmaß R_{adj}^2 deuten auf einen solchen Effekt hin.²⁰

3.6.8 Angewandte statistische Tests

In der Regressionsanalyse ist die Klärung, welche Faktoren im Modell $E(Y) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$ relevant sind, von fundamentaler Bedeutung.

t – Test

Die Nullhypothesen der Form $H_0 : \beta_j = 0, j = 1, \dots, k$ werden getestet, d. h. jede unabhängige Variable wird unter der Voraussetzung, dass alle anderen Prädiktoren gleichbleiben, geprüft. Kann die Nullhypothese für ein β_j nicht verworfen werden, ist daraus

¹⁹ Literaturverzeichnis [11]

²⁰ Literaturverzeichnis [13]

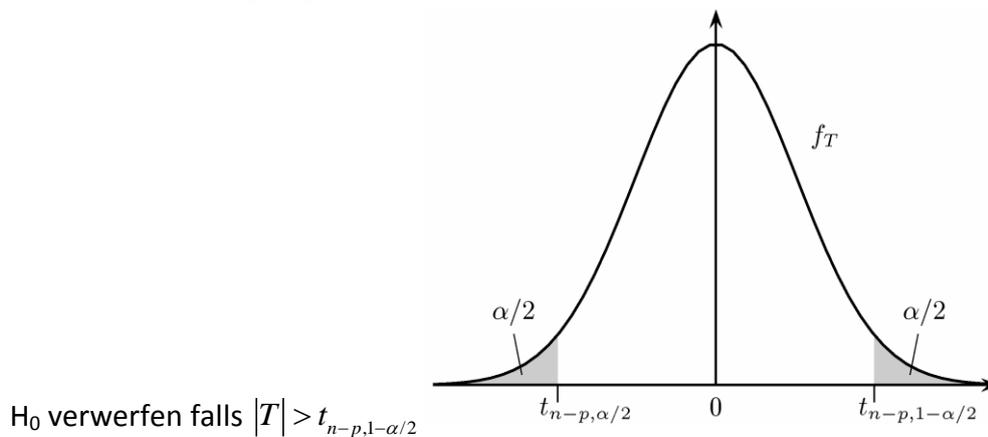
zu folgern, dass die j – te unabhängige Variable einen nicht zu beachtenden Einfluss auf die Response ausübt, und die dazugehörige Variable ist nicht in das Modell aufzunehmen.²¹

Teststatistik:

$$T = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 m_{jj}}} \stackrel{H_0}{\sim} t_{n-p}$$

wobei $(X^T X)^{-1}_{jj} = m_{jj}$.

Verwerfungsregel:



3.6.9 Residuenanalyse

Aufgrund dessen, dass statistische Fehler ε_i nicht beobachtet werden können, werden zur Prüfung der Annahme $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ die Residuen $R_i = Y_i - \hat{\mu}_i$ herangezogen, wobei für die Residuen folgendes gilt:

$$R_i \sim N(0, \sigma \sqrt{1 - h_{ii}}) \quad 22$$

H_{ii} ... Diagonalelement der Hatmatrix H

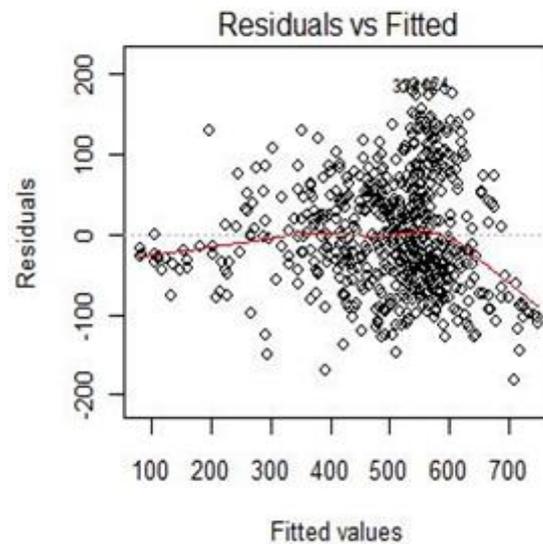
²¹ Literaturverzeichnis [5], [6], [12]

²² Siehe 3.6.4

Konstante Varianz (Residuen gegen Vorhersagewerte)

Hierzu trägt man die (standardisierten) Residuen R_i gegen die Vorhersagewerte $\hat{\mu}_i$ auf. Liegen die Residuen in zufälliger Anordnung in einem horizontalen Band, das symmetrisch um Null ist, so besteht keine Abhängigkeit zwischen Residuen und Gruppenmittelwerten. Muster, in Form von Trichtern oder Punktverdichtungen im Scatterplot, deuten aber sehr wohl auf eine solche Abhängigkeit hin. In einem solchen Fall liegt eine Inhomogenität der Varianzen vor.²³

Beispiel:



Eine Stabilisierung der Varianzen kann möglicherweise durch eine Transformation der abhängigen Variablen erreicht werden.

²³ Literaturverzeichnis [13]

Distanz von Cook

Hebelpunkte sind Datenrealisierungen, die die Parameterschätzung einer Regressionsanalyse negativ beeinflussen.

Da $\hat{\mu} = X(X^t X)^{-1} X^t y = Hy$ gilt (siehe 3.6.4):

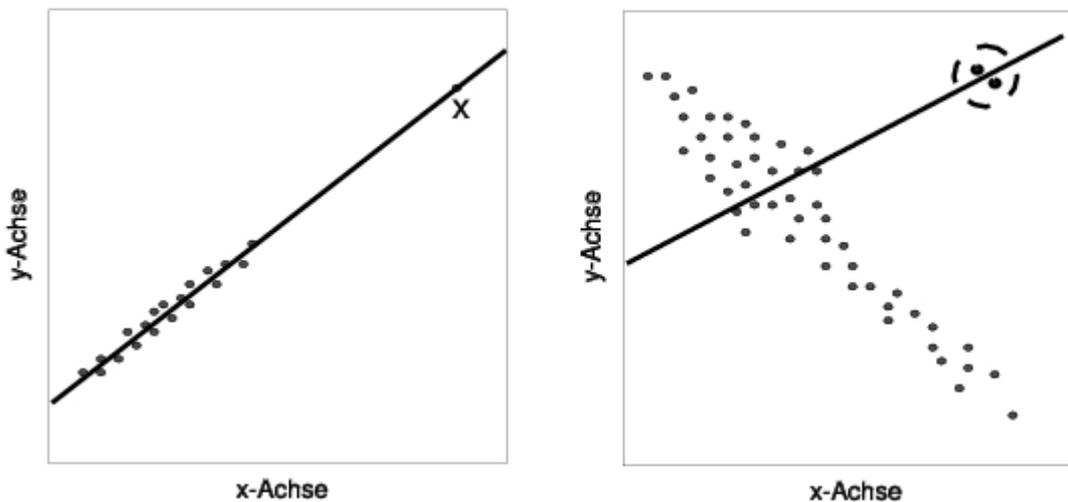
$$\hat{\mu}_i = \sum_{j=1}^n h_{ij} y_j = h_{ii} y_i + \sum_{j \neq i} h_{ij} y_j, \quad i = 1, \dots, n$$

weil
$$\frac{\partial \hat{\mu}_i}{\partial y_i} = h_{ii}, \quad i = 1, \dots, n.$$

können die Diagonalelemente h_{ii} als Maß für den Einfluss von y_i auf den Vorhersagewert $\hat{\mu}_i$ betrachtet werden.

Aufgrund der Idempotenz und Symmetrie der Hatmatrix H gilt zusätzlich

$$h_{ii} = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2, \quad i = 1, \dots, n \quad \Rightarrow \quad 0 \leq h_{ii} \leq 1$$



Hebelpunkte sind mit einem Cook-Distanz-Plot zu lokalisieren. Cook-Distanzen sind Werte, die Veränderungen aller angepassten Werte zu einer Zahl zusammenfassen, die beim Weglassen einer Beobachtung entstehen. Sind Hebelpunkte vorhanden, so sind diese Datenrealisierungen zu entfernen, und die Regressionsanalyse zu wiederholen.²⁴

²⁴ Literaturverzeichnis [6], [13]

Beachte: Nicht jede einflussreiche Beobachtung hat große Hebelwirkung.

Definition: Die Cook-Distanz gibt die skalierte Distanz $b_{(i)}$ zu b wieder.

$$C_i = \frac{(b_{(i)} - b)^t X^t X (b_{(i)} - b)}{p \hat{\sigma}^2} = \frac{\hat{r}_i^2}{p (1 - h_{ii})}, \quad i = 1, \dots, n$$

wobei $\hat{r}_i = r_i / (\hat{\sigma} \sqrt{1 - h_{ii}})$... standardisierte Residuen

Je größer der Wert C_i ausfällt, desto größer wird die Parameterschätzung, durch die i -te Datenrealisierung, beeinflusst. D. h. C_i ist groß, wenn die i -te Beobachtung und/oder \hat{r}_i groß ist.²⁵

Normalverteilung

Die Normalverteilungsannahme der statistischen Fehler kann mittels QQ-Plot, bezüglich der standardisierten Residuen, überprüft werden. Dazu trägt man die Quantile einer Standard-Normalverteilung gegen die sortierten Werte \hat{r}_i auf. Bilden die daraus resultierenden Punkte ein lineares Muster, so liegt eine Normalverteilung der Residuen vor.

3.6.10 Varianzstabilisierende Transformationen

Liegt der Verdacht nahe, dass sich die Varianzen aller Ausprägungen nicht signifikant unterscheiden und/oder die Annahme einer Normalverteilung verletzt wird, so kann möglicherweise eine Transformation der abhängigen Variablen Abhilfe schaffen.

Definition: Sei Y eine Zufallsvariable mit $E(Y) = \mu$, $\text{Var}(Y) = \sigma^2$ proportional zu $\mu^{2\alpha}$ und

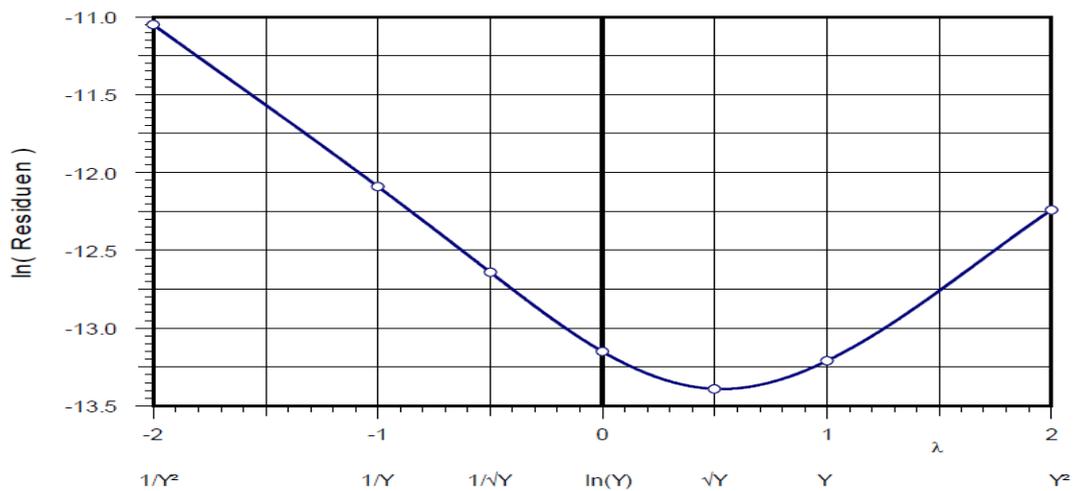
$$Y^* = T(Y) = \begin{cases} Y^\lambda & \text{für } \lambda \neq 0 \\ \log Y & \text{für } \lambda = 0 \end{cases}$$

wobei für $T(Y)$ $\lambda = 1 - \alpha$ eine konstante Varianz ist.

Häufig angewandte Transformationen sind:

²⁵ Literaturverzeichnis [12]

λ	$Y^* = T(Y)$
1	$Y^* = Y$
$\frac{1}{2}$	$Y^* = \sqrt{Y}$
0	$Y^* = Y$
$-\frac{1}{2}$	$Y^* = 1/\sqrt{Y}$
-1	$Y^* = 1/Y$



Analytische Bestimmung der Transformation

Ein allgemeiner Ansatz ist die **Box-Cox-Transformation**²⁶

$$Y^{(\lambda)} = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & \text{für } \lambda \neq 0 \\ \log Y & \text{für } \lambda = 0 \end{cases}$$

3.6.11 Vertrauens- und Prognoseintervalle für Parameter

Das Vertrauens- oder auch Konfidenzintervall umfasst alle Parameterwerte, die basierend auf das Ergebnis eines bestimmten statistischen Tests, nicht abgelehnt werden. Das ist dann der Fall, wenn der Absolutbetrag der Testgröße kleiner als das entsprechende Quantil ist.

²⁶ Literaturverzeichnis [6], [13]

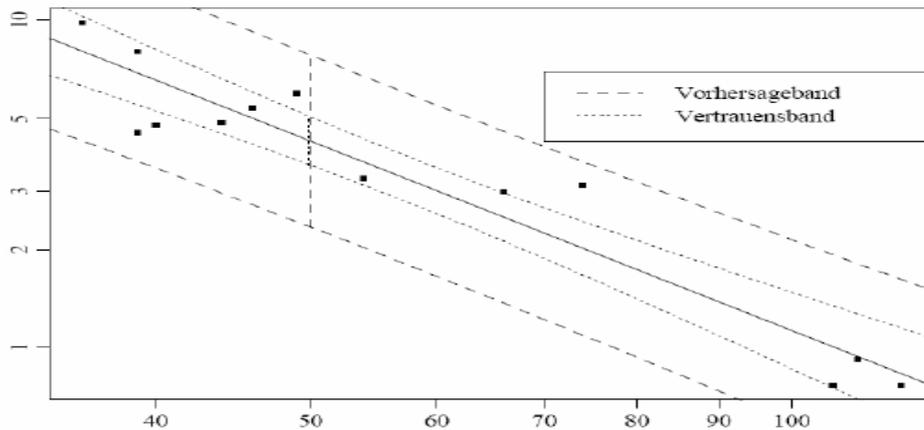
(1 - α) – KIV für den Erwartungswert

$$\hat{\mu} \pm \sqrt{\hat{\sigma}^2 x'(X'X)^{-1}x} t_{n-2,1-\alpha/2}$$

Jener Bereich, wo künftige Beobachtungen zu liegen kommen, nennt man hingegen Prognoseintervall. Dieses Intervall muss breiter als das Vertrauensintervall sein, da diesbezüglich die Zufallsabweichungen der künftigen Datenrealisierungen, zusätzlich berücksichtigt werden müssen.

Prognoseintervall (Vorhersageintervall)²⁷

$$\hat{\mu} \pm \sqrt{\hat{\sigma}^2 (1 + x'(X'X)^{-1}x)} t_{n-2,1-\alpha/2}$$



²⁷ Literaturverzeichnis [10], [13]

Schlussbemerkung

Beide in dieser Arbeit behandelten Themen wurden mit Daten durchgeführt, die sich in ihrer zeitlichen als auch örtlichen Realisierung zur Gänze unterscheiden. Die für die Mischanalyse verwendeten Daten stammen von einem Autobahnabschnitt in der Nähe von Nürnberg. Datenverunreinigungen sind aufgrund der Sensorpositionierungen und der geographischen Gegebenheiten so gut wie auszuschließen.

Daten aus urbanen Gebieten, wie der Stadt Graz, erfordern, um aussagekräftige Analysen durchführen zu können, hinsichtlich ihrer Handhabung sorgfältige Bereinigungen. Erst dadurch sind qualitativ hochwertige Schlussfolgerungen möglich.

Viele Personen assoziieren mit einer möglichst hohen Bluetoothquote bessere Hochrechnungsmöglichkeiten und daher auch realitätsnahe Aussagen über den tatsächlichen Verkehr. Basierend auf den Erkenntnissen der im Kapitel 3 durchgeführten Regressionsanalysen kann man das nicht generell sagen. Nicht nur die Höhe der Quote ist ausschlaggebend für gute Ergebnisse, sondern auch deren Verlauf. Starke Fluktuationen, lineare oder quadratische Steigungen, erfordern individuelle Analysemethoden. Deshalb kann ein genereller Leitfaden auch nicht zu allgemein gültigen Ergebnissen führen.

Aufgrund der vorhandenen Daten (Erhebung einer Stichprobe mittels BLIDS – Sensoren, Erhebung der Verkehrsfrequenz mittels Radarsensoren, Zeit und Ort der Erfassungen) ist die Durchführung einer Regressionsanalyse möglich. Da man für eine Mischverteilungsanalyse lediglich die Anzahl der Pärchen und deren Erfassungszeitpunkte benötigt, sind beide Analysen (Regressions- und Mischverteilungsanalyse) mit demselben Datensatz durchführbar. Basierend auf diesen Informationen, ist die Erstellung von umfangreichen Verkehrskonzepten in Ballungsräumen möglich. Ein Ansteigen der Durchdringungsrate der Bluetoothtechnologie bringt natürlich eine Verkürzung der Analysezeiträume mit sich und damit sind dynamische Berechnungen vorstellbar.

Verknüpfungen zwischen den Informationen einer Reisezeitanalyse und beispielsweise einer Luftgüteüberwachung sind durchaus denkbar, um den mittel- oder unmittelbaren Einfluss der verschiedenen Geschwindigkeiten des Gesamtverkehrs auf die Schadstoffemissionen, zu überprüfen.

Ein weiteres Anwendungsgebiet der BLIDS – Technologie wäre auch das aktive Suchen nach bestimmten Bluetoothdevices, um bei Erkennen einer solchen ID die Grünphasen von Ampelanlagen direkt zu steuern. Eine solche Anwendung wäre für Einsatzkräfte, die so rasch wie möglich am eigentlichen Geschehen sein müssen, von Vorteil.

Des weiteren kann man aufgrund der Zeitspannen zwischen Erst- und Letzterfassungen bei ein und demselben Sensor, auf die Verweildauer von Personen schließen. Bestückt man einen abgegrenzten Raum mit mehreren Geräten und wird dieselbe ID – Nummer in der Schnittstelle aller Sensoren registriert, kann man neben der Bleibedauer auch den ungefähren Standort ermitteln. Damit könnte man die Ansprechbarkeit von Werbungen oder Informationssendungen auf Personen ermitteln.

Der Empfang von bestimmten Bluetoothsignalen kann auch für die Verwaltung von Zugangsbeschränkungen verwendet werden. Erst bei Übereinstimmung einer ID – Nummer mit gespeicherten Devices, ist ein Eintritt in sensible Bereiche möglich.

Ein Pendant zur Verkehrsstromanalyse, ist die Personenstromanalyse. Damit könnte man das Verhaltensmuster von Menschenansammlungen bei Großveranstaltungen, wie Konzerten oder Ausstellungen, klären.

Literaturverzeichnis

- [1] www.statoek.wiso.uni-goettingen.de, Mischverteilungen, [August 2010]
- [2] Samuel Kotz & Norman L. Johnson, *Encyclopedia of statistical sciences, Volume 5, Lindeberg Condition – Multitrait-Multimethod Matrices*, S. 559 ff, *Mixture Distributions* John Wiley, New York, 1985
- [3] Thorn Thaler, *Mischverteilungsmodelle für die Feinstaubanalyse im Grazer Stadtgebiet*, Bakkalaureatsprojekt, Institut für Statistik, Technische Universität Graz, 2008
- [4] www.mathematik.uni-ulm.de/.../ss04/.../node9.html, Autor: Ursa Pantle 2004, *Gammaverteilung und χ^2 - Verteilung*, [August 2010]
- [5] www.viles.zef.uni-oldenburg.de, *Hypothesentests*, [August 2010]
- [6] Ernst Stadlober, *Angewandte Statistik, Vorlesungsskriptum*, Institut für Statistik, Technische Universität Graz, 2004
- [7] P. D. M. Macdonald, P. E. J. Green, *User's Guide to Programm MIX, An Interactive Program for Fitting Mixtures of Distributions*, ICHTU DATA SYSTEMS, 1988
- [8] www.math.mcmaster.ca/peter/mix/mix.html, P.D.M. Macdonald, *Package 'mixdist'*, [August 2010]
- [9] Konrad Maitisch, Ernst Stadlober, Kurt Fallast, Winfried Höpfl, *BLIDS2 Bluetooth-Based Traffic Data Collection System Demonstrator*, c.c.com Andersen & Moser GmbH, Grambach, 2006
- [10] David G. Kleinbaum, Lawrence L. Kupper, Keith E. Muller, Azhar Nizam, *Applied Regression Analysis and Other Multivariable Methods*, Pacific Grove, 3. Auflage 1998
- [11] www.de.wikipedia.org/wiki/Bestimmtheitsmaß, [August 2010]
- [12] www.thomaskaelin.ch/media/archive1/zusammenfassungen *Angewandte Statistik und Datenanalyse, Zusammenfassung v1.1*, Thomas Kälin, MSE ICT, HS 2008/2009
- [13] www.wiwi.uni-bielefeld.de, *Statistische Intervalle*, Andreas Handl, [August 2010]
- [14] c.c.com Anderson & Moser GmbH, *Installations- und Konfigurationsbeschreibung*, Grambach, 2010