

# Zeitlicher Zusammenhang zwischen Influenza- und Meningokokkenausbrüchen

Markus Breitenacher, Michael Obermeier, Michael Schomaker

22.Oktober 2004

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>2</b>
<b>2</b>	<b>Beschreibung des Datensatzes</b>	<b>4</b>
2.1	Allgemeine deskriptive Beschreibung . . . . .	4
2.2	Beschreibung der Daten mit Hilfe von Zeitreihendiagrammen .	7
2.3	Analyse der Daten mit Hilfe von Autokorrelationsfunktionen .	12
<b>3</b>	<b>Regressionsanalyse</b>	<b>17</b>
3.1	Lineare Modelle . . . . .	17
3.1.1	Lineares Modell mit nicht-transformierten Meningo- kokkendaten . . . . .	17
3.1.2	Lineares Modell mit wurzeltransformierten Meningo- kokkendaten . . . . .	22
3.1.3	Lineares Modell mit nicht saisonbereinigten Influenzada- ten . . . . .	25
3.1.4	Lineares Modell mit Meningokokken als Einflussgröße .	26
3.2	Generalisierte lineare Modelle . . . . .	27
3.2.1	Poissonregression . . . . .	27
3.2.2	Quasi-Poissonregression . . . . .	30
3.3	Regressionsanalyse für spezielle Untergruppen . . . . .	31
<b>4</b>	<b>Abschließende Bewertung</b>	<b>35</b>
4.1	Zusammenfassung . . . . .	35
4.2	Ergebnisse anderer Studien . . . . .	36
<b>A</b>	<b>S-Plus Code</b>	<b>39</b>

# Kapitel 1

## Einleitung

Die folgende Projektarbeit beschäftigt sich mit dem zeitlichen Zusammenhang zwischen Influenza- und Meningokokkenerkrankungen. Es wird vermutet, dass im Abstand von einer bzw. zwei Wochen nach einem Anstieg der Influenza- auch die Anzahl der Meningokokkenerkrankungen zunimmt. Dieser zeitliche Zusammenhang soll vorerst deskriptiv untersucht und im weiteren Verlauf anhand von geeigneten Regressionsmodellen statistisch analysiert werden.

Zunächst eine kurze Charakterisierung der beiden Krankheiten: Grippeerkrankungen sind eine akute Infektion der Atemwege durch Influenza-Viren. Man unterscheidet grundsätzlich zwischen A, B und C-Viren. Übertragen werden sie durch Tröpfcheninfektion, die Inkubationszeit beträgt zwischen einem und fünf Tagen. Die Grippe kann als leichte Erkältungskrankheit verlaufen, aber auch mit starkem Krankheitsgefühl, Fieber, Husten und Schnupfen einhergehen. Das Fieber hält ca. zwei bis drei Tage an und wenn keine Komplikationen auftreten, klingen die Erscheinungen nach ca. acht Tagen ab. Typischerweise kommt es jährlich in der kalten Jahreszeit zu Grippeepidemien. [1]

Bei Meningokokken handelt es sich um gramnegative Bakterien, die in 13 Serogruppen unterteilt werden können. Übertragungsweg ist die Tröpfcheninfektion, die Inkubationszeit liegt hier zwischen zwei und zehn Tagen. Untersuchungen haben ergeben, dass bei bis zu 40 Prozent bestimmter Altersgruppen und Populationen die Schleimhäute im Nasen-Rachenraum mit Meningokokken infiziert sind, was bedeutet, dass bei den meisten infizierten Personen die Krankheit nicht zum Ausbruch kommt. Generell sind Meningokokkenerkrankungen wesentlich seltener als Influenzafälle. Im Jahr 2003 beispielsweise gab es in Deutschland 8482 gemeldete Grippefälle und 772 Fälle von Meningokokkenerkrankungen.

Unter bestimmten Voraussetzungen können die Bakterien durch die Schleimhäute

eindringen und schwerwiegende Krankheiten auslösen. Meningokokken verursachen hauptsächlich zwei Arten von Krankheitsbildern, die einzeln oder gemeinsam auftreten können: Hirnhautentzündung (Meningokokken-Meningitis) und Blutvergiftung (Meningokokken-Sepsis). Die Letalität beträgt zwischen fünf und zehn Prozent. Zu den Voraussetzungen, die zum Ausbruch der Krankheit führen können, zählt auch die Schädigung der Schleimhäute durch virale Infektionen. [2,3] Das führt wiederum zu der Fragestellung für dieses Projekt: Inwieweit besteht ein zeitlich-kausaler Zusammenhang zwischen Grippeepidemien und einer anschließend erhöhten Zahl an Meningokokkenkrankungen? Kommt es nach einer Grippewelle immer auch zu einem signifikanten Anstieg der Krankheitsfälle mit Meningokokken?

Benutzt wurden die Daten des Robert Koch Instituts für Deutschland von Januar 2001 bis August 2004, sortiert nach den einzelnen Wochen. [4] Jede Zeile des Datensatzes stellt somit eine Woche dar. Da das Robert Koch Institut mit 53 Meldewochen pro Jahr arbeitet, besteht die erste und letzte in der Regel nicht aus vollen 7 Tagen. Deshalb fassten wir die 53. Woche im Jahr  $t$  mit der ersten im Jahr  $t + 1$  zu einer kompletten Woche zusammen. Hiermit wurde garantiert, dass nun jede Meldewoche aus genau 7 Tagen besteht. Im Jahr 2001 endet die 53. Kalenderwoche mit einem Montag. Dieser wurde somit zu den sechs Tagen der ersten Woche 2002 hinzugezählt. Damit besteht nun jedes Jahr aus genau 52 Meldewochen.

Die Tatsache, dass sich ein Jahr in nicht exakt 52 Wochen aufteilen lässt, kann hier vernachlässigt werden, da nur drei Jahresübergänge zu beachten sind, bei denen dieses Problem ohne Auswirkungen übergangen werden konnte. Die erste Woche im ersten Beobachtungsjahr 2001 beginnt zudem mit einem Montag, sodass auch hier kein Problem auftritt.

Die Variablen, also die Spalten, sind unterteilt in die Gesamtanzahl der Influenza- bzw. Meningokokkenfälle, die jeweiligen Fälle pro Alterskategorie, Bundesland, Erregertyp, Geschlecht und Diagnoseverfahren. Die Alterskategorien unterteilen sich wie folgt: Von null bis vier Jahren gibt es für jedes Jahr eine Kategorie, von fünf bis 29 sind die Kategorien im 5-Jahresabstand, danach im 10-Jahresabstand.

# Kapitel 2

## Beschreibung des Datensatzes

### 2.1 Allgemeine deskriptive Beschreibung

Zunächst sollen hier die Daten im Hinblick auf allgemeine Gesichtspunkte beschrieben werden.

Geschlechtsspezifische Unterschiede fielen bei Betrachtung der Gesamtfälle kaum auf: Der Anteil der männlichen Krankheitsfälle lag jeweils geringfügig höher als der der weiblichen. Bei Influenza beträgt das Verhältnis 52 % zu 48 %, bei Meningokokken 53 % zu 47 %.

Die beiden Kreisdiagramme aus 2.1 zeigen die Anteile der jeweiligen Alterskategorien zu den gemeldeten Fällen in den Jahren 2001 bis zur 34. Kalenderwoche 2004.

Es fällt auf, dass von beiden Krankheitstypen jüngere Menschen überproportional oft betroffen sind. Bei den Meningokokkenfällen wird es besonders deutlich: Die weitaus größte Gruppe ist die der 0 bis 4-jährigen mit einem Anteil von mehr als einem Drittel, wobei hier wiederum die 0-Jährigen (12,1 %) die größte Kategorie bilden. Die zweitgrößte Gruppe bildet die der 15 bis 19-Jährigen (21,8 %). Insgesamt sind drei Viertel der Meningokokken-Patienten jünger als 20 Jahre.

Auch bei den Influenzafällen wird deutlich, wenn auch weniger stark, dass überproportional oft jüngere Menschen betroffen sind. Mit 26,5 % der Fälle stellen die 0 bis 4-Jährigen wieder die größte Gruppe dar, gefolgt von den 5 bis 9-Jährigen (16,9 %) und den 10 bis 14-Jährigen (16,4 %). Anders als bei den Meningokokkenerkrankungen gibt es hier keinen zweiten Gipfel bei den 15 bis 19-Jährigen, diese Gruppe ist weit weniger oft von Influenza betroffen als die vorangegangenen Altersklassen. Etwa zwei Drittel (68,3 %) der Influenzafälle verteilen sich auf die Altersgruppen 0 bis 19 Jahre.

Bei den Erkrankungszahlen handelt es sich jedoch um Absolutzahlen. So-

mit sind die hier verwendeten Anteilswerte nicht um die Altersstruktur der Bevölkerung korrigiert. Dies ist bei der Interpretation zu berücksichtigen.

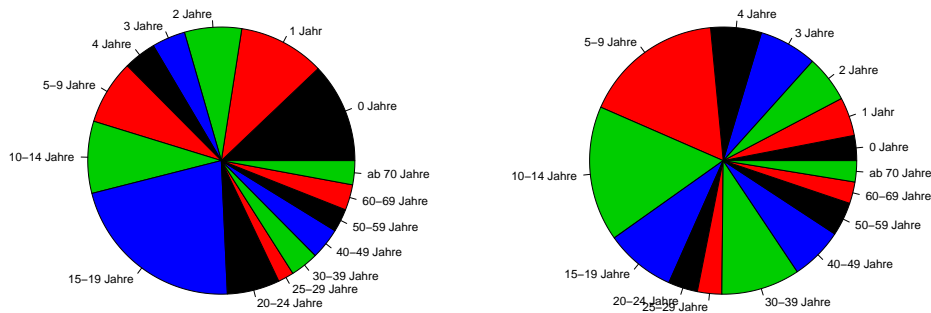


Abbildung 2.1: Altersverteilung der Meningokokken- und Influenzafälle

Die darauffolgenden Diagramme aus 2.2 stellen die Häufigkeit der einzelnen Erregertypen dar. Wie bereits in der Einleitung erwähnt, sind die Grippeviren grundsätzlich in A, B und C-Virentypen unterteilt. Die meisten der Erkrankungen gehen auf den Influenza-Erreger der Gruppe A zurück (58,9 %), die zweitgrößte Gruppe bilden die Fälle, in denen sowohl A als auch B-Virus auftraten (25,7 %). B-Virustyp allein (6,5 %) und C-Virustyp (7,6 %) kamen relativ selten vor, wobei sich die Ausbrüche des C-Virus beinahe ausschließlich auf die Grippewelle des Jahres 2002 beschränkte, während der B-Typ hauptsächlich im Jahr 2003 eine Rolle spielte.

Bei den Meningokokkenfällen fällt zunächst auf, dass für viele Fälle die Information über die Serogruppe nicht vorliegt (27,5 %). Die meisten Erkrankungsfälle verursachte die Serogruppe B (46,7 %), gefolgt von der Serogruppe C (18,9 %). Die Serogruppen A (0,7 %), W 135 (1,6 %) und Y (1,4 %) spielten eine untergeordnete Rolle.

Zur räumlichen Verteilung der Krankheiten innerhalb Deutschlands bieten die beiden Schaubilder des Robert-Koch-Instituts aus 2.3 eine Übersicht. [4] Dargestellt sind die Inzidenzwerte je Bundesland, ermittelt über alle vier Jahre. Inzidenzen stellen die Anzahl der Erkrankungen je 100.000 Einwohner dar.

Bei den Grippefällen fällt dabei auf, dass im Nordwesten die Erkrankungsquote wesentlich niedriger ist als in den übrigen Teilen Deutschlands. Ausnahme ist dabei der Stadtstaat Bremen mit einer Inzidenz von 38,0 Fällen

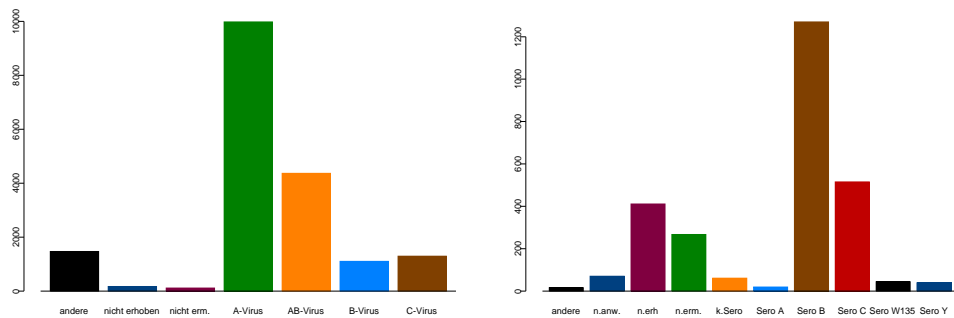


Abbildung 2.2: Virentypen bei Influenza, Serotypen bei Meningokokken

je 100.000 Einwohner. Die mit Abstand höchsten Quoten wurden in Sachsen (69,9) und Sachsen-Anhalt (49,5) beobachtet. Zum Vergleich dazu lag die Quote in den nordwestlichen Bundesländern generell unter 15. Die niedrigsten Inzidenzquoten lagen in Hamburg (3,4) und im Saarland (6,2).

Bei den Meningokokkeninzidenzen wird ein starkes Nord- Südgefälle deutlich, wobei hier zu beachten ist, dass die Quoten natürlich generell niedriger sind, also keine so hohen Unterschiede wie bei den Influenzadaten zwischen den Bundesländern auftreten. Im Süden liegen die Werte zwischen 2.6 in Hessen und 2.8 in Bayern. Die Spitzenwerte liegen im Norden Deutschlands in Mecklenburg-Vorpommern mit 5.1 und in Hamburg mit 4.7. Besonders interessant für die hier zu untersuchende Hypothese erscheinen die Werte von Hamburg, das die mit Abstand niedrigste Inzidenzquote bei Influenza, aber die zweithöchste Meningokokkenquote aufweist. Bei dem vermuteten kausalen Zusammenhang zwischen beiden Krankheiten würde man dagegen vor allem in Gebieten mit hoher Influenzarate eine ebenfalls erhöhte Anzahl Meningokokkenfälle erwarten.

Ein stellvertretender Vergleich zwischen Bundesländern stellt die darauffolgende Graphik 2.4 dar, die den zeitlichen Verlauf der Influenzaerkrankungen über die Jahre 2001 bis 2004 in den Ländern Nordrhein-Westfalen und Sachsen gegenüberstellt. Zunächst ist auffällig, dass das eigentlich einwohnerstärkere Bundesland Nordrhein-Westfalen eine geringere Anzahl an Krankheitsfällen aufweist. Desweiteren erscheint der Verlauf der beiden Kurven interessant. Man erkennt eine zeitliche Verschiebung bei den jährlichen Influenzaausbrüchen. Die Höchstwerte im Osten folgen den Spitzen im Westen um ca. ein bis drei Wochen. Die Grippewellen breiteten sich also generell in diesen Jahren von West nach Ost aus. Ebenso lässt sich eine deutliche Saisonalität in den Wintermonaten feststellen, die im Folgenden näher be-

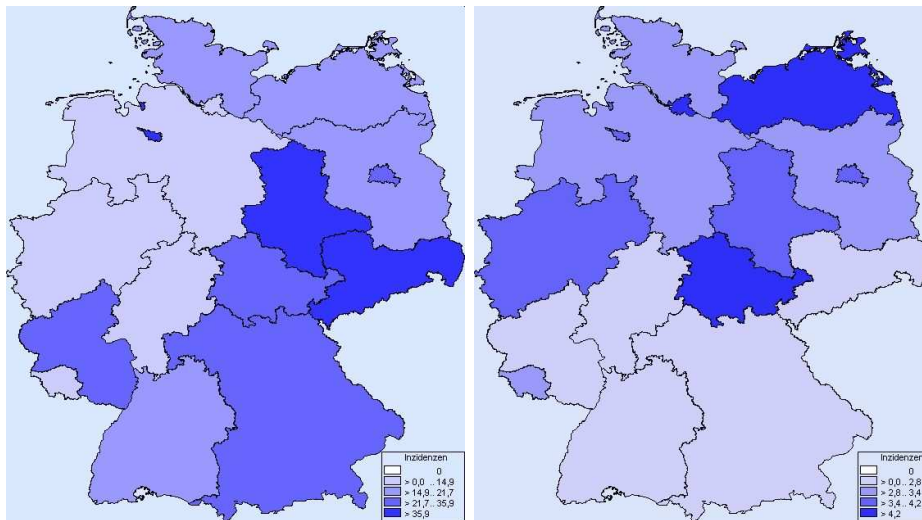


Abbildung 2.3: Inzidenzrate Influenza und Meningokokken je Bundesland

trachtet wird.

## 2.2 Beschreibung der Daten mit Hilfe von Zeitreihendiagrammen

Da die allgemeine Fragestellung die nach einem zeitlichen Zusammenhang zwischen Influenza- und Meningokokkenerkrankungen ist, erscheint es zunächst sinnvoll, eine Graphik zu betrachten, in der die beiden Zeitreihen miteinander verglichen werden. Das folgende Schaubild 2.5 stellt die übereinander geplotteten Verlaufskurven der jeweiligen Krankheitsfälle dar. Zunächst fällt der bereits erwähnte große Unterschied in den Fallzahlen auf. Die Erkrankungszahlen der Influenza sind mit 16.929 gemeldeten Fällen bedeutend größer als die der Meningokokkenerkrankungen mit lediglich 2719. Ein zeitlicher Zusammenhang scheint in dieser Graphik auf den ersten Blick nicht erkennbar, den Influenzawellen folgt kein deutlicher Anstieg der Meningokokkenzahlen. Da es sich hier jedoch um sehr unterschiedliche Absolutzahlen handelt, fällt ein sinnvoller Vergleich bei diesem Maßstab schwer.

Um die Daten auf ein vergleichbares Niveau zu bringen, war es notwendig, sie auf geeignete Art und Weise zu transformieren. Eine Wurzeltransformation erscheint dabei sinnvoll.

Da es sich hier um Zähldaten handelt, kann von einer Poissonverteilung ausgegangen werden. Bei poissonverteilten Zufallsvariablen entspricht die Vari-

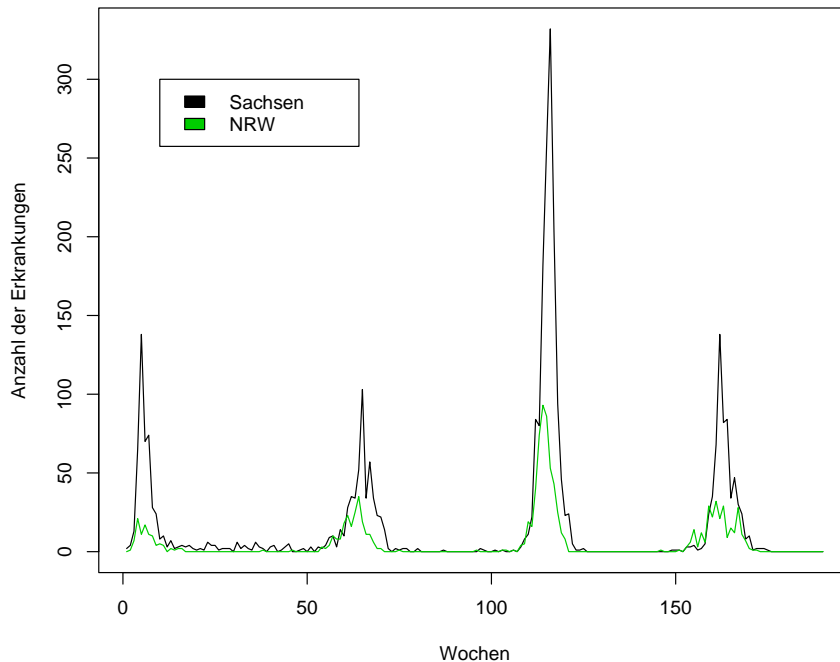


Abbildung 2.4: Influenzadaten im West-Ost Vergleich

anz dem Erwartungswert. Da dieser theoretisch unendlich sein kann, sollte eine varianzstabilisierende Transformation durchgeführt werden. Bei einer Wurzeltransformation beträgt die approximative Varianz ein Viertel. [5] Die Graphik 2.6 zeigt die übereinander geplotteten Verlaufskurven der transformierten Daten. Der gesuchte zeitliche Zusammenhang lässt sich lediglich im ausbruchsstärksten Jahr 2003 vermuten: Das Maximum der Meningokokkenerkrankungen folgt dem der Influenza mit einer zeitlichen Verzögerung von einer Woche.

Um die bereits erwähnte Saisonalität der Krankheiten näher zu analysieren, betrachten wir in 2.7 die übereinandergelegten jährlichen Verlaufskurven. Bei Influenza zeichnet sich eine sehr starke Saisonalität ab, die Ausbrüche liegen in etwa zwischen der ersten und 15ten Kalenderwoche. Der größte lässt sich für das Jahr 2003 beobachten. Zwischen der 20ten und 45ten Kalenderwoche treten nur vereinzelte Grippefälle auf.

Bei den Meningokokken ist die Saisonalität bedeutend geringer ausgeprägt. Der Unterschied in den Fallzahlen zwischen Winter- und Sommermonaten ist

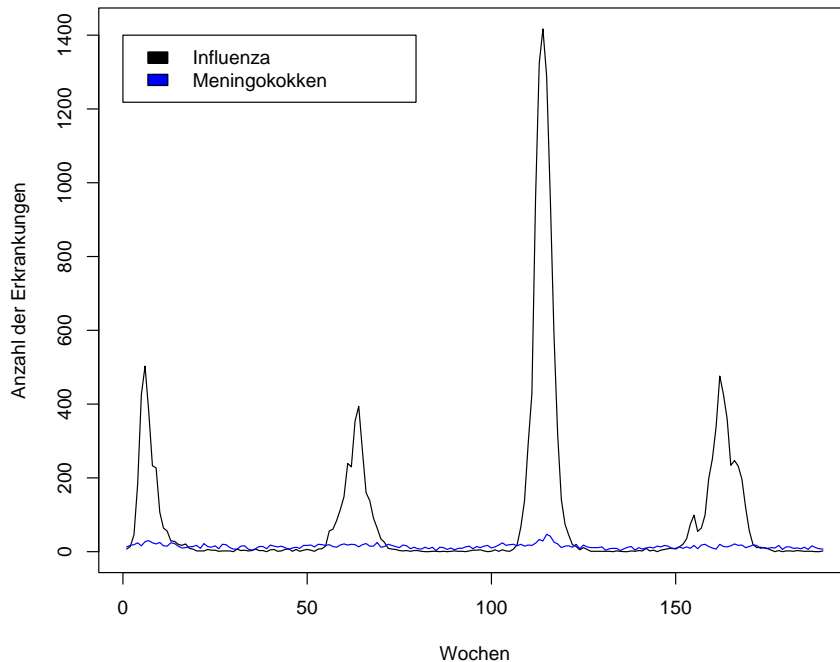


Abbildung 2.5: Vergleich Zeitreihen Influenza - Meningokokken

in der Relation wesentlich kleiner als bei den Grippefällen. Die vorhandene schwache Saisonalität besitzt allerdings in etwa die gleiche zeitliche Struktur wie die der Influenza. Allerdings bezieht sich diese Saisonalität vor allem auf die Jahre 2001 und 2003.

Vergleicht man die Meningokokken- und Influenzaverlaufskurven, so stellt man für die Jahre 2001 bis 2003 eine zeitliche Verzögerung der Meningokokkenspitzen gegenüber denen der Infuenza fest. Im Jahr 2001 liegt die Meningokokkenspitze in der 7ten Woche und läuft somit der der Infuenza eine Woche nach. Im darauffolgenden Jahr liegt der höchste Ausbruch bei Infuenza in der 12ten und bei Meningokokken in der 14ten und 17ten Woche. Im Jahr 2003 lässt sich eine zeitliche Verzögerung der Meningokokkenkurve in der 11ten Woche gegenüber der Infuenza in der 10ten Woche feststellen. Für das Jahr 2004 kann eine Verzögerung dieser Art nicht aus den Graphen abgelesen werden.

Betrachtet man nun die zeitliche Verzögerung bei Influenza- und Meningokokkenerkrankungen altersspezifisch (Abbildung 2.8), so lässt sich diese nur

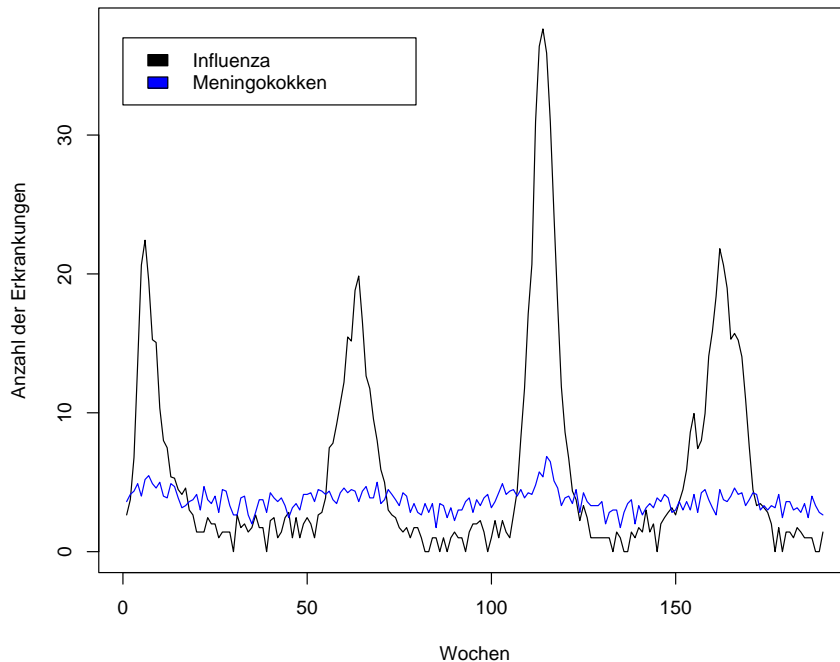


Abbildung 2.6: Vergleich Zeitreihen Influenza - Meningokokken, transformiert

bedingt feststellen. Nur bei der Alterskategorie der 0-Jährigen ist teilweise solch eine Verzögerung erkennbar. In den Jahren 2002 bzw. 2003 ist eine ein bis zwei wöchig nachlaufende Meningokokkenspitze zu sehen. 2004 sind die Spitzen in der gleichen Woche, 2001 ist eine Interpretation aufgrund der sehr geringen Fallzahlen nur schwer möglich.

Generell läßt sich in den anderen Alterskategorien die Verzögerung von ein bis zwei Wochen nur sehr selten erkennen. Am Beispiel der 15 bis 19-Jährigen, die speziell bei den Meningokokkenerkrankungen eine gefährdete Gruppe darstellen, kann man dies deutlich sehen. Nur das Jahr 2003 würde für die oben genannte These sprechen.

Als weitere Möglichkeit differenzierter Betrachtung bietet es sich an, die Meningokokkenfälle seperat mit den einzelnen Virentypen der Influenza zu vergleichen. Die Grafik 2.9 zeigt die Verlaufskurven der Influenzaerreger A, B, C, und AB jeweils zusammen mit der Gesamtverlaufskurve der Meningokokken. Eine zeitliche Verzögerung lässt sich lediglich im Jahr 2003 bei den Virentypen A bzw. AB erkennen. In den übrigen Jahren und bei den anderen

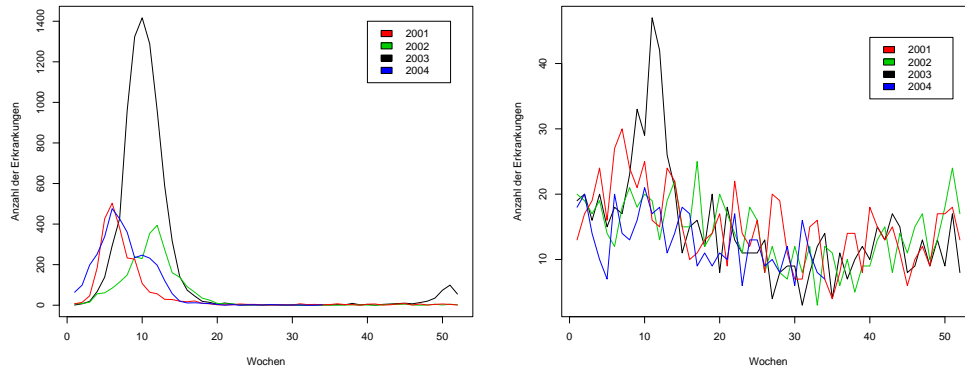


Abbildung 2.7: Zeitreihen Influenza, Meningokokken

Virentypen bestätigt sich die These graphisch nicht. Erwähnt sei hier noch, dass der Virentypus C im Jahr 2004 nicht erhoben wurde und im Datensatz als "missing value" dokumentiert ist. Außerdem ist bei den Grafiken erkennbar, dass der A-Virus im Jahr 2002 alleine bedeutend seltener auftrat, als in den übrigen Jahren. Zusammen mit dem B-Virus trat er kaum im Jahr 2001 auf, der C-Virus trat lediglich im Jahr 2003 verstärkt in Erscheinung.

Im folgenden wurden aufgrund der starken Saisonalität der Influenzawerte die bereits wurzeltransformierten Daten saisonbereinigt. Dabei wurden die Daten mit Hilfe von gleitenden Durchschnitts mit einem Zeitfenster von sechs Wochen geglättet und als transformierte Werte die Differenzen dieser Kurve zur Originalkurve betrachtet. Die folgende Formel zeigt die Funktion, mit deren Hilfe die Daten geglättet wurden.

$$\bar{X}_t = \frac{1}{2}X_{t-3} + \sum_{i=-2}^2 X_{t+i} + \frac{1}{2}X_{t+3} \quad t = 4, 5, \dots, 187 \quad (2.1)$$

Diese Saisonbereinigung wird sichtbar in den beiden Graphen aus Bild 2.10. Es sind bei den Meningokokkenwerten kaum mehr regelmäßige und somit von der Saison beeinflusste Schwankungen erkennbar. Auch die Fallzahlen der Influenza wurden bereinigt, allerdings zeigt sich, dass der Einfluss der Saison nicht völlig verschwunden ist. Im Großen und Ganzen stellt das beschriebene Verfahren aber eine hinreichend gute Saisonbereinigung dar.

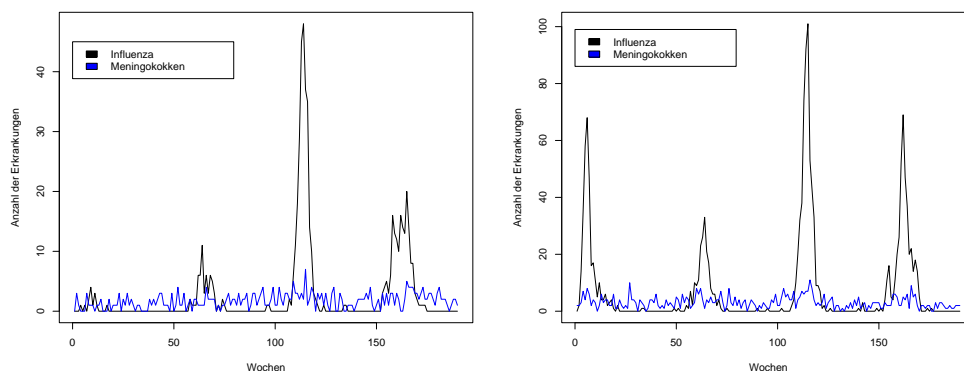


Abbildung 2.8: Zeitreihen der "Alterskategorie 0" bzw. "Alterskategorie 15 bis 19"

## 2.3 Analyse der Daten mit Hilfe von Autokorrelationsfunktionen

Zur tiefergehenden Untersuchung eines zeitlichen Zusammenhanges haben wir im Folgenden die Autokorrelationsfunktion benutzt:

$$\gamma(h) = \text{cov}(X_{t+h}, X_t) \quad (2.2)$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(h)}{\sigma^2}, \quad h = 0, 1, 2, \dots \quad (2.3)$$

Hierbei bezeichnet  $\gamma(h)$  die Kovarianzfunktion. Sie mißt den Zusammenhang zwischen  $X_t$  und  $X_{t+h}$ .  $\rho(h)$  ist somit die auf  $[-1; 1]$  standardisierte Kovarianzfunktion. Sie wird (Auto-)Korrelationsfunktion genannt.

Wird der Zusammenhang zweier verschiedener Zeitreihen  $X$  und  $Y$  untersucht, misst  $\gamma(h)$  die Kovarianz zwischen  $X_t$  und  $Y_{t+h}$ . Dementsprechend ist hier  $\rho(h)$  die Kreuzkorrelation dieser Reihen. Als natürliche Schätzer für die Kovarianzfunktionen verwendet man:

$$g_{xx}(k) = \frac{1}{n} \sum_{t=k+1}^n (x_t - \bar{x})(x_{t-k} - \bar{x}) \quad (2.4)$$

$$g_{xy}(k) = \frac{1}{n} \sum_{t=k+1}^n (x_t - \bar{x})(y_{t-k} - \bar{y}), \quad k \geq 0 \quad (2.5)$$

Die natürlichen Schätzer  $r(k)$  der Autokorrelationsfunktion bzw. Kreuzkorrelationsfunktion sind somit die standardisierten Werte dieser Funktionen.

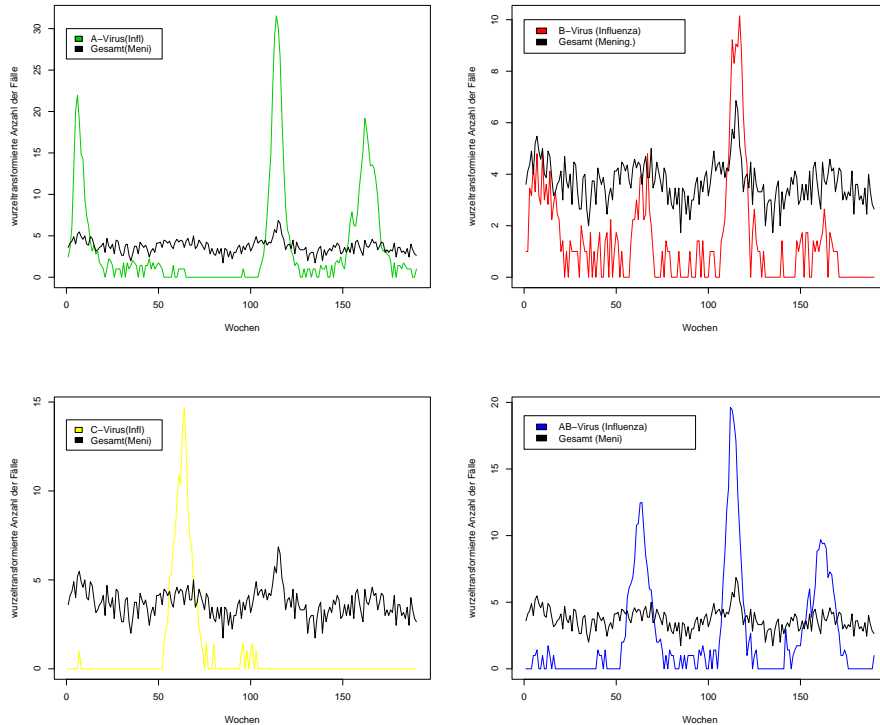


Abbildung 2.9: Zeitreihe Influenza, Meningokokken nach den Virentypen A, B, C und AB

Für große  $k$  und damit für große Zeitdifferenzen (lags) strebt der Wert der beiden Funktionen  $g_{xx}(k)$  und  $g_{xy}(k)$  gegen 0.

Die folgende Abbildung 2.11 zeigt die Auto- und Kreuzkorrelationen der wurzeltransformierten und nicht saisonbereinigten Daten.

Deutlich zu sehen sind die saisonalen Schwankungen der Influenza. Werte im Abstand von etwa einem Jahr scheinen einen deutlichen positiven Zusammenhang zu besitzen. Zurückzuführen ist das auf die jährlich zum gleichen Zeitpunkt auftretenden Erkrankungswellen.

Auch bei den Werten der Meningokokken im Bild rechts unten ist eine jährliche Struktur zu erkennen, die hier um einiges deutlicher zu Tage tritt als bei den übereinandergeplotteten Zeitreihen.

Für die Suche nach einem zeitlichen Zusammenhang, bei dem einer Influenzawelle ein Anstieg der Meningokokkenfälle folgt, ist der Graph links unten von Bedeutung. Hier wird die Kreuzkorrelation berechnet, wobei  $X_t$  die Meningokokken- und  $Y_{t+h}$  die Influenzawerte sind. Durch die negativen Wer-

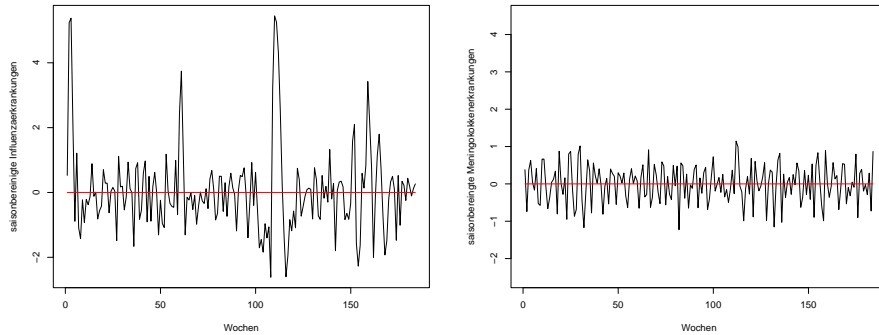


Abbildung 2.10: Saisonbereinigte Influenza- und Meningokokkendaten

te von  $h$  werden die Korrelationen zwischen vorauslaufenden Grippe- und nachfolgenden Meningokokkendaten berechnet.

Auch hier ist aufgrund der sehr starken Saisonkomponente eine Interpretation im Hinblick auf die zu untersuchende Fragestellung nicht möglich. Mögliche Effekte werden durch die Saisonschwankungen überdeckt. Das linke obere Schaubild, bei dem die Korrelationen zwischen Meningokokken- und nachlaufenden Influenzawerten abgebildet werden, ist ebenfalls aus diesem Grund nicht interpretierbar.

Deswegen werden in der folgenden Abbildung 2.12 die Auto- und Kreuzkorrelationen für die saisonbereinigten Daten betrachtet.

Zunächst ist beim Autokorrelationsdiagramm der Influenza links oben der Saisoneinfluss wesentlich schwächer ausgeprägt. Allerdings sind noch Saisonkomponenten erkennbar: So sind beispielsweise die Daten für den lag eins deutlich positiv miteinander korreliert und auch bei einem Abstand von ein oder zwei Jahren zeigen sich saisontypische, wenn auch kleinere, Spitzen der Korrelationswerte.

Im Schaubild der Autokorrelationswerte für die Meningokokken zeigt sich, dass die Daten wesentlich besser saisonbereinigt wurden, sie pendeln ohne erkennbare Saisonstruktur um die Nulllinie.

In den Graphen der Kreuzkorrelationen sind für unsere Fragestellung die lags zwischen  $-3$  und  $0$ , sowie  $0$  und  $3$  interessant. In den Werten wird dabei eine Asymmetrie für die negativen bzw. positiven lags deutlich: Im linken unteren Bild, bei dem die Meningokokkendaten denen der Influenza nachlaufen, liegen die betreffenden Korrelationen im positiven, wenn auch nicht signifikanten Bereich. Sie wären signifikant, wenn sie über der eingezeichneten 95% Konfidenzgrenze liegen würden. Folgen die Werte der Influenza denen

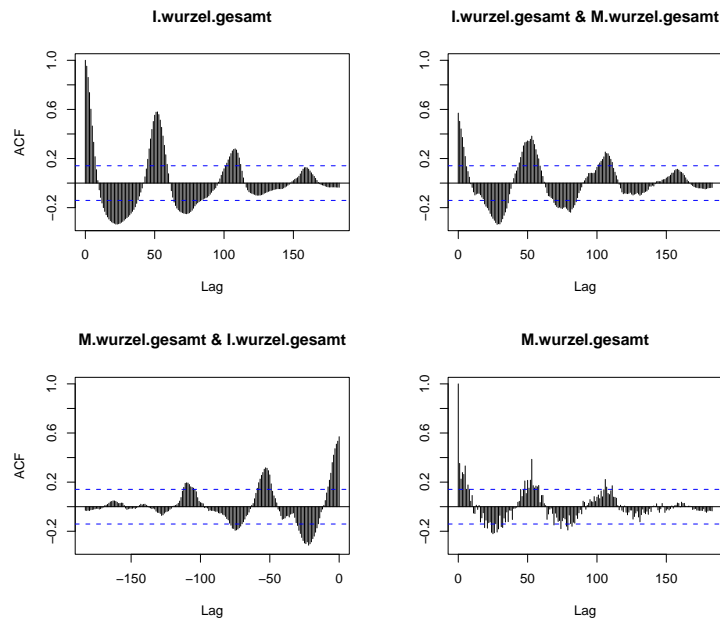


Abbildung 2.11: Auto- und Kreuzkorrelationen bei nicht saisonbereinigten Daten

der Meningokokken (rechts oben), so liegen die Werte für die lags eins bis drei sogar leicht im negativen, wenn auch hier nicht signifikanten Bereich. Es wird also bei nachlaufenden Meningokokkendenaten ein positiver, während bei vorauslaufenden ein negativer Zusammenhang mit der Influenza angedeutet. Im Großen und Ganzen entsprechen die gefundenen Strukturen unserer Hypothese: Für vorauslaufende Influenza- und nachfolgende Meningokokkenwerte werden positive Korrelationen, also Zusammenhänge, berechnet, während sich bei umgedrehten Voraussetzungen ein völlig anderes Bild ergibt, bei dem die Korrelationswerte sogar ins Negative gehen.

### Zusammenfassung der deskriptiven Untersuchungen

Die Untersuchungshypothese, ob einer erhöhten Influenzaausbruchsrates in einem Abstand von ein bis drei Wochen eine ebenfalls erhöhte Anzahl an Meningokokkenerkrankungen folgt, hat sich deskriptiv nur ansatzweise bestätigt. In den Zeitreihendiagrammen wurde der vermutete Zusammenhang quer durch alle Untergruppen vor allem im ausbruchsstärksten Jahr 2003 deutlich. Für die anderen Jahren bestätigte sich die Vermutung nur teilweise. Deutlichere Resultate ergaben sich bei der Auto- bzw. Kreuzkorrelation, bei

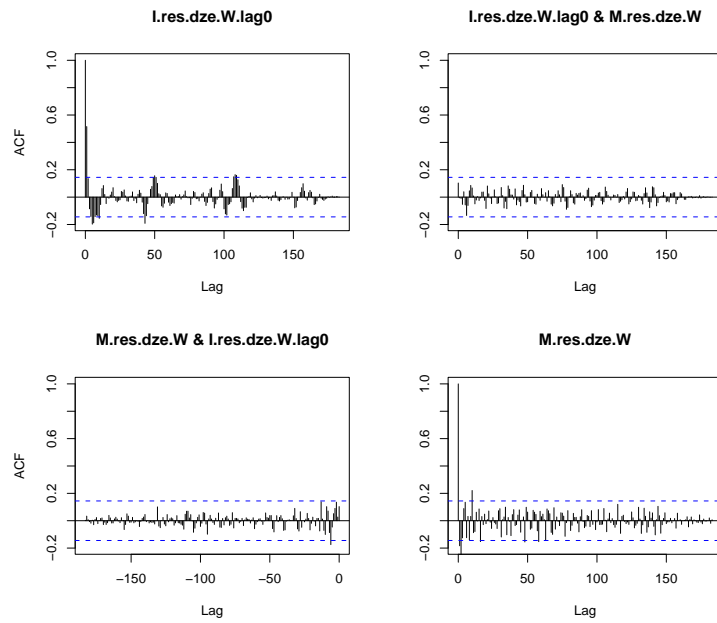


Abbildung 2.12: Auto- und Kreuzkorrelationen bei saisonbereinigten Daten

der sich eine Asymmetrie in den Werten zeigte, die unserer Hypothese entspricht. Allerdings sind diese Ergebnisse nicht signifikant. Der rein deskriptiven Betrachtung der Daten soll im Folgenden eine tieferegehende statistische Untersuchung mit Hilfe von Regressionsmodellen folgen.

# Kapitel 3

## Regressionsanalyse

### 3.1 Lineare Modelle

Als geeignetes Regressionsmodell bietet sich zunächst das Modell der multiplen linearen Regression an. Beschreibe  $Y$  die Anzahl der Meningokokkenkrankungen und  $X$  die Anzahl der gemeldeten Influenzafälle, so empfiehlt sich folgender Ansatz:

$$Y_t = \alpha_0 + \sum_{k=1}^n \{\beta_k \cos(\omega_k t) + \gamma_k \sin(\omega_k t)\} + \delta X_{t-l} + \epsilon_t \quad (3.1)$$

Hierbei beschreibt die Summe aus Cosinus- und Sinustermen die Saisonkomponente,  $\omega_k$  sei  $\frac{2k\pi}{T}$  mit Periodendauer  $T$  ( $=52$ ) und  $\epsilon_t$  widerspiegelt einen unabhängigen und normalverteilten Fehler mit Erwartungswert 0. Für das hier vorliegende Problem genügt es, sich auf  $n=4$  zu beschränken, da sich nahezu sicher eine größere Anzahl an Sinus- und Cosinustermen nicht mehr signifikant auswirkt. Die zu schätzenden Parameter des Modells sind  $\alpha_0$  (Intercept),  $\beta_k$  und  $\gamma_k$  (Saisonkomponente) und  $\delta$  (Einfluss von Influenza).

#### 3.1.1 Lineares Modell mit nicht-transformierten Meningokokkendaten

In der folgenden Tabelle 3.1 sind die Ergebnisse des linearen Regressionsmodells mit nicht-transformierten Meningokokkenwerten dargestellt. Sie sollen anhand der saisonbereinigten Influenzadaten erklärt werden. Für die Saisonbereinigung wurden, wie bereits beschrieben, die Differenzen zwischen

den wurzeltransformierten Originaldaten und einer mit Hilfe von gleitenden Durchschnitten geglätteten Kurve genommen. Eine positive zeitliche Verschiebung (lag) beschreibt dabei ein Vorauslaufen der Influenzadaten gegenüber denen der Meningokokken. Zunächst wird aufgrund unserer Fragestellung ein lag von plus zwei Wochen näher untersucht.

Dabei wird deutlich, dass lediglich im Modell mit nur einer Sinus- und Cosi-

	Wert	Standardfehler	t-Wert	$P(>  t )$		Wert	Standardfehler	t-Wert	$P(>  t )$
<i>Intercept</i>	14.0803	0.3572	39.4235	0.0000	<i>Intercept</i>	14.0749	0.3580	39.3153	0.0000
$\sin(\frac{2\pi t}{T})$	3.3719	0.5074	6.6452	0.0000	$\sin(\frac{2\pi t}{T})$	3.3689	0.5086	6.6237	0.0000
$\cos(\frac{2\pi t}{T})$	3.4040	0.5061	6.7256	0.0000	$\cos(\frac{2\pi t}{T})$	3.3921	0.5069	6.6916	0.0000
$\sin(\frac{4\pi t}{T})$	0.6581	0.5145	1.2791	0.2026	$\sin(\frac{4\pi t}{T})$	0.6238	0.5153	1.2106	0.2277
$\cos(\frac{4\pi t}{T})$	-0.3512	0.5082	-0.6911	0.4904	$\cos(\frac{4\pi t}{T})$	-0.3385	0.5093	-0.6645	0.5072
$\sin(\frac{6\pi t}{T})$	0.6571	0.5047	1.3020	0.1947	$\sin(\frac{6\pi t}{T})$	0.6881	0.5031	1.3677	0.1732
$\cos(\frac{6\pi t}{T})$	-0.3899	0.5207	-0.7487	0.4551	$\cos(\frac{6\pi t}{T})$	-0.4126	0.5210	-0.7920	0.4294
$\sin(\frac{8\pi t}{T})$	0.0172	0.5049	0.0340	0.9729	<i>Influenza</i>	1.2548	0.2858	4.3898	0.0000
$\cos(\frac{8\pi t}{T})$	-0.8591	0.5100	-1.6845	0.0939					
<i>Influenza</i>	1.1688	0.2905	4.0230	0.0001					

	Wert	Standardfehler	t-Wert	$P(>  t )$		Wert	Standardfehler	t-Wert	$P(>  t )$
<i>Intercept</i>	14.0911	0.3583	39.3310	0.0000	<i>Intercept</i>	14.0894	0.3580	39.3599	0.0000
$\sin(\frac{2\pi t}{T})$	3.3632	0.5092	6.6049	0.0000	$\sin(\frac{2\pi t}{T})$	3.3623	0.5075	6.6252	0.0000
$\cos(\frac{2\pi t}{T})$	3.3896	0.5075	6.6789	0.0000	$\cos(\frac{2\pi t}{T})$	3.4498	0.5048	6.8336	0.0000
$\sin(\frac{4\pi t}{T})$	0.6276	0.5155	1.2175	0.2250	<i>Influenza</i>	1.3867	0.2699	5.1379	0.0000
$\cos(\frac{4\pi t}{T})$	-0.2566	0.5072	-0.5059	0.6135					
<i>Influenza</i>	1.3048	0.2773	4.7051	0.0000					

Tabelle 3.1: geschätzte Parameter im linearen Modell (lag 2, Meningokokkendaten nicht-transformiert, Influenzawerte saisonbereinigt)

nusschwingung die geschätzten Parameter signifikant sind. Wenn weitere Sinus- und Cosinusparameter ins Modell einfließen, stellen sie aufgrund fehlender Signifikanz keine Verbesserung des Modells dar. Auch der Wert der F-Statistik sinkt mit steigender Anzahl von Parametern von 44.06 auf 15.56 bei vier Sinus-Cosinustermen. Für alle weiteren Regressionsmodelle, lineare als auch generalisierte, waren ebenfalls nur der erste Sinus- bzw. Cosinusparameter signifikant, sodass wir uns in der Folge in diesem Bericht auf Modelle mit nur jeweils einem Schwingungsparameter beschränken.

Somit sind die Werte der Tabelle links unten für unsere Fragestellung von Bedeutung, wobei der am meisten interessierende Wert, der Einfluss der Influenzadaten ist. Mit einem Wert von 1.3867 und einer Standardabweichung

von 0.2699 befindet sich der Parameter auch eindeutig im positiven Bereich. Die zu untersuchende Hypothese wird durch diesen Wert also bestätigt: Es scheint einen positiven Zusammenhang bei einem lag von zwei Wochen zu geben; bei einem saisonbereinigten Anstieg der wurzeltransformierten Influenzawerte um eins erhöht sich laut diesem Modell somit die Anzahl der Meningokokkenwerte zwei Wochen später um etwa 1.4. Mit einem t-Wert von 5.1379 ist der geschätzte Parameter hochsignifikant.

Als Gütekriterium für dieses Modell liefert der S-Plus Output ein multiples  $R^2$  von 0.4261. Der Anteil der erklärten Streuung an der Gesamtstreuung beträgt somit 42.61%.

Um den zeitlichen Zusammenhang genauer zu untersuchen, wurde das oben beschriebene Regressionsmodell mit verschiedenen lags von 4 bis -4 berechnet. In der nachfolgenden Tabelle 3.2 sind die entscheidenden Werte dieser Modelle aufgeführt.

Die zu untersuchende Hypothese wird durch diese Werte bestätigt: den stärk-

	Parameter	Standard Fehler	t-Wert	p-Wert	$R^2$	F-Statistik
lag 4	0.4679	0.2830	1.6536	0.1000	0.3383	29.99
lag 3	1.0056	0.2780	3.6177	0.0004	0.3806	36.25
lag 2	1.3867	0.2699	5.1379	0.0000	0.4267	44.06
lag 1	1.1895	0.2746	4.3322	0.0000	0.4027	40.24
lag 0	1.1909	0.2746	4.3367	0.0000	0.4066	41.12
lag -1	0.5471	0.2852	1.9186	0.0566	0.3636	34.09
lag -2	0.1830	0.3004	0.6092	0.5432	0.3473	31.57
lag -3	-0.4211	0.3150	-1.3368	0.1830	0.3525	32.12
lag -4	-0.4633	0.3176	-1.4586	0.1464	0.3508	31.70

Tabelle 3.2: Werte des linearen Regressionsmodells mit nicht-transformierten Meningokokkenwerten bei verschiedenen lags

sten und signifikantesten Zusammenhang gibt es im Bereich von null bis drei Wochen. Die Anzahl der Influenzaerkrankungen scheint für diese lags also einen positiven Einfluss auf die der Meningokokken zu haben, während für die übrigen Abstände dieser Einfluss immer geringer bzw. weniger signifikant wird. Für einen lag von -3 bzw. -4, was ein Vorauslaufen der Meningokokkendaten gegenüber denen der Influenza um drei bzw. vier Wochen bedeutet, dreht sich gar die Richtung des Einflusses um, was einen negativen Zusammenhang suggerieren würde. Allerdings sind die Werte im negativen Bereich in keinem Fall signifikant. Für die positiven lags gilt: ab einer zeitlichen Ver-

schiebung um mehr als zwei Wochen werden die geschätzten Parameter wieder niedriger und weniger signifikant.

Die Standardabweichungen unterscheiden sich nur geringfügig, zeigen aber das gleiche Muster wie die Parameterwerte, was bedeutet, dass sie für die Wochen null bis zwei am geringsten sind (diese Werte lassen sich somit am genauesten schätzen) und sich für die anderen lags vergrößern, was die mangelnde Sicherheit der jeweiligen geschätzten Parameter zeigt.

Derselbe Trend zeigt sich in den Werten des  $R^2$  und der F-Statistik. Maximale Erklärungskraft besitzt das Modell mit einer zeitlichen Verzögerung von zwei Wochen. Ab einem lag von drei nach oben bzw. -1 nach unten nimmt die Güte des Modells deutlich ab.

Vergleicht man die t-Werte der unterschiedlichen lags separat in einer Graphik (3.1), so sieht man sehr deutlich den beschriebenen Trend: Die hochsignifikanten Werte liegen bei den lags von null bis drei mit dem Maximum bei zwei. Bei einem Vorlaufen der Influenzawerte um vier Wochen kann man nicht mehr von einem signifikanten Einfluss dieser Daten sprechen (die blauen Geraden stellen die Schranken für die 95 % Signifikanz dar).

Lässt man die Meningokokkenwerte vorauslaufen, so fallen die t-Werte mit größer werdendem negativen lag stark ab und werden ab einem lag von -3 negativ.

Unter der Annahme, dass kein zeitlich-verzögerter Zusammenhang zwischen Influenza- und Meningokokkenerkrankungen bestünde, wäre ein um den lag null symmetrischer Graph zu erwarten. Bei diesem Schaubild kann man also sagen, dass das lineare Regressionsmodell unsere These des zeitlichen Zusammenhangs zwischen beiden Krankheiten unterstützt.

## **Bewertung des Modells**

Um die Güte des Modells bewerten zu können, soll zunächst in der Abbildung 3.2 die Anpassung des Modells an die Originaldaten betrachtet werden. Im Streudiagramm wurden diese gegen die von unserem Modell mit einem lag von zwei geschätzten Daten geplottet. Eine perfekte Modellierung wäre gegeben, wenn die Punkte alle auf der Referenzlinie liegen würden. Wie der Wert des Gütekriteriums  $R^2$  (0.4261) bereits vermuten ließ, ist diese Anpassung auch graphisch nur auf mittlerem Niveau. Ausserdem streuen die Werte speziell bei größeren gefitteten Werten tendenziell etwas mehr um die Referenzlinie. Im Besonderen fallen drei Ausreisser oberhalb der Geraden auf. Die Modellierung der Daten scheint also noch verbesserungswürdig.

Desweiteren muss die Annahme der Unabhängigkeit und der Normalverteilung der Residuen untersucht werden. Die nachfolgenden Graphiken aus Schaubild 3.3 zeigen den Plot der standardisierten Residuen und den Q-Q-

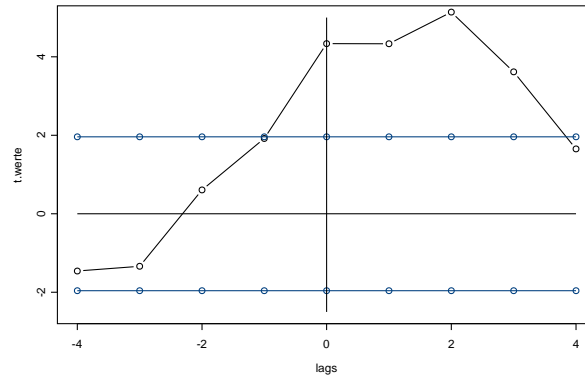


Abbildung 3.1: t-Werte für verschiedene lags beim linearen Modell mit nicht-transformierten Meningokokkdaten

Plot wiederum bei einem lag von zwei.

Der Q-Q-Plot im linken Bild ist ein graphisches Mittel, um die Normalverteilungsannahme zu untersuchen. Es werden die Quantile der Standardnormalverteilung gegen die Residuenwerte geplottet. Die Residuen wären normalverteilt, wenn alle Punkte näherungsweise auf der Referenzgeraden liegen würden. Bis auf drei Ausreisser nach oben scheint dies hier der Fall zu sein. Gegen Ende liegen die Residuen jedoch tendenziell eher oberhalb der Referenzlinie. Dennoch bestätigt dieser Graph wohl im Großen und Ganzen die Annahme normalverteilter Residuen.

In der zweiten Graphik sieht man die Streuung der standardisierten Residualwerte. Bei kleineren gefitteten Werten ist keine deutliche Struktur in den Residuen zu erkennen. Sie schwanken in etwa im Intervall  $[-2, 2]$ . Bei größeren prognostizierten Werten scheinen sie jedoch auch eine größere Streuung zu besitzen. Somit wäre die Modellannahme der Homoskedastizität der Residuen verletzt, die Modellaussagen verlieren damit an Glaubwürdigkeit. Aufgrund dessen scheint es im Weiteren sinnvoll, die Response-Daten geeignet zu transformieren, um dieses Problem auszuschalten. Wie bereits in Kapitel 2.2 erwähnt, stabilisiert eine Wurzeltransformation die Varianz bei poissonverteilten Daten. In der Folge betrachten wir somit das obige lineare Regressionsmodell mit wurzeltransformierten Meningokokkdaten.

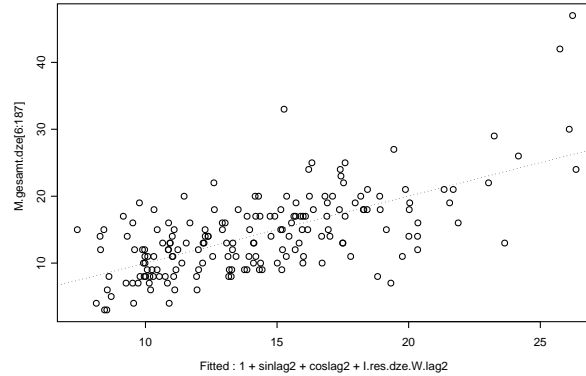


Abbildung 3.2: Güte der Modellierung

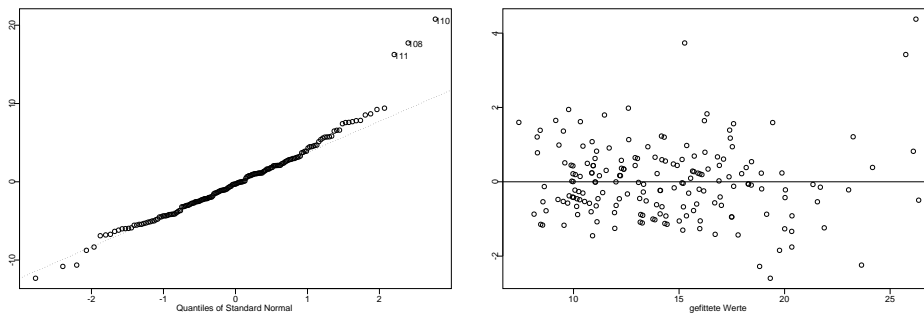


Abbildung 3.3: Plot der standardisierten Residuen und Q-Q-Plot mit lag 2 im linearen Modell mit nicht-wurzeltransformierten Meningokokkendenaten

### 3.1.2 Lineares Modell mit wurzeltransformierten Meningokokkendenaten

Die folgende Tabelle 3.3 ist wie im vorherigen Kapitel eine Zusammenfassung der für unsere Fragestellung interessanten Werte über die lags von -4 bis 4.

Es ergeben sich keine wesentlichen Unterschiede zu den Aussagen des vorherigen Modells. Wiederum ist für die lags null bis drei ein positiver signifikanter Zusammenhang zwischen Influenza und Meningokokken zu erkennen. Allerdings sind die t-Werte nicht mehr so hoch wie beim Modell mit den nichttransformierten Meningokokkendenaten. Bei den übrigen lags fällt der prognostizierte Wert wiederum stark ab und ist hier in keinem Fall signifikant. Bei einem Wochenunterschied von -3 bzw. -4 widerspiegelt er erneut einen

Parameter	Standard Fehler	t-Wert	p-Wert	$R^2$	F-Statistik
lag 4	0.0609	1.7228	0.0867	0.3649	33.71
lag 3	0.1050	3.0012	0.0031	0.3918	38.01
lag 2	0.1428	4.1580	0.0000	0.4229	43.48
lag 1	0.1011	2.8821	0.0044	0.3947	38.90
lag 0	0.1156	3.3199	0.0011	0.4073	47.22
lag -1	0.0460	1.2929	0.1977	0.3842	37.22
lag -2	0.0156	0.4203	0.6748	0.3738	35.43
lag -3	-0.0432	-1.1047	0.2708	0.3782	35.89
lag -4	-0.0307	-0.7761	0.4387	0.3728	34.87

Tabelle 3.3: Werte des Regressionsmodells mit wurzeltransformierten Meningokokkdaten bei verschiedenen lags

negativen Einfluss der Influenza.

Wie in Graphik 3.1 zeigt auch das Schaubild 3.4 die t-Werte für die verschiedenen lags. Die Asymmetrie um den lag null lässt sich auch hier gut erkennen. Die Bewertung des Graphen bleibt die gleiche wie beim vorherigen Modell. Somit wird die Hypothese des zeitlich-kausalen Zusammenhangs erneut verstärkt.

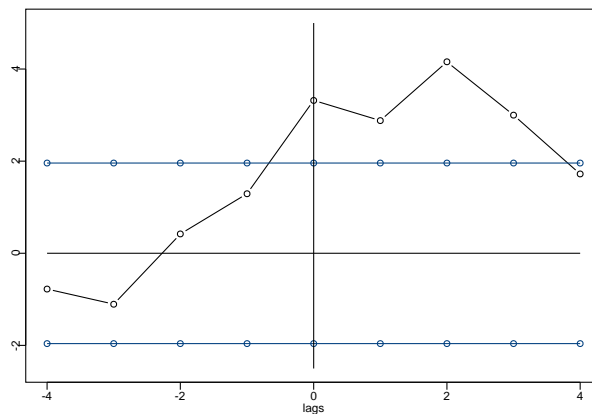


Abbildung 3.4: t-Werte für verschiedene lags beim linearen Modell mit wurzeltransformierten Meningokokkdaten

## Bewertung des Modells

Zur graphischen Bewertung der Güte des hier verwendeten Modells dient wie beim Modell ohne wurzeltransformierte Meningokokkdaten ein Streudiagramm, bei dem die gefitteten Daten den originalen gegenübergeplottet werden.

Wie in Graph 3.5, der wiederum das Modell mit einem lag von zwei zeigt, gut zu sehen ist, sind die Ausreisser nach oben bei größeren gefitteten Werten nicht mehr so deutlich wie noch beim Modell mit nichttransformierten Meningokokkdaten. Die Werte streuen gleichmäßiger um die Referenzlinie, was eine bessere Anpassung des Modells an die Daten vermuten lässt.

Zur Betrachtung der Residuen dienen in Anlehnung an die Graphen aus Ab-

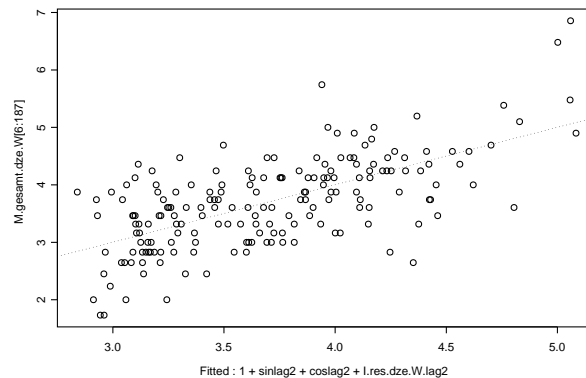


Abbildung 3.5: Güte der Modellierung

bildung 3.3 die folgenden Darstellungen 3.6. Beim Q-Q-Plot im linken Bild wird die Verbesserung des Modells durch die Transformation der Meningokokkdaten gut deutlich. Die drei Ausreisser nach oben sind nicht mehr ganz so extrem und auch im Allgemeinen liegen die Daten gegen Ende nicht mehr so deutlich über der Referenzlinie. Insgesamt scheint eine gute Annäherung an die Werte der Normalverteilung gelungen zu sein.

Die Systematik bei der Streuung der Residuenwerte ist im rechten Graph im Vergleich zu dem des vorhergehenden Modells nahezu verschwunden. Die Werte streuen ohne starken Trend um die Nulllinie, die Ausreisser nach oben sind auch hier wesentlich schwächer.

Insgesamt lässt sich eine deutliche Verbesserung gegenüber dem Modell mit nicht-transformierten Meningokokkdaten erkennen. Auch wenn sich an den Ergebnissen der Parameterschätzungen und somit an den Aussagen über die

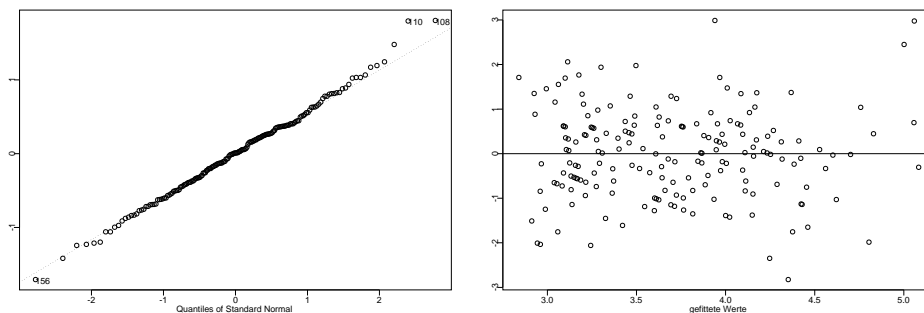


Abbildung 3.6: Plot der standardisierten Residuen und Q-Q-Plot mit lag 2 im linearen Modell mit wurzeltransformierten Meningokokkdaten

Arbeitshypothese nichts Wesentliches ändert, scheint dieses Modell doch um einiges besser geeignet, die Daten zu beschreiben und einen Einfluss der Influenzawerte auf die nachfolgenden Meningokokkenerkrankungen nachzuweisen.

### 3.1.3 Lineares Modell mit nicht saisonbereinigten Influenzadaten

Im folgenden soll der Vollständigkeit halber noch das lineare Modell mit lediglich wurzeltransformierten und somit nicht saisonbereinigten Influenzadaten betrachtet werden. Als Response dienen wiederum die wurzeltransformierten Meningokokkenwerte.

Zunächst sei erwähnt, dass die geschätzten Parameter für die Sinus- und Cosinuswerte zum großen Teil von Beginn an nicht mehr signifikant sind. Die modellierte Saisonkomponente spielt bei nicht saisonbereinigten Daten keine Rolle mehr.

Die t-Werte für die verschiedenen lags in Graphik 3.7 zeigen wie schon in den beiden vorhergehenden Modellen den gesuchten Zusammenhang: Bei einem positiven lag, sprich bei einem Vorlaufen der Influenzafälle, sind die t-Werte bedeutend größer als bei den negativen lags. Allerdings liegt hier die Spitze der t-Werte bei einem Wochenunterschied von einer Woche, anschließend fallen sie kontinuierlich nach links und rechts ab. Auch gehen sie bei den negativen lags nicht in den negativen Bereich, die geschätzten Einflussparameter sind also immer positiv. Generell sind die t-Werte größer als die der vorhergehenden Modelle und auch für die lags -1 und -2 signifikant.

Die Aussagekraft dieses Modells wird allerdings stark von der Annahme beeinflusst, dass der Zusammenhang zwischen den beiden Krankheiten nicht

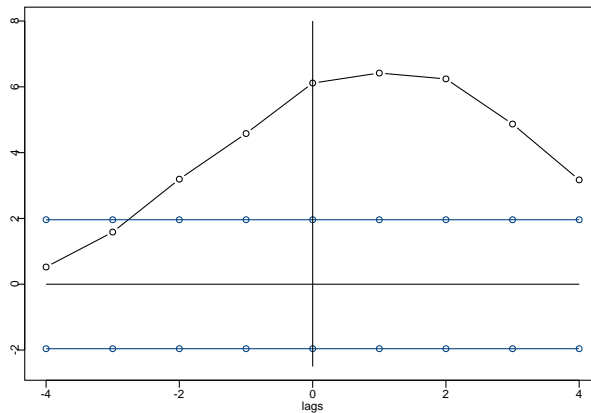


Abbildung 3.7: t-Werte für verschiedene lags beim linearen Modell mit wurzeltransformierten Meningokokken- und nicht saisonbereinigten Influenzadaten

darauf zurückzuführen ist, dass sie beide den gleichen Saisonverläufen folgen. Es wäre theoretisch möglich, dass beide Krankheitstypen unabhängig voneinander dieselbe saisonale Struktur besitzen, was auf scheinbare, tatsächlich aber nicht vorhandene Zusammenhänge schließen lassen würde.

Im Modell mit den saisonbereinigten Daten wird das im Prinzip ausgeschlossen: Lediglich saisonuntypische Schwankungen der Influenzaraten gehen in die erklärende Variable ein. Folgen auf derartige Erhöhungen ebenfalls erhöhte Meningokokkenwerte, so kann mit größerer Wahrscheinlichkeit von einem zeitlich-kausalen Zusammenhang gesprochen werden.

Das Modell mit nicht saisonbereinigten Influenzawerten sei somit nur der Vollständigkeit halber erwähnt. Bei der Interpretation der Daten stützen wir uns lediglich auf die Modelle, in denen die Saisonkomponente eigens berechnet wird.

### 3.1.4 Lineares Modell mit Meningokokken als Einflussgröße

Zur weiteren Ergänzung sei noch das lineare Regressionsmodell erwähnt, in dem Regressor und Regressand vertauscht wurden. Als Einflussgröße gelten hier nun also die Daten der Meningokokken, die die der Influenza erklären sollen. Ziel hierbei soll es sein, die asymmetrischen t-Werte-Plots der vorangegangenen Plots zu validieren. Ein in die andere Richtung asymmetrischer Graph der t-Werte bei diesem Modell würde die vorhergehenden Ergebnisse

bestätigen, ein anderes Muster neue Fragen aufwerfen. Verwendet wurden die wurzeltransformierten Grippedaten und die saisonbereinigten Werte der ebenfalls wurzeltransformierten Meningokokkenfälle.

Im dargestellten Graphen 3.8 zeigt sich kein überraschendes Bild: Die Asym-

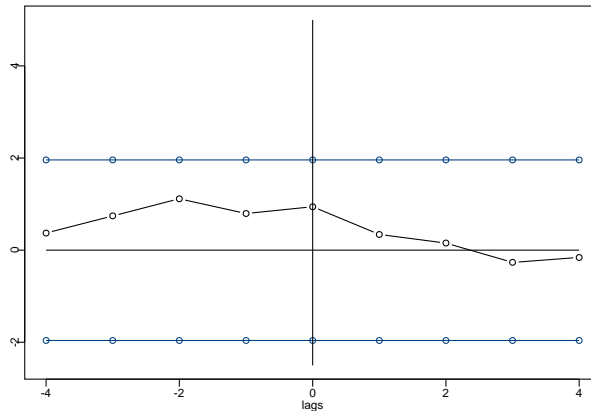


Abbildung 3.8: t-Werte für verschiedene lags beim linearen Modell mit vertauschten Regressor und Regressand

metrie geht dieses Mal in die andere Richtung, wobei die t-Werte generell deutlich niedriger sind und nie im signifikanten Bereich liegen. Dies lässt sich wohl auf die niedrigen Regressorwerte (Meningokokken) zurückführen, die die weitaus höheren Influenzawerte erklären sollen.

## 3.2 Generalisierte lineare Modelle

### 3.2.1 Poissonregression

Im folgenden Abschnitt sollen geeignete generalisierte lineare Modelle zur Beschreibung der zeitlichen Verzögerung betrachtet werden. Da Zähldaten sehr häufig mit Hilfe der Poissonverteilung modelliert werden, bietet sich ein Poisson-Regressionsansatz an:

$$E(Y_t) = \exp\left(\alpha_0 + \sum_{k=1}^n \{\beta_k \cos(\omega_k t) + \gamma_k \sin(\omega_k t)\} + \delta X_{t-l}\right) \quad (3.2)$$

Y steht hierbei für die nicht-transformierte Anzahl der Meningokokkenerkrankungen und X für die Menge der saisonbereinigten und wurzeltransformierten Influenzafälle.

In der anschließenden Tabelle 3.4 sind die verschiedenen errechneten Parameter des Influenzaeinflusses für alle lags aufgeführt.

Die bereits bei den linearen Modellen aufgefallene Struktur wird auch hier

	Parameter	t-Wert	Devianz	Freiheitsgrade	AIC
lag 4	0.025	1.86	277.52	176	285.52
lag 3	0.052	4.07	269.80	177	277.80
lag 2	0.070	5.60	258.36	178	266.36
lag 1	0.060	4.75	267.26	179	275.26
lag 0	0.060	4.74	268.72	180	276.72
lag -1	0.027	2.04	282.88	179	290.88
lag -2	0.007	0.50	284.94	178	292.94
lag -3	-0.028	-1.84	281.42	177	289.42
lag -4	-0.030	-1.96	280.33	176	288.33

Tabelle 3.4: geschätzter Parameter, t-Wert, Devianz und AIC bei verschiedenen lags

wieder deutlich: Den größten positiven Einfluss gibt es bei einem Unterschied von zwei Wochen, dieser ist mit einem t-Wert von 5.60 auch am signifikantesten. Bei einem lag von -3 und -4 scheint sich der Einfluss sogar wieder ins Negative umzukehren. Allerdings sind diese Werte nicht hochsignifikant.

Die Werte der Devianz bzw. des AIC als Gütemaße zeigen den gleichen Verlauf: Am aussagekräftigsten ist demnach das Modell mit dem lag zwei mit den niedrigsten Werten, am wenigsten aussagekräftig das mit dem lag -2, was auch schon in der geringen Signifikanz des t-Wertes zu Ausdruck kam. Die unterschiedliche Anzahl an Freiheitsgraden bei den verschiedenen Modellen ist darauf zurückzuführen, dass mit zunehmendem lag immer weniger Datenpaare bei der Berechnung in Betracht gezogen werden konnten.

Zur besseren Veranschaulichung ein Blick auf den Plot der t-Werte der verschiedenen lags (Abbildung 3.9).

Es zeigt sich wieder sehr gut die bereits erwähnte Asymmetrie zwischen den Werten: bei den lags null bis drei ist der Einfluss der Influenza positiv und deutlich signifikant, das Maximum liegt hier wieder beim Abstand von zwei Wochen. Bei negativen lags sinken die t-Werte deutlich ab, bis sie bei -3 und

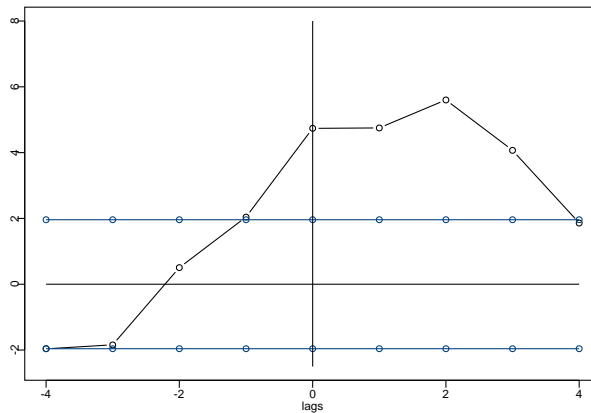


Abbildung 3.9: t-Werte für verschiedene lags beim Poissonregressionsmodell mit nicht transformierten Meningokokken- und saisonbereinigten Influenzadaten

-4 ins Negative abrutschen.

### Bewertung des Modells

Als Gütemaß dient hier die Devianz, die zusammen mit den Freiheitsgraden ein Kriterium für Überdispersion darstellt: Ist das Verhältnis deutlich größer als eins, was bedeutet, dass die Devianz größer als die Anzahl der Freiheitsgrade ist, so spricht man von Über-, bei einem Verhältnis kleiner als eins von Unterdispersion. Bei Überdispersion ist die in den Daten beobachtete Varianz höher als die vom Modell vorgesehene, bei Unterdispersion geringer. Die vom Modell vorgesehene Varianz ist bei der Poissonregression gleich dem Erwartungswert.

Bei diesem Modell ergibt sich beispielsweise für den lag zwei eine Devianz von 258.36 bei 178 Freiheitsgraden, das Verhältnis ist deutlich größer als eins. Die Modellwahl scheint somit fragwürdig und zur weiteren Untersuchung eine Quasi-Poissonregression angebracht, bei der dieses Problem der Überdispersion in die Signifikanz der Parameter mit einfließt.

Zur weiteren Beurteilung dienen, wie schon bei den linearen Modellen, der Q-Q-Plot, sowie das Streudiagramm der standardisierten Residuenwerte (Abbildung 3.10). Die Residuen streuen insgesamt relativ gleichmäßig, vielleicht scheint es eine etwas größere Streuung bei größeren gefitteten Werten zu geben, die auf eine Systematik in den Residuenwerten hindeuten könnte, die

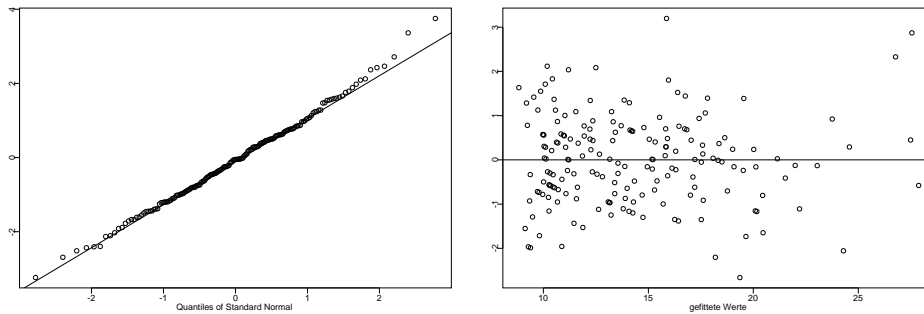


Abbildung 3.10: Plot der standardisierten Residuen und Q-Q-Plot mit lag 2 im Poissonregressionsmodell

für größere gefittete Werte größere Residuen bedeutet.

Auch im Q-Q-Plot wird zunächst eine ziemlich gute Anpassung an die Quantile der Standardnormalverteilung deutlich. Allerdings werden die Abweichungen von der Referenzlinie gegen Ende systematischer, die Werte liegen zum Schluss deutlich über der Geraden.

### 3.2.2 Quasi-Poissonregression

Bei diesem Modell der generalisierten linearen Regressionsrechnung wird das Problem der Überdispersion vom statistischen Programmpaket R durch Anpassung des Dispersionsparameters an die Daten gelöst. Verwendet wurden wieder die gleichen Daten wie beim normalen Poissonregressionsmodell.

Im Folgenden (Abbildung 3.11) ist wieder die Graphik der t-Werte des Modells zu sehen.

An der Grundaussage des vorhergehenden Modells ändert sich nichts: Es ist die gleiche asymmetrische Struktur mit dem Maximum bei lag zwei zu erkennen. Es fällt nur eine konservativere Struktur der Werte auf: Der Maximalwert liegt bei 4.639, der Minimalwert bei -1.538 und ist somit nicht im signifikant-negativen Bereich. Bei einem lag von minus eins ist, im Gegensatz zum reinen Poissonmodell der t-Wert nicht mehr zum 95%-Niveau signifikant.

Zum besseren Vergleich der beiden Poissonregressionsmodelle dient die folgende Tabelle 3.5. Es werden die aufgrund der Anpassung des Dispersionsparameters vorsichtigeren Schätzungen bei der Quasi-Poissonregression gut deutlich: Der berechnete Einfluss der Influenza wird weniger signifikant, die t-Werte sind deutlich konservativer. Die Parameterschätzungen für den Einfluss der Influenza ändern sich selbst jedoch nicht, der Überdispersion im

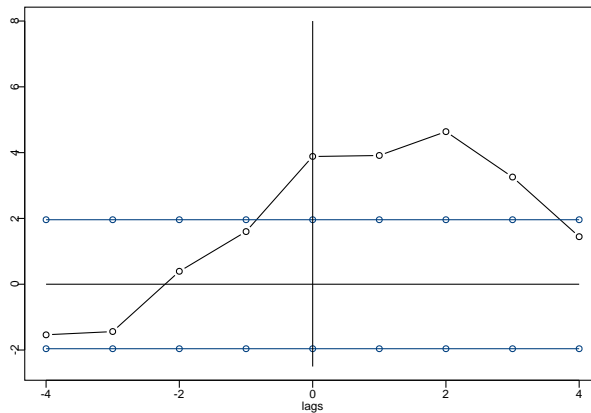


Abbildung 3.11: t-Werte für verschiedene lags beim Quasi-Poissonregressionsmodell mit nicht transformierten Meningokokken- und saisonbereinigten Influenzadaten

Modell wird nur dadurch Rechnung getragen, dass sich die Signifikanz hin zu einer vorsichtigeren Aussage verändert.

Auch bei den Graphen der Residuen ändert sich nichts, die Plots für die Poissonregression gelten auch für die Quasi-Poissonregression.

### 3.3 Regressionsanalyse für spezielle Untergruppen

In der Folge soll die Hypothese nicht nur für die Gesamtdaten, sondern im Speziellen für einzelne Untergruppen überprüft werden. Dabei ergibt sich das Problem, dass sich je nach Unterteilung die Fallzahlen mehr oder weniger stark verringern und sich somit die Güte des Modells verschlechtert. Dargestellt werden in der Folge die Untersuchungsergebnisse der jeweiligen Untergruppen im linearen Regressionsmodell mit wurzeltransformierten Meningokokken- und saisonbereinigten Influenzadaten. Untersuchungen mit Hilfe der Quasi-Poissonregression ergaben im Wesentlichen die gleichen Resultate, werden hier aber aufgrund der Übersichtlichkeit weggelassen.

Als erste Spezifizierung des Datensatzes bietet sich eine nach dem Geschlecht differenzierte Betrachtung an, wobei auch die Einflussgröße, also die Influenzadaten, geschlechtsspezifisch in das Modell mit eingehen. Wie bereits in Kapitel 2.1 beschrieben, gibt es jeweils etwas mehr männliche Fälle als weibliche. Es ist zu untersuchen, ob es einen geschlechtsspezifischen Unterschied

	Parameter	t-Wert	t-Wert (Quasi-Poisson)
	lag 4	0.025	1.86
	lag 3	0.052	4.07
	lag 2	0.070	5.60
	lag 1	0.060	4.75
	lag 0	0.060	3.88
	lag -1	0.027	2.04
	lag -2	0.007	0.50
	lag -3	-0.028	-1.84
	lag -4	-0.030	-1.96

Tabelle 3.5: geschätzte Parameter, t-Werte der Poisson- und der Quasipoissonregression bei verschiedenen lags

im Zusammenhang zwischen den Krankheiten gibt.

Die folgenden Graphiken 3.12 zeigen die t-Werte für die lags von -4 bis 4 für das jeweilige Geschlecht.

Es sind unterschiedliche Strukturen in den Abbildungen zu erkennen: Während

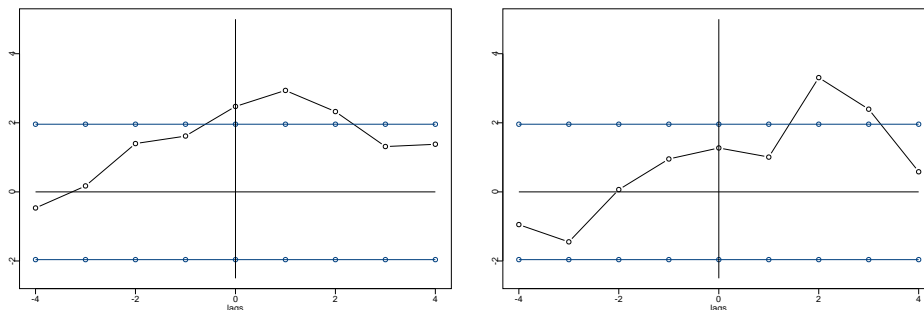


Abbildung 3.12: t-Werte für verschiedene lags beim linearen Regressionsmodell mit wurzeltransformierten Meningokokken- und saisonbereinigten Influenzadaten, aufgesplittet nach Geschlecht (links: weiblich, rechts: männlich)

es bei den Männern eine sehr deutliche Asymmetrie um den lag null ergibt und das Maximum bei dem lag zwei liegt, ist bei den Frauen die Asymmetrie schwächer ausgeprägt. Das Maximum der t-Werte liegt hier bei einem lag von eins, während bei den männlichen Fällen bei diesem lag der Einfluss der Influenza nicht einmal mehr zum 95% Niveau signifikant ist.

Es scheint also einen Unterschied in den zeitlichen Auswirkungen der Influenza

enza zwischen den Geschlechtern zu geben. Allerdings sind diese Werte mit Vorsicht zu genießen: Die reduzierten Fallzahlen wirken sich auf die Aussagekraft der Modelle aus. So ist die Zahl der Wochen, in denen keine Meningokokkenerkrankungen gemeldet wurden, deutlich angestiegen. Das  $R^2$  verringert sich beispielsweise beim wichtigen lag zwei bei den Männern von 0.4229 auf 0.3253, bei den Frauen gar auf 0.2149. Generell zeigt sich die schlechtere Erklärung der Daten auch in den im Vergleich zu den Gesamtdaten niedrigeren t-Werten.

Dieses Problem tritt natürlich auch bei anderen Unterteilungen auf. Aufgrund der Vermutung, dass bei einer regionalen Eingrenzung die Meningokokken durch die entsprechenden Influenzadaten des selben Gebiets besser erklärt werden, wurde im Folgenden eine geographische Aufteilung vorgenommen. Die Diagramme aus Schaubild 3.13 zeigen die t-Werte für die verschiedenen lags bei einer Aufspaltung in Süd, Nord und Ost. Im Süden wurden dabei die Fälle aus Bayern, Saarland, Rheinland-Pfalz, Hessen und Baden-Württemberg zusammengefasst, im Norden aus Hamburg, Niedersachsen, Schleswig-Holstein und Nordrhein-Westfalen, sowie im Osten aus Sachsen, Sachsen-Anhalt, Thüringen, Brandenburg, Mecklenburg-Vorpommern und Berlin. Für den Stadtstaat Bremen lagen teilweise keine Daten vor, so dass auch die vorhandenen Werte nicht miteinbezogen wurden. Generell ist in allen drei Schaubildern die bereits bekannte Struktur, wenn auch nicht so stark wie in den ursprünglichen Graphen, zu erkennen. Wie bereits bei der geschlechtsspezifischen Untersuchung beobachtet, sind auch hier deutlich weniger t-Werte signifikant. Im Süden und Norden ist lediglich der lag zwei zum 95%-Niveau signifikant, während dies im Osten auch die lags null und eins sind.

Die aufgrund der geringeren Fallzahlen geringere Aussagekraft zeigt sich erneut an den Werten des Bestimmtheitsmaßes  $R^2$ , die in Tabelle 3.6 aufgelistet sind. Im Vergleich zu denen der Gesamtdaten sind diese hier deutlich niedriger. Auch die gleichmäßige Struktur, mit einem Maximum bei lag zwei und anschließend nach links und rechts stetig abfallenden Werten, ist nur noch in Teilen vorhanden. So sind beispielsweise für den Norden bei negativen lags höhere Bestimmtheitsmaße als für die lags null und eins zu erkennen.

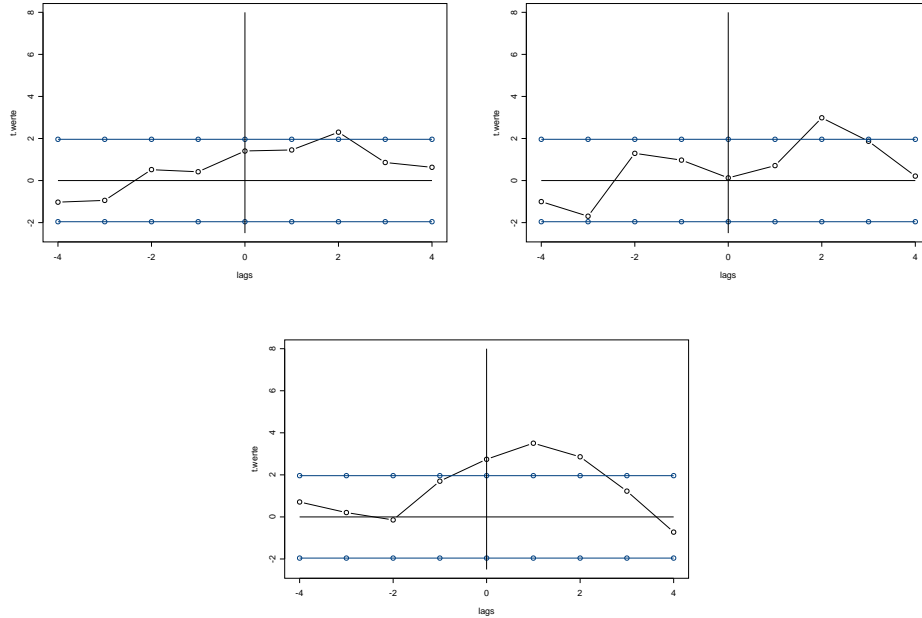


Abbildung 3.13: t-Werte für verschiedene lags beim linearen Regressionsmodell mit wurzeltransformierten Meningokokken- und saisonbereinigten Influenzadaten, geographische Aufspaltung (von links nach rechts: Süd, Nord, Ost)

	$R^2_{Sued}$	$R^2_{Nord}$	$R^2_{Ost}$
lag -4	0.1825	0.2448	0.1914
lag -3	0.1856	0.2502	0.1938
lag -2	0.1848	0.2444	0.1895
lag -1	0.1865	0.2465	0.2055
lag 0	0.1941	0.2294	0.2255
lag 1	0.2008	0.2264	0.2405
lag 2	0.2155	0.2670	0.2197
lag 3	0.1902	0.2408	0.1852
lag 4	0.1821	0.2214	0.1742

Tabelle 3.6:  $R^2$  für verschiedene Lags und geographische Gebiete

# Kapitel 4

## Abschließende Bewertung

### 4.1 Zusammenfassung

Die Untersuchungshypothese, ob einer erhöhten Influenzaerkrankungsrate in einem Abstand von ein bis drei Wochen eine ebenfalls erhöhte Anzahl an Meningokokkenfällen folgt, konnte teilweise bestätigt werden.

Im Zuge der deskriptiven Analyse kann in den verschiedenen Zeitreihendiagrammen ein Anstieg der Meningokokkenerkrankungen in der Folge eines Anstiegs der Influenzaerkrankungen vermutet werden. Am deutlichsten wird dies stets im ausbruchstärksten Jahr 2003. Untersucht man die Hypothese jedoch in verschiedenen Untergruppen, wie etwa in einer bestimmten Alterskategorie, für eine spezielle Erregergruppe oder in einem speziellen Bundesland, so sind entsprechende Ergebnisse nur ansatzweise zu finden. Oft kann die Untersuchungshypothese nur in vereinzelt Jahren, speziell im Jahr 2003, nachgewiesen werden.

Bei den Werten der Kreuzkorrelationsfunktion zwischen beiden Krankheiten bestätigte sich, wenn auch nicht signifikant, die Vermutung eines positiven Zusammenhangs bei vorauslaufenden Influenzawerten und eher negativer Korrelationswerte bei nachlaufenden.

Im Zuge der linearen Regressionsanalyse kristallisierte sich heraus, dass mit wurzeltransformierten Meningokokkendaten gearbeitet werden sollte. Ein Vergleich mit den Modellen der nicht transformierten Meningokokkendaten ergab, dass die Wurzeltransformation die Güte der Modelle verbessert. Speziell einige Residual- und QQ-Plots bestätigten dies. Es stellte sich heraus, dass immer jeweils nur ein Sinus- bzw. Cosinusterm, der den Saisoneinfluß in den Modellen repräsentieren soll, signifikant war.

Stets war der geschätzte Parameter des Influenzaeinflusses für die Lags von null bis drei positiv und signifikant. Dies bestätigte unsere Hypothese aber-

mals. Ein Anstieg der Meningokokkenerkrankungen folgt also dem der Influenza. Im Modell mit wurzeltransformierten Meningokokken- und saisonbereinigten Influenzadaten, das sich als das beste lineare Modell herausstellte, konnte bei einem Lag von zwei die geringste Standardabweichung, der höchste t-Wert und das größte  $R^2$  beobachtet werden.

Im linearen Modell, bei dem Regressor und Regressand vertauscht wurden, somit also die Daten der Meningokokken als Einflußgröße eingehen, liegt der höchste t-Wert bei einem Lag von minus zwei. Alle t-Werte liegen im nicht-signifikanten Bereich, was mit der geringen Anzahl der Meningokokkenerkrankungen erklärt werden kann.

Versucht man nun die Daten mit Hilfe der Poissonregression zu beschreiben, so ergeben sich ähnliche Ergebnisse. Hier ist der geschätzte Parameter des Influenzaeinflusses für die Lags von minus eins bis plus drei positiv und signifikant, allerdings nicht ganz so groß wie in den vorherigen Modellen. Bei einem Lag von zwei lässt sich erneut der höchste t-Wert messen, auch der AIC ist hier am geringsten. Allerdings kann man eine Überdispersion für alle Lags feststellen. Aus diesem Grunde bietet es sich an, das Poisson-Regressionsmodell mit dem eines Quasi-Poisson-Regressionsmodells zu ersetzen. Dadurch wird dieses Problem der Überdispersion gelöst. An den Werten ändert sich nichts, außer dass sich die t-Werte konservativer verhalten und damit der Lag von minus eins nicht mehr signifikant ist.

## 4.2 Ergebnisse anderer Studien

Mit der Analyse, ob ein zeitlich-kausaler Zusammenhang zwischen Influenza- und Meningokokkenerkrankungswellen besteht, beschäftigten sich auch zwei weitere Studien.

Eine der Untersuchungen wurde in Paris vom "Institut National de la Santé et de la Recherche Médicale" durchgeführt. [6] Als Beobachtungszeitraum dienten die Jahre von 1985 bis 1990. Im Vergleich zu unserer Studie ist ein signifikanter Zusammenhang sogar über sechs Wochen bei einem lag von Null bis fünf feststellbar. Der größte t-Wert liegt auch hier bei einem lag von zwei. Es ist zu bemerken, dass in weiten Teilen die gleiche Asymmetrie um den lag null, wenn auch mit signifikanteren t-Werten, erkennbar ist.

In der anderen Studie beschäftigten sich Mitarbeiter verschiedener medizinischer Institute aus Dänemark mit dieser Fragestellung. [7] Der Datensatz basiert auf den in North Jutland County von 1980 bis 1999 registrierten Influenza- und Meningokokkenerkrankungen. In diesem Zeitraum wurden insgesamt 413 Meningokokkenfälle gemeldet. In Vergleich zu unserer

Studie, in der in nicht einmal vier Jahren 2719 Fälle registriert wurden, sind hier deutlich geringere Fallzahlen zu beobachten. Der mögliche Zusammenhang zwischen den Krankheiten wurde mit einem log-linearen Poisson-Regressionsmodell analysiert. Es wurde ein positiver, jedoch nicht signifikanter, Zusammenhang zwischen Grippe- und Meningokokkenerkrankungen in derselben Woche, also bei einem lag von Null entdeckt. Bei einem Lag von ein oder zwei Wochen konnte dagegen ein solcher Zusammenhang nicht mehr festgestellt werden.

# Literaturverzeichnis

- [1] Universität Leipzig, <http://www.medizin.uni-leipzig.de/krankheitsdatenbank/>
- [2] Bildungsministerium Österreich, <http://www.bmbwk.gv.at/schulen/unterricht/>
- [3] Robert Koch-Institut, *Epidemiologisches Bulletin*, 2.7.2004, Nr.27
- [4] Robert Koch-Institut, SurvStat, <http://www3.rki.de/SurvStat>, Datenstand: 08.09.2004
- [5] Schlittgen, Streiberg, 1991, *Zeitreihenanalyse*, 4. Auflage
- [6] Hubert, Watier, Garberin, Richardson, *Meningococcal Disease and Influenza-like Syndrome: A New Approach to an Old Question* aus "The journal of Infection Diseases" 1992;166:542-545
- [7] Jensen, Lundbye-Christensen, Smuelsson, Sørensen, Schönheyder, *A 20-year old ecological study of the temporal association between influenza and meningococcal disease* aus "European Journal of Epidemiology" 2004, 19:181-187

# Anhang A

## S-Plus Code

```
‡PRAKTIKUM SOMMER 2004

‡ Einlesen und Bearbeiten des Datensatzes

influenza <- read.table("C:/Eigene Dateien/Eigene Dokumente/Uni/influenza2.txt")
zeile <- c(seq(1:193))
variablen <- c("woche", "IAlter0", "IAlter1", "IAlter2", "IAlter3", "IAlter4", "IAlter5.9", "IAlter10.14", "IAlter15.19",
"IAlter20.24", "IAlter25.29", "IAlter30.39", "IAlter40.49", "IAlter50.59", "IAlter60.69", "IAlter70.99", "Isonstige", "I.m",
"I.w", "I.kBL", "I.BaWu", "I.Bay", "I.Ber", "I.Brand", "I.Bre", "I.HH", "I.Hes", "I.MeVo", "I.Nieder", "I.NoWe", "I.RhPf",
"I.Saar", "I.Sachs", "I.SaAn", "I.ScHo", "I.Thu", "I.epidem", "I.labor", "I.gesamt", "MAlter0", "MAlter1", "MAlter2",
"MAlter3", "MAlter4", "MAlter5.9", "MAlter10.14", "MAlter15.19", "MAlter20.24", "MAlter25.29", "MAlter30.39", "MAlter40.49",
"MAlter50.59", "MAlter60.69", "MAlter70.99", "M.n.erm", "M.m", "M.w", "M.BaWu", "M.Bay", "M.Ber",
"M.Brand", "M.Bre", "M.HH", "M.Hes", "M.MeVo", "M.Nieder", "M.NoWe", "M.RhPf", "M.Saar", "M.Sachs", "M.SaAn",
"M.ScHo", "M.Thu", "M.epidem", "M.labor", "M.gesamt", "Jahr", "M.andere", "M.n.a", "M.n.erm", "M.n.ero",
"M.poli", "M.seroA", "M.seroB", "M.seroC", "M.seroW135", "M.seroX", "M.seroY", "I.andere", "I.n.aw", "I.n.erho",
"I.n.ermi", "I.Avirus", "I.ABvirus", "I.Bvirus", "I.Cvirus")
dimnames(influenza) <- list(zeile,variablen)
attach(influenza)

‡ Erstellen neuer Variablen

‡Das Jahr hat 52 Wochen...

I.gesamt.dze <- dze(I.gesamt)
M.gesamt.dze <- dze(M.gesamt)
I.Alter0.dze <- dze(IAlter0)
I.Alter15.19.dze <- dze(IAlter15.19)
M.Alter0.dze <- dze(MAlter0)
M.Alter15.19.dze <- dze(MAlter15.19)
I.Sachs.dze <- dze(I.Sachs)
M.Sachs.dze <- dze(M.Sachs)
I.NoWe.dze <- dze(I.NoWe)
M.NoWe.dze <- dze(M.NoWe)

‡Wurzeltransformation mit 52 Wochen

I.gesamt.dze.W <- sqrt(I.gesamt.dze)
M.gesamt.dze.W <- sqrt(M.gesamt.dze)
I.Alter0.dze.W <- sqrt(I.Alter0.dze)
I.Alter15.19.dze.W <- sqrt(I.Alter15.19.dze)
M.Alter0.dze.W <- sqrt(M.Alter0.dze)
M.Alter15.19.dze.W <- sqrt(M.Alter15.19.dze)
I.Sachs.dze.W <- sqrt(I.Sachs.dze)
M.Sachs.dze.W <- sqrt(M.Sachs.dze)
I.NoWe.dze.W <- sqrt(I.NoWe.dze)
M.NoWe.dze.W <- sqrt(M.NoWe.dze)

I.gesamt.2001.dze <- I.gesamt.dze [Jahr==2001]
I.gesamt.2002.dze <- I.gesamt.dze [Jahr==2002]
I.gesamt.2003.dze <- I.gesamt.dze [Jahr==2003]
I.gesamt.2004.dze <- I.gesamt.dze [Jahr==2004]

M.gesamt.2001.dze <- M.gesamt.dze [Jahr==2001]
M.gesamt.2002.dze <- M.gesamt.dze [Jahr==2002]
M.gesamt.2003.dze <- M.gesamt.dze [Jahr==2003]
M.gesamt.2004.dze <- M.gesamt.dze [Jahr==2004]

‡Glätten der Daten mit 52 Wochen
```

```

#Glätten der Wurzeldaten mit Fensterbreite 6 für lag 1
I.glatt6.dze.W.lagEins <- glaetten6lagEins(I.gesamt.dze.W)
I.glatt6.0.dze.W.lagEins <- glaetten6lagEins(I.Alter0.dze.W)
I.glatt6.Sachs.dze.W.lagEins <- glaetten6lagEins(I.Sachs.dze.W)
I.glatt6.NoWe.dze.W.lagEins <- glaetten6lagEins(I.NoWe.dze.W)

#Berechnen der Residuen mit lag 1
I.res.dze.W.lagEins <- I.gesamt.dze.W[4:186] - I.glatt6.dze.W.lagEins
I.res.0.dze.W.lagEins <- I.Alter0.dze.W[4:186] - I.glatt6.0.dze.W.lagEins
I.Sachs.res.dze.W.lagEins <- I.Sachs.dze.W[4:186] - I.glatt6.Sachs.dze.W.lagEins
I.NoWe.res.0.dze.W.lagEins <- I.NoWe.dze.W[4:186] - I.glatt6.NoWe.dze.W.lagEins

#analog für andere lags

#Glätten der Wurzeldaten
I.wurzel.glatt <- glaetten6(I.wurzel.gesamt)
M.wurzel.glatt <- glaetten6(M.wurzel.gesamt)
I.wurzel.glatt.0 <- glaetten6(I.wurzel.0)
M.wurzel.glatt.0 <- glaetten6(M.wurzel.0)

#Berechnen der Residuen
I.wurzel.res <- I.wurzel.gesamt - I.wurzel.glatt
M.wurzel.res <- M.wurzel.gesamt - M.wurzel.glatt
I.wurzel.res.0 <- I.wurzel.0 - I.wurzel.glatt.0
M.wurzel.res.0 <- M.wurzel.0 - M.wurzel.glatt.0

# Funktion, zur Bereinigung des 53-Wochen Problems
dze <- function(var)
var[53] <- var[53] + var[54]
for (i in 54:104)
var[i] <- var[i+1]
var[105] <- var[106] + var[107]
for (i in 106:156)
var[i] <- var[i+2]
var[157] <- var[159] + var[160]
for (i in 158:190)
var[i] <- var[i+3]
print(var[1:190])#wegen Differenzenbildung mit glaetten6

# Funktion um gleitende Durchschnitte zu berechnen

# für 6 Wochen
glaetten6 <- function(var)
x <- c(rep(0,190)) for (i in (4:187))
x[i] <- (1/2*var[i-3]+var[i-2]+var[i-1]+var[i]+var[i+1]+var[i+2]+1/2*var[i+3])/6
#tsplot(var)
#tslines(x, col=2)
print(x)

# Funktionen, um Anteilswerte zu berechnen

# anteilI <- function(var)
# {
# {x <- sum(var)/sum(I.gesamt)
# }
# print(x)}

# anteilM <- function(var)
# {
# {x <- sum(var)/sum(M.gesamt)
# }
# print(x)}

# Verteilung des Alters bei Meningokokken und Influenza

altersverteilungM <- c(sum(MAlter0), sum(MAlter1), sum(MAlter2), sum(MAlter3), sum(MAlter4), sum(MAlter5.9),
sum(MAlter10.14), sum(MAlter15.19), sum(MAlter20.24), sum(MAlter25.29), sum(MAlter30.39), sum(MAlter40.49), sum(MAlter50.59),
sum(MAlter60.69), sum(MAlter70.99))

altersverteilungI <- c(sum(IAlter0), sum(IAlter1), sum(IAlter2), sum(IAlter3), sum(IAlter4), sum(IAlter5.9), sum(IAlter10.14),
sum(IAlter15.19), sum(IAlter20.24), sum(IAlter25.29), sum(IAlter30.39), sum(IAlter40.49), sum(IAlter50.59), sum(IAlter60.69),
sum(IAlter70.99))
alternamen <- c("0 Jahre", "1 Jahr", "2 Jahre", "3 Jahre", "4 Jahre", "5-9 Jahre", "10-14 Jahre", "15-19 Jahre", "20-24
Jahre", "25-29 Jahre", "30-39 Jahre", "40-49 Jahre", "50-59 Jahre", "60-69 Jahre", "ab 70 Jahre")

pie(altersverteilungM, names=alternamen, col=2:12)
title(main="Altersverteilung der Meningokokkenfälle")
pie(altersverteilungI, names=alternamen, col=2:12)
title(main="Altersverteilung der Influenzafälle")

# Zeitreihenplots

```

```

## Vergleich Influenza- vs. Meningokokkenfälle

tsplot(I.gesamt.dze, xlab="Wochen", ylab="Anzahl der Erkrankungen")
tslines(M.gesamt.dze, col=2)
legend(0,1400, c("Influenza", "Meningokokken"), fill=c(1,2))

## Zeitreihe Influenza, aufgesplittet nach Jahren
tsplot(I.gesamt.dze[105:156], xlab="Wochen", ylab="Anzahl der Erkrankungen")
tslines(I.gesamt.dze[53:104], col=3)
tslines(I.gesamt.dze[1:52], col=2)
tslines(I.gesamt.dze[157:190], col=4)
legend(40,1400, c("2001", "2002", "2003", "2004"), fill=c(2,3,1,4))

## Zeitreihe Meningokokken, aufgesplittet nach Jahren
tsplot(M.gesamt.dze[105:156], xlab="Wochen", ylab="Anzahl der Erkrankungen")
tslines(M.gesamt.dze[53:104], col=3)
tslines(M.gesamt.dze[1:52], col=2)
tslines(M.gesamt.dze[157:190], col=4)
legend(40,44, c("2001", "2002", "2003", "2004"), fill=c(2,3,1,4))

# Zeitreihenplots mit transformierten Daten

## Wurzeltransformation der Erregerdaten (Influenza)
trans.I.Avirussqrt <- sqrt(I.Avirus)
trans.I.ABvirussqrt <- sqrt(I.ABvirus)
trans.I.Bvirussqrt <- sqrt(I.Bvirus)
trans.I.Cvirussqrt <- sqrt(I.Cvirus)

## Zeitreihe Influenza - Meningokokken (wurzeltransformiert)
tsplot(I.gesamt.dze.W, xlab="Wochen", ylab="Anzahl der Erkrankungen")
tslines(M.gesamt.dze.W, col=2)
legend(0,37, c("Influenza", "Meningokokken"), fill=c(1,2))

# Zeitreihenplot mit West-Ost Vergleich

## zwischen Sachsen und Nordrhein-Westfalen
tsplot(I.Sachs.dze, xlab="Wochen", ylab="Anzahl Erkrankungen")
tslines(I.NoWe.dze, col=3)
legend(10, 320, c("Sachsen", "NRW"), fill = c(1, 3))

# Plotten der transformierten Erregerdaten

## wurzeltransformierte Verlaufskurven von Influenza B-Virus und Meningokokken
tsplot(dze(trans.I.Bvirussqrt), xlab="Wochen", ylab="wurzeltransformierte Anzahl der Fälle", col=2))
tslines(Mwurzel, col=1)
legend(0,10, c("B-Virus(Inf)", "Gesamt(Meni)"), fill = c(2, 1))

## wurzeltransformierte Verlaufskurven von Influenza A-Virus und Meningokokken
tsplot(dze(trans.I.Avirussqrt, xlab="Wochen", ylab="wurzeltransformierte Anzahl der Fälle", col=8))
tslines(Mwurzel, col=1)
legend(0, 30, c("A-Virus(Inf)", "Gesamt(Meni)"), fill = c(8,1))

## wurzeltransformierte Verlaufskurven von Influenza AB-Virus und Meningokokken
tsplot(dze(trans.I.ABvirussqrt, xlab="Wochen", ylab="wurzeltransformierte Anzahl der Fälle", col=4))
tslines(Mwurzel, col=1)
legend(0, 19, c("AB-Virus(Inf)", "Gesamt(Meni)"), fill = c(4, 1))

## wurzeltransformierte Verlaufskurven von Influenza C-Virus und Meningokokken
tsplot(dze(trans.I.Cvirussqrt, xlab="Wochen", ylab="wurzeltransformierte Anzahl der Fälle", col=5))
tslines(Mwurzel, col=1)
legend(127, 14, c("C-Virus(Inf)", "Gesamt(Meni)"), fill = c(5, 1))

# Diagramm Influenza: Anteil der Erreger
## es fehlt: "I.n.anw" wegen fehlender Werte und
## geringer Fallzahl
erregersummeI <- c(sum(I.andere), sum(I.n.erho), sum(I.n.ermi), sum(I.Avirus), sum(I.ABvirus), sum(I.Bvirus), sum(I.Cvirus))
erregernamenI <- c("andere", "nicht erhoben", "nicht erm.", "A-Virus", "AB-Virus", "B-Virus", "C-Virus")
barplot(erregersummeI, names = erregernamenI, col=1:7)

# Diagramm Meningokokken: Anteil der Serotypen
## es fehlen: "M.poli", "M.seroX" wegen fehlender Werte
## und geringer Fallzahlen
erregersummeI <- c(sum(M.andere), sum(M.n.a), sum(M.n.erh), sum(M.n.ermi), sum(M.n.sero), sum(M.seroA), sum(M.seroB),
sum(M.seroC), sum(M.seroW135), sum(M.seroY))
erregernamenI j- c("andere", "nicht anwendbar", "nicht erhoben", "nicht ermittelbar", "kein Serotyp", "Serotyp A", "Se-
rototyp B", "Serotyp C", "Serogruppe W135", "Serogruppe Y")
barplot(erregersummeI, names = erregernamenI, col=1:8)

# Zeitreihen Influenza-Meningokokken

```

```

## Alterskategorie 0
tsplot(IAAlter0.dze, xlab="Wochen", ylab="Anzahl der Faelle")
tslines(MAlter0.dze, col=2)
legend(0,47,c("Influenza","Meningokokken"),fill=c(1,2))

## Alterskategorie 15-19
tsplot(IAAlter15.19.dze, xlab="Wochen", ylab="Anzahl der Faelle")
tslines(MAlter15.19.dze, col=2)
legend(0,100, c("Influenza","Meningokokken"), fill=c(1,2))

## Autokorrelation- und Kreuzkorrelationsfunktion

## Bilden relevanter Vektoren
I.M.res.dze <- cbind(M.res.dze,I.res.dze)
I.M.dze <- cbind(M.dze,I.dze)

##ACF für saisonbereinigte Daten
I.M.res.dze.j- cbind(I.res.dze.W.lag0, M.res.dze.W)
acf(I.M.res.dze, lag.max=183)

## ACF für nicht saisonbereinigte Wurzeldaten
I.M.dze.j- cbind(I.wurzel.gesamt, M.wurzel.gesamt)
acf(I.M.dze, lag.max=183)

Regressionsanalyse

##Cosinus und Sinus Werte berechnen
t <- c(1:190)
a <- -cos((1 * 2 * pi * t)/52)
b <- -sin((1 * 2 * pi * t)/52)
d <- -cos((2 * 2 * pi * t)/52)
e <- -sin((2 * 2 * pi * t)/52)
f <- -cos((3 * 2 * pi * t)/52)
g <- -sin((3 * 2 * pi * t)/52)
k <- -cos((4 * 2 * pi * t)/52)
l <- -sin((4 * 2 * pi * t)/52)

##Intervallbereich dem entsprechenden lag und Glättenfunktion angepasst coslagEins <- a[4:186]
sinlagEins <- b[4:186]
coslag2 <- a[4:185]
sinlag2 <- b[4:185]
coslag3 <- a[4:184]
sinlag3 <- b[4:184]
coslag4 <- a[4:183]
sinlag4 <- b[4:183]
coslag0 <- a[4:187]
sinlag0 <- b[4:187]
coslagMinus <- a[5:187]
sinlagMinus <- b[5:187]
coslagMinus2 <- a[6:187]
sinlagMinus2 <- b[6:187]
coslagMinus3 <- a[7:187]
sinlagMinus3 <- b[7:187]
coslagMinus4 <- a[8:187]
sinlagMinus4 <- b[8:187]

## für weitere Sinus-,Cosinusterme entsprechende Variablenbildung

##lineare Modelle mit nichttransformierten Meningokokkendenaten

##lineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag 1
LM.lagEins <- lm(M.gesamt.dze[5:187] +1 + sinlagEins +coslagEins + I.res.dze.W.lagEins)
LM4.lagEins <- lm(M.gesamt.dze[5:187] +1 + sinlagEins + coslagEins + sinlagEinsa + coslagEinsa + sinlagEinsb +
coslagEinsb + coslagEinsc + sinlagEinsc + I.res.dze.W.lagEins)
summary(LM.lagEins)
plot(LM.lagEins)
summary(LM4.lagEins)
plot(LM4.lagEins)

##lineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag 2
LM.lag2 <- lm(M.gesamt.dze[6:187] +1 + sinlag2 + coslag2 + I.res.dze.W.lag2)
summary(LM.lag2)
plot(LM.lag2)

##lineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag 3
LM.lag3 <- lm(M.gesamt.dze[7:187] +1 + sinlag3 + coslag3 + I.res.dze.W.lag3)
summary(LM.lag3)

```

```

plot(LM.lag3)

#lineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag 4
LM.lag4 <- lm(M.gesamt.dze[8:187] +1 + sinlag4 + coslag4 + I.res.dze.W.lag4)
summary(LM.lag4)

#lineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag 0
LM.lag0 <- lm(M.gesamt.dze[4:187] +1 + sinlag0 + coslag0 + I.res.dze.W.lag0)
summary(LM.lag0)
plot(LM.lag0)

#lineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag -1
LM.lagMinus <- lm(M.gesamt.dze[4:186] +1 + sinlagMinus + coslagMinus + I.res.dze.W.lagMinus)
summary(LM.lagMinus)
plot(LM.lagMinus)

#lineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag -2
LM.lagMinus2 <- lm(M.gesamt.dze[4:185] +1 + sinlagMinus2 + coslagMinus2 + I.res.dze.W.lagMinus2)
summary(LM.lagMinus2)
plot(LM.lagMinus2)

#lineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag -3
LM.lagMinus3 <- lm(M.gesamt.dze[4:184] +1 + sinlagMinus3 + coslagMinus3 + I.res.dze.W.lagMinus3)
summary(LM.lagMinus3)
plot(LM.lagMinus3)

#lineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag -4
LM.lagMinus4 <- lm(M.gesamt.dze[4:183] +1 + sinlagMinus4 + coslagMinus4 + I.res.dze.W.lagMinus4)
summary(LM.lagMinus4)
plot(LM.lagMinus4)

#t-Werte Plot, lineares Modell mit Normaldaten
lags <- c(-4,-3,-2,-1,0,1,2,3,4)
t.werte <- c(-1.4586,-1.3368, 0.6092, 1.9186, 4.3367, 4.3322, 5.1379, 3.6177, 1.6536)
nulllinie <- c(rep(0,9))
niveauO <- c(rep(1.96,9))
niveauU <- c(rep(-1.96,9))
a <- c(1e-006, 1e-007)
b <- c(-4,5)
plot(lags,t.werte, type="b", ylim=c(-2.5,5))
lines(lags,nulllinie)
lines(lags, niveauO, type="o", col=2)
lines(lags,niveauU,type="o", col=2)
lines(a,b)

#Lineare Modelle mit wurzeltransformierten Meningokokkenwerten

#lineares Modell Y=Meningokokken,Wurzeldaten, X=Influenza, saisonbereinigt, lag 1
LM.lagEins.W <- lm(M.gesamt.dze.W[5:187] +1 + sinlagEins + coslagEins + I.res.dze.W.lagEins)
summary(LM.lagEins.W)
plot(LM.lagEins.W)

#für andere lags analog wie oben

#t-Werte Plot, lineares Modell mit Wurzeldaten
lags <- c(-4,-3,-2,-1,0,1,2,3,4)
t.werte <- c(-0.7761, -1.1047, 0.4203, 1.2929, 3.3199, 2.8821, 4.1580, 3.0012, 1.7228)
nulllinie <- c(rep(0,9))
niveauO <- c(rep(1.96,9))
niveauU <- c(rep(-1.96,9))
a <- c(1e-006, 1e-007)
b <- c(-2.5,5)
plot(lags,t.werte, type="b", ylim=c(-2.5,5))
lines(lags,nulllinie)
lines(lags, niveauO, type="o", col=2)
lines(lags,niveauU,type="o", col=2)
lines(a,b)

#Influenza-Daten nicht-saisonbereinigt
I.gesamt.dze.W.lag4 <- I.gesamt.dze.W[4:183]
I.gesamt.dze.W.lag3 <- I.gesamt.dze.W[4:184]
I.gesamt.dze.W.lag2 <- I.gesamt.dze.W[4:185]
I.gesamt.dze.W.lagEins <- I.gesamt.dze.W[4:186]
I.gesamt.dze.W.lag0 <- I.gesamt.dze.W[4:187]
I.gesamt.dze.W.lagMinus <- I.gesamt.dze.W[5:187]
I.gesamt.dze.W.lagMinus2 <- I.gesamt.dze.W[6:187]
I.gesamt.dze.W.lagMinus3 <- I.gesamt.dze.W[7:187]
I.gesamt.dze.W.lagMinus4 <- I.gesamt.dze.W[8:187]

#lineares Modell Y=Meningokokken Wurzeldaten, X=Influenza Wurzeldaten, lag 1

```

```

LM.lag1.XW <- lm(M.gesamt.dze[5:187] +1 + sinlag1 + coslag1 + I.gesamt.dze.W.lag1)
summary(LM.lag1.XW)
plot(LM.lag1.XW)

#für andere lags analog wie oben

#t-Werte Plot, lineares Modell mit Wurzelndaten, Influenza nicht saisonbereinigt
lags <- c(-4,-3,-2,-1,0,1,2,3,4)
t.werte <- c(0.5228, 1.5848, 3.1925, 4.5815, 6.1148, 6.4187, 6.2417, 4.8690, 3.1718)
nulllinie <- c(rep(0,9))
niveauO <- c(rep(1.96,9))
niveauU <- c(rep(-1.96,9))
a <- c(1e-006, 1e-007)
b <- c(-2.5,8)
plot(lags,t.werte, type="b", ylim=c(-2.5,8))
lines(lags,nulllinie)
lines(lags, niveauO, type="o", col=2)
lines(lags,niveauU,type="o", col=2)
lines(a,b)

#Poissonregression

#generellineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag 1
GLM.lagEins <- glm(M.gesamt.dze[5:187] +1 + sinlagEins + coslagEins + I.res.dze.W.lagEins, family="poisson")
summary(GLM.lagEins)
step(GLM.lagEins, direction="backward")
plot(GLM.lagEins)

#für andere lags analog wie oben

#t-Werte Plot für Poissonregression
lags <- c(-4,-3,-2,-1,0,1,2,3,4)
t.werte <- c(-1.9625,-1.8438,0.5047,2.040,4.7370,4.7508,5.6011,4.0688,1.8609)
nulllinie <- c(rep(0,9))
niveauO <- c(rep(1.96,9))
niveauU <- c(rep(-1.96,9))
a <- c(1e-006, 1e-007)
b <- c(-2.5,8)
plot(lags,t.werte, type="b", ylim=c(-2.5,8))
lines(lags,nulllinie)
lines(lags, niveauO, type="o", col=2)
lines(lags,niveauU,type="o", col=2)
lines(a,b)

#Quasi-Poisson

#generellineares Modell Y=Meningokokken, X=Influenza, saisonbereinigt, lag 1
QGLM.lagEins <- glm(M.gesamt.dze[5:187] +1 + sinlagEins + coslagEins + I.res.dze.W.lagEins, family="quasipoisson")
summary(QGLM.lagEins)
step(GLM.lagEins, direction="backward")
plot(GLM.lagEins)

#für andere lags analog wie oben

#t-Werte Plot für Quasi-Poissonregression
lags <- c(-4,-3,-2,-1,0,1,2,3,4)
t.werte <- c(-1.538,-1.439,0.391,1.600,3.883,3.912,4.639,3.257,1.444)
nulllinie <- c(rep(0,9))
niveauO <- c(rep(1.96,9))
niveauU <- c(rep(-1.96,9))
a <- c(1e-006, 1e-007)
b <- c(-2.5,8)
plot(lags,t.werte, type="b", ylim=c(-2.5,8))
lines(lags,nulllinie)
lines(lags, niveauO, type="o", col=2)
lines(lags,niveauU,type="o", col=2)
lines(a,b)

#Vertauschen der Response- und abhängigen Variable

#Glätten der Daten mit 52 Wochen
#Glätten der Wurzelndaten mit Fensterbreite 6 für lag 1
M.glatt6.dze.W.lagEins <- glaetten6lagEins(M.gesamt.dze.W)
M.glatt6.0.dze.W.lagEins <- glaetten6lagEins(M.Alter0.dze.W)
M.glatt6.Sachs.dze.W.lagEins <- glaetten6lagEins(M.Sachs.dze.W)
M.glatt6.NoWe.dze.W.lagEins <- glaetten6lagEins(M.NoWe.dze.W)

#analog für andere lags

```

```

#Berechnen der Residuen mit lag 1
M.res.dze.W.lagEins <- M.gesamt.dze.W[4:186] - M.glatt6.dze.W.lagEins
M.res.0.dze.W.lagEins <- M.Alter0.dze.W[4:186] - M.glatt6.0.dze.W.lagEins
M.Sachs.res.dze.W.lagEins <- M.Sachs.dze.W[4:186] - M.glatt6.Sachs.dze.W.lagEins
M.NoWe.res.0.dze.W.lagEins <- M.NoWe.dze.W[4:186] - M.glatt6.NoWe.dze.W.lagEins

#analog für andere lags

#lineares Modell Y=Influenza,Wurzeldaten, X=Meningokokken, saisonbereinigt, lag 1
LM.lagEins.W <- lm(I.gesamt.dze.W[5:187] +1 + sinlagEins + coslagEins + M.res.dze.W.lagEins)
summary(LM.lagEins.W)
plot(LM.lagEins.W)

#für andere lags analog wie oben

#t-Werte Plot für Quasi-Poissonregression
lags <- c(-4,-3,-2,-1,0,1,2,3,4)
t.werte <- c(0.3726 , 0.7445, 1.1167, 0.7970, 0.9437, 0.3413, 0.1544,-0.2650, -0.1577)
nulllinie <- c(rep(0,9))
niveauO <- c(rep(1.96,9))
niveauU <- c(rep(-1.96,9))
a <- c(1e-006, 1e-007)
b <- c(-2.5,8)
plot(lags,t.werte, type="b", ylim=c(-2.5,8))
lines(lags,nulllinie)
lines(lags, niveauO, type="o", col=2)
lines(lags,niveauU,type="o", col=2)
lines(a,b)

#Regressionsanalyse für spezielle Untergruppen

#Geographische Untersuchung
I.sued <- (I.Bay + I.BaWu)
I.west <- (I.Hes + I.RhPf + I.Saar + I.NoWe)
I.nord <- (I.HH + I.Nieder + I.ScHo + I.MeVo)
I.ost <- (I.Sachs + I.SaAn + I.Thu + I.Ber + I.Brand)

M.sued <- (M.Bay + M.BaWu)
M.west <- (M.Hes + M.RhPf + M.Saar + M.NoWe)
M.nord <- (M.HH + M.Nieder + M.ScHo + M.MeVo)
M.ost <- (M.Sachs + M.SaAn + M.Thu + M.Ber + M.Brand)

#Westen, lag 1
LMlagEins.west <- lm(M.west.dze.W [5:187] +1 + sinlagEins + coslagEins + I.west.res.dze.W.lagEins)
summary(LMlagEins.west)
plot(LMlagEins.west)

#analog für andere lags und Bundesländer

#t-Wert Plot Westen
lags <- c(-4,-3,-2,-1,0,1,2,3,4)
t.werte <- c(-1.1315 , -0.8288 , -0.0933 , 1.7039 , -0.9220 ,0.8294 ,3.4664 ,0.6080 ,0.6629 )
nulllinie <- c(rep(0,9))
niveauO <- c(rep(1.96,9))
niveauU <- c(rep(-1.96,9))
a <- c(1e-006, 1e-007)
b <- c(-2.5,8)
plot(lags,t.werte, type="b", ylim=c(-2.5,8))
lines(lags,nulllinie)
lines(lags, niveauO, type="o", col=2)
lines(lags,niveauU,type="o", col=2)
lines(a,b)

# analog für andere Länder

#Geschlechtsspezifische Untersuchung

#Jahr mit 52 Meldewochen
I.m.dze <- dze(I.m)
M.m.dze <- dze(M.m)
I.w.dze <- dze(I.w)
M.w.dze <- dze(M.w)

#Wuzeltransformation mit 52 Wochen
I.m.dze.W <- sqrt(I.m.dze)
M.m.dze.W <- sqrt(M.m.dze)
I.w.dze.W <- sqrt(I.w.dze)
M.w.dze.W <- sqrt(M.w.dze)

```

```

#Glätten der Daten mit 52 Wochen
#Glätten der Wurzeldaten mit Fensterbreite 6 für lag 1
I.glatt6.m.dze.W.lagEins <- glaetten6lagEins(I.m.dze.W)
I.glatt6.w.dze.W.lagEins <- glaetten6lagEins(I.w.dze.W)

#analog für andere Lags

#Berechnen der Residuen
#Berechnen der Residuen mit lag 1
I.m.res.dze.W.lagEins <- I.m.dze.W[4:186] - I.glatt6.m.dze.W.lagEins
I.w.res.dze.W.lagEins <- I.w.dze.W[4:186] - I.glatt6.w.dze.W.lagEins

# analog für andere lags

# Lineare Modelle für männliche Population

#lineares Modell Y=Meningokokken,Wurzeldaten,männlich, X=Influenza, saisonbereinigt,männlich, lag 1
LM.lagEins.W.m <- lm(M.m.dze.W[5:187] +1 + sinlagEins + coslagEins + I.m.res.dze.W.lagEins)
summary(LM.lagEins.W.m)
plot(LM.lagEins.W.m)

#t-Wert Plot Männer
lags <- c(-4,-3,-2,-1,0,1,2,3,4)
t.werte <- c(-4,-3,-2,-1,0,1,2,3,4)
t.werte <- c(-0.9469, -1.4458, 0.0679, 0.9531, 1.2730, 1.0087, 3.3087, 2.3942, 0.5834)
nulllinie <- c(rep(0,9))
niveauO <- c(rep(1.96,9))
niveauU <- c(rep(-1.96,9))
a <- c(1e-006, 1e-007)
b <- c(-2.5,8)
plot(lags,t.werte, type="b", ylim=c(-2.5,8))
lines(lags,nulllinie)
lines(lags, niveauO, type="o", col=2)
lines(lags,niveauU,type="o", col=2)
lines(a,b)

#analog für andere lags und Frauen

```