Zyklische Modellierung von Mottenanzahlen

Statistische Methoden für circuläre Daten

Diplomarbeit

im Fachgebiet Mathematik

UNIKASSEL VERSITÄT

CENTRO INTERNACIONAL DE LA PAPA CIP

vorgelegt von:	Agnieszka Wenska
am:	28. Dezember 2010
Studienbereich:	Mathematik
Erstgutachterin:	Prof. Dr. Christine Müller
Zweitgutachter:	Prof. Dr. Wolfram Koepf

Zyklische Modellierung von Mottenanzahlen
Statistische Methoden für circuläre Daten

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

0 Einleitung

0.1 Motivation

Das Programm ILCYM (Insect Life Cycle Modeling) [Sporleder u. a. 2009] wurde durch das Internationale Kartoffelzentrum (Centro Internacional de la Papa, CIP), Lima, Peru entwickelt. Dieses Software erlaubt temperaturabhängige Modellbildung von Kartoffelmotten (Potato Tuber Moth, PTM). Da die Schädlinge Kaltblüter sind, können sie ihre eigene Temperatur nicht innerlich regeln. Folglich hängt ihre Entwicklung von der Temperatur ab, zu der sie in der Umwelt ausgestellt werden. Sie verlangen eine gewisse Menge der Hitze, um sich von einer Stufe in ihrem Leben-Zyklus zu einer anderen, z.B von Ei bis Larven, zu entwickeln. Wegen jährlicher Veränderungen in Wetter sind Kalender-Daten nicht eine gute Grundlage, um Plage-Bevölkerungswachstum und Ausbrüche vorauszusagen und Management-Entscheidungen zu machen. Wenn man jedoch über die im Laufe der Zeit angesammelten Temperaturdaten vefügt, ist das ein wertvolles Werkzeug, um Phänologie-Modelle zu bekommen, die die Zeit von Ereignissen in einer Entwicklung von Organismen voraussagen. Phänologie-Modelle für Insekt-Plagen sind wichtige analytische Werkzeuge für Voraussagen, Auswerten, und Verstehen der Dynamik von Plage-Bevölkerungen in Agrarökosystemen unter einer Vielfalt von Umweltbedingungen. CIP hat von Temperatur getriebenen Phänologie-Modell für die Kartoffelmotte, Phthorimaea Operculella (Zeller) entwickelt, der gut Lebensparameter für verschiedene landwirtschaftliche Regionen voraussagt. Verbunden mit geographischem Informationssystem (Geographic Information System, GIS) und atmosphärischer Temperatur erlaubt das Modell, diese Gefahr-Indexe auf einer weltweiten Skala vorzutäuschen oder es zu gebrauchen, um zukünftige Änderungen dieser Indexe wegen Erderwärmung vorauszusagen. Das Modell, das am Anfang für Phthorimaea Operculella gedacht war, kann auch für andere Insekten-Spezies verwendet werden. Das vom CIP entwickelte Software ILCYM stellt analytische Werkzeuge bereit, um die Insektenphänologie besser zu verstehen und zu studieren.

0.2 Ziel der Arbeit

Die Entwicklung einer Kartoffelknolle-Motte hängt aber nicht nur von der atmosphärischen Temperatur ab. Zu den wichtigsten Faktoren gehören:

- Niederschlag
- relative Feuchtigkeit
- Sonnenstrahlung
- Wind

Das Ziel ist ein möglichst gutes Modell zu entwickeln, das den wirklichen Umweltbedingungen entspricht. In dieser Arbeit soll das ILCYM Programm um eine Komponente ergänzt werden, um die Vorhersage der PTM Populationsentwicklung zu verbessern. Anhand der Motten-Pheromonfallen Fängen aus Huancayo (Peru) soll ein Modell gefunden werden, das die Flugaktivität der Kartoffelmotte beschreibt.

0.3 Aufbau der Arbeit

Im ersten Kapitel findet man zunächst eine Einführung in die Thematik und einen Datensatz, auf dessen Grundlage eine Regressionskurve gesucht wird (Kapitel 5 und Kapitel 5.4). Es wird auch getestet, ob die Daten korreliert sind und wie man die Korrelation vergrössern kann.

Kapitel 2 liefert Einführung in die Notation und Schreibweise der circulären Daten[1]. Dieses Kapitel baut grundsätzlich auf den Büchern von Edward Batschelet, Kanti V. Mardia und Peter E. Jupp ([Batschelet 1981], [Mardia u. Jupp 2000], [Mardia 1972]) auf. Nach grundlegenden Definitionen und Darstellungsarten werden wichtige deskriptive Methoden gezeigt und bewiesen.

In Kapitel 3 werden die wichtigsten circulären Verteilungen vorgestellt, u.a. die Von Mises Verteilung und einige gewickelte Verteilungen. Die Grundlage hierfür bilden vor allem der Artikel von J.A.Greenwood, E.J.Gumbel und D.Durand ([J.A.Greenwood u. a. 1953]) und das Buch von Kanti V. Mardia und Peter E. Jupp ([Mardia u. Jupp

[1] Eine Übertragung auf den 3-dimensionalen Fall ist z.B. in [Mardia u. Jupp 2000] und in [Batschelet 1981] zu finden.

2000]). Die Form der charakteristischen Funktion für die Gitter- und die Stetige Gleichverteilung wird bewiesen. Die Dichte der gewickelten Cauchy-Verteilung wird hergeleitet.

Lokations- und Streuungsmaße einer circulären Zufallsgröße werden in Kapitel 4 erläutert. Es werden der Erwartungswert und die Varianz definiert und es wird der Erwartungswert für einige Verteilungen ausgerechnet.

In Kapitel 5 wird ein Regressionsmodell für circuläre Daten vorgestellt. In Abschnitt 5.3 werden schließlich Regressionskurven gefunden, die den Zusammenhang zwischen der Anzahl der Männchen der Kartoffelmotte und der Regenmenge beschreiben.

Im ersten Kapitel wurde festgestellt, dass die Verschiebung der Regenmenge um fünf Monate die größte Korrelation zwischen der Mottenanzahl und der Regenmenge liefert. Demzufolge werden in Abschnitt 5.4 Modelle für den verschobenen Datensatz gefunden. Um zu sehen, ob die Verschiebung was gebracht hat, wird ein um fünf Messwerte reduzierter Datensatz gebildet. Für den Datensatz werden auch die Mottenanzahlen modelliert und mit denen für den verschobenen Datensatz verglichen. Anschließend werden die Ergebnisse der Modellfindung diskutiert.

Am Ende des Kapitels sind noch zwei Weltkarten[2] zu sehen. Auf der ersten werden Gebiete markiert, wo ähliche Wetterbedingungen zu denen in Huancayo herrschen und Symmetrischema Tangolias vorkommen könnte. Auf der zweiten wird noch zusätzlich markiert, wo Kartoffeln überhaupt angebaut werden.

[2]Diese zwei Weltkarten sind in dem Internationalen Kartoffelzentrum (Centro Internacional de la Papa, CIP), Lima, Peru von Diana Giraldo angefertigt worden.

1 Pheromon-Fänge in der Station 'La Victoria' - Symmetrischema Tangolias, Teil 1

Wir verfügen über Datensätze, die 2 Jahre (1998-2000) in verschiedenen Orten gesammelt wurden. Es wurden Pheromon-Fallen verwendet, um die Anzahl der Männchen von Motten zu schätzen. Die Motten wurden jede Woche gesammelt und gezählt (siehe A.1.1). Die erste Spalte enthält immer den Tag, an dem evaluiert wurde, die zweite den entsprechenden Monat, die dritte den Regenfall und die vierte die Anzahl der Männchen der Kartoffelmotte (Potato Tuber Moth, PTM).

Man will untersuchen, wie die Anzahl der Männchen der Kartofellmotte von dem Niederschlag abhängt. Abbildung 1.1 stellt die Fluktuation der Adulten in Abhängigkeit vom Regenfall dar.

Man beobachtet periodische Schwankungen. In der Regenzeit (November - März) wird der Niederschlag sichtbar höher. Bei vielen Niederschlagstagen nimmt die Anzahl der gefangenen Motten deutlich ab, um später wieder zuzunehmen, wenn die Regenzeit vorbei ist.

Es stellt sich die Frage, ob die Daten überhaupt korreliert sind und ob es Sinn macht, die zu untersuchen. Die Tabelle wird geladen und es wird überprüft, ob die Anzahl der Motten vom Niederschlag abhängt.

```
> library(RODBC)
> library(MASS)
> channel <- odbcConnectExcel("LaVictoria.xls")
> VicST <- sqlFetch(channel, "VicST")

> cor(VicST[,3:4])
```

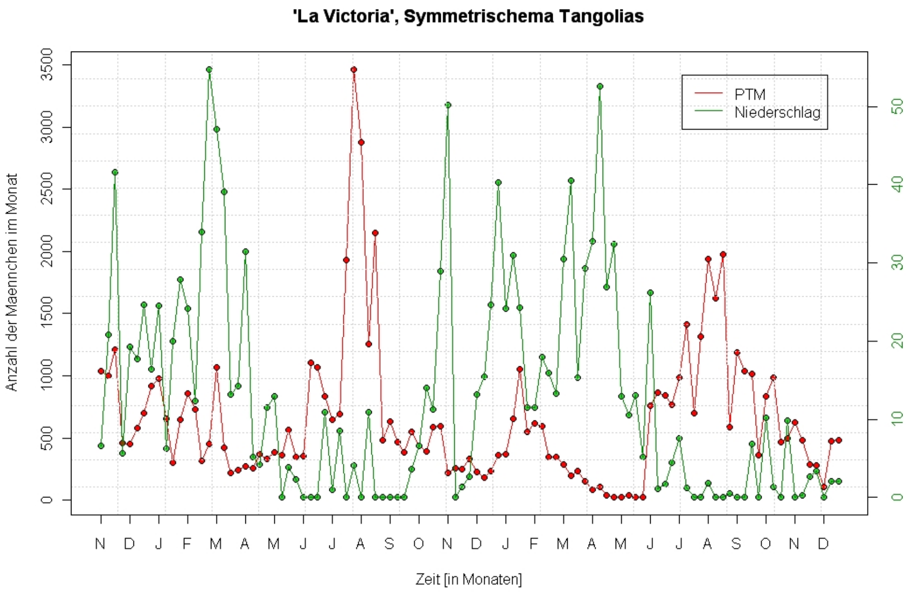

Abbildung 1.1: Flugaktivität von Symmetrischema Tangolias in der CIP Station 'La Victoria' und der entsprechende Niederschlag.

```
            Rainfall        PTM
Rainfall   1.0000000 -0.2743055
PTM       -0.2743055  1.0000000
```

Man sieht, dass die Korrelation zwar besteht, aber nicht so groß ist. Um die zu vergrössern, werden die Wochenfänge zu Monatsfängen zusammengefasst, um die Reaktion der Population auf die Wetterbedingungen wiederzuspiegeln.

Tabelle 1.1: Flugaktivität von Symmetrischema Tangolias an der CIP-Station 'La Victoria' in Huancayo von November 1998 bis Oktober 2000 und der entsprechende Niederschlag. Der Datensatz RainM.ST.

Year	Month	Rainfall	PTM
1998	11	74.7	3712
1998	12	102.5	3627
1999	1	78.4	2462
1999	2	148.1	2562
1999	3	103.3	1407

Fortsetzung auf der nächsten Seite

Flugaktivität von Symmetrischema Tangolias – Fortsetzung

Year	Month	Rainfall	PTM
1999	4	28.6	1447
1999	5	6.2	2372
1999	6	20.4	5169
1999	7	15	9741
1999	8	3.6	2519
1999	9	60.8	2014
1999	10	54.2	1056
1999	11	117.8	1371
1999	12	78.3	2869
2000	1	77.6	1573
2000	2	170.5	768
2000	3	82.8	124
2000	4	45.5	1668
2000	5	15	4702
2000	6	1.8	6849
2000	7	7.3	3824
2000	8	7.3	3145
2000	9	6.3	1670
2000	10	4.1	1062

Quelle: [Keller 2003]

Jetzt wird die Korrelation für den Datensatz RainM.ST getestet.

```
> cor(RainM.ST[,3:4])
           Rainfall       PTM
Rainfall   1.000000 -0.396392
PTM       -0.396392  1.000000
```

Infolge der Zusammenfassung hat sich die Korrelation zwischen PTM und Rainfall vergrössert. PTM und Rainfall sind negativ korreliert. Es stimmt mit den früheren Überlegungen überein.

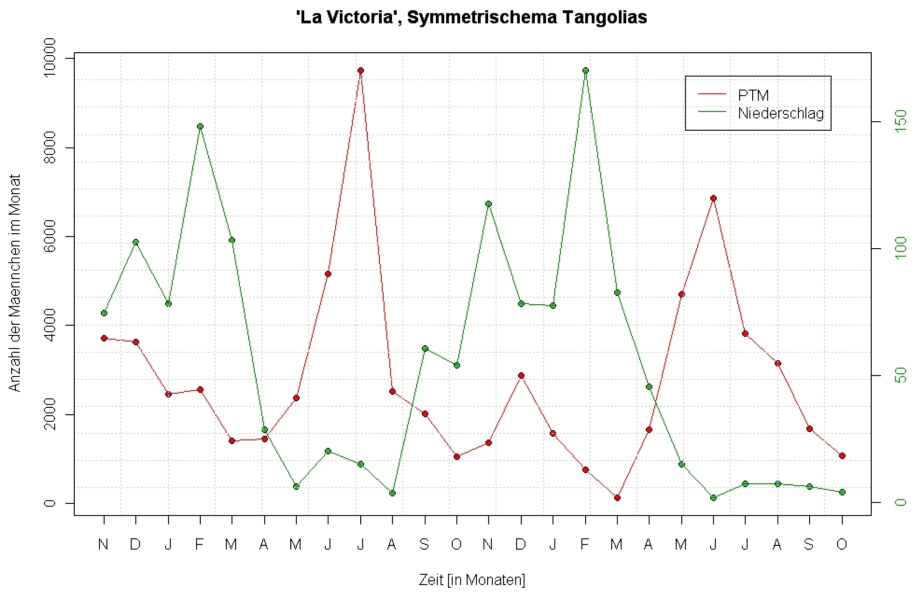

Abbildung 1.2: Flugaktivität von Symmetrischema Tangolias in der CIP Station 'La Victoria' und der entsprechende Niederschlag.

Man kann sich noch überlegen, ob man eine Verschiebung ermitteln könnte, die noch größere Korrelation liefert. Intuitiv würde man die PTM-Kurve um 5 Monate verschieben, sodass die PTM- und die Regenkurve gleichzeitig steigen. Um die Vermutung zu bestätigen bzw. zu widerlegen, wird die Kreuzkorrelation benutzt. Die vergleicht die Korrelation zweier Vektoren bei unterschiedlichen Zeitverschiebungen τ.

```
> ccf(RainM.ST$PTM, RainM.ST$Rainfall, lag.max = 11,
type = c("correlation", "covariance"),plot = TRUE,
na.action = na.fail,main="", xlim=c(-0.3,11))
```

Das Korrelationsdiagramm (Abb. 1.3) zeigt, dass die Behauptung richtig war. Der Koeffizient τ_5 hat den größten Wert, was darauf hinweist, dass die Korrelation zwischen der PTM- und der Regen-Folge am größten für um 5 Einheiten gegeneinander verschobene Paare ist.

Man will natürlich die neue Korrelation ausrechnen. Um dies zu tun, muss der ursprügliche Datensatz um 5 Monate verschoben werden. Der sieht nun folgendermaßen aus:

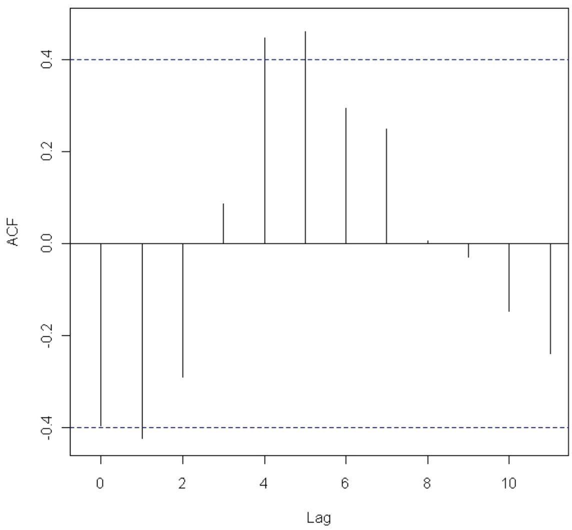

Abbildung 1.3: Kreuzkorrelation des Datensatzes RainM.ST.

Tabelle 1.2: Flugaktivität von Symmetrischema Tangolias nach der Verschiebung. Der Datensatz month.ptm.rain.

Year	Month	PTM	YearR	MonthR	Rainfall
1999	4	1447	1998	11	74.7
1999	5	2372	1998	12	102.5
1999	6	5169	1999	1	78.4
1999	7	9741	1999	2	148.1
1999	8	2519	1999	3	103.3
1999	9	2014	1999	4	28.6
1999	10	1056	1999	5	6.2
1999	11	1371	1999	6	20.4
1999	12	2869	1999	7	15.0
2000	1	1573	1999	8	3.6
2000	2	768	1999	9	60.8
2000	3	124	1999	10	54.2
2000	4	1668	1999	11	117.8
2000	5	4702	1999	12	78.3

Fortsetzung auf der nächsten Seite

Flugaktivität von Symmetrischema Tangolias – Fortsetzung

Year	Month	PTM	YearR	MonthR	Rainfall
2000	6	6849	2000	1	77.6
2000	7	3824	2000	2	170.5
2000	8	3145	2000	3	82.8
2000	9	1670	2000	4	45.5
2000	10	1062	2000	5	15.0

Man berechnet die Korrelation für den verschobenen Datensatz (Quellcode siehe A.2.2) und plottet die (Abb. 1.4).

```
> cor(month.ptm.rain[,c(3,6)])
               PTM  Rainfall
PTM      1.0000000 0.5484252
Rainfall 0.5484252 1.0000000
```

Die Korrelation ρ(PTM,Rainfall) beträgt nun 0.5484252.

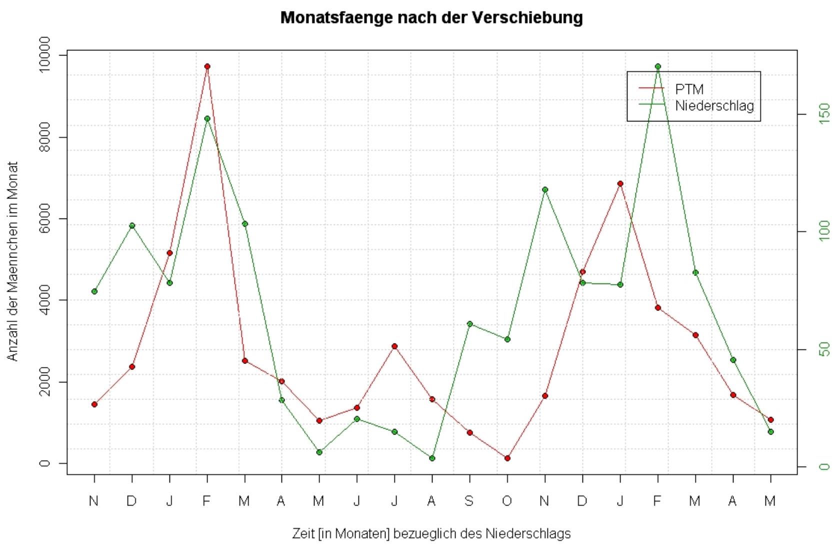

Abbildung 1.4: Der Datensatz month.ptm.rain.

2 Grundbegriffe und deskriptive Methoden der Statistik von Richtungsdaten

2.1 Einführung

Circuläre Daten entstehen auf verschiedene Weise. Die beiden wichtigsten Möglichkeiten entsprechen den beiden wichtigsten zyklischen Messgeräten, dem Kompass und der Uhr. Typische Beobachtungen gemessen mit dem Kompass sind Windrichtung und Richtungen von Zugvögeln. Typische Beobachtungen mit der Uhr gemessen umfassen die Ankunftszeiten (im 24 Stunden-Format) der Patienten in der Notaufnahme in einem Krankenhaus [Mardia u. Jupp 2000].

Eine circuläre Beobachtung kann als ein Punkt am Einheitskreis oder ein Einheitsvektor (Richtunsvektor) in der Ebene gesehen werden. Am Anfang jedes Experiments wird der Urspung festgelegt. Dann kann jede circuläre Beobachtung durch den Winkel zwischen dem Urspung und dem entsprechenden Punkt auf dem Einheitskreis angegeben werden. Circuläre Daten werden in Grad von $0°$ bis $360°$ oder äquivalent in Radian von 0 bis 2π gemessen. Auf der anderen Seite muss die Summe oder Differenz zweier Winkel immer Modulo $360°$ gerechnet werden.

Circuläre Variablen treten auch im Bereich der biologischen Rhythmen auf [Batschelet 1981]. Die 24-Stunden-Periode entspricht einer vollen Umdrehung von 360 Grad. Folglich kann eine Stunde mit einem Winkel von 15 Grad verglichen werden. Analog kann man einen Monat mit 30 Grad identifizieren.

2.2 Graphische Darstellung

2.2.1 Rohdaten

Die einfachste Darstellung der circulären Daten ist ein Plot der circulären Rohdaten, wo jede Beobachtung als Punkt auf dem Einheitskreis erscheint.

Die Abbildung 2.1 (für den Quellcode siehe A.2.3) zeigt 100 generierte Beobachtungen von der Von-Mises-Verteilung mit der mittleren Richtung $\vartheta = 3$ und dem Konzentrationsparameter $\kappa = 0$ (für circuläre Verteilungen siehe Kapitel 4).

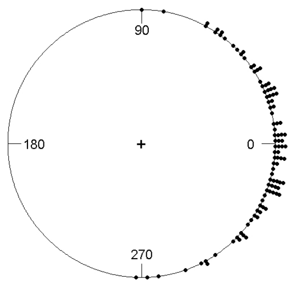

Abbildung 2.1: 100 generierte Daten von der Von-Mises-Verteilung mit der mittleren Richtung $\vartheta = 3$ und dem Konzentrationsparameter $\kappa = 0$.

2.2.2 Gruppierte Daten

Gruppierte circuläre Daten können genauso wie im linearen Fall mithilfe von Histogrammen dargestellt werden. Es gibt zwei Typen: circuläre und lineare Histogramme (vgl.[Fisher 2000]).

Lineares Histogramm Übliches Histogramm mit einem Unterschied: in Abhängigkeit von der Lage des Ursprungs kann ein Histogramms desselben Datensatzes unterschiedlich aussehen (siehe Abbildung 2.2).

Circuläres Histogramm Circuläres Histogramm ist analog zu linearem Histogramm. Die Balken des Histogramms befinden sich aber auf einem Kreis statt auf der Linie.

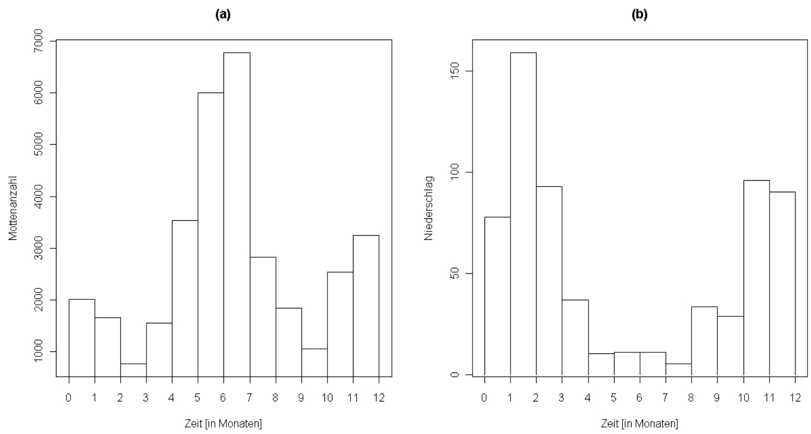

Abbildung 2.2: Lineares Histogramm zu dem Datensatz 1.1: (a) Mottenanzahlen, (b) Regenmenge.

Windrose-Diagramm Eine Variante des circulären Histrogramms ist Windrose-Diagramm. Die Balken werden durch Sektoren ersetzt, deren Fläche proportional zu der Häufigkeit in der entsprechender Gruppe ist. Um dies zu erreichen, soll der Radius jedes Sektors dem Wurzel der relativen Häufigkeit gleich sein. Nicht alle Authoren folgen diese Konvention.

Die Abbildung 2.3 zeigt die Verteilung der Kartoffelmotten und der Regenmenge im Jahr in Anlehnung an den Datensatz 1.1 (R-Quellcode siehe A.2.4). Um diese Darstellung zu bekommen, wurden die Mittelwerte gebildet.

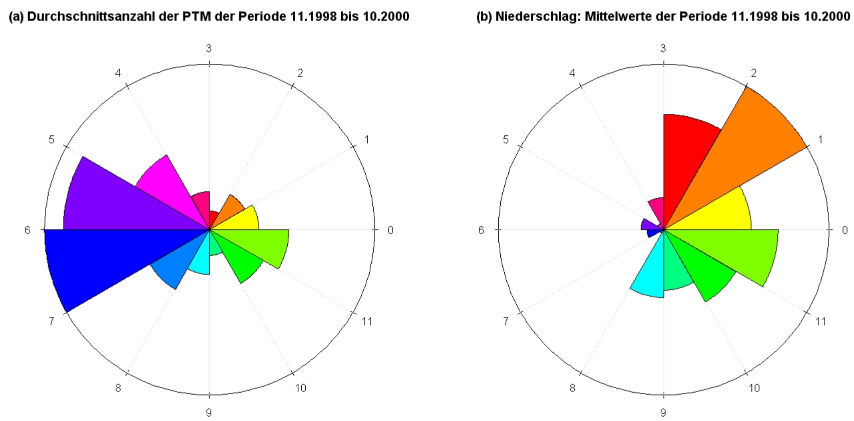

Abbildung 2.3: Windrose-Diagramm zum Datensatz 1.1: (a) Mottenanzahlen, (b) Regenmenge.

2.3 Notation und Schreibweise

Eine circuläre Beobachtung x ist ein Punkt auf dem Einheitskreis und kann folglich durch einen Winkel θ oder eine komplexe Zahl z dargestellt werden

$$x = (\cos\theta, \sin\theta)^T \quad \text{und} \quad z = \mathrm{e}^{\imath\theta} = \cos\theta + \imath \sin\theta$$

Diese Darstellungen sind nicht eindeutig: sie hängen von der Auswahl der Ursprung und der Orientation ab (siehe Abbildung 2.4).

2.4 Lokations-, Konzentration- und Streuungsmaße

Beispiel 2.4.1. Gegeben sei eine Stichprobe, die aus 3 durch Winkel gegebenen Richtungen besteht

$$\theta_1 = 70°, \quad \theta_2 = 340°, \quad \theta_1 = 40°.$$

Man will die mittlere Richtung bestimmen. Offensichtich soll die irgendwo zwischen 0° und 40° liegen. Man rechnet das arithmetische Mittel aus:

$$\frac{1}{3}(\theta_1 + \theta_2 + \theta_3) = 150°.$$

Man sieht, dass dieses Resultat inakzeptabel ist. Zur Bestimmung der mittleren Richtung wird der Massenmittelpunkt $(\bar{C}, \bar{S})$ bestimmt.

Definition 2.4.2 (Mittlere Länge, mittlere Richtung, Hauptvektor)**.** *Es seien circuläre Beobachtungen $x_1, \ldots, x_n$ mit dazugehörigen Winkeln θ_j,*
$j = 1, \ldots, n$ (d.h. $x_j = (\cos\theta_j, \sin\theta_j)^T$ für $j = 1, \ldots, n$) gegeben. Dann heißen

$$\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2} \tag{2.1}$$

die mittlere Länge und

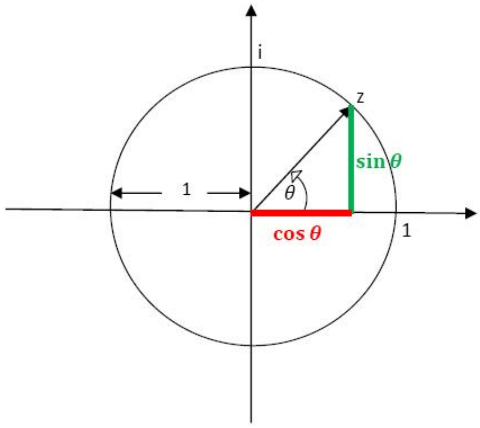

Abbildung 2.4: Darstellung der Richtung x durch den Winkel θ und durch die komplexe Zahl $z = \cos\theta + \imath \sin\theta$.

$$\bar{\theta} = \begin{cases} \arctan(\bar{S}/\bar{C}), & \textit{falls } \bar{C} > 0, \\ \arctan(\bar{S}/\bar{C}) + \pi, & \textit{falls } \bar{C} < 0, \\ \frac{1}{2}\pi, & \textit{falls } \bar{C} = 0 \textit{ und } \bar{S} > 0, \\ \frac{3}{2}\pi, & \textit{falls } \bar{C} = 0 \textit{ und } \bar{S} < 0, \\ \textit{nicht definiert}, & \textit{falls } \bar{C} = 0 \textit{ und } \bar{S} = 0, \end{cases} \tag{2.2}$$

die mittlere Richtung, wobei $\bar{C}$ *und* $\bar{S}$ *folgendermaßen definiert sind*

$$\bar{C} = \frac{1}{n}\sum_{j=1}^{n} \cos\theta_j, \qquad \bar{S} = \frac{1}{n}\sum_{j=1}^{n} \sin\theta_j. \tag{2.3}$$

[Mardia u. Jupp 2000]

Bemerkung 2.4.3. Die in 2.2 definierte mittlere Richtung $\bar{\theta}$ ist Lösung der Gleichungen

$$\bar{C} = \bar{R}\cos\bar{\theta}, \qquad \bar{S} = \bar{R}\sin\bar{\theta}. \tag{2.4}$$

Die mittlere Länge bestimmt zusammen mit der mittleren Richtung den Hauptvektor, m.

Beweis. Die Behauptung folgt sofort aus Definition 2.4.2. □

Bemerkung 2.4.4. Sind die Beobachtungen $x_1, \dots, x_n$ Einheitsvektoren, dann gilt

(i) $0 \leq \bar{R} \leq 1$

(ii) $\bar{R} \xrightarrow[\forall i,j |\theta_i - \theta_j| \to 0]{} 1.$

Beweis. (i) Nach Definition 2.4.2 ist die mittlere Länge gegeben durch $\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2}$, wobei $\bar{C}$ und $\bar{S}$ haben folgende Gestalt $\bar{C} = \frac{1}{n}\sum_{j=1}^{n} \cos\theta_j$ und $\bar{S} = \frac{1}{n}\sum_{j=1}^{n} \sin\theta_j$. Es folgt sofort, dass $\bar{R} \geq 0$ ist.

Sei nun $n \in \mathbb{N}$ beliebig. Dann gilt mit dem Additionstheorem $\cos(\theta_i - \theta_j) = \cos\theta_i \cos\theta_j + \sin\theta_i \sin\theta_j$

$$
\begin{aligned}
\bar{R}^2 &= \left(\frac{1}{n}\sum_{j=1}^{n} \cos\theta_j\right)^2 + \left(\frac{1}{n}\sum_{j=1}^{n} \sin\theta_j\right)^2 \\
&= \frac{1}{n^2}\left(\underbrace{\sum_{j=1}^{n} \cos^2\theta_j + \sum_{j=1}^{n} \sin^2\theta_j}_{=n} + 2\sum_{\substack{i=1 \\ j>i}}^{n} \cos\theta_i \cos\theta_j + 2\sum_{\substack{i=1 \\ j>i}}^{n} \sin\theta_i \sin\theta_j\right) \\
&= \frac{1}{n^2}\left(n + 2\underbrace{\sum_{\substack{i=1 \\ j>i}}^{n} \underbrace{\cos(\theta_i - \theta_j)}_{\leq 1}}_{\leq \frac{n(n-1)}{2}}\right) \\
&\leq \frac{1}{n^2}\left(n + 2\,\frac{n(n-1)}{2}\right) \\
&= \frac{1}{n^2}\,(n + n^2 - n) \\
&= 1.
\end{aligned}
$$

$\Rightarrow \bar{R} \leq 1$

(ii) Folgt aus (i).

□

Die mittlere Länge $\bar{R}$ kann also als Konzentration der Stichprobe gesehen werden (Abbildung 2.5, [Batschelet 1981]).

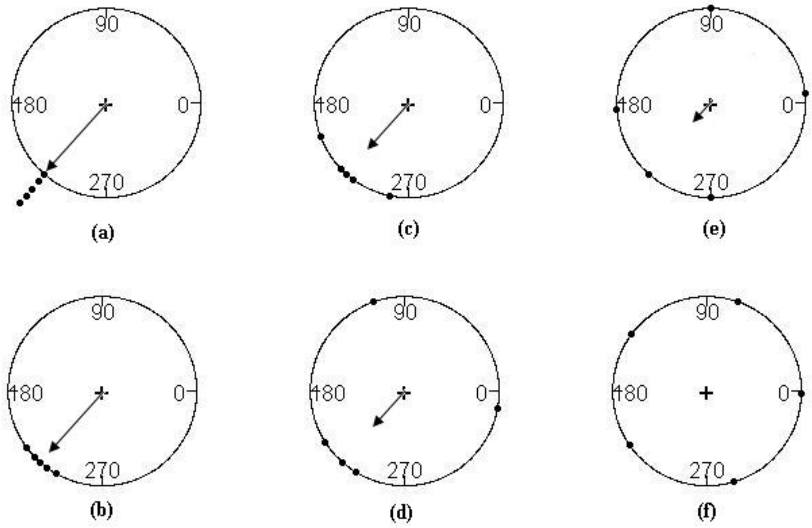

Abbildung 2.5: Die mittlere Länge nimmt die Werte aus dem Intervall $[0, 1]$ an. Je gleichmäßiger die Beobachtungen auf dem Kreis verteilt sind, desto kleiner ist die mittlere Länge : a) $\bar{R} = 1$, b) $\bar{R} = 0.95$, c) $\bar{R} = 0.85$, d) $\bar{R} = 0.5$, e) $\bar{R} = 0.1$, f) $\bar{R} = 0$. In f) ist die mittlere Richtung $\bar{\theta}$ nicht definiert, da $\bar{C} = \bar{S} = 0$ sind (siehe Def. 2.2).

Folgerung 2.4.5. *Es gilt*

$$\frac{1}{n}\sum_{j=1}^{n}\cos(\theta_j - \bar{\theta}) = \bar{R} \tag{2.5}$$

und für $\bar{R} > 0$

$$\sum_{j=1}^{n}\sin(\theta_j - \bar{\theta}) = 0. \tag{2.6}$$

Beweis. Die Additionstheoreme liefern

$$\cos\left(\theta_j - \bar{\theta}\right) = \cos\theta_j\cos\bar{\theta} + \sin\theta_j\sin\bar{\theta}$$

und

$$\sin\left(\theta_j - \bar{\theta}\right) = \sin\theta_j \cos\bar{\theta} - \cos\theta_j \sin\bar{\theta}.$$

Aus 2.3 und 2.4 folgt insbesondere

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n}\cos\left(\theta_j - \bar{\theta}\right) &= \cos\bar{\theta}\ \frac{1}{n}\sum_{j=1}^{n}\cos\theta_j + \sin\bar{\theta}\ \frac{1}{n}\sum_{j=1}^{n}\sin\theta_j \\
&= \left(\frac{\bar{C}}{\bar{R}}\right)\bar{C} + \left(\frac{\bar{S}}{\bar{R}}\right)\bar{S} \\
&= \frac{1}{\bar{R}}\underbrace{\left(\bar{C}^2 + \bar{S}^2\right)}_{=\bar{R}^2} \\
&= \bar{R}.
\end{aligned}$$

Und analog für die Gleichung 2.6

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n}\sin\left(\theta_j - \bar{\theta}\right) &= \cos\bar{\theta}\ \frac{1}{n}\sum_{j=1}^{n}\sin\theta_j - \sin\bar{\theta}\ \frac{1}{n}\sum_{j=1}^{n}\cos\theta_j \\
&= \left(\frac{\bar{C}}{\bar{R}}\right)\bar{S} - \left(\frac{\bar{S}}{\bar{R}}\right)\bar{C} \\
&= \frac{1}{\bar{R}}\left(\bar{C}\bar{S} - \bar{S}\bar{C}\right) \\
&= 0.
\end{aligned}$$

$$\Rightarrow \sum_{j=1}^{n}\sin\left(\theta_j - \bar{\theta}\right) = 0 \tag{2.7}$$

□

Bemerkung 2.4.6. Die Gleichung 2.7 entspricht $\sum_{j=1}^{n}(x_j - \bar{x}) = 0$ im linearen Fall.

Soll eine Kenngröße beschreiben, wie stark die Daten streuen, so sollte diese Kenngröße unveändert bleiben, wenn der Ursprung um den Winkel α gedreht wird. Diese Eigenschaft wird Rotationsinvarianz genannt und kann analog zu Lokationsinvarianz definiert werden (vgl.[Small 1996]).

Definition 2.4.7 (Rotationsinvarianz)**.** *Eine Schätzfunktion $\hat{\rho} : \mathbb{R}^n \to \mathbb{R}$ heißt rotationsinvariant, wenn für alle $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ und alle $\alpha \in [0°, 360°]$ gilt:*

$$\hat{\rho}(x_1 + \alpha, \ldots, x_n + \alpha) = \hat{\rho}(x).$$

Will man dagegen die Richtung des Datensatzes durch eine Kenngröße angeben, so sollte die zugehörige Richtungschätzfunktion folgende Eigenschaft besitzen: Verschiebt man alle Daten um den Winkel α, so sollte sich die Richtung-Kenngröße auch um diesen Betrag ändern. Diese Eigenschaft wird Rotationsäquivarianz genannt und äquivalent zu Skaleninvarianz definiert (vgl.[Small 1996]).

Definition 2.4.8 (Rotationsäquivarianz)**.** *Eine Schätzfunktion $\hat{\rho} : \mathbb{R}^n \to \mathbb{R}$ heißt rotationsäquivariant, wenn für alle $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ und alle $\alpha \in [0°, 360°]$ gilt:*

$$\hat{\rho}(x_1 + \alpha, \ldots, x_n + \alpha) = \hat{\rho}(x) + \alpha.$$

Lemma 2.4.9 (Eigenschaften des Hauptvektors)**.** *Sei m der Hauptvektor einer Stichprobe $x_1, \ldots, x_n$. Dann hat m folgende Eigenschaften:*

(i) die mittlere Länge des Hauptvektors ist rotationsinvariant

(ii) die mittlere Richtung des Hauptvektors ist rotationsäquivariant

Beweis. Es sei ein neuer Ursprung festgelegt, der mit dem ursprünglichen einen Winkel α bildet. Das heißt, die zu Beobachtungen $x_1, \ldots, x_n$ zugehörige Winkeln haben im neuen Koordinatensystem folgende Darstellung

$$\theta'_j = \theta_j - \alpha, \qquad j = 1, \ldots, n.$$

Setze

$$\bar{C}' = \frac{1}{n}\sum_{j=1}^{n} \cos\theta'_j, \qquad \bar{S}' = \frac{1}{n}\sum_{j=1}^{n} \sin\theta'_j.$$

$$\begin{aligned}
\Rightarrow \bar{C}' &= \frac{1}{n}\sum_{j=1}^{n} \cos\theta'_j \\
&= \frac{1}{n}\sum_{j=1}^{n} \cos\left(\theta_j - \alpha\right) \\
&= \cos\alpha\ \frac{1}{n}\sum_{j=1}^{n} \cos\theta_j + \sin\alpha\ \frac{1}{n}\sum_{j=1}^{n} \sin\theta_j \\
&\overset{(2.3)}{=} \cos\alpha\ \bar{C} + \sin\alpha\ \bar{S} \\
&\overset{(2.4)}{=} \cos\alpha\ \bar{R}\cos\bar{\theta} + \sin\alpha\ \bar{R}\sin\bar{\theta} \\
&= \bar{R}\cos\left(\bar{\theta} - \alpha\right)
\end{aligned}$$

Analog erhält man

$$\bar{S}' = \bar{R}\sin\left(\bar{\theta} - \alpha\right) \tag{2.8}$$

Andererseits haben $\left(\bar{C}', \bar{S}'\right)$ die Polarkoordinaten

$$\bar{C}' = \bar{R}'\sin\bar{\theta}', \qquad \bar{S}' = \bar{R}\sin\bar{\theta}'. \tag{2.9}$$

Vergleich von 2.8 und 2.9 ergibt

$$\bar{\theta}' = \left(\bar{\theta} - \alpha\right) \mod 2\pi, \qquad \bar{R}' = \bar{R}. \tag{2.10}$$

□

Analog zu reellem Median (vgl.[Sachs u. Hedderich 2006]) kann man einen circulären Median einführen.

Definition 2.4.10 (Richtungsmedian). *Der Richtungsmedian für die Daten* $(x_1, \theta_1), \ldots, (x_n, \theta_n)$ *ist ein Winkel* $\tilde{\theta} \in (0, 2\pi]$, *für den gilt:*

$$\#\{j; \theta_j \in [\tilde{\theta}; \tilde{\theta} + \pi)\} \geq \frac{n}{2} \leq \#\{j; \theta_j \in (\tilde{\theta} - \pi; \tilde{\theta}]\}.$$

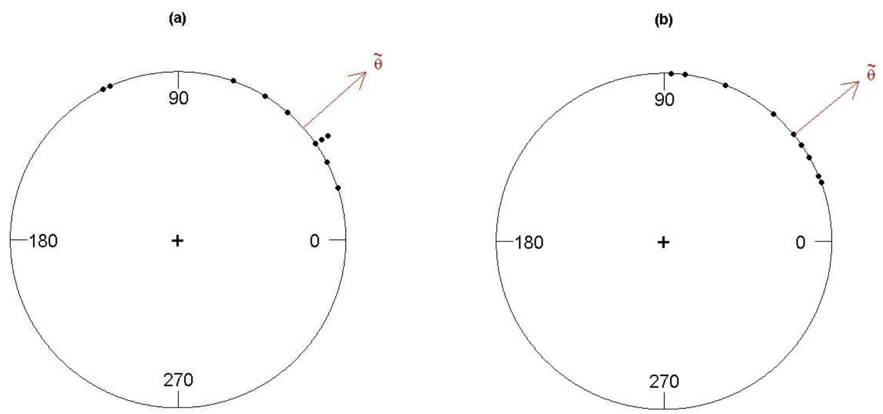

Abbildung 2.6: Veranschaulichung des Richtungsmedians. (a) Gerader Stichprobenumfang (b)Ungerader Stichprobenumfang

Bemerkung 2.4.11. Der Median ist nur für vernünftige unimodale Daten eindeutig. Für multimodale oder isotrope Daten kann der Median nicht eindeutig sein (siehe Abbildung 2.7). Ist der Stichprobenumfang n ungerade, so ergibt sich der zu der in der Mitte liegenden Beobachtung zugehörige Winkel des geordneten Datensatzes als Median. Ist n gerade, so ist der zu dem Mittelpunkt zwischen zwei geeigneten Beobachtungen zugehörige Winkel der Median.

Beispiel 2.4.12. Man kann sich die Frage stellen, ob man für den Datensatz 1.1 den Median ausrechnen kann. Da man im Fall der Mottenanzahlen mit multimodalen Daten zu tun hat, ist es möglich, dass der Median nicht eindeutig ist. Da Monate jedes Jahr nicht unbedingt gleich viele Wochen haben, und es wenig Sinn machen würde, die unterschiedliche Anzahl der Wochen zusammenzufassen, wird der Median nur für das Jahr 1999 ausgerechnet. Gegeben sind:

$$n = 34989 - \text{ die Anzahl der Motten im Jahr.}$$

$$\Rightarrow \frac{n}{2} = 17495.$$

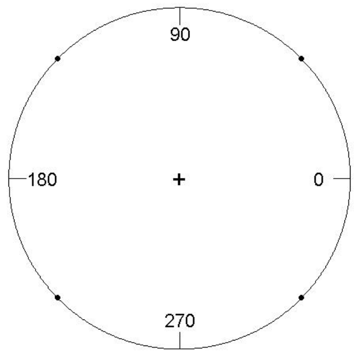

Abbildung 2.7: Die Beobachtungen sind gleichmäßig verteilt. Der Richtungsmedian ist in diesem Fall nicht eindeutig.

Zu bestimmen ist also ein Winkel $\tilde{\theta} \in (0, 2\pi]$, für den gilt

$$\#\{j; \theta_j \in [\tilde{\theta}; \tilde{\theta} + \pi)\} \geq 17495 \leq \#\{j; \theta_j \in (\tilde{\theta} - \pi; \tilde{\theta}]\}.$$

Die Wochen im Jahr werden mit Winkeln identifiziert, d.h. die erste Woche $\stackrel{\wedge}{=} 0°$, die zweite $\stackrel{\wedge}{=} \frac{360°}{52} \approx 6.923° \approx 7°$ u.s.w. Der Datensatz hat nun folgende Gestalt:

$$\underbrace{0°, \ldots, 0°}_{\text{656-mal}}, \underbrace{7°, \ldots, 7°}_{\text{298-mal}}, \ldots, \underbrace{182°, \ldots, 182°}_{\text{1932-mal}}, \underbrace{189°, \ldots, 189°}_{\text{3465-mal}},$$

$$\underbrace{196°, \ldots, 196°}_{\text{2874-mal}}, \ldots, \underbrace{357°, \ldots, 357°}_{\text{616-mal}}.$$

Der Winkel, der obige Anforderungen erfüllt, ist gleich 189°, also die erste Juliwoche.

Bisher wurde die Konzentration der Beobachtungen eines Datensatzes untersucht. Man ist aber auch an der Streuung der einzelnen Werte um die mittlere Richtung interessiert, d.h. wie dicht an diese Richtungmaßzahl die Beobachtungen liegen. In circulärer Statistik wird die Abweichung von der mittleren Richtung in $\sin(\theta_j - \bar{\theta})$ und $\cos(\theta_j - \bar{\theta})$ gemessen (siehe Abb. 2.8). Der Abstand zwischen zwei Winkeln θ_i und θ_j kann man also mit $1 - \cos(\theta_i - \theta_j)$ ausdrücken ([Batschelet 1981]).

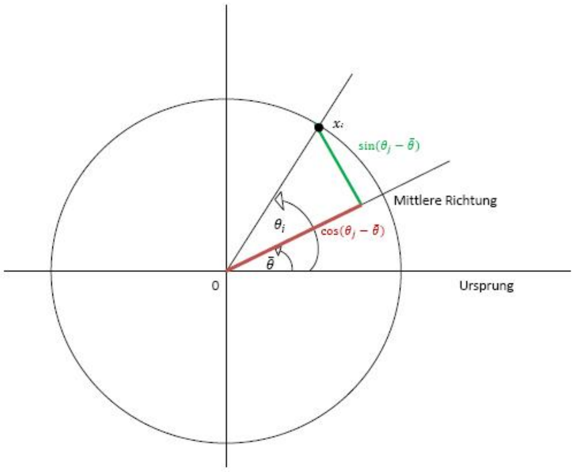

Abbildung 2.8: In circulärer Statistik wird die Abweichung von der mittleren Richtung in $\sin(\theta_j - \bar{\theta})$ und $\cos(\theta_j - \bar{\theta})$ gemessen.

Die einfachste Masszahl für die Streuung der Daten ist die empirische Varianz. Die kann man auf zwei verschiedene Weisen definieren. Die erste intuitive Definition der Varianz basiert auf der mittleren Länge.

Definition 2.4.13 (Richtungsvarianz)**.** *Es seien circuläre Beobachtungen $x_1, \ldots, x_N$ mit dazugehörigen Winkeln θ_j,*
$j = 1, \ldots, N$ (d.h. $x_j = (\cos\theta_j, \sin\theta_j)^T$ für $j = 1, \ldots, N$) gegeben. Dann heißt

$$V = 1 - \bar{R}$$

empirische circuläre Varianz, wobei $\bar{R}$ die mittlere Länge (2.1) bezeichne ([Mardia u. Jupp 2000]).

Folgerung 2.4.14. *Sind die Beobachtungen $x_1, \ldots, x_n$ Einheitsvektoren, dann gilt*

$$0 \leq V \leq 1.$$

Beweis. Die Behauptung folgt sofort aus Bemerkung 2.4.4. □

Die empirische Varianz hängt also mit der mittleren Länge zusammen. Je größeren Wert $\bar{R}$ hat, desto kleiner ist die Varianz. Die andere Idee, die Varianz zu bestimmen, basiert auf Minimierung der Summe der Abstände aller Beobachtungen. Man geht also

von dem Abstand zwischen dem Winkel θ_j $(j = 1, \dots, n)$ und einem gegeben Winkel α aus, man summiert über alle Winkel und dann minimiert man diese Summe. Sei $\bar{R} \in (0, 1]$. Im Fall $\bar{R} = 0$ ist die circuläre Varianz gleich 1.
Der Abstand zwischen den Winkeln θ und α ist gegeben durch

$$1 - \cos(\theta - \alpha).$$

Um die Streuung von den Winkeln θ_j $(j = 1, \dots, n)$ um die gegebene Richtung α zu messen, wird der Mittelwert gebildet

$$S(\alpha) = \frac{1}{n} \sum_{j=1}^{n} \left(1 - \cos(\theta_j - \alpha)\right). \tag{2.11}$$

Differenzieren nach α und Nullsetzen ergibt

$$\begin{aligned} S'(\alpha) &= \frac{1}{n} \sum_{j=1}^{n} (-\sin(\theta_j - \alpha)) = 0 \\ \Longleftrightarrow & \sum_{j=1}^{n} \sin(\theta_j - \alpha) = 0 \\ \overset{\text{Fol.2.4.5}}{\Longrightarrow} & \ \alpha = \bar{\theta}. \end{aligned}$$

Die zweite Ableitung bei $\alpha = \bar{\theta}$

$$S''(\alpha) = \frac{1}{n} \sum_{j=1}^{n} \cos(\theta_j - \alpha)|_{\alpha = \bar{\theta}} = \bar{R}$$

ist für $\bar{R} \neq 0$ größer Null. Somit hat die Gleichung 2.11 (für $\bar{R} \neq 0$) eine Minimalstelle bei $\alpha = \bar{\theta}$.

Es gilt also für die Streuung um die mittlere Richtung $\bar{\theta}$:

$$\begin{aligned} S(\bar{\theta}) &= \frac{1}{n}\sum_{j=1}^{n}\big(1-\cos(\theta_j-\bar{\theta})\big) \\ &= 1-\underbrace{\frac{1}{n}\sum_{j=1}^{n}\cos(\theta_j-\bar{\theta})}_{=\bar{R}} \\ &= V. \end{aligned}$$

Circuläre Spannweite ist der kleinste Winkel, der alle Beobachtungen enthält. In [Mardia u. Jupp 2000] wird sie folgendermaßen definiert.

Definition 2.4.15 (Circuläre Spannweite). *Seien $x_1, x_2, \ldots, x_n$ n Beobachtungen mit den dazugehörigen Winkeln $\theta_1, \theta_2, \ldots, \theta_n$, $\theta_j \in [0, 2\pi]$ für alle $j \in \{1, \ldots, n\}$ gegeben. Sei $\theta_{(1)}, \ldots, \theta_{(n)}$ die geordnete Stichprobe. Definiere*

$$T_j = \theta_{(j+1)} - \theta_{(j)}, \qquad j = 1, \ldots, n-1, \qquad T_n = 2\pi - \theta_{(n)} + \theta_{(1)}.$$

Dann ist die circuläre Spannweite gegeben durch

$$\omega = 2\pi - \max(T_1, \ldots, T_n).$$

Im stetigen Fall ist der Träger normalerweise $(0, 2\pi)$. Die Spannweite beträgt also 2π.

Die Definition 2.4.15 macht Sinn, denn zuerst werden die Abstände $T_1, \ldots, T_n$ zwischen den n Beobachtungen berechnet und dann wird der größte Abstand von 2π abgezogen. Es ist klar, dass in dem Winkel ω alle Beobachtungen liegen müssen, d.h. ω ist die circuläre Spannweite. ω ist auch rotationsinvariant. Man kann sich dies an einem kleinen Beispiel veranschaulichen.

Beispiel 2.4.16. Sei $\epsilon > 0$. Es seien 2 Beobachtungen x_1, x_2 mit den dazugehörigen Winkeln $\theta_1 = \epsilon, \theta_2 = 2\pi - \epsilon$ gegeben. Die Stichprobe ist bereits geordnet, d.h. $\theta_{(1)} = \theta_1$ und $\theta_{(2)} = \theta_2$. Es gilt

$$\begin{aligned} T_1 &= \theta_{(2)} - \theta_{(1)} = 2\pi - \epsilon - \epsilon = 2\pi - 2\epsilon \\ T_n &= T_2 = 2\pi - \theta_{(2)} + \theta_{(1)} = 2\pi - 2\pi + \epsilon + \epsilon = 2\epsilon \\ \omega &= 2\pi - \max(T_1, T_2) = 2\pi - (2\pi - 2\epsilon) = 2\epsilon \end{aligned}$$

Die circuläre Spannweite ω beträgt also 2ϵ.
Es sei ein neuer Ursprung festgelegt, der mit dem ursprünglichen einen Winkel α, $|\alpha| > 0$ bildet. Die Beobachtungen x_1 und x_2 haben im neuen Koordinatensystem die Darstellung

$$\begin{aligned} \theta'_1 &= \theta_1 + \alpha = \alpha + \epsilon \\ \theta'_2 &= \theta_2 + \alpha = 2\pi - \epsilon + \alpha = \alpha - \epsilon \end{aligned}$$

Die Ordnung der neuen Stichprobe hat sich geändert, nämlich

$$\begin{aligned} \theta'_{(1)} &= \theta'_2 = \alpha - \epsilon \\ \theta'_{(2)} &= \theta'_1 = \alpha + \epsilon \end{aligned}$$

Bestimmung von T'_1 und T'_2

$$\begin{aligned} T'_1 &= \theta'_{(2)} - \theta'_{(1)} = \epsilon + \alpha - \alpha + \epsilon = 2\epsilon \\ T'_2 &= 2\pi - \theta'_{(2)} + \theta'_{(1)} = 2\pi - \epsilon - \alpha + \alpha - \epsilon = 2\pi - 2\epsilon \\ \omega' &= 2\pi - \max(T'_1, T'_2) = 2\pi - (2\pi - 2\epsilon) = 2\epsilon \end{aligned}$$

$$\Rightarrow \omega' = \omega$$

Also ist die circuläre Spannweite rotationsinvariant.

Die in Definition 2.4.2 definierten Momente

$$\bar{C} = \frac{1}{n}\sum_{j=1}^{n}\cos\theta_j, \qquad \bar{S} = \frac{1}{n}\sum_{j=1}^{n}\sin\theta_j$$

spielen Schlüsselrolle in Bestimmung der empirischen mittleren Richtung.

$$m_1' = \bar{C} + \imath\bar{S} = \bar{R}\mathrm{e}^{\imath\bar{\theta}} \tag{2.12}$$

heißt erstes trigonometrisches Moment bezüglich des Ursprungs.

Zweites trigonometrisches Moment bezüglich des Ursprungs bekommt man, indem man für die doppelten Winkel $\theta_1, \ldots, \theta_n$ die mittlere Länge $\bar{R}_2$ und die mittlere Richtung $\bar{\theta}_2$ ausrechnet:

$$m_2' = a_2 + \imath b_2 = \bar{R}_2\mathrm{e}^{\imath\bar{\theta}_2}, \tag{2.13}$$

wo

$$a_2 = \frac{1}{n}\sum_{j=1}^{n}\cos 2\theta_j, \quad b_2 = \frac{1}{n}\sum_{j=1}^{n}\sin 2\theta_j.$$

Allgemein kann man p-tes trigonometrisches Moment bezüglich des Ursprungs definieren:

$$m_p' = a_p + \imath b_p = \bar{R}_p\mathrm{e}^{\imath\bar{\theta}_p}, \tag{2.14}$$

wo

$$a_p = \frac{1}{n}\sum_{j=1}^{n}\cos p\theta_j, \quad b_p = \frac{1}{n}\sum_{j=1}^{n}\sin p\theta_j.$$

Und schließlich p-tes zentriertes trigonometrisches Moment:

$$m_p = \bar{a}_p + \imath\bar{b}_p = \bar{R}_p\mathrm{e}^{\imath(\bar{\theta}_p - p\bar{\theta})}$$

mit

$$\bar{a}_p = \frac{1}{n}\sum_{j=1}^{n}\cos p(\theta_j - \bar{\theta}), \quad \bar{b}_p = \frac{1}{n}\sum_{j=1}^{n}\sin p(\theta_j - \bar{\theta}).$$

Mithilfe des ersten und zweiten zentrierten trigonometrischen Moments werden circuläre Schiefe und Wölbung ermittelt.

Bemerkung 2.4.17. Speziell für m_1 gilt:

$$m_1 = \frac{1}{n} \sum_{j=1}^{n} \cos(\theta_j - \bar{\theta}) = \bar{R},$$

da der Sinus-Term nach Folgerung 2.4.5 für $\bar{R} > 0$ verschwindet:

$$\sum_{j=1}^{n} \sin(\theta_j - \bar{\theta}) = 0.$$

Die in Definition 2.4.13 definierte circuläre Varianz hängt also linear von dem ersten zentrierten Moment ab.

3 Circuläre Verteilungen

3.1 Einführung

In diesem Kapitel werden die wichtigsten Verteilungen vorgestellt. Circuläre Verteilungen kann man in 3 Gruppen unterteilen

- gewickelte Verteilungen - sie entstehen durch Wickeln linearer Verteilungen um den Einheitskreis
- projizierte Verteilungen - sie entstehen durch radiales Projizieren linearer Verteilungen
- von Mises Verteilung.

In dieser Arbeit werden nur gewickelte Verteilungen und die von Mises Verteilung betrachtet.

Der Träger von linearen Verteilungen kann endlich, wie bei der Binomial Verteilung, oder auch unendlich, wie bei der Poisson Verteilung, sein. Der Träger von circulären Verteilungen ist dagegen immer endlich, nämlich $T_\Theta = [0, 2\pi]$ bzw. $T_\Theta = [0°, 360°]$.

Um Wahrscheinlichkeitsverteilungen definieren zu können, wird es an die Begriffe wie Wahrscheinlichkeitsmaß, σ-Algebra und Wahrscheinlichkeitsraum erinnert (vgl.[Krengel 2005]).

Definition 3.1.1. *(i) Eine nichtleere Menge Ω heißt Ereignisraum*

(ii) Ein System $\mathscr{A}$ von Teilmengen von Ω, also $\mathscr{A} \subset \mathscr{P}(\Omega)$, heißt σ-Algebra in Ω, falls

- $\Omega \in \mathscr{A}$
- $A \in \mathscr{A} \Rightarrow \bar{A} \in \mathscr{A}$

- $A_1, A_2, \ldots \in \mathscr{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$

(iii) Sei $\Omega \neq \emptyset$ und $\mathscr{A}$ eine σ-Algebra in Ω. Eine Abbildung $P : \mathscr{A} \ni A \to P(A) \in [0,1]$ heißt ein Wahrscheinlichkeitsmaß auf $\mathscr{A}$, falls

- $P(A) \geq 0$, *für alle* $A \in \mathscr{A}$
- $P(\Omega) = 1$
- $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, *für alle* $(A_i)_{i=1,2,\ldots} \in A, \quad A_i \cap A_j = \emptyset, \ i \neq j$

Das Tupel $(\Omega, \mathscr{A})$ heißt messbarer Raum. Das Tripel $(\Omega, \mathscr{A}, P)$ heißt Wahrscheinlichkeitsraum.

Definition 3.1.2. *Seien $(\Omega, \mathscr{A})$ und $(\Omega', \mathscr{B})$ messbare Räume.*

- *Eine Abbildung Θ: $\Omega \ni \omega \to \Theta(\omega) \in \Omega'$ heißt $\mathscr{A}$-$\mathscr{B}$-messbar $\Leftrightarrow$*

$$\Theta^{-1}(B) := \{\omega \in \Omega | \ \Theta(\omega) \in B\} \in \mathscr{A}, \quad \text{für alle } B \in \mathscr{B}$$

- *Eine Abbildung Θ: $\Omega \ni \omega \to \Theta(\omega) \in \Omega'$ heißt Zufallsvariable auf $(\Omega, \mathscr{A})$ mit Werten in $(\Omega', \mathscr{B})$, falls $(\Omega, \mathscr{A}, P)$ ein Wahrscheinlichkeitsraum ist und Θ $\mathscr{A}$-$\mathscr{B}$-messbar ist.*

Definition 3.1.3 (Zufallsgröße, Zufallsvektor)**.** *Sei $\mathscr{B}$ bzw. $\mathscr{B}_n$ die Borel-σ-Algebra über $(0, 2\pi]$ bzw. $(0, 2\pi]^n$. Eine Zufallsvariable Θ auf $(\Omega, \mathscr{A}, P)$ mit Werten in $((0, 2\pi], \mathscr{B})$ bzw. in $((0, 2\pi]^n, \mathscr{B}_n)$ heißt Zufallsgröße bzw. Zufallsvektor.*

Definition 3.1.4 (Verteilungsfunktion)**.** *Ist Θ eine Zufallsgröße, dann heißt*

$$F_\Theta : (0, 2\pi] \ni \theta \to F_\Theta(\theta) := P^\Theta((0, \theta]) = P(0 < \Theta \leq \theta) \in [0,1]$$

die Verteilungsfunktion von Θ.

Bemerkung 3.1.5. Die Definition 3.1.4 bleibt gültig für alle $\theta \in \mathbb{R}\backslash(0, 2\pi]$ mit folgender Vereinbarung

(i) Fall 1: $\theta \leq 0$

$$F_\Theta(\theta) := F_\Theta(k2\pi + \theta) \in [0,1], \text{für ein geeignetes } k \in \mathbb{N}, \text{ so dass } k2\pi + \theta \in (0, 2\pi]$$

(ii) Fall 2: $\theta > 2\pi$

$$F_\Theta(\theta) := F_\Theta(\theta \bmod 2\pi) \in [0,1].$$

Satz 3.1.6 (Eigenschaften der Verteilungsfunktion einer circulären Zufallsgröße)**.** *Ist* $F_\Theta : (0, 2\pi] \ni \theta \to F_\Theta(\theta) \in [0,1]$ *eine Verteilungsfunktion, dann gilt:*

(i) F_Θ *ist monoton wachsend, d.h.* $\theta_1 < \theta_2 \Rightarrow F_\Theta(\theta_1) \leq F_\Theta(\theta_2)$

(ii) F_Θ *ist rechtsstetig, d.h.* $\theta_n \downarrow \theta \Rightarrow lim_{n\to\infty} F_\Theta(\theta_n) = F_\Theta(\theta)$

(iii) $\lim_{\theta \to 2\pi} F_\Theta(\theta) = 1$

(iv) $\lim_{\theta \to 0} F_\Theta(\theta) = 0$

(v) Für $0 < \theta_1 < \theta_2 \leq 2\pi$ *ist* $F_\Theta(\theta_1) - F_\Theta(\theta_2)$ *unabhängig von Auswahl des Ursprungs*

(vi) Der Wert der F_Θ *hängt von Auswahl des Ursprungs ab*

Beweis. (i)-(iv) Siehe [Krengel 2005],

(v) Folgt aus (vi),

(vi) Klar, da Verschiebung des Ursprungs den Wert von F_Θ um eine Konstante ändert.

□

Man kann die trigonometrischen Momente als Terme der charakteristischen Funktion interpretieren (vgl.[Lee 2010], [Fisher 1993]). Die Dichte einer circulären Variablen Θ ist periodisch. Daraus folgt, dass Θ und $\Theta + 2\pi$ die gleiche Verteilung haben:

$$\varphi_\Theta(t) = E(\mathrm{e}^{\imath t\Theta}) = E(\mathrm{e}^{\imath t(\Theta + 2\pi)}) = \mathrm{e}^{\imath t2\pi} E(\mathrm{e}^{\imath t\Theta}).$$

Die obige Gleicheit gilt für $\varphi_\Theta(t) = 0$ oder für $\mathrm{e}^{\imath t2\pi} = 1$. D.h., es macht Sinn, die charakteristische Funktion circulärer Zufallsgrößen nur für $t \in \mathbb{Z}$ zu betrachten.

Definition 3.1.7 (Charakteristische Funktion)**.** *Sei Θ eine circuläre Zufallsgröße. Die Funktion $\varphi_\Theta : \mathbb{Z} \to \mathbb{C}$ gegeben durch*

$$\varphi_\Theta(t) := E(\mathrm{e}^{\imath t\Theta}) = E(cos(t\Theta)) + \imath E(sin(t\Theta)) := \alpha_t + \imath\beta_t \tag{3.1}$$

heißt charakteristische Funktion der (Verteilung der) Zufallsgröße Θ ([Mardia u. Jupp 2000]).

Satz 3.1.8. *Die charakteristische Funktion von Θ lässt sich auch in Integralform schreiben*

$$\begin{align} \varphi_\Theta(t) &= \int_0^{2\pi} \mathrm{e}^{\imath t\theta} f_\Theta(\theta) d\theta \tag{3.2} \\ &= \int_0^{2\pi} \cos t\theta f_\Theta(\theta) d\theta + \imath \int_0^{2\pi} \sin t\theta f_\Theta(\theta) d\theta, \tag{3.3} \end{align}$$

wobei f_Θ die Dichte von Θ. Dann gilt

$$\varphi_\Theta(0) = 1, \quad \bar{\varphi}_\Theta(t) = \varphi_\Theta(-t), \quad |\varphi_\Theta(t)| \leq 1. \tag{3.4}$$

Beweis. Die Eigenschaft 3.2 folgt aus Definition des Erwartungswertes. Nun zu den Eigenschaften 3.4:

$$\varphi_\Theta(0) = \int_0^{2\pi} \mathrm{e}^{\imath 0\theta} f_\Theta(\theta) d\theta = \int_0^{2\pi} f_\Theta(\theta) d\theta = 1,$$

da f_Θ eine Dichte.
Mit den Rechenregeln für komplexe Zahlen gilt

$$\bar{\varphi}_\Theta(t) \overset{\text{komplex konj.}}{=} \int_0^{2\pi} \mathrm{e}^{-\imath t\theta} f_\Theta(\theta) d\theta = \int_0^{2\pi} \mathrm{e}^{\imath(-t)\theta} f_\Theta(\theta) d\theta = \varphi_\Theta(-t),$$

und

$$\begin{aligned} |\varphi_\Theta(t)| &= |\int_0^{2\pi} \mathrm{e}^{\imath t\theta} f_\Theta(\theta) d\theta| \\ &\leq \int_0^{2\pi} \underbrace{|\mathrm{e}^{\imath t\theta}|}_{=1} \underbrace{|f_\Theta(\theta)|}_{=f_\Theta(\theta),\ \text{da}\ f_\Theta(\theta)\geq 0\ \text{f.a.}\ \theta\in\Theta} d\theta = \int_0^{2\pi} f_\Theta(\theta) d\theta = 1. \end{aligned}$$

$\square$

3.2 Verteilungen

Im Folgenden werden die wichtigsten Verteilungen eingeführt (vgl.[Mardia u. Jupp 2000]). Die von Mises Verteilung spielt in der Kreisstatistik eine große Rolle. Man kann sie als Approximation der gewickelten Normalverteilung ansehen, die wieder von enormer Bedeutung in linearer Statistik ist.

3.2.1 Einpunktverteilung $\mathcal{E}_{\mu_0}$

- Träger: $T_\Theta = \{\mu_0\}$
- Zähldichte: $p_\Theta(\mu_0) = 1$
- Charakteristische Funktion: $\varphi_\Theta(t) = \mathrm{e}^{\imath t \mu_0}$

3.2.2 Gitter-Verteilung $\mathcal{G}[\nu, m]$

- Dichte:

$$p_\Theta(\nu + \frac{2\pi r}{m}) = p_r, \qquad \nu \in [0, 2\pi m), \quad m \text{ beliebig, fest,}$$
$$r = 0, \dots, m-1, \quad p_r \geq 0, \quad \sum_{r=0}^{m-1} p_r = 1$$

- Bemerkung: Die Punkte $\nu + \frac{2\pi r}{m}$ sind äquidistant und bilden dementsprechend Ecken eines regelmäßigen in einen Einheitskreis eingeschriebenen m-Eckes.
- Charakteristische Funktion: Für ν gleich Null ist die charakteristische Funktion gegeben durch

$$\varphi_\Theta(t) = \sum_{r=0}^{m-1} p_r \mathrm{e}^{\imath t \frac{2\pi r}{m}}$$

Für $p_1 = \ldots = p_{m-1} = \frac{1}{m}$ heißt die Gitter-Verteilung auch diskrete Gleichverteilung. In diesem Fall vereinfacht sich die charakteristische Funktion zu

$$\varphi_\Theta(t) = \begin{cases} 1 & \text{falls } t = 0 \bmod m, \\ 0 & \text{sonst,} \end{cases}$$

da

1. Sei $t = km, k \in \mathbb{N}$. Dann gilt

$$\begin{aligned}
\varphi_\Theta(km) &= \sum_{r=0}^{m-1} \frac{1}{m} \mathrm{e}^{2\pi\imath rk} \overset{k \in \mathbb{N}}{=} \frac{1}{m} \sum_{r=0}^{m-1} \mathrm{e}^{2\pi\imath r} \\
&= \frac{1}{m} (\mathrm{e}^0 + \mathrm{e}^{2\pi\imath} + \mathrm{e}^{4\pi\imath} + \ldots + \mathrm{e}^{2\pi\imath(m-1)}) \\
&= \frac{1}{m} m \\
&= 1.
\end{aligned}$$

2. Sei $t = km + a, k \in \mathbb{N}, a \in \{1, \ldots, m-1\}$. Dann gilt

$$\begin{aligned}
\varphi_\Theta(km + a) &= \sum_{r=0}^{m-1} \frac{1}{m} \mathrm{e}^{\imath(km+a)\frac{2\pi r}{m}} \\
&= \frac{1}{m} \sum_{r=0}^{m-1} \mathrm{e}^{2\pi\imath rk} \mathrm{e}^{\imath a \frac{2\pi r}{m}} \\
&\overset{k \in \mathbb{N}, a \in \mathbb{N}}{=} \frac{1}{m} \sum_{r=0}^{m-1} \mathrm{e}^{2\pi\imath r} \mathrm{e}^{\imath \frac{2\pi r}{m}} \\
&= \frac{1}{m} (\mathrm{e}^0 \mathrm{e}^0 + \mathrm{e}^{2\pi\imath} \mathrm{e}^{\frac{2\pi\imath}{m}} + \mathrm{e}^{4\pi\imath} \mathrm{e}^{\frac{4\pi\imath}{m}} + \ldots + \mathrm{e}^{2\pi\imath(m-1)} \mathrm{e}^{\frac{(m-1)\pi\imath}{m}}) \\
&= \frac{1}{m} (\underbrace{\mathrm{e}^0 + \mathrm{e}^{\frac{2\pi\imath}{m}} + \mathrm{e}^{\frac{4\pi\imath}{m}} + \ldots + \mathrm{e}^{\frac{(m-1)\pi\imath}{m}}}_{\text{die m-ten Einheitswurzeln}}) \\
&= \frac{1}{m} 0 \\
&= 0.
\end{aligned}$$

3.2.3 Stetige Gleichverteilung $\mathcal{U}[0, 2\pi]$

- Dichte: $f_\Theta(\theta) = \frac{1}{2\pi}, \qquad 0 < \theta \leq 2\pi$
- Verteilungsfunktion: $F_\Theta(\theta) = \frac{\theta}{2\pi}$
- Charakteristische Funktion:

$$\varphi_\Theta(t) = \begin{cases} 1 & \text{falls } t = 0, \\ 0 & \text{falls } t \neq 0, \end{cases}$$

da

1. Sei $t = 0$. Dann gilt

$$\begin{aligned}\varphi_\Theta(0) &= \int_0^{2\pi} \mathrm{e}^{\imath 0 \theta} f_\Theta(\theta) d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\theta = 1\end{aligned}$$

2. Sei $t \neq 0$. Dann gilt

$$\begin{aligned}\varphi_\Theta(t) &= \int_0^{2\pi} \mathrm{e}^{\imath t \theta} \frac{1}{2\pi} d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} \cos(t\theta) d\theta + \imath \frac{1}{2\pi} \int_0^{2\pi} \sin(t\theta) d\theta \\ &= \frac{1}{2\pi} \left(\int_0^{2\pi} \cos(t\theta) d\theta + \imath \int_0^{2\pi} \sin(t\theta) d\theta \right) \\ &= \frac{1}{2\pi} \left(\left[\frac{1}{t} \sin(t\theta) \right]_0^{2\pi} + \imath \left[-\frac{1}{t} \cos(t\theta) \right]_0^{2\pi} \right) \\ &= \frac{1}{2\pi} \left(\frac{1}{t} \sin(2\pi t) - \frac{1}{t} \sin(0) - \imath \frac{1}{t} \cos(2\pi t) + \imath \frac{1}{t} \cos(0) \right) \\ &\overset{t \in \mathbb{Z}}{=} \frac{1}{2\pi} \left(0 - 0 - \imath \frac{1}{t} + \imath \frac{1}{t} \right) \\ &= 0.\end{aligned}$$

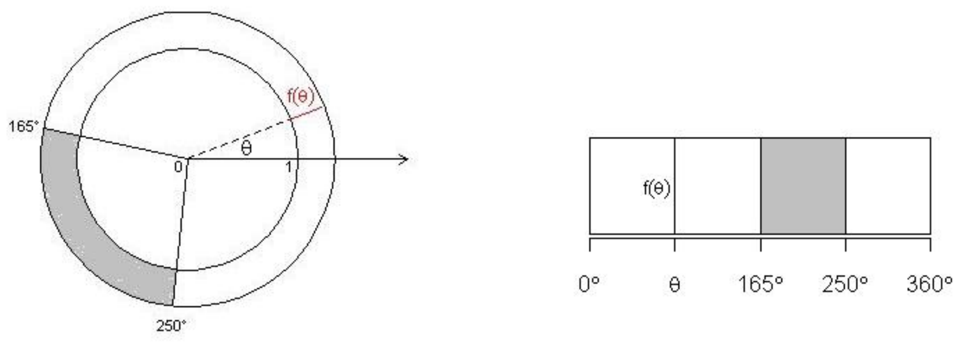

Abbildung 3.1: Die stetige Gleichverteilung. Links: circulärer Plot, rechts: linearer Plot. Die Wahrscheinlichkeit $P(165° \leq \Theta \leq 250°)$ ist schattiert ([Batschelet 1981]).

3.2.4 Gewickelte Verteilungen

Eine gewickelte Verteilung ist eine stetige Verteilung auf dem Kreis, die entsteht, wenn man eine lineare Verteilung um den Umfang des Einheitskreises wickelt (vgl.[Mardia 1972]).

Definition 3.2.1. *Sei X eine lineare Zufallsvariable, mit Dichte f_X und Verteilungsfunktion F_X. Dann ist die zugehörige Zufallsvariable Θ der gewickelten Verteilung, ihre Dichte f_Θ und Verteilungsfunktion F_Θ gegeben durch*

$$X^w := \Theta := X \bmod 2\pi, \tag{3.5}$$

$$f_X^w(\theta) := f_\Theta(\theta) := \sum_{k=-\infty}^{\infty} f_X(\theta + 2\pi k) \tag{3.6}$$

und

$$F_X^w(\theta) := F_\Theta(\theta) := \sum_{k=-\infty}^{\infty} \{F_X(\theta + 2\pi k) - F_X(2\pi k)\}, \qquad 0 \leq \theta \leq 2\pi. \tag{3.7}$$

Satz 3.2.2 (Eigenschaften gewickelter Verteilungen)**.** *Seien X, Y zwei lineare Zufallsgößen und X^w, Y^w die daraus abgeleiteten circulären Zufallsgrößen. Dann gilt*

(i) $(X + Y)^w = X^w + Y^w$

(ii) Ist $\varphi_X(t)$ die charakteristische Funktion von X, dann ist die charakteristische Funktion $\varphi_\Theta(t)$ von $\Theta = X \bmod 2\pi$ gegeben durch

$$\varphi_\Theta(t) = \varphi_X(t). \tag{3.8}$$

(iii) Die gewickelte Dichte f_Θ lässt sich durch eine unendliche Fourierreihe darstellen

$$f_\Theta(\theta) = \frac{1}{2\pi}\left(1 + 2\sum_{t=1}^{\infty}(\alpha_t \cos t\theta + \beta_t \sin t\theta)\right),$$

wobei α_t bzw. β_t der Realteil bzw. der Imaginärteil der charakteristischen Funktion $\varphi_X(t)$ sind.

(iv) Zu jeder Verteilung auf dem Kreis gibt es eine Verteilung in $\mathbb{R}$, die gewickelt, diese Verteilung ergibt.

Beweis.

(i) Die Aussage ist äquivalent zu $X + Y \bmod 2\pi = X \bmod 2\pi + Y \bmod 2\pi$, die aus den Rechenregeln für Division mit Rest folgt.

(ii)

$$\begin{aligned}\varphi_\Theta(t) &\overset{\text{Bem.3.1.8}}{=} \int_0^{2\pi} \mathrm{e}^{\imath t\theta} f_\Theta(\theta) d\theta \\ &\overset{\text{Def.3.2.1}}{=} \int_0^{2\pi} \mathrm{e}^{\imath t\theta} \sum_{k=-\infty}^{\infty} f_X(\theta + 2\pi k) d\theta \\ &\overset{\text{Linearität}}{=} \sum_{k=-\infty}^{\infty} \int_0^{2\pi} \mathrm{e}^{\imath t\theta} f_X(\theta + 2\pi k) d\theta \\ &= \sum_{k=-\infty}^{\infty} \int_{2\pi k}^{2\pi(k+1)} \mathrm{e}^{\imath t\theta} f_X(\theta) d\theta \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{\imath tx} f_X(x) dx \\ &= \varphi_X(t).\end{aligned}$$

(iii) Dichte einer circulären Variablen $f_\Theta(\theta)$ ist eine 2π-periodische Funktion. $f_\Theta(\theta)$ ist auf dem Intervall $[0, 2\pi]$ beschränkt, und stetig als Summe von stetigen Dichten (Def.3.2.1). Fourierreihe für eine Funktion f mit Periode 2π lautet ([Weber u. Ulrich 2007], [Heuser 2008]):

$$f_\Theta(\theta) = a_0 + \sum_{t=1}^{\infty} (a_t \cos(t\theta) + b_t \sin(t\theta)). \tag{3.9}$$

Den Koeffizienten a_0 der Dichte $f_\Theta(\theta)$ erhält man durch Integration der obigen Fourierreihe über das Intervall $[0, 2\pi]$, d.h. über eine volle Periode

$$\underbrace{\int_0^{2\pi} f_\Theta(\theta) d\theta}_{=1,\text{ da } f_\Theta \text{ Dichte}} = \int_0^{2\pi} a_0 d\theta + \sum_{t=1}^{\infty} \left(\underbrace{\int_0^{2\pi} a_t \cos(t\theta) d\theta}_{=0} + \underbrace{\int_0^{2\pi} b_t \sin(t\theta)) d\theta}_{=0} \right) = a_0 2\pi$$

$$\Rightarrow a_0 = \frac{1}{2\pi}.$$

Um die Fourierkoeffizienten a_t bzw. b_t für $t > 0$ zu bestimmen, wird für die Gleichung 3.9 vorübergehend k als Summationsindex gewählt, dann mit $\cos(t\theta)$ bzw. $\sin(t\theta)$ ($t \in \mathbb{Z}$) multipliziert und anschließend über eine Periode integriert:

$$\int_0^{2\pi} f_\Theta(\theta)\cos(t\theta)d\theta = a_0 \underbrace{\int_0^{2\pi} \cos(t\theta)d\theta}_{=0} + \sum_{k=1}^{\infty}\left(a_k \int_0^{2\pi} \cos(k\theta)\cos(t\theta)d\theta + b_k \int_0^{2\pi} \sin(k\theta)\cos(t\theta)d\theta\right) \tag{3.10}$$

Da $\int_0^{2\pi} \cos(k\theta)\cos(t\theta)d\theta = \pi$ für $k = t$ und sonst 0 ist, und $\int_0^{2\pi} \sin(k\theta)\cos(t\theta)d\theta = 0$ für alle k und t ist, folgt

$$\int_0^{2\pi} f_\Theta(\theta)\cos(t\theta)d\theta = a_t\pi$$

$$\Rightarrow a_t = \frac{1}{\pi}\int_0^{2\pi} f_\Theta(\theta)\cos(t\theta)d\theta$$

Nach Definition 3.1.7 ist $\int_0^{2\pi} f_\Theta(\theta)\cos(t\theta)d\theta = \alpha_t$. Damit ergibt sich für den Koeffizienten a_t:

$$a_t = \frac{1}{\pi}\alpha_t$$

Durch Multiplikation der Gleichung 3.9 (nach vorheriger Umbennenung des Summationsindexes) mit $\sin(t\theta)$ und anschließender Integration über eine volle Periode erhält man

$$\int_0^{2\pi} f_\Theta(\theta)\sin(t\theta)d\theta = a_0 \underbrace{\int_0^{2\pi} \sin(t\theta)d\theta}_{=0} + \sum_{k=1}^{\infty}\left(a_k \int_0^{2\pi} \cos(k\theta)\sin(t\theta)d\theta + b_k \int_0^{2\pi} \sin(k\theta)\sin(t\theta)d\theta\right) \tag{3.11}$$

Da $\int_0^{2\pi} \sin(k\theta)\sin(t\theta)d\theta = \pi$ für $k = t$ und sonst 0 ist, und $\int_0^{2\pi} \cos(k\theta)\sin(t\theta)d\theta = 0$ für alle k und t ist, folgt

$$\int_0^{2\pi} f_\Theta(\theta)\sin(t\theta)d\theta = b_t\pi \qquad \overset{\text{Def.3.1.7}}{\Longrightarrow} b_t = \frac{1}{\pi}\beta_t$$

Also folgt insgesamt die Darstellung für die Dichte $f_\Theta(\theta)$:

$$f_\Theta(\theta) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{t=1}^{\infty}(\alpha_t\cos(tx) + \beta_t\sin(tx)).$$

(iv) Sei Θ eine circuläre Zufallsgröße mit Dichte g_Θ und sei X eine Zufallsgröße mit Dichte f_X gegeben durch

$$f_X(x) = \alpha_z g_\Theta(x), \qquad 2\pi z < x \leq 2\pi(z+1), \quad z \in \mathbb{Z}, \quad \alpha_z \in [0,1], \quad \sum_{z=-\infty}^{\infty}\alpha_z = 1.$$

Behauptung: $f_X^w = g_\Theta$
Beweis:

$$\begin{aligned} f_X^w(\theta) &= \sum_{k=-\infty}^{\infty} f_X(\theta + 2\pi k) && (3.12)\\ &= \sum_{k=-\infty}^{\infty} \alpha_k \underbrace{g_\Theta(\theta + 2\pi k)}_{=g_\Theta(\theta)} && (3.13)\\ &= g_\Theta(\theta)\sum_{k=-\infty}^{\infty}\alpha_k && (3.14)\\ &= g_\Theta(\theta). && (3.15)\end{aligned}$$

□

3.2.4.1 Gewickelte Normalverteilung $\mathcal{WN}(\mu,\rho)$

- Dichte:

$$f_\Theta(\theta) = \frac{1}{\sigma\sqrt{2\pi}}\sum_{k=-\infty}^{\infty}\exp\left(\frac{-(\theta-\mu+2\pi k)^2}{2\sigma^2}\right), \qquad 0 \leq \theta < 2\pi, \quad 0 \leq \rho \leq 1$$

- Charakteristische Funktion:

$$\varphi_\Theta(t) = \mathrm{e}^{i\mu t - \frac{t^2\sigma^2}{2}}$$

.

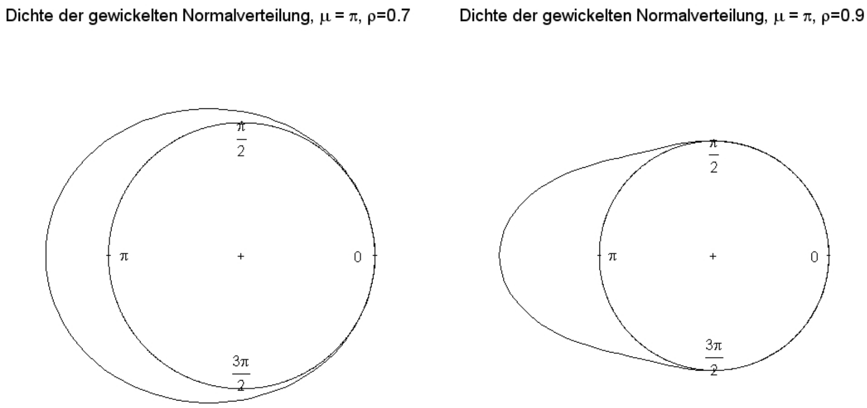

Abbildung 3.2: Dichte der gewickelten Normalverteilung $\mathcal{WN}(\pi, \rho)$ für $\rho = 0.7, 0.9$.

3.2.4.2 Gewickelte Cauchy-Verteilung $\mathcal{WC}(\mu, \rho)$

Betrachte zunächst die Cauchy-Verteilung in $\mathbb{R}$

$$f_X(x) = \frac{1}{\pi}\frac{\sigma}{\sigma^2 + (x-\mu)^2}, \qquad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+.$$

Mit Satz 3.2.2 wird die zu $\Theta = X^w$ zugehörige gewickelte Dichte gefunden

$$f_\Theta(\theta) = \frac{1}{2\pi}\left(1 + 2\sum_{t=1}^{\infty}(\alpha_t \cos(t\theta) + \beta_t \sin(t\theta))\right). \tag{3.16}$$

Mit der charakteristischen Funktion der Cauchy-Verteilung

$$\varphi_X(t) = \mathrm{e}^{\imath t\mu - \sigma|t|}$$

bekommt man die Koeffizienten α_t und β_t

$$\begin{aligned}\varphi_X(t) &= \mathrm{e}^{-\sigma|t|}\mathrm{e}^{\imath t\mu} \\ &= \mathrm{e}^{-\sigma|t|}\cos(t\mu) + \imath\mathrm{e}^{-\sigma|t|}\sin(t\mu)\end{aligned}$$

$$\Rightarrow \qquad \alpha_t = \mathrm{e}^{-\sigma|t|}\cos(t\mu), \quad \beta_t = \mathrm{e}^{-\sigma|t|}\sin(t\mu).$$

Einsetzen in 3.16 ergibt mit dem Additionstheorem

$$f_\Theta(\theta) = \frac{1}{2\pi}\left(1 + 2\sum_{t=1}^{\infty}(\mathrm{e}^{-\sigma|t|}\cos(t\mu)\cos(t\theta) + \mathrm{e}^{-\sigma|t|}\sin(t\mu)\sin(t\theta))\right) \tag{3.17}$$

$$\overset{t\geq 1}{=} \frac{1}{2\pi}\left(1 + 2\sum_{t=1}^{\infty}\mathrm{e}^{-\sigma t}\cos t(\theta-\mu)\right) \tag{3.18}$$

Man beachte, dass $\sum_{t=1}^{\infty}\mathrm{e}^{-\sigma t}\cos t(\theta-\mu)$ der Realteil der geometrischen Reihe $\sum_{t=1}^{\infty}\mathrm{e}^{-\sigma t}\mathrm{e}^{\imath t(\theta-\mu)}$ ist. Mithilfe dieser Reihe rechnet man weiter

$$\begin{aligned}\sum_{t=1}^{\infty}\mathrm{e}^{-\sigma t}\cos t(\theta-\mu) &= \Re\left(\sum_{t=1}^{\infty}\mathrm{e}^{-\sigma t}\mathrm{e}^{\imath t(\theta-\mu)}\right) \\ &= \Re\left(\mathrm{e}^{-\sigma}\mathrm{e}^{\imath(\theta-\mu)} + \mathrm{e}^{-2\sigma}\mathrm{e}^{2\imath(\theta-\mu)} + \mathrm{e}^{-3\sigma}\mathrm{e}^{3\imath(\theta-\mu)} + \ldots\right) \\ &\overset{(*)}{=} \Re\left(\frac{\mathrm{e}^{-\sigma}\mathrm{e}^{\imath(\theta-\mu)}}{1-\mathrm{e}^{-\sigma}\mathrm{e}^{\imath(\theta-\mu)}}\right) \\ &= \Re\left(\frac{\mathrm{e}^{-\sigma}\mathrm{e}^{\imath(\theta-\mu)}}{1-\mathrm{e}^{-\sigma}\cos(\theta-\mu) - \imath\mathrm{e}^{-\sigma}\sin(\theta-\mu)}\right. \\ &\qquad \left.\frac{1-\mathrm{e}^{-\sigma}\cos(\theta-\mu) + \imath\mathrm{e}^{-\sigma}\sin(\theta-\mu)}{1-\mathrm{e}^{-\sigma}\cos(\theta-\mu) + \imath\mathrm{e}^{-\sigma}\sin(\theta-\mu)}\right) \\ &= \Re\left(\frac{\mathrm{e}^{-\sigma}\cos(\theta-\mu) + \imath\mathrm{e}^{-\sigma}\sin(\theta-\mu) - \mathrm{e}^{-2\sigma}\cos^2(\theta-\mu)}{1-2\mathrm{e}^{-\sigma}\cos(\theta-\mu) + \mathrm{e}^{-2\sigma}\cos^2(\theta-\mu) + \mathrm{e}^{-2\sigma}\sin^2(\theta-\mu)}\right. \\ &\qquad + \frac{-\imath\mathrm{e}^{-2\sigma}\sin(\theta-\mu)\cos(\theta-\mu) + \imath\mathrm{e}^{-2\sigma}\sin(\theta-\mu)\cos(\theta-\mu)}{1-2\mathrm{e}^{-\sigma}\cos(\theta-\mu) + \mathrm{e}^{-2\sigma}\cos^2(\theta-\mu) + \mathrm{e}^{-2\sigma}\sin^2(\theta-\mu)} \\ &\qquad \left.+ \frac{-\mathrm{e}^{-2\sigma}\sin^2(\theta-\mu)}{1-2\mathrm{e}^{-\sigma}\cos(\theta-\mu) + \mathrm{e}^{-2\sigma}\cos^2(\theta-\mu) + \mathrm{e}^{-2\sigma}\sin^2(\theta-\mu)}\right) \\ &= \frac{\mathrm{e}^{-\sigma}\cos(\theta-\mu) - \mathrm{e}^{-2\sigma}}{1+\mathrm{e}^{-2\sigma} - 2\mathrm{e}^{-\sigma}\cos(\theta-\mu)}\end{aligned} \tag{3.19}$$

Einsetzen in 3.17 liefert

$$\begin{aligned} f_\Theta(\theta) &= \frac{1}{2\pi}\left(1 + 2\frac{\mathrm{e}^{-\sigma}\cos(\theta-\mu) - \mathrm{e}^{-2\sigma}}{1 + \mathrm{e}^{-2\sigma} - 2\mathrm{e}^{-\sigma}\cos(\theta-\mu)}\right) \\ &= \frac{1}{2\pi}\frac{2\mathrm{e}^{-\sigma}\cos(\theta-\mu) - 2\mathrm{e}^{-2\sigma} + 1 + e^{-2\sigma} - 2\mathrm{e}^{-\sigma}\cos(\theta-\mu)}{1 + \mathrm{e}^{-2\sigma} - 2\mathrm{e}^{-\sigma}\cos(\theta-\mu)} \\ &= \frac{1}{2\pi}\frac{1 - \mathrm{e}^{-2\sigma}}{1 + \mathrm{e}^{-2\sigma} - 2\mathrm{e}^{-\sigma}\cos(\theta-\mu)} \end{aligned}$$

Beweis von $(*)$:
Die geometrische Reihe konvergiert, da der Quotient $|\mathrm{e}^{-\sigma}\mathrm{e}^{\imath(\theta-\mu)}| < 1$ ist

$$|\mathrm{e}^{-\sigma}\mathrm{e}^{\imath(\theta-\mu)}| = |\mathrm{e}^{-\sigma}|\underbrace{|\mathrm{e}^{\imath(\theta-\mu)}|}_{=1} = |\mathrm{e}^{-\sigma}| \overset{\sigma>0}{<} 1.$$

Zusammenfassung der Ergebnisse

- Dichte:
$$f_\Theta(\theta) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)}, \qquad 0 \le \theta < 2\pi$$
wobei $\rho = \mathrm{e}^{-\sigma}$ von der zugrunde liegenden Cauchy-Verteilung mit den Parametern μ und σ kommt

- Charakteristische Funktion:
$$\varphi_\Theta(t) = \mathrm{e}^{\imath t\mu - \sigma|t|}$$

- Bemerkung: Für $\rho = 0$ ergibt sich die stetige Gleichverteilung.

3.2.5 Von Mises Verteilung $\mathcal{M}(\mu, \kappa)$

- Dichte:
$$f_\Theta(\theta) = \frac{1}{2\pi I_0(\kappa)}\mathrm{e}^{\kappa\cos(\theta-\mu)}, \qquad 0 \le \theta < 2\pi$$
wobei I_0 die modifizierte Bessel-Funktion der nullten Ordnung erster Gattung ist und folgendermaßen definiert ist

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{\kappa\cos\theta} d\theta.$$

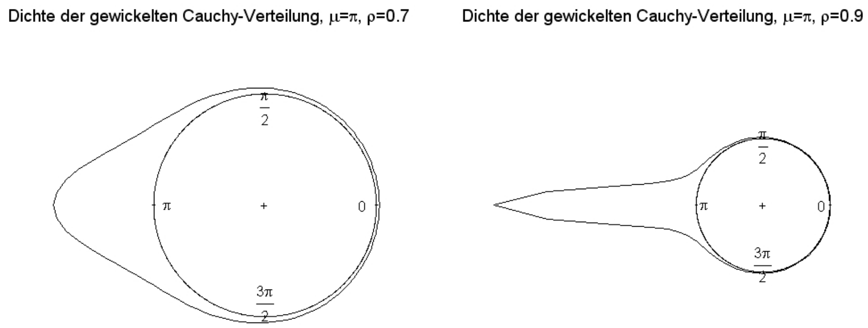

Abbildung 3.3: Dichte der gewickelten Cauchy-Verteilung $\mathcal{WC}(\pi, \rho)$ für $\rho = 0.7, 0.9$.

- Verteilungsfunktion: Die Verteilungsfunktion von $\Theta \sim \mathcal{M}(0, \kappa)$ ist gegeben durch

$$F_\Theta(\theta) = \frac{1}{2\pi I_0(\kappa)} \int_0^\theta \mathrm{e}^{\kappa \cos(\theta - \mu)} du.$$

- Bemerkung: μ ist die mittlere Richtung und κ der Konzentrationsparameter. Für $\kappa = 0$ ergibt sich die stetige Gleichverteilung (siehe Abbildung 3.4). Nimmt κ zu, wird die Streuung kleiner.

- Anwendung: Wichtige Verteilung in circulärer Statistik. Sie wurde von Gumbel, Greenwood und Durand 1953 [J.A.Greenwood u. a. 1953] circuläre Normalverteilung genannt.

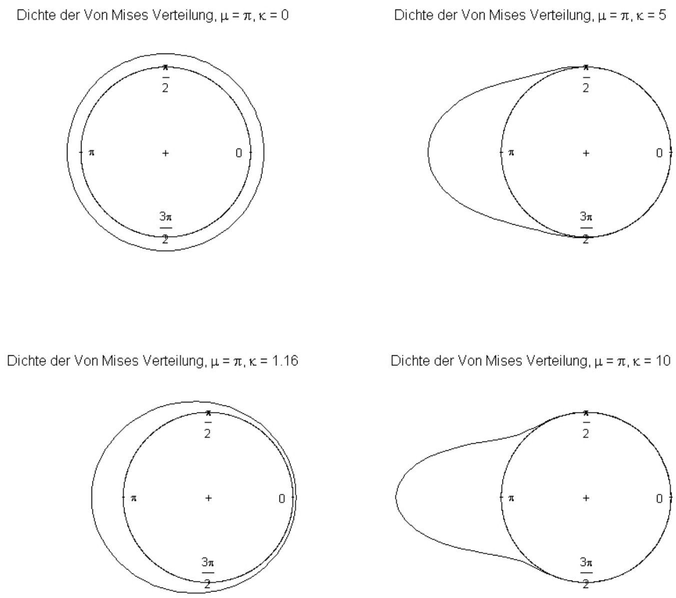

Abbildung 3.4: Dichte der Von Mises Verteilung $\mathcal{M}(\pi, \kappa)$ für $\kappa = 0, 1.16, 5, 10$.

4 Maßzahlen für Verteilugen - Lokations- und Streuungsmaße

Gegenstand des Kapitels 2 war Aufbereitung und Beschreibung der Daten. In diesem Kapitel werden Maßzahlen einer Verteilung betrachtet.

4.1 Lokationsmaße

Motivation 4.1.1. Sei Θ eine circuläre Zufallsgröße und o.B.d.A. absolut stetig. Einsetzen in (4.4) liefert

$$\begin{aligned}E(\Theta) &= \int_0^{2\pi} \mathrm{e}^{\imath\theta} f_\Theta(\theta) d\theta \\ &= \int_0^{2\pi} \cos\theta f_\Theta(\theta) d\theta + \imath \int_0^{2\pi} \sin\theta f_\Theta(\theta) d\theta \\ &= \alpha_t + \imath\beta_t \\ &\overset{\text{Bem.3.1.8}}{=} \varphi_\Theta(1).\end{aligned}$$

Das heißt also, dass der Erwartungswert einer circulärer Zufallsgröße gleich der charakteristischen Funktion $\varphi_\Theta(t)$ für $t = 1$ und dem ersten trigonometrischen Moment ist.
In 2.4 wurden die empirischen trigonometrischen Momente eingeführt. Nun werden trigonometrische Momente einer Zufallsgröße definiert. P-tes trigonometrisches Moment einer Zufallsgröße Θ bezüglich des Ursprungs ist gegeben durch (vgl.[Fisher 2000]):

$$\phi'_p = \rho_p \mathrm{e}^{\imath\mu'_p} = \rho_p \cos(\mu'_p) + \imath\rho_p \sin(\mu'_p) \tag{4.1a}$$
$$= \int_0^{2\pi} \cos(p\theta) f_\Theta(\theta) d\theta + \imath \int_0^{2\pi} \sin(p\theta) f_\Theta(\theta) d\theta \tag{4.1b}$$
$$= E(\cos(p\theta)) + \imath E(\sin(p\theta)) \tag{4.1c}$$
$$= \alpha_p + \imath\beta_p. \tag{4.1d}$$

Im Fall $p = 1$ wird einfach ρ anstatt ρ_1 und μ anstatt μ_1 geschrieben, nämlich

$$\phi'_1 = \rho \mathrm{e}^{\imath\mu},$$

wo μ mittlere Richtung und ρ mittlere Länge der Zufallsgröße Θ sind.

Analog zu dem p-ten empirischen zentrierten Moment 2.14 kann man das p-te zentrierte trigonometrische Moment folgendermaßen definieren

$$\phi_p = \rho_p \mathrm{e}^{\imath\mu_p} = \rho_p \cos(\mu_p) + \imath\rho_p \sin(\mu_p) \tag{4.2a}$$
$$= \int_0^{2\pi} \cos p(\theta - \mu) f_\Theta(\theta) d\theta + \imath \int_0^{2\pi} \sin p(\theta - \mu) f_\Theta(\theta) d\theta \tag{4.2b}$$
$$= E(\cos p(\theta - \mu)) + \imath E(\sin p(\theta - \mu)) \tag{4.2c}$$
$$= \bar{\alpha}_p + \imath\bar{\beta}_p. \tag{4.2d}$$

Definition 4.1.2 (Erwartungswert für diskrete Zufallsgrößen)**.** *Ist Θ eine diskrete circuläre Zufallsgröße mit $\sum_\theta |\mathrm{e}^{\imath\theta}| p_\Theta(\theta) < \infty$, dann heißt*

$$E(\Theta) := \sum_\theta \mathrm{e}^{\imath\theta} p_\Theta(\theta) \tag{4.3}$$

Erwartungswert von Θ.

Definition 4.1.3 (Erwartungswert für absolut stetige Zufallsgrößen)**.** *Ist Θ eine absolut stetige circuläre Zufallsgröße mit $\int |\mathrm{e}^{\imath\theta}| f_\Theta(\theta) d\theta < \infty$, dann heißt*

$$E(\Theta) := \int \mathrm{e}^{\imath\theta} f_\Theta(\theta) d\theta \tag{4.4}$$

Erwartungswert von Θ.

Beispiel 4.1.4. Es werden Erwartungswerte von einigen Verteilungen berechnet:

(i) von Mises-Verteilung : Gilt $\Theta \sim \mathcal{M}(\mu, \kappa)$, so gilt

$$\begin{aligned}
E(\Theta) &= \int_0^{2\pi} \mathrm{e}^{\imath\theta} f_\Theta(\theta) d\theta \\
&= \int_0^{2\pi} \cos(\theta) \frac{\mathrm{e}^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)} d\theta + \imath \int_0^{2\pi} \sin(\theta) \frac{\mathrm{e}^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)} d\theta \\
&\overset{\text{Subst.}\tilde{\theta} := \theta - \mu}{=} \int_{-\mu}^{2\pi-\mu} \cos(\tilde{\theta}+\mu) \frac{\mathrm{e}^{\kappa\cos(\tilde{\theta})}}{2\pi I_0(\kappa)} d\tilde{\theta} + \imath \int_{-\mu}^{2\pi-\mu} \sin(\tilde{\theta}+\mu) \frac{\mathrm{e}^{\kappa\cos(\tilde{\theta})}}{2\pi I_0(\kappa)} d\tilde{\theta} \\
&\overset{\text{Additionstheorem}}{=} \cos(\mu) \int_{-\mu}^{2\pi-\mu} \cos(\tilde{\theta}) \frac{\mathrm{e}^{\kappa\cos(\tilde{\theta})}}{2\pi I_0(\kappa)} d\tilde{\theta} - \sin(\mu) \underbrace{\int_{-\mu}^{2\pi-\mu} \sin(\tilde{\theta}) \frac{\mathrm{e}^{\kappa\cos(\tilde{\theta})}}{2\pi I_0(\kappa)} d\tilde{\theta}}_{=0,\ \text{wegen } (*)} \\
&\qquad + \imath \cos(\mu) \underbrace{\int_{-\mu}^{2\pi-\mu} \sin(\tilde{\theta}) \frac{\mathrm{e}^{\kappa\cos(\tilde{\theta})}}{2\pi I_0(\kappa)} d\tilde{\theta}}_{=0,\ \text{wegen } (*)} + \imath \sin(\mu) \int_{-\mu}^{2\pi-\mu} \cos(\tilde{\theta}) \frac{\mathrm{e}^{\kappa\cos(\tilde{\theta})}}{2\pi I_0(\kappa)} d\tilde{\theta} \\
&= \frac{\cos(\mu)}{2\pi I_0(\kappa)} \int_{-\mu}^{2\pi-\mu} \cos(\tilde{\theta}) \mathrm{e}^{\kappa\cos(\tilde{\theta})} d\tilde{\theta} + \imath \frac{\sin(\mu)}{2\pi I_0(\kappa)} \int_{-\mu}^{2\pi-\mu} \cos(\tilde{\theta}) \mathrm{e}^{\kappa\cos(\tilde{\theta})} d\tilde{\theta} \\
&= \cos(\mu) \frac{I_1(\kappa)}{I_0(\kappa)} + \imath \sin(\mu) \frac{I_1(\kappa)}{I_0(\kappa)} \\
&= \mathrm{e}^{\imath\mu} \frac{I_1(\kappa)}{I_0(\kappa)},
\end{aligned}$$

wobei I_1 die modifizierte Bessel-Funktion der ersten Ordnung ist

$$I_1(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(\theta) \mathrm{e}^{\kappa\cos(\theta)} d\theta = \frac{1}{2\pi} \int_{-\mu}^{2\pi-\mu} \cos(\theta) \mathrm{e}^{\kappa\cos(\theta)} d\theta.$$

Zu $(*)$:

$$\int_{-\mu}^{2\pi-\mu} \sin(\theta) \frac{\mathrm{e}^{\kappa\cos(\theta)}}{2\pi I_0(\kappa)} d\theta = \int_0^{2\pi} \sin(\theta) \underbrace{\frac{\mathrm{e}^{\kappa\cos(\theta)}}{2\pi I_0(\kappa)}}_{=f_\Theta(\theta)} d\theta = 0,$$

wegen der Symmetrie von f_Θ und Antisymmetrie von sin.

(ii) Einpunktverteilung : Gilt $\Theta \sim \mathcal{E}_{\mu_0}$, so gilt

$$\begin{aligned} E(\Theta) &= \sum_{\theta} \mathrm{e}^{\imath\theta} p_\Theta(\theta) = \mathrm{e}^{\imath\mu_0} p_\Theta(\mu_0) \\ &= \mathrm{e}^{\imath\mu_0}. \end{aligned}$$

(iii) Stetige Gleichverteilung : Gilt $\Theta \sim \mathcal{U}[0, 2\pi]$, so gilt

$$\begin{aligned} E(\Theta) &= \int_0^{2\pi} \mathrm{e}^{\imath\theta} \frac{1}{2\pi} d\theta \\ &= \frac{1}{2\pi} \left(\int_0^{2\pi} \cos\theta d\theta + \imath \int_0^{2\pi} \sin\theta d\theta \right) \\ &= \frac{1}{2\pi} \left([\sin\theta]_0^{2\pi} + \imath[-\cos\theta]_0^{2\pi} \right) \\ &= \frac{1}{2\pi} (0 - 0 - \imath(1-1)) \\ &= 0. \end{aligned}$$

4.2 Streuungsmaße

Analog zu empirischer Varianz wird Varianz einer circulären Zufallsgröße Θ definiert. Aus 4.1 a) und b) folgt für ϕ_1' (vgl.[Fisher 2000]):

$$\int_0^{2\pi} \cos(\theta) f_\Theta(\theta) d\theta = \rho \cos(\mu) \tag{4.5a}$$

$$\int_0^{2\pi} \sin(\theta) f_\Theta(\theta) d\theta = \rho \sin(\mu). \tag{4.5b}$$

Damit folgt mit den Additionstheoremen für ϕ_1 mit 4.2 b):

$$\begin{aligned} \int_0^{2\pi} \cos(\theta - \mu) f_\Theta(\theta) d\theta &= \cos(\mu) \underbrace{\int_0^{2\pi} \cos(\theta) f_\Theta(\theta) d\theta}_{=\rho\cos(\mu),\ \text{wegen 4.5a)}} + \sin(\mu) \underbrace{\int_0^{2\pi} \sin(\theta) f_\Theta(\theta) d\theta}_{=\rho\sin(\mu),\ \text{wegen 4.5b)}} \\ &= \rho\cos^2(\mu) + \rho\sin^2(\mu) \\ &= \rho, \end{aligned}$$

und

$$\begin{aligned}\int_0^{2\pi} \sin(\theta-\mu) f_\Theta(\theta) d\theta &= \cos(\mu) \underbrace{\int_0^{2\pi} \sin(\theta) f_\Theta(\theta) d\theta}_{=\rho\sin(\mu)} - \sin(\mu) \underbrace{\int_0^{2\pi} \cos(\theta) f_\Theta(\theta) d\theta}_{=\rho\cos(\mu)} \\ &= \rho\cos(\mu)\sin(\mu) - \rho\sin(\mu)\cos(\mu) \\ &= 0,\end{aligned}$$

und damit

$$\Rightarrow \phi_1 = \rho = E(\cos(\theta-\mu)).$$

Nun kann man die circuläre Varianz einer Zufallsgröße angeben:

Definition 4.2.1. *Sei Θ eine circuläre Zufallsgröße.*

$$\nu = 1 - \phi_1 = 1 - \rho = 1 - E\left(\cos(\theta-\mu)\right)$$

heißt Varianz von Θ, falls $E\left(\cos(\theta-\mu)\right) < \infty$.

Ähnlich wird auch Streuung von Θ um einen gegebenen Winkel bestimmt.

Definition 4.2.2.

$$\nu(\alpha) = 1 - E\left(\cos(\theta-\alpha)\right)$$

heißt Varianz von Θ um den Winkel α, falls $E\left(\cos(\theta-\alpha)\right) < \infty$.

5 Periodische Regression

5.1 Einführung

In diesem Kapitel werden Beziehungen zwischen circulären Variablen untersucht. In 5.2 wird die abhängige Variable linear und die unabhängige circulär sein. Der Fall, wenn beide Variablen circulär sind, wird hier nicht betrachtet.
Sei $\beta \in \mathbb{R}^R$ im Folgenden stets ein unbekannter R dimensionaler Parameter, $t_1, \ldots, t_N \in \mathcal{T}$ bekannte Versuchsbedingungen in dem Design Region $\mathcal{T}$, $x : \mathcal{T} \longrightarrow \mathbb{R}^R$ unabhängige und $y = (y_1, \ldots, y_N)$ abhängige Variablen. Das Regressionsmodell kann wie folgt geschrieben werden

$$y = \mathbf{X}_d \beta + z, \tag{5.1}$$

wobei

$\mathbf{X}_d = (x(t_1), \ldots, x(t_N))^\top$ - die Planungsmatrix,

$z = (z_1, \ldots, z_N)^\top$ - der Fehlervektor,

$\beta = (\beta_0, \ldots, \beta_{r-1})^\top \in \mathbb{R}^R$ - die Regressionsparameter.

Definition 5.1.1 (Planungsmatrix)**.** *Es seien N Beobachtungen $x(t_1), \ldots, x(t_N)$ und der Versuchsplan $d = (t_1, \ldots, t_N) \in \mathbb{R}^{N \times q}$ gegeben. Dann heißt*

$$\mathbf{X}_d = (x(t_1), \ldots, x(t_N))^\top \in \mathbb{R}^{N \times R}$$

die Planungsmatrix (Designmatrix).

5.2 Linear-circuläres Verhältnis

Definition 5.2.1. *Das circuläre Regressionsmodell ist gegeben durch*

$$y_n = \beta + \alpha \; cos(\omega t_n - \theta) + z_n \quad \text{mit} \quad z_n \sim \mathcal{N}(0, \sigma^2) \tag{5.2}$$

für alle $n = 1, \ldots, N$. Die Gleichung enthält folgende Parameter:

> t_n *- stellt normalerweise Zeit in einem bestimmten Zeitraum dar,*
>
> β *- mittlere Höhe,*
>
> α *- Amplitude ($\alpha \geq 0$),*
>
> ω *- Kreisfrequenz ($\omega = 2\pi/T$ oder $\omega = 360°/T$)*
>
> θ *- Phasenverschiebung*

([Batschelet 1981])

Bemerkung 5.2.2. Die Gleichung 5.2 kann auch in folgender Form geschrieben werden:

$$y_n = \beta + A \; \cos \omega t_n + B \; \sin \omega t_n + z_n \tag{5.3}$$

mit

$$A = \alpha \; \cos \theta, \qquad B = \alpha \; \sin \theta. \tag{5.4}$$

Das Modell 5.3 ist in dem geschätzten Parameter ω nicht linear. Wenn aber ω gegeben ist, wird dieses Modell linear.

Beispiel 5.2.3. Man betrachte die Regressionsgleichung 5.3:

$$y_n = \beta + A \; \cos \omega t_n + B \; \sin \omega t_n + z_n,$$

d.h. die multiple Regression mit $r = 3$. Man kann dann die Gleichung auch in folgender Form $y = \mathbf{X}_d \beta + z$ schreiben, d.h. es sind $t_n \in \mathbb{R}^q$, $x(t) = (1, \cos(\omega t), \sin(\omega t)) \in \mathbb{R}^{N \times q}$ und $\beta = (\beta_0, \; A = \alpha \cos \theta, \; B = \alpha \sin \theta) \in \mathbb{R}^3$.

Definition 5.2.4. *Es seien Beobachtungen* $(t_1, y_1), (t_2, y_2), \ldots, (t_N, y_N)$ *gegeben.* $\hat{\beta}(y, \mathbf{X}_d)$ *heißt Kleinste-Quadrat-Summen-Schätzung für* β *bei* y *und* $\mathbf{X}_d$*, falls gilt*

$$\hat{\beta}(y, \mathbf{X}_d) \in \arg\min_{\beta \in \mathbb{R}^r} \sum_{n=1}^{N} \left(y_n - x(t)^\top \beta\right)^2 = \arg\min_{\beta \in \mathbb{R}^r} (y - \mathbf{X}_d\beta)^\top (y - \mathbf{X}_d\beta).$$

(vgl. z.B [Zucchini 2009])

Da die Planungsmatrix $\mathbf{X_d} \in \mathbb{R}^{n \times m}$ manchmal keinen vollen Rang besitzt, ist die Matrix $\mathbf{X_d}^\top \mathbf{X_d} \in \mathbb{R}^{m \times m}$ nicht invertierbar, denn es gilt $\operatorname{rg}\left(\mathbf{X_d}^\top \mathbf{X_d}\right) < m$. Man betrachtet in diesem Fall den Begriff der verallgemeinerten inversen Matrix. Da in dieser Arbeit dieser Fall nicht auftritt, wird dies nicht behandelt.

Satz 5.2.5. $\hat{\beta}(y, \mathbf{X}_d)$ *ist Kleinste-Quadrat-Summen-Schätzung für* β *bei* y *und* $\mathbf{X}_d$ *genau dann, wenn gilt:*

$$\mathbf{X}_d^\top \mathbf{X}_d \hat{\beta}(y, \mathbf{X}_d) = \mathbf{X}_d^\top y.$$

(vgl. z.B. [Kohn 2005])

Beweis. Zu minimieren ist die Summe der quadratischen Abständen. Setze also $g(\beta) := \sum_{n=1}^{N} \left(y_n - x(t)^\top \beta\right)^2 = (y - \mathbf{X}_d\beta)^\top (y - \mathbf{X}_d\beta)$ und berechne die erste Ableitung von $g(\beta)$ an der Stelle β

$$\begin{aligned} \frac{\partial g(\beta)}{\partial \beta} &= \frac{\partial}{\partial \beta} \left(y^\top y - 2\beta^\top \mathbf{X}_d^\top y + \beta^\top \mathbf{X}_d^\top \mathbf{X}_d \beta\right) \\ &= -2\mathbf{X}_d^\top y + 2\mathbf{X}_d^\top \mathbf{X}_d \beta \end{aligned}$$

Setzen gleich null ergibt

$$\mathbf{X}_d^\top \mathbf{X}_d \beta = \mathbf{X}_d^\top y.$$

Die zweite Ableitung der Funktion g wird berechnet

$$\frac{\partial^2 g(\beta)}{(\partial \beta)^2} = 2\mathbf{X}_d^\top \mathbf{X}_d.$$

Die hinreichende Bedingung für ein Minimum ist, dass die zweite Ableitung größer null ist. Die folgende Gleichung wird erfüllt, wenn $\mathbf{X}_d^\top \mathbf{X}_d$ positiv definit ist.

$$2\mathbf{X}_d^\top \mathbf{X}_d > 0$$

□

Beispiel 5.2.6. (Fortsetzung des Beispiels 5.2.3)
Man setzt

$$c_n = \cos \omega t_n, \qquad s_n = \sin \omega t_n \tag{5.5}$$

da $t_1, \ldots, t_N$ und ω bekannt sind. Dann hat die Planungsmatrix $\mathbf{X}_d$ folgende Gestalt

$$\mathbf{X}_d = \begin{pmatrix} 1 & c_1 & s_1 \\ 1 & c_2 & s_2 \\ \vdots & \vdots & \vdots \\ 1 & c_N & s_N \end{pmatrix} \in \mathbb{R}^{N \times 3}$$

$$\Rightarrow \mathbf{X}_d \mathbf{X}_d^\top = \begin{pmatrix} 1 & c_1 & s_1 \\ 1 & c_2 & s_2 \\ \vdots & \vdots & \vdots \\ 1 & c_N & s_N \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & \cdots & 1 \\ c_1 & c_2 & \cdots & c_N \\ s_1 & s_2 & \cdots & s_N \end{pmatrix} = \begin{pmatrix} N & \sum_{i=1}^N c_i & \sum_{i=1}^N s_i \\ \sum_{i=1}^N c_i & \sum_{i=1}^N c_i^2 & \sum_{i=1}^N s_i c_i \\ \sum_{i=1}^N s_i & \sum_{i=1}^N s_i c_i & \sum_{i=1}^N s_i^2 \end{pmatrix}$$

$$\Rightarrow \mathbf{X}_d \mathbf{X}_d^\top \beta = \begin{pmatrix} \beta_0 N + A \sum_{i=1}^N c_i + B \sum_{i=1}^N s_i \\ \beta_0 \sum_{i=1}^N c_i + A \sum_{i=1}^N c_i^2 + B \sum_{i=1}^N s_i c_i \\ \beta_0 \sum_{i=1}^N s_i + A \sum_{i=1}^N s_i c_i + B \sum_{i=1}^N s_i^2 \end{pmatrix}$$

und

$$\mathbf{X}_d^\top y = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ c_1 & c_2 & \cdots & c_N \\ s_1 & s_2 & \cdots & s_N \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N c_i y_i \\ \sum_{i=1}^N s_i y_i \end{pmatrix}$$

Das heißt, zu lösen ist ein lineares Gleichungsystem

$$\sum_{i=1}^{N} y_i = N\beta_0 + A\sum_{i=1}^{N} c_i + B\sum_{i=1}^{N} s_i \tag{5.6a}$$

$$\sum_{i=1}^{N} c_i y_i = \beta_0\sum_{i=1}^{N} c_i + A\sum_{i=1}^{N} c_i^2 + B\sum_{i=1}^{N} c_i s_i \tag{5.6b}$$

$$\sum_{i=1}^{N} s_i y_i = \beta_0\sum_{i=1}^{N} s_i + A\sum_{i=1}^{N} c_i s_i + B\sum_{i=1}^{N} s_i^2 \tag{5.6c}$$

Die Güte der Anpassung wird in dieser Arbeit mithilfe des Akaike-Kriteriums (siehe z.B. [Sachs u. Hedderich 2006]) gemessen.

Definition 5.2.7 (Das Akaike-Informationskriterium)**.** *Seien die Residuen des Regressionsmodells normalverteilt, dann heißt*

$$AIC = N\ln\left(\frac{\sum \epsilon_i^2}{N}\right) + 2K$$

das Akaike-Informationskriterium (AIC), wobei K die Anzahl der im Regressionsmodell zu schätzenden Parameter ist, ϵ_i die Residuen im gegebenen Modell und N Anzahl der Messwerte([Burnham u. Anderson 2002]).

5.3 Pheromon-Fänge an der Station 'La Victoria' - Symmetrischema Tangolias, Teil 2

Zuerst soll für den Datensatz 1.1 ein Modell

$$y_t = \beta + \gamma r_t + \alpha\cos(\omega t - \theta) + z_t \tag{5.7}$$

gefunden werden, wobei

y_t - die Mottenanzahl,

r_t - die Regenmenge,

im Monat t darstellt. Zu schätzen sind die Werte α, β, θ und γ.

Danach, in Kapitel 5.4, wird ein Modell für den um fünf Monate verschobenen Datensatz (Tabelle 1.2) gefunden.

Gegeben sind $N = 2 \cdot 12 = 24$ äquidistante Punkte, die den 12 Monaten entsprechen. Das heißt, die Zeitabschnitte können mit den Winkeln $0, \frac{1}{6}\pi, \frac{1}{3}\pi, \dots, \frac{11}{6}\pi$ identifiziert werden. Die Periode T dauert ein Jahr oder äquivalent 2π

$$\Rightarrow \omega = \frac{2\pi}{T} = 1.$$

Nach Bemerkung 5.2.2 kann das Modell 5.7 in ein äquivalentes lineares umgeformt werden:

$$\begin{aligned} y_t &= \beta + \gamma r_t + \alpha \cos(\omega t - \theta) \\ &\overset{\omega=1}{=} \beta + \gamma r_t + \underbrace{\alpha \cos(\theta)}_{=:A} \cos(t) + \underbrace{\alpha \sin(\theta)}_{=:B} \sin(t) \\ &= \beta + \gamma r_t + A\cos(t) + B\sin(t) \end{aligned} \tag{5.8}$$

Lineare Regression Folgende lineare Modelle sind für den Datensatz 1.1 möglich.

Modell	Regressionskurve
`mod.lm.1`	$y_t = \beta + A\cos(t) + B\sin(t)$
`mod.lm.2`	$y_t = \beta_0 + \beta_1 t + A\cos(t) + B\sin(t)$
`mod.lm.3`	$y_t = \beta_0 + \beta_1 t$
`mod.lm.R.1`	$y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$
`mod.lm.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$
`mod.lm.R.3`	$y_t = \beta + \gamma r_t$
`mod.lm.R.4`	$y_t = \beta_0 + \beta_1 t + \gamma r_t$

In dem ersten Modell (`mod.lm.1`) werden nur drei Parameter geschätzt, nämlich β, A und B, in dem zweiten Modell (`mod.lm.2`) ein Parameter mehr - β_1. Das dritte Modell `mod.lm.3` ist eine einfache lineare Regression. Die Modelle `mod.lm.R.1` bis `mod.lm.R.4` werden um eine erklärende Variable `r` (Regenmenge) ergänzt. Das Regressionsmodell `mod.lm.R.3` ist eine einfache lineare Regression mit nur einer Variablen `r`.

Der Datensatz wird eingelesen (für den R-Quellcode siehe A.2.6.2).

```
> channel <- odbcConnectExcel("LaVictoriaM.xls")
> RainM.ST <- sqlFetch(channel, "LV_ST_RainM")
```

Zuerst werden Modelle ohne den Niederschlag betrachtet.

Mit dem Shapiro-Wilk-Test werden die Modelle `mod.lm.1` - `mod.lm.3` auf die Normalverteilung getestet. Die P-Werte werden aufgelistet:

Modell	Regressionskurve	Shapiro-Wilk P-Wert
`mod.lm.1`	$y_t = \beta + A\cos(t) + B\sin(t)$	0.01537650
`mod.lm.2`	$y_t = \beta_0 + \beta_1 t + A\cos(t) + B\sin(t)$	0.0158398
`mod.lm.3`	$y_t = \beta_0 + \beta_1 t$	0.001152614

In jedem Fall wird die Normalverteilung verworfen, was heißt, dass bei den ANOVA-Tests Vorsicht geboten wird.

```
> bartlett.test(PTM~Month,data=RainM.ST)
Bartlett's K-squared = 18.9864, df = 11, p-value = 0.06134
```

Der ANOVA-Test wird erzeugt. Das größte Modell `mod.lm.2` $y_t = \beta_0 + \beta_1 t + A\cos(t) + B\sin(t)$ wird untersucht.

```
> anova(lm(PTM~cos(Month)+sin(Month)+Month,data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
cos(Month)  1 49718144 49718144 18.4963 0.0003483 ***
sin(Month)  1   242637   242637  0.0903 0.7669396
Month       1  1037753  1037753  0.3861 0.5413917
Residuals  20 53760144  2688007
```

Die Terme `sin(Month)` ($P = 0.7669396$) und `Month` ($P = 0.5413917$) haben keinen signifikanten Effekt auf die Zielgröße. Es wird nun die umgekehrte Reihenfolge getestet.

```
> anova(lm(PTM~sin(Month)+cos(Month)+Month,data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
sin(Month)  1   107210   107210  0.0399 0.8437231
cos(Month)  1 49853571 49853571 18.5467 0.0003435 ***
Month       1  1037753  1037753  0.3861 0.5413917
Residuals  20 53760144  2688007

> anova(lm(PTM~Month+sin(Month)+cos(Month),data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value   Pr(>F)
Month       1  1940734  1940734  0.7220 0.405547
sin(Month)  1   965592   965592  0.3592 0.555668
cos(Month)  1 48092207 48092207 17.8914 0.000411 ***
Residuals  20 53760144  2688007
```

Die Änderung der Reihenfolge bestätigt, dass die Terme `sin(Month)` und `Month` keinen signifikanten Effekt auf die Zielgröße haben. Sie werden daher aus dem Modell ausgeschlossen. Da die Normalverteilung auch in diesem Fall nicht vorliegt, können bei den ANOVA-Tests keine sicheren Ergebnisse garantiert werden.

```
> shapiro.test(aov(PTM~cos(Month),data=RainM.ST)$residuals)$p.value
[1] 0.009226567
```

Mit `summary(lm(...))` wird getestet, ob der Vektor (β, A) sich signifikant von Null unterscheidet.

```
> summary(lm(PTM~cos(Month),data=RainM.ST))
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2920.6      323.6   9.024 7.55e-09 ***
cos(Month)    2091.4      469.1   4.458 0.000197 ***
---
Residual standard error: 1582 on 22 degrees of freedom
Multiple R-squared: 0.4746,Adjusted R-squared: 0.4507
F-statistic: 19.87 on 1 and 22 DF,  p-value: 0.0001971
```

Die letzte Zeile ($P = 0.0001971$) zeigt, dass das der Fall ist. Dieses Modell wird `mod.lm.4` genannt.

```
> mod.lm.4<-lm(PTM~cos(Month),data=RainM.ST)
```

Man kann noch zusätzlich zum Vergleich die 3 ursprünglichen Modelle bilden und ihre Anpassung vergleichen.

```
> mod.lm.1<-lm(PTM~cos(Month)+sin(Month),data=RainM.ST)
> mod.lm.2<-lm(PTM~cos(Month)+sin(Month)+Month,data=RainM.ST)
> mod.lm.3<-lm(PTM~Month,data=RainM.ST)
```

Die geschätzten Parameter und die AIC-Werte werden für die ersten vier Modelle zusammengefasst.

Modell	Regressionskurve	Intercept	A	B	β_1	γ	AIC
`mod.lm.1`	$y_t = \beta + A\cos(t) + B\sin(t)$	2922.2	2094.8	138.9	-	-	427.4957
`mod.lm.2`	$y_t = \beta_0 + \beta_1 t + A\cos(t) + B\sin(t)$	2491.29	2066.45	267.52	66.29	-	429.0368
`mod.lm.3`	$y_t = \beta_0 + \beta_1 t$	2285.93	-	-	82.38	-	440.5990
`mod.lm.4`	$y_t = \beta + A\cos(t)$	2921	2091	-	-	-	425.6017

Das Modell `mod.lm.4` hat den kleinsten Wert für das AIC. Die Regressionskurven werden in Abbildung 5.1 gezeigt. Aus der Grafik kann man ablesen, dass eine volle Periode sechs Monate und nicht zwölf dauert. Das liegt daran, dass der Datensatz bimodal ist (siehe z.B. Abbildung 2.3).

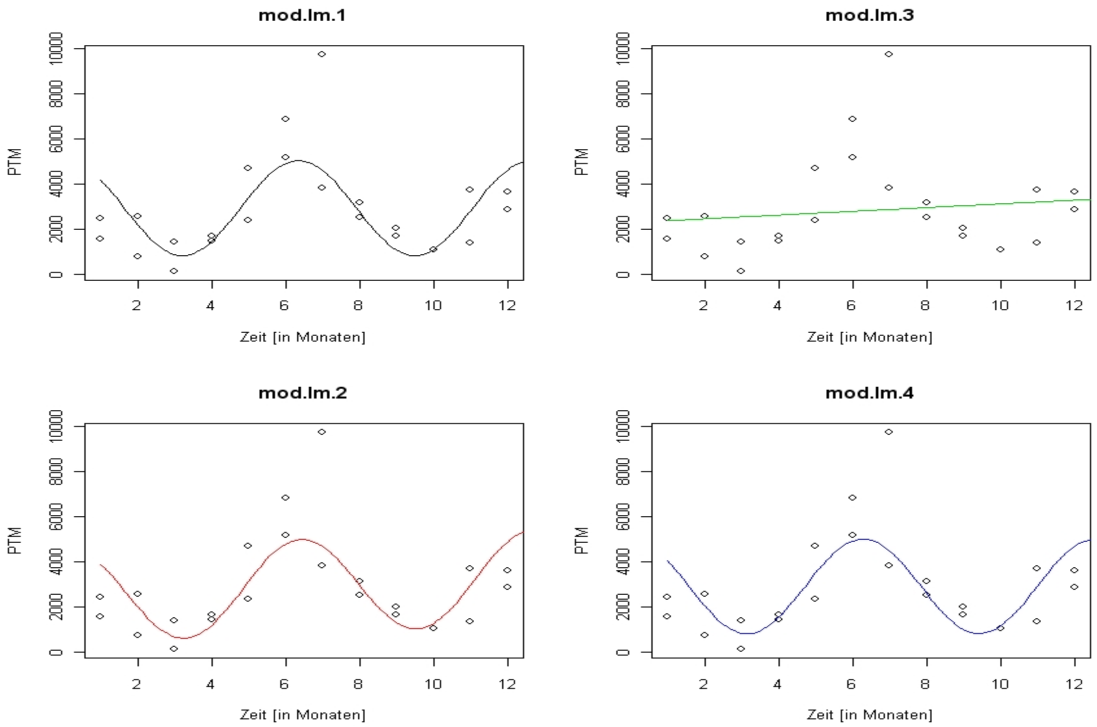

Abbildung 5.1: Vergleich der linearen Modelle ohne Regenmenge, ohne Verschiebung.

Nun wird auch der Niederschlag zu der Regression aufgenommen, d.h. die Anzahl der Männchen der Kartoffelmotte wird jetzt in Abhängigkeit von dem Monat und von der Regenmenge geschätzt.

Wie in den Modellen ohne Regenmenge wird die Hypothese "die Daten sind normalverteilt" abgelehnt. Die Ergebnisse des Shapiro-Wilk-Tests werden in der nachfolgenden Tabelle aufgelistet:

Modell	Regressionskurve	Shapiro-Wilk P-Wert
`mod.lm.R.1`	$y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$	0.01120242
`mod.lm.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$	0.01046258
`mod.lm.R.3`	$y_t = \beta + \gamma r_t$	0.01660329
`mod.lm.R.4`	$y_t = \beta_0 + \beta_1 t + \gamma r_t$	0.01556918

Das größte Modell `mod.lm.R.2` $y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$ wird mit dem ANOVA-Test getestet.

```
> anova(lm(PTM~cos(Month)+Rainfall+sin(Month)+Month,data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
cos(Month)  1 49718144 49718144 20.7630 0.0002156 ***
Rainfall    1  8539502  8539502  3.5662 0.0743321 .
sin(Month)  1   770855   770855  0.3219 0.5770992
Month       1   233653   233653  0.0976 0.7581594
Residuals  19 45496524  2394554
```

Die Variablen `Rainfall`, `sin(Month)` und `Month` zeigen keinen bedeutenden Effekt. Es wird geprüft, ob die Reihenfolge eine Rolle spielt. Der Term `sin(Month)` vergrößert in keiner Konstellation seinen Effekt, `Rainfall` dagegen schon (siehe A.2.6.1).

```
> anova(lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
Rainfall    1 16460380 16460380  6.8741 0.0167832 *
cos(Month)  1 41797266 41797266 17.4551 0.0005104 ***
sin(Month)  1   770855   770855  0.3219 0.5770992
Month       1   233653   233653  0.0976 0.7581594
Residuals  19 45496524  2394554
```

Mit `summary(lm(...))` wird getestet, ob der Vektor (β, A, γ) sich in dem reduzierten Modell wesentlich von Null unterscheidet.

```
> summary(lm(PTM~Rainfall+cos(Month),data=RainM.ST))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3597.983    460.103   7.820 1.18e-07 ***
```

```
Rainfall     -12.537      6.384  -1.964 0.062931 .
cos(Month)  1944.768    447.626   4.345 0.000285 ***
---
Residual standard error: 1488 on 21 degrees of freedom
Multiple R-squared: 0.5561,Adjusted R-squared: 0.5138
F-statistic: 13.15 on 2 and 21 DF,  p-value: 0.0001979
```

(β, A, γ) unterscheidet sich signifikant von Null. Der Effekt der Regenmenge ist in diesem Modell nicht signifikant. Es kann daran liegen, dass die Studie nur zwei Jahre gedauert hat und für jeden Monat nur zwei Messwerte vorliegen. Zu testen bleibt noch, ob die einzelnen Variablen einen signifikanten Effekt haben. Das Modell (`mod.lm.4`) mit nur der Variablen `cos(Month)` wurde bereits untersucht.

```
> summary(lm(PTM~Rainfall,data=RainM.ST))
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 3758.221    617.445   6.087 3.98e-06 ***
Rainfall     -17.162      8.475  -2.025   0.0552 .
---
Residual standard error: 2003 on 22 degrees of freedom
Multiple R-squared: 0.1571,Adjusted R-squared: 0.1188
F-statistic: 4.101 on 1 and 22 DF,  p-value: 0.05516
```

Das Ergebnis weist nicht auf einen signifikanten Effekt hinsichtlich der Regenmenge hin ($P = 0.0552$). Das Modell wird zum Vergleich der AIC-Werte `mod.lm.R.5` genannt:

```
> mod.lm.R.5<-lm(PTM~cos(Month)+Rainfall,data=RainM.ST)
```

Auch in diesem Fall liegt keine Normalverteilung vor.

```
> shapiro.test(aov(PTM~cos(Month)+Rainfall,data=RainM.ST)$residuals)$p.value
[1] 0.005433883
```

Es werden noch die ursprünglichen Modelle gebildet und mit dem Modell `mod.lm.R.5` verglichen.

```
> mod.lm.R.1<-lm(PTM~Rainfall+cos(Month)+sin(Month),data=RainM.ST)
> mod.lm.R.2<-lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,data=RainM.ST)
> mod.lm.R.3<-lm(PTM~Rainfall,data=RainM.ST)
> mod.lm.R.4<-lm(PTM~Rainfall+Month,data=RainM.ST)
```

Vergleich der Anpassung und geschätzte Parameter:

Modell	Intercept	A	B	β_1	γ	AIC
`mod.lm.R.1` $y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$	3627.44	1945.17	249.75	-	-13.03	425.1542
`mod.lm.R.2` $y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$	3398.80	1935.91	308.51	31.98	-12.64	427.0313
`mod.lm.R.3` $y_t = \beta_0 + \gamma r_t$	3758.22	-	-	-	-17.16	436.9453
`mod.lm.R.4` $y_t = \beta_0 + \beta_1 t + \gamma r_t$	3542.35	-	-	29.21	-16.21	438.8823
`mod.lm.R.5` $y_t = \beta + \gamma r + A\cos(t)$	3597.98	1944.77	-	-	-12.54	423.5554

Obwohl der Effekt der Regenmenge nicht signifikant war, zeigt das Modell `mod.lm.R.5` die beste Anpassung (AIC=423.5554). Alle fünf Kurven werden in Abbildung 5.2 geplottet (R-Quellcode siehe A.2.6.1).

Alle Werte für das AIC werden in einer Tabelle zusammengefasst.

Modell	Regressionskurve	AIC
`mod.lm.1`	$y_t = \beta + A\cos(t) + B\sin(t)$	427.4957
`mod.lm.2`	$y_t = \beta_0 + \beta_1 t + A\cos(t) + B\sin(t)$	429.0368
`mod.lm.3`	$y_t = \beta_0 + \beta_1 t$	440.5990
`mod.lm.4`	$y_t = \beta + A\cos(t)$	425.6017
`mod.lm.R.1`	$y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$	425.1542
`mod.lm.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$	427.0313
`mod.lm.R.3`	$y_t = \beta + \gamma r_t$	436.9453
`mod.lm.R.4`	$y_t = \beta_0 + \beta_1 t + \gamma r_t$	438.8823
`mod.lm.R.5`	$y_t = \beta + A\cos(t) + \gamma r_t$	423.5554

Der Vergleich der AIC-Werte zeigt, dass das Modell `mod.lm.R.5` $y_t = \beta + A\cos(t) + \gamma r_t$ mit $AIC = 423.5554$ das angebrachte Modell ist, obwohl die Regenmenge keinen signifikanten Effekt hatte. Nach dem AIC-Kriterium müsste man den Niederschlag in das Modell reinnehmen.

Nichtlineare Regression Man nehme nun an, ω sei in dem Regressionsmodell 5.7 unbekannt[1]. Dadurch wird das Modell in dem Parameter ω nicht mehr linear. Die Parameter werden nun mit der Funktion `nls2()` (die genau in [Ritz u. Streibig 2008] beschrieben ist) geschätzt. Möglich sind folgende Modelle:

Modell	Regressionskurve
`mod.nlm.1`	$y_t = \beta + A\cos(\omega t) + B\sin(\omega t)$
`mod.nlm.2`	$y_t = \beta_0 + \beta_1 t + A\cos(\omega t) + B\sin(\omega t)$
`mod.nlm.R.1`	$y_t = \beta + \gamma r_t + A\cos(\omega t) + B\sin(\omega t)$
`mod.nlm.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(\omega t) + B\sin(\omega t)$

Für alle Modelle werden Funktionen und Gitter konstruiert, die für die Modellbildung mit Hilfe der Funktion `nls2()` notwendig sind (für den R-Quellcode siehe A.2.6.2). Die Vorgehensweise wird am Beispiel des Modells `mod.nlm.1` für die Regressionskurve $y_t = \beta + A\cos(\omega t) + B\sin(\omega t)$ vorgeführt.

```
> fct1.nlm<-function(beta,A,B,omega,t){
+ beta+A*cos(omega*t)+B*sin(omega*t)}

> grid1.nlm<-expand.grid(list(beta=seq(2921,2923,by=0.1),A=seq(2093,
+ 2095,by=0.1),B=seq(138,140,by=0.1),omega=seq(0.9,1.1,by=0.1)))
```

Anschließend wird das Gitter `grid1.nlm` mit den Startwerten zusammen mit der Funktion `fct1.nlm` für die Modellbildung benutzt.

[1] Diese Annahme macht zwar keinen Sinn, da die Periode - 1 Jahr - gegeben ist, es ist aber interessant zu sehen, ob die geschätzte Periode wegen der kurzen Laufzeit der Studie viel von der gegebenen abweicht.

```
> mod.nlm.1<-nls2(PTM~fct1.nlm(beta,A,B,omega,Month),data=RainM.ST,
+ start=grid1.nlm)
```

Die geschätzten Parameter und die Werte für das AIC, in denen die Regenmenge nicht aufgenommen wurde: `mod.nlm.1` und `mod.nlm.2`, werden in der nachfolgenden Tabelle aufgelistet.

Modell	Intercept	A	B	β_1	ω	γ	AIC
`mod.nlm.1` $y_t = \beta + A\cos(\omega t) + B\sin(\omega t)$	3275.84	1725.80	-1841.70	-	0.86	-	421.685
`mod.nlm.2` $y_t = \beta_0 + \beta_1 t + A\cos(\omega t) + B\sin(\omega t)$	2819.43	1757.75	-1797.59	70.82	0.86	-	422.870

In beiden Fällen weicht die geschätzte Periode ω sehr von der ab, die man eigentlich erwarten würde. Der Wert 0.86 würde heißen, dass der Zyklus

$$T = \frac{2\pi}{0.86} \cong 2.33\pi$$

dauert, was ungefähr 14 Monaten entspricht.

Den kleineren Wert für das AIC hat das Modell `mod.nlm.1`, d.h. man kann vermuten, dass der Term `Month` ausgeschlossen werden sollte. Mit der Funktion `summary()` werden die Eigenschaften des Modells `mod.nlm.2` übersichtlich dargestellt.

```
> summary(mod.nlm.2)
Formula: PTM ~ fct2.nlm(beta0, A, B, omega, beta1, Month)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta0  2.819e+03  6.438e+02   4.380 0.000322 ***
A      1.758e+03  6.734e+02   2.610 0.017211 *
B     -1.798e+03  6.417e+02  -2.801 0.011396 *
omega  8.562e-01  4.328e-02  19.782 3.89e-14 ***
beta1  7.082e+01  8.733e+01   0.811 0.427404
---
Residual standard error: 1419 on 19 degrees of freedom
```

In der Tat zeigt die Variable `Month` in dem Modell `mod.nlm.2` keinen signifikanten Einfluss auf die Zielgröße. Das Ausschließen der Variablen `Month` war richtig:

```
> summary(mod.nlm.1)
Formula: PTM ~ fct1.nlm(beta, A, B, omega, Month)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta   3.276e+03  3.068e+02  10.676 1.04e-09 ***
A      1.726e+03  6.738e+02   2.561   0.0186 *
B     -1.842e+03  6.265e+02  -2.939   0.0081 **
omega  8.585e-01  4.273e-02  20.092 9.90e-15 ***
---
Residual standard error: 1407 on 20 degrees of freedom
```

Im Vergleich zu den entsprechenden linearen Modellen (wo der Parameter ω nicht geschätzt war) hat sich die Bedeutung der Terme `cos(Month)` und `sin(Month)` geändert: Die Vergrösserung der Periode (ω wurde 0.86 geschätzt) bewirkt, dass die Regressionskurve sich mehr einer Sinuskurve ähnelt (siehe Abbildung 5.3).

Man könnte ein äquivalentes nichlineares Modell mit nur einer Einflussgröße `cos(Month)` bilden und mit dem linearen Modell `mod.lm.4` vergleichen.

```
> mod.nlm.3<-nls2(PTM~fct3.nlm(beta,A,omega,Month),data=RainM.ST,
+ start=grid3.nlm)
> summary(mod.nlm.3)
Formula: PTM ~ fct3.nlm(beta, A, omega, Month)
Parameters:
       Estimate Std. Error t value Pr(>|t|)
beta  3.020e+03  3.256e+02   9.278 7.08e-09 ***
A     2.242e+03  4.742e+02   4.729 0.000114 ***
omega 9.613e-01  2.608e-02  36.863  < 2e-16 ***
---
Residual standard error: 1553 on 21 degrees of freedom
```

Modell		Intercept	A	B	β_1	ω	γ	AIC
`mod.nlm.3`	$y_t = \beta + A\cos(\omega t)$	3020.37	2242.37	-	-	0.96	-	425.5904

Im Modell `mod.nlm.3` mit nur einer Einflussgröße `cos(Month)` wird ω beinah 1 geschätzt.

In den Modellen `mod.nlm.R.1` und `mod.nlm.R.2` wird zusätzlich die Regenmenge aufgenommen. Die Eigenschaften der beiden Modelle werden mithilfe der Funktion `summary()` geprüft:

```
> summary(mod.nlm.R.1)
Formula: PTM ~ fct1.nlm.R(beta,gamma,A,B,omega,Month,Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta   3.095e+03  5.169e+02   5.987 9.23e-06 ***
gamma  4.559e+00  1.011e+01   0.451   0.6571
A      1.605e+03  7.828e+02   2.050   0.0544 .
B     -2.177e+03  9.098e+02  -2.393   0.0272 *
omega  8.388e-01  5.504e-02  15.241 4.16e-12 ***
---
Residual standard error: 1436 on 19 degrees of freedom

> summary(mod.nlm.R.2)
Formula: PTM ~ fct2.nlm.R(beta0,gamma,A,B,omega,beta1,Month,Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta0  2116.0388   968.8305   2.184   0.0424 *
gamma    10.9112    11.2600   0.969   0.3454
A      1457.8060   842.1212   1.731   0.1005
B     -2570.0132   888.5385  -2.892   0.0097 **
omega     0.8124     0.0518  15.684  6.1e-12 ***
beta1   111.3664    94.7568   1.175   0.2552
---
Residual standard error: 1424 on 18 degrees of freedom
```

Die Aufnahme den Variablen `Rainfall` und `Month` hat den Effekt der anderen Einflussgrößen verkleinert. Der Term `cos(Month)` hat jetzt sogar keinen signifikanten Effekt auf die Zielgröße. Analog zu dem Fall ohne den Niederschlag wird auch hier ein äquivalentes nichtlineares Modell mit zwei Einflussgrößen: `cos(Month)` und `Rainfall` erzeugt.

```
> mod.nlm.R.3<-nls2(PTM~fct3.nlm.R(beta,gamma,A,omega,Month,Rainfall),
+ data=RainM.ST,start=grid3.nlm.R)
> summary(mod.nlm.R.3)
Formula: PTM ~ fct3.nlm.R(beta, gamma, A, omega, Month, Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta  3573.74070  463.77172   7.706 2.07e-07 ***
gamma  -10.92295    6.72332  -1.625 0.119896
A     2045.84535  468.55056   4.366 0.000299 ***
omega    0.97336    0.02903  33.528  < 2e-16 ***
---
Residual standard error: 1496 on 20 degrees of freedom
```

Für die Parameter haben sich folgende Werte ergeben:

Modell	Intercept	A	B	β_1	ω	γ	AIC
`mod.nlm.R.1` $y_t = \beta + \gamma r_t$ $A\cos(\omega t) + B\sin(\omega t)$	3094.84	1604.80	-2176.89	-	0.84	4.56	423.4542
`mod.nlm.R.2` $y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(\omega t) + B\sin(\omega t)$	2116.04	1457.80	-2570.01	111.37	0.81	10.91	423.7603
`mod.nlm.R.3` $y_t = \beta_0 + \gamma r_t + A\cos(\omega t)$	3573.74	2045.85	-	-	0.97	-10.92	424.6283

Nach der Aufnahme des Niederschlags in die Modelle wurde ω noch kleiner geschätzt: 0.84 bzw. 0.81.

Die Werte für das AIC von nichtlinearen Modellen werden in einer Tabelle zusammengefasst.

Modell	Regressionskurve	AIC
`mod.nlm.1`	$y_t = \beta + A\cos(\omega t) + B\sin(\omega t)$	421.6850
`mod.nlm.2`	$y_t = \beta_0 + \beta_1 t + A\cos(\omega t) + B\sin(\omega t)$	422.8700
`mod.nlm.3`	$y_t = \beta_0 + A\cos(\omega t)$	425.5904
`mod.nlm.R.1`	$y_t = \beta + \gamma r_t + A\cos(\omega t) + B\sin(\omega t)$	423.4542
`mod.nlm.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(\omega t) + B\sin(\omega t)$	423.7603
`mod.nlm.R.3`	$y_t = \beta_0 + \gamma r_t + A\cos(\omega t)$	424.6283

Als bestes unter den nichtlinearen Modellen hat sich das Modell `mod.nlm.1` mit dem AIC-Wert 421.685 ergeben. Es sei noch einmal betont, dass die nichtlinearen Modelle in diesem Fall keinen Sinn machen, da $\omega = 1$ gegeben war. Die Regressionskurven werden in Abbildung 5.3 visualisiert.

Zusammenfassung Die AIC-Werte von allen Modellen werden aufgelistet:

Lineare Modelle		
Modell	Regressionskurve	AIC
`mod.lm.1`	$y_t = \beta + A\cos(t) + B\sin(t)$	427.4957
`mod.lm.2`	$y_t = \beta_0 + \beta_1 t + A\cos(t) + B\sin(t)$	429.0368
`mod.lm.3`	$y_t = \beta_0 + \beta_1 t$	440.5990
`mod.lm.4`	$y_t = \beta + A\cos(t)$	425.6017
`mod.lm.R.1`	$y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$	425.1542
`mod.lm.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$	427.0313
`mod.lm.R.3`	$y_t = \beta + \gamma r_t$	436.9453
`mod.lm.R.4`	$y_t = \beta_0 + \beta_1 t + \gamma r_t$	438.8823
`mod.lm.R.5`	$y_t = \beta + \gamma r_t + A\cos(t)$	423.5554
Nichtlineare Modelle		
Modell	Regressionskurve	AIC
`mod.nlm.1`	$y_t = \beta + A\cos(\omega t) + B\sin(\omega t)$	421.6850
`mod.nlm.2`	$y_t = \beta_0 + \beta_1 t + A\cos(\omega t) + B\sin(\omega t)$	422.8700
`mod.nlm.3`	$y_t = \beta_0 + A\cos(\omega t)$	425.5904
`mod.nlm.R.1`	$y_t = \beta + \gamma r_t + A\cos(\omega t) + B\sin(\omega t)$	423.4542
`mod.nlm.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(\omega t) + B\sin(\omega t)$	423.7603
`mod.nlm.R.3`	$y_t = \beta_0 + \gamma r_t + A\cos(\omega t)$	424.6283

Das beste unter den linearen Modellen ist das Modell `mod.lm.R.5`

$$y_t = 3597.98 - 12.54 r_t + 1944.77\cos(t)$$

mit dem AIC-Wert 423.5554. Da eigentlich eine Regressionsgleichung der Form

$$y_t = \beta + \gamma r_t + \alpha \cos(\omega t - \theta)$$

gesucht war (vgl. Gleichung 5.7), muss die Kurve $y_t = \beta + \gamma r_t + A\cos(t)$ zurücktransformiert werden. Da kein Sinusterm vorhanden ist, unterscheidet sich die gesuchte Darstellung kaum von der bereits existierenden:

$$y_t = 3597.98 - 12.54 r_t + 1944.77 \cos(1 \cdot t - 0)\,. \tag{5.9}$$

Die Phasenverschiebung ist also gleich Null. Natürlich erfüllt jedes $\tilde{\theta} = k\frac{2\pi}{\omega} = k2\pi$, $k \in \mathbb{N}$ die Gleichung 5.9. Das heißt, dass alle sechs Monate ein Wachstum der Anzahl der Kartoffelmotte beobachtet wird.

Für das beste Modell unter den nichtlinearen `mod.nlm.1`

$$y_t = 3094.84 + 1725.8\cos(0.84t) - 1841.7\sin(0.84t)$$

ergibt sich als Lösung (vgl. Definition 2.4.2 und Bemerkung 2.4.3)

```
> alpha <- sqrt((1725.8)^2+(-1841.7)^2)
> alpha
[1] 2523.934
> theta <- atan(-1841.7/1725.8)/pi
> theta
[1] -0.2603375
```

$$y_t = 3094.84 + 2523.934\cos(0.84t + 0.2603375\pi). \tag{5.10}$$

In diesem Fall heißt es, dass ungefähr alle sieben Monate ein Wachstum der Mottenanzahl beobachtet wird.

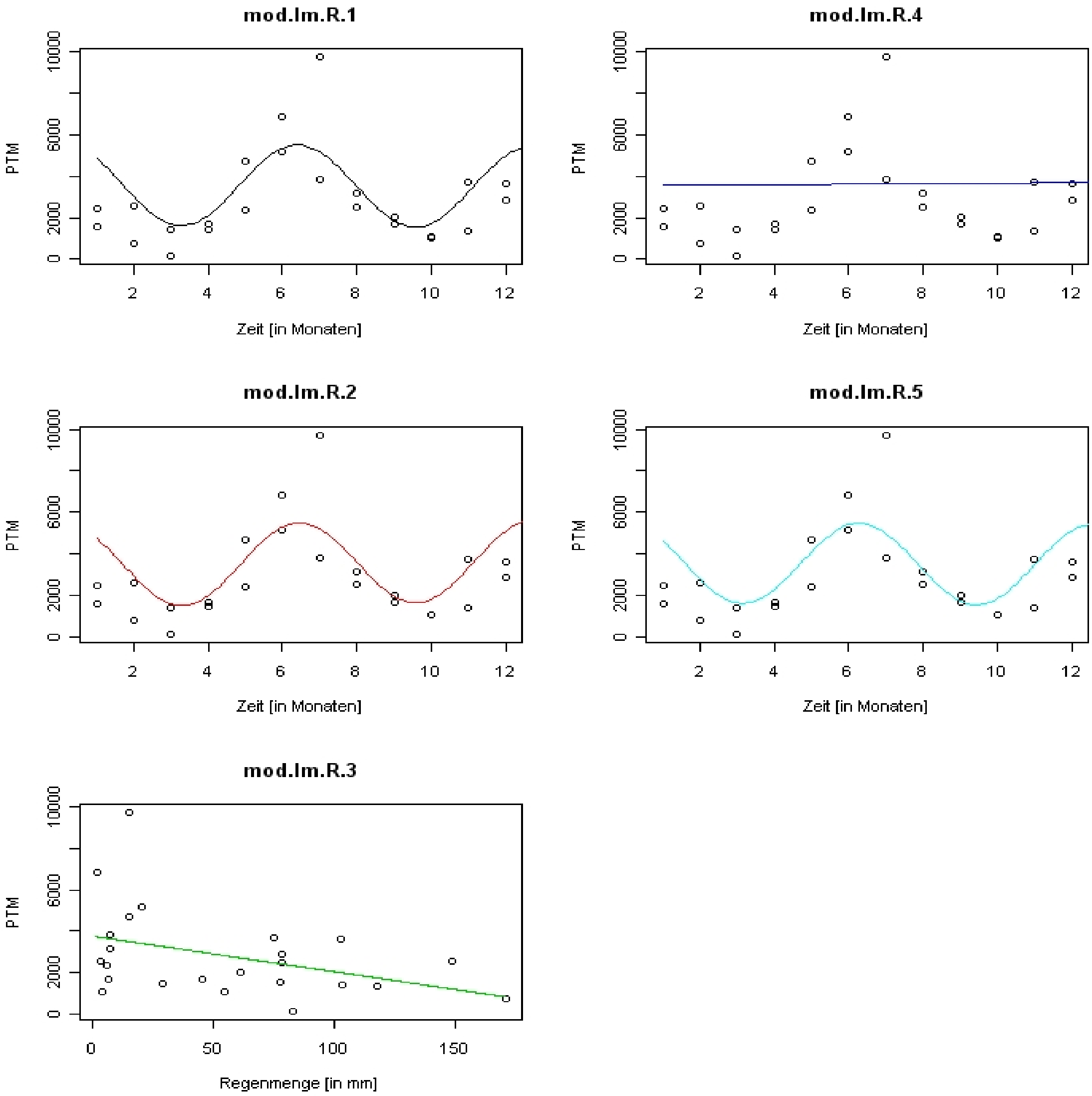

Abbildung 5.2: Vergleich der Modelle mit Regenmenge, ohne Verschiebung, lineare Regression.

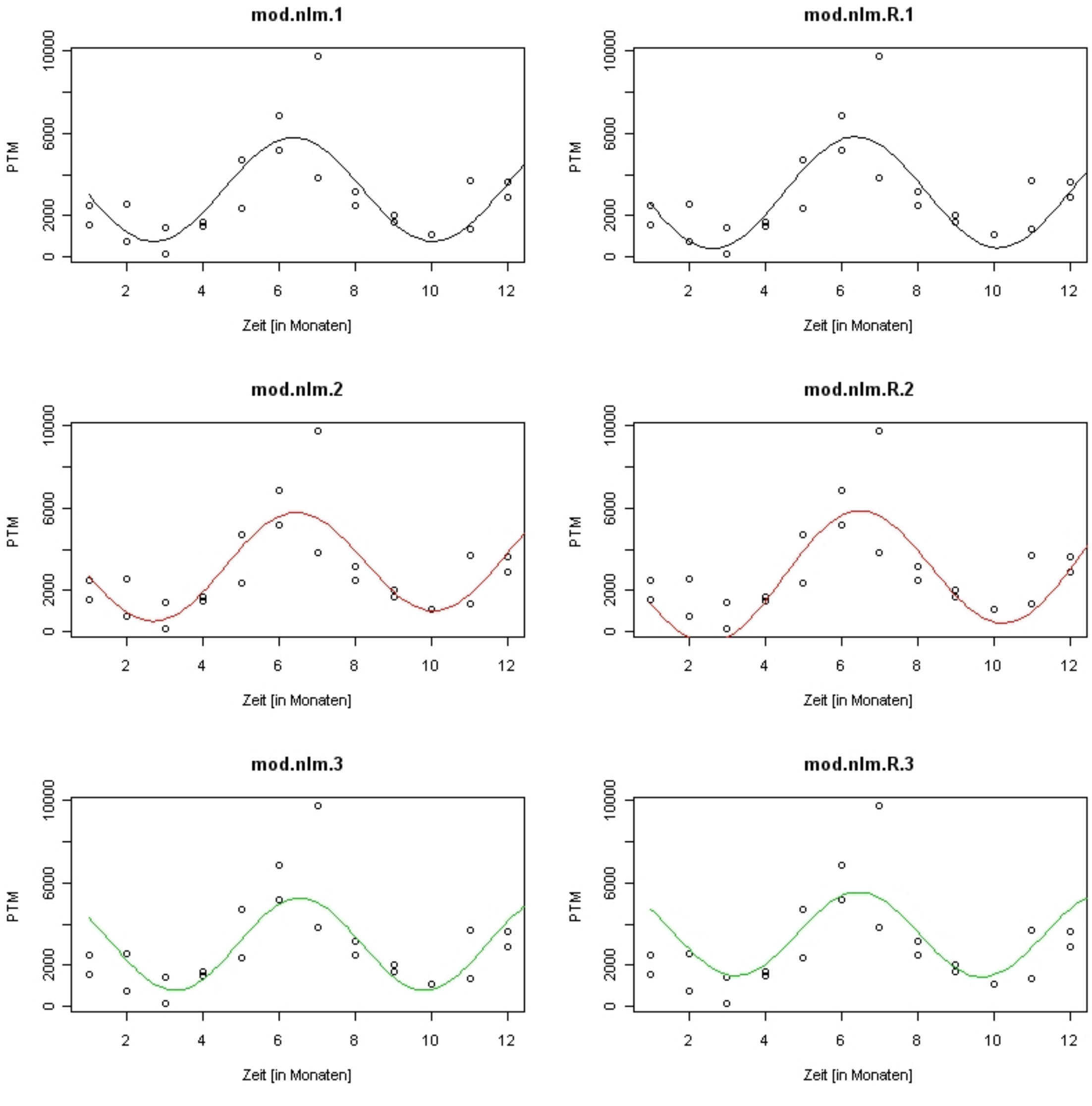

Abbildung 5.3: Vergleich der nichtlinearen Modelle ohne Verschiebung, Symmetrischema Tangolias.

5.4 Pheromon-Fänge an der Station 'La Victoria' - Symmetrischema Tangolias, Teil 3

In Abschnitt 5.3 hat sich für den ursprünglichen Datensatz A.1.2 das Modell `mod.lm.R.5` $y_t = \beta + A\cos(t) + \gamma r_t$ als angebrachtes Modell mit dem AIC-Wert 423.5554 ergeben. In Kapitel 1 hat die Verschiebung um 5 Monate die größte Korrelation geliefert. Nun wird es getestet, ob die größte Korrelation auch den niedrigsten AIC-Wert zufolge hat.

Tabelle 5.1: Flugaktivität von Symmetrischema Tangolias nach der Verschiebung um fünf Monate. Der Datensatz month.ptm.rain.

Year	Month	PTM	YearR	MonthR	Rainfall
1999	4	1447	1998	11	74.7
1999	5	2372	1998	12	102.5
1999	6	5169	1999	1	78.4
1999	7	9741	1999	2	148.1
1999	8	2519	1999	3	103.3
1999	9	2014	1999	4	28.6
1999	10	1056	1999	5	6.2
1999	11	1371	1999	6	20.4
1999	12	2869	1999	7	15.0
2000	1	1573	1999	8	3.6
2000	2	768	1999	9	60.8
2000	3	124	1999	10	54.2
2000	4	1668	1999	11	117.8
2000	5	4702	1999	12	78.3
2000	6	6849	2000	1	77.6
2000	7	3824	2000	2	170.5
2000	8	3145	2000	3	82.8
2000	9	1670	2000	4	45.5
2000	10	1062	2000	5	15.0

Es wird also für den Datensatz in Tabelle 5.1 folgendes Modell gesucht

$$y_t = \beta + \gamma r_{t-5} + A\cos(t) + B\sin(t),$$

wobei β, A, B und γ zu schätzende Parameter sind. Jetzt wegen der Verschiebung um fünf Monate sind es nur 24-5=19 statt 24 Beobachtungen. Zuerst wird es wie im letzten Abschnitt mit linearen Modellen begonnen.

Lineare Regression Da die Regression ohne den Niederschlag jetzt keinen Sinn macht, werden hier nur folgende Modelle untersucht:

Modell	Regressionskurve
`mod.lm.R.V.1`	$y_t = \beta + \gamma r_{t-5} + A\cos(t) + B\sin(t)$
`mod.lm.R.V.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_{t-5} + A\cos(t) + B\sin(t)$
`mod.lm.R.V.3`	$y_t = \beta + \gamma r_{t-5}$
`mod.lm.R.V.4`	$y_t = \beta_0 + \beta_1 t + \gamma r_{t-5}$

Mit dem Shapiro-Wilk-Test werden die Modelle `mod.lm.R.V.1` - `mod.lm.R.V.4` auf die Normalverteilung getestet (für den R-Quellcode siehe A.2.7). Die P-Werte werden in folgender Tabelle aufgelistet:

Modell	Regressionskurve	Shapiro-Wilk P-Wert
`mod.lm.R.V.1`	$y_t = \beta + \gamma r_{t-5} + A\cos(t) + B\sin(t)$	0.3783714
`mod.lm.R.V.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_{t-5} + A\cos(t) + B\sin(t)$	0.6351482
`mod.lm.R.V.3`	$y_t = \beta + \gamma r_{t-5}$	0.1266351
`mod.lm.R.V.4`	$y_t = \beta_0 + \beta_1 t + \gamma r_{t-5}$	0.02677987

Die Normalverteilung wird nur in dem letzten Modell verworfen. Der Bartlett-Test kann nicht angewandt werden, denn in manchen Monaten gibt es nach der Verschiebung nur eine Beobachtung.

```
> bartlett.test(PTM~Month,data=month.ptm.rain)
Error in bartlett.test.default(c(1447,2372,5169,9741,2519,2014,1056, :
  there must be at least 2 observations in each group
```

Das größte Modell `mod.lm.R.V.2` wird mit dem ANOVA-Test getestet.

```
> anova(lm(PTM~cos(Month)+Rainfall+sin(Month)+Month,data=month.ptm.rain))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
cos(Month)  1 51645771 51645771 19.9503 0.0005322 ***
Rainfall    1 10203962 10203962  3.9417 0.0670501 .
sin(Month)  1    42964    42964  0.0166 0.8993259
Month       1  2978866  2978866  1.1507 0.3015434
Residuals  14 36242015  2588715
```

Aus dem Test folgt, dass die Terme `Rainfall`, `sin(Month)` und `Month` keinen signifikanten Effekt auf die Zielgröße haben. Die umgekehrte Reihenfolge von `Rainfall` und `cos(Month)` wird untersucht.

```
> anova(lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,data=month.ptm.rain))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value   Pr(>F)
Rainfall    1 30411947 30411947 11.7479 0.004083 **
cos(Month)  1 31437786 31437786 12.1442 0.003643 **
sin(Month)  1    42964    42964  0.0166 0.899326
Month       1  2978866  2978866  1.1507 0.301543
Residuals  14 36242015  2588715
```

`sin(Month)` und `Month` werden aus dem Modell ausgeschlossen. Mit `summary(lm(...))` wird getestet, ob (β, A, γ) sich signifikant von Null unterscheidet.

```
> summary(lm(PTM~Rainfall+cos(Month),data=month.ptm.rain))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1811.351    676.691   2.677  0.01654 *
Rainfall      16.941      8.308   2.039  0.05831 .
cos(Month)  1974.527    551.662   3.579  0.00251 **
---
Residual standard error: 1567 on 16 degrees of freedom
Multiple R-squared: 0.6117,Adjusted R-squared: 0.5631
F-statistic:  12.6 on 2 and 16 DF,  p-value: 0.000517
```

Das ist der Fall ($P = 0.000517$). Der Effekt des Niederschlags ist in diesem Modell nicht signifikant. Es kann daran liegen, dass der Datensatz nur neunzehn Messwerte enthält. Zu testen bleibt noch, ob die einzelnen Variablen einen signifikanten Einfluss haben.

```
> summary(lm(PTM~cos(Month),data=month.ptm.rain))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2978.9      392.7   7.585 7.48e-07 ***
cos(Month)    2369.6      562.5   4.213 0.000585 ***
---
Residual standard error: 1706 on 17 degrees of freedom
Multiple R-squared: 0.5108,Adjusted R-squared: 0.482
F-statistic: 17.75 on 1 and 17 DF,  p-value: 0.0005851

> summary(lm(PTM~Rainfall,data=month.ptm.rain))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   989.55     828.68   1.194   0.2488
Rainfall       27.38      10.13   2.704   0.0150 *
---
Residual standard error: 2039 on 17 degrees of freedom
Multiple R-squared: 0.3008,Adjusted R-squared: 0.2596
F-statistic: 7.312 on 1 and 17 DF,  p-value: 0.01504
```

Das Ergebnis weist auf einen signifikanten Effekt hinsichtlich der Regenmenge ($P = 0.01504$) und des `cos(Month)` ($P = 0.0005851$) auf die Anzahl der Männchen der Kartoffelmotte hin. Das Modell wird `mod.lm.R.V.5` genannt:

```
> mod.lm.R.V.5<-lm(PTM~cos(Month)+Rainfall,data=month.ptm.rain)
```

In diesem Fall liegt die Normalverteilung vor.

```
> shapiro.test(aov(PTM~Rainfall+cos(Month),data=month.ptm.rain)$residuals)
+ $p.value
[1] 0.3435277
```

Es werden noch die ursprünglichen Modelle gebildet und mit dem Modell `mod.lm.R.5` verglichen.

```
> mod.lm.R.V.1<-lm(PTM~Rainfall+cos(Month)+sin(Month),
+ data=month.ptm.rain)
> mod.lm.R.V.2<-lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,
+ data=month.ptm.rain)
> mod.lm.R.V.3<-lm(PTM~Rainfall,data=month.ptm.rain)
> mod.lm.R.V.4<-lm(PTM~Rainfall+Month,data=month.ptm.rain)
```

Die geschätzten Parameter und die AIC-Werte werden in folgender Tabelle aufgelistet:

Modell	Intercept	A	B	β_1	γ	AIC
`mod.lm.R.V.1` $y_t = \beta + \gamma r_{t-5} + A\cos(t) + B\sin(t)$	1829.67	1978.42	68.03	-	16.71	340.1850
`mod.lm.R.V.2` $y_t = \beta_0 + \beta_1 t + \gamma r_{t-5} + A\cos(t) + B\sin(t)$	759.87	1946.10	149.16	137.75	18.92	340.6842
`mod.lm.R.V.3` $y_t = \beta_0 + \gamma r_{t-5}$	989.55	-	-	-	27.38	347.3809
`mod.lm.R.V.4` $y_t = \beta_0 + \beta_1 t + \gamma r_{t-5}$	-250.76	-	-	158.75	30.04	348.2596
`mod.lm.R.V.5` $y_t = \beta + \gamma r_{t-5} + A\cos(t)$	1811.35	1974.53	-	-	16.94	338.2058

Das letzte Modell `mod.lm.R.V.5` zeigt die beste Anpassung an den Datensatz `month.ptm.rain` ($AIC = 338.2058$). Alle fünf Kurven werden in Abbildung 5.4 geplottet.

Datensatz RainM.ST mit 19 Beobachtungen, lineare Regression In Kapitel 1 wurden die Mottenanzahlen gegen die Regenmenge um fünf Monate verschoben. Man könnte aber auch einen nichtverschobenen, um fünf Messwerte reduzierten Datensatz zum Vergleich heranziehen und sehen, ob sich die Verschiebung gelohnt hat (für den R-Quellcode siehe A.2.8).

Der reduzierte Datensatz wird gebildet und ist in Tabelle 5.2 zu sehen.

```
> RainM.ST_19<-RainM.ST[6:24,]
```

Tabelle 5.2: Der Datensatz RainM.ST_19.

Year	Month	Rainfall	PTM
1999	4	28.6	1447
1999	5	6.2	2372
1999	6	20.4	5169
1999	7	15	9741
1999	8	3.6	2519
1999	9	60.8	2014
1999	10	54.2	1056
1999	11	117.8	1371
1999	12	78.3	2869
2000	1	77.6	1573
2000	2	170.5	768
2000	3	82.8	124
2000	4	45.5	1668
2000	5	15	4702
2000	6	1.8	6849
2000	7	7.3	3824
2000	8	7.3	3145
2000	9	6.3	1670
2000	10	4.1	1062

Folgende Modelle werden betrachtet:

Modell	Regressionskurve
`mod.lm_19.R.1`	$y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$
`mod.lm_19.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$
`mod.lm_19.R.3`	$y_t = \beta + \gamma r_t$
`mod.lm_19.R.4`	$y_t = \beta_0 + \beta_1 t + \gamma r_t$

Weil die Residuen nicht normalverteilt sind, wird bei den ANOVA-Test Vorsicht geboten.

Modell	Regressionskurve	Shapiro-Wilk P-Wert
`mod.lm_19.R.1`	$y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$	0.02823879
`mod.lm_19.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$	0.02687639
`mod.lm_19.R.3`	$y_t = \beta + \gamma r_t$	0.01940782
`mod.lm_19.R.4`	$y_t = \beta_0 + \beta_1 t + \gamma r_t$	0.01725636

Das größte Modell `mod.lm_19.R.2` $y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$ wird untersucht.

```
> anova(lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,data=RainM.ST_19))
Analysis of Variance Table

Response: PTM
           Df   Sum Sq  Mean Sq F value   Pr(>F)
Rainfall    1 21432778 21432778  8.0867 0.013010 *
cos(Month)  1 41286833 41286833 15.5777 0.001461 **
sin(Month)  1  1192173  1192173  0.4498 0.513337
Month       1    96523    96523  0.0364 0.851394
Residuals  14 37105271  2650376
```

Wie in den vorherigen Tests werden `sin(Month)` und `Rainfall` aus dem Modell ausgeschlossen.

```
> summary(lm(PTM~Rainfall+cos(Month),data=RainM.ST_19))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3695.385    488.312   7.568 1.13e-06 ***
Rainfall     -17.247      8.028  -2.148 0.047355 *
cos(Month)  2157.770    520.201   4.148 0.000757 ***
---
Residual standard error: 1549 on 16 degrees of freedom
Multiple R-squared: 0.6203,Adjusted R-squared: 0.5728
F-statistic: 13.07 on 2 and 16 DF,  p-value: 0.0004321
```

Der Vektor (β, γ, A) unterscheidet sich wesentlich von Null ($P = 0.0004321$). Es werden noch Tests auf Signifikanz der einzelnen Variablen durchgeführt:

```
> summary(lm(PTM~cos(Month),data=RainM.ST_19))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2978.9      392.7   7.585 7.48e-07 ***
cos(Month)    2369.6      562.5   4.213 0.000585 ***
---
Residual standard error: 1706 on 17 degrees of freedom
Multiple R-squared: 0.5108,Adjusted R-squared: 0.482
F-statistic: 17.75 on 1 and 17 DF,  p-value: 0.0005851

> summary(lm(PTM~Rainfall,data=RainM.ST_19))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3834.90     680.84   5.633 2.98e-05 ***
Rainfall      -23.56      11.02  -2.138   0.0473 *
---
Residual standard error: 2165 on 17 degrees of freedom
Multiple R-squared: 0.212,Adjusted R-squared: 0.1656
F-statistic: 4.573 on 1 and 17 DF,  p-value: 0.0473
```

In beiden Fällen sind die Effekte signifikant. Dieses Modell wird `mod.lm_19.R.5` genannt.

```
> mod.lm_19.R.5<-lm(PTM~Rainfall+cos(Month),data=RainM.ST_19)
```

Die Residuen des Modells sind auch nicht normalverteilt.

```
> shapiro.test(aov(PTM~cos(Month)+Rainfall,data=RainM.ST_19)$residuals)
+ $p.value
[1] 0.007708397
```

Modell	Regressionskurve	Shapiro-Wilk P-Wert
`mod.lm_19.R.5`	$y_t = \beta + \gamma r_t + A\cos(t)$	0.007708397

Die ursprünglichen vier Modelle werden gebildet:

```
> mod.lm_19.R.1<-lm(PTM~Rainfall+cos(Month)+sin(Month),data=RainM.ST_19)
> mod.lm_19.R.2<-lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,
+ data=RainM.ST_19)
> mod.lm_19.R.3<-lm(PTM~Rainfall,data=RainM.ST_19)
> mod.lm_19.R.4<-lm(PTM~Rainfall+Month,data=RainM.ST_19)
```

In folgender Tabelle werden die Anpassung und die geschätzten Parameter verglichen:

Modell	Intercept	A	B	β_1	γ	AIC
`mod.lm_19.R.1` $y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$	3721.47	2144.94	351.12	-	-17.62	339.1808
`mod.lm_19.R.2` $y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$	3544.36	2152.28	369.33	24.61	-17.29	341.1314
`mod.lm_19.R.3` $y_t = \beta_0 + \gamma r_t$	3834.90	-	-	-	-23.56	349.6526
`mod.lm_19.R.4` $y_t = \beta_0 + \beta_1 t + \gamma r_t$	4045.00	-	-	-29.05	-23.93	351.6192
`mod.lm_19.R.5` $y_t = \beta + \gamma r_t + A\cos(t)$	3695.38	2157.77	-	-	-17.25	337.7801

Die Regressionskurven sind in Abbildung 5.5 zu sehen.

Nichtlineare Regression Folgende Kurven werden untersucht (für den R-Quellcode siehe A.2.9).

Modell	Regressionskurve
`mod.nlm.V.R.1`	$y_t = \beta + \gamma r_{t-5} + A\cos(\omega t) + B\sin(\omega t)$
`mod.nlm.V.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_{t_5} + A\cos(\omega t) + B\sin(\omega t)$

Für die Parameter und für das AIC haben sich folgende Werte ergeben:

Modell	Intercept	A	B	β_1	γ	ω	AIC
`mod.nlm.V.R.1` $y_t = \beta + \gamma r_{t_5} + A\cos(\omega t) + B\sin(\omega t)$	3311.63	1653.95	-2455.80	-	-2.71	0.82	337.1963
`mod.nlm.V.R.2` $y_t = \beta_0 + \beta_1 t + \gamma r_{t_5} + A\cos(\omega t) + B\sin(\omega t)$	2451.56	1714.23	-2276.65	94.66	0.53	0.83	338.3128

Auch hier weicht die geschätzte Periode ω sehr von der ab, die man eigentlich erwartet. Mit der Funktion `summary()` werden die Eigenschaften der beiden Modelle dargestellt.

```
> summary(mod.nlm.V.R.1)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta   3.312e+03  9.019e+02   3.672  0.00251 **
gamma -2.709e+00  1.175e+01  -0.231  0.82094
A      1.654e+03  9.736e+02   1.699  0.11145
B     -2.456e+03  1.040e+03  -2.361  0.03325 *
omega  8.240e-01  5.812e-02  14.177 1.07e-09 ***
---
Residual standard error: 1468 on 14 degrees of freedom

> summary(mod.nlm.V.R.2)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
```

```
beta0  2451.5557  1418.7615   1.728   0.1077
gamma     0.5323    12.5486   0.042   0.9668
A      1714.2260   951.9936   1.801   0.0950 .
B     -2276.6486  1107.3224  -2.056   0.0604 .
omega     0.8328     0.0624  13.346 5.78e-09 ***
beta1    94.6593   120.0893   0.788   0.4447
---
Residual standard error: 1488 on 13 degrees of freedom
```

Man beobachtet, ähnlich wie in Abschnitt 5.3, keinen signifikanten Effekt der Regenmenge und und des `cos(Month)` auf die Zielgröße. Es ist klar, dass der lineare Term `Month` ausgeschlossen werden soll. Auch hier wird zum Vergleich ein nichtlineares Modell mit zwei Einflussgrößen: `cos(Month)` und `Rainfall` erzeugt.

```
> mod.nlm.V.R.3<-nls2(PTM~fct3.nlm.V.R(beta,gamma,A,omega,Month,Rainfall),
+ data=month.ptm.rain,start=grid3.nlm.V.R)

> summary(mod.nlm.V.R.3)
Formula: PTM ~ fct3.nlm.V.R(beta,gamma,A,omega,Month,Rainfall)
Parameters:
       Estimate Std. Error t value Pr(>|t|)
beta  2.102e+03  7.692e+02   2.733   0.0154 *
gamma 1.329e+01  9.491e+00   1.401   0.1816
A     2.122e+03  6.029e+02   3.519   0.0031 **
omega 9.684e-01  3.568e-02  27.139 3.64e-14 ***
---
Residual standard error: 1591 on 15 degrees of freedom
```

Der Effekt des Niederschlags ist wie in allen anderen nichtlinearen Modellen nicht wesentlich.

Modell	Intercept	A	B	β_1	γ	ω	AIC
`mod.nlm.V.R.3`	2102.19	2121.83	-	-	13.29	0.97	339.5731
$y_t = \beta + \gamma r_{t-5} + A\cos(\omega t)$							

Die Regressionskurven werden in Abbildung 5.6 visualisiert.

Der reduzierte Datensatz RainM.ST mit nur 19 Beobachtungen, nichtlineare Regression Nun werden nichtlineare Modelle gebildet, die auf dem reduzierten Datensatz `RainM.ST_19` basieren (Für den R-Quellcode siehe A.2.10). Folgende Regressionskurven werden betrachtet.

Modell	Regressionskurve
`mod.nlm_19.R.1`	$y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$
`mod.nlm_19.R.2`	$y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$
`mod.nlm_19.R.3`	$y_t = \beta_0 + \gamma r_t + A\cos(t)$

Mit der Funktion `summary()` werden die Eigenschaften der drei Modelle aufgelistet:

```
> summary(mod.nlm_19.R.1)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta   2.990e+03  5.993e+02   4.990 0.000198 ***
gamma  3.213e+00  1.228e+01   0.262 0.797427
A      1.589e+03  1.055e+03   1.506 0.154296
B     -2.494e+03  1.056e+03  -2.362 0.033181 *
omega  8.190e-01  6.589e-02  12.429 5.95e-09 ***
---
Residual standard error: 1467 on 14 degrees of freedom

> summary(mod.nlm_19.R.2)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta0  1.930e+03  1.192e+03   1.620   0.1293
gamma  8.803e+00  1.323e+01   0.665   0.5175
A      1.369e+03  1.152e+03   1.188   0.2562
B     -2.903e+03  1.019e+03  -2.850   0.0137 *
omega  8.000e-01  6.176e-02  12.952 8.32e-09 ***
beta1  1.253e+02  1.225e+02   1.023   0.3249
---
Residual standard error: 1465 on 13 degrees of freedom
```

```
> summary(mod.nlm_19.R.3)
Formula: PTM ~ fct3.nlm_19.R(beta, gamma, A, omega, Month, Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta  3630.44192  492.71997   7.368 2.34e-06 ***
gamma  -14.57802    8.46796  -1.722 0.105703
A     2278.14952  534.93951   4.259 0.000687 ***
omega    0.96605    0.03115  31.016 5.08e-15 ***
---
Residual standard error: 1546 on 15 degrees of freedom
```

Die Ergebnisse zeigen typisches Verhalten für nichtlineare Regression mit Regenmenge: In den ersten zwei Modellen bewirkt die Vergrösserung der Periode, dass die Regressionskurve sich mehr einer Sinuskurve ähnelt - der Parameter A ist nicht signifikant. Auch der Parameter γ hat keinen wesentlichen Effekt auf die Zielgröße. Das ändert sich in dem reduzierten Modell `mod.nlm_19.R.3` nicht.

Die geschätzten Parameter und die Anpassung:

Modell	Intercept	A	B	β_1	γ	ω	AIC
`mod.nlm_19.R.1` $y_t = \beta + \gamma r_t + A\cos(t) + B\sin(t)$	2990.40	1588.52	-249.79	-	3.21	0.82	337.1810
`mod.nlm_19.R.2` $y_t = \beta_0 + \beta_1 t + \gamma r_t + A\cos(t) + B\sin(t)$	1930.00	1368.83	-2903.36	125.31	8.80	0.80	337.7135
`mod.nlm_19.R.3` $y_t = \beta_0 + \gamma r_t + A\cos(t)$	3630.44	2278.15	-2903.36	125.31	-14.58	0.96	338.4756

Die Regressionskurven sind in Abbildung 5.7 zu sehen.

Zusammenfassung Die AIC-Werte von allen Modellen, die nach der Verschiebung des Datensatzes bzw. nach Entnehmung fünf Messwerte entstanden sind, werden in einer Tabelle zusammengefasst: In der ersten Spalte steht die Regressionsfunktion, in der zweiten der AIC-Wert für den Datensatz `month.ptm.rain` und in der letzten das AIC für den reduzierten Datensatz `RainM.ST_19`.

Lineare Modelle		
	month.ptm.rain $r = r_{t-5}$	RainM.ST_19 $r = r_t$
Regressionsfunktion	AIC	AIC
1. $y_t = \beta + \gamma r + A\cos(t) + B\sin(t)$	340.1850	339.1808
2. $y_t = \beta_0 + \beta_1 t + \gamma r + A\cos(t) + B\sin(t)$	340.6842	341.1314
3. $y_t = \beta + \gamma r$	347.3809	349.6526
4. $y_t = \beta_0 + \beta_1 t + \gamma r$	348.2596	351.6192
5. $y_t = \beta + \gamma r + A\cos(t)$	338.2058	337.7801
Nichtlineare Modelle		
	month.ptm.rain $r = r_{t-5}$	RainM.ST_19 $r = r_t$
Regressionsfunktion	AIC	AIC
1. $y_t = \beta + \gamma r + A\cos(\omega t) + B\sin(\omega t)$	337.1963	337.1810
2. $y_t = \beta_0 + \beta_1 t + \gamma r + A\cos(\omega t) + B\sin(\omega t)$	338.3128	337.7135
3. $y_t = \beta + \gamma r + A\cos(\omega t)$	339.5731	338.4756

Der hier untersuchte Datensatz enthält Mottenfänge, die nur über zwei Jahre gesammelt wurden. Aus diesem Grund lässt sich nicht sagen, welches Verhalten der Kartoffelmotte normal ist: die relativ hohen Werte im Juli/August 1998 oder die niedrigen im Winter 1999? Mann kann folglich keine dieser Beobachtungen als Ausreißer bezeichnen. Durch die Verschiebung des Datensatzes nahmen Mottenanzahlen und Regenmenge zu gleicher Zeit zusammen zu und zusammen wieder ab. Durch den Vergleich des verschobenen Datensatzes mit dem reduzierten wurde klar, dass die Verschiebung eigentlich nichts gebracht hat. Das Modell `mod.lm.R.V.5`

$$y_t = 1811.35 + 16.94 r_{t-5} + 1974.53\cos(t-0)$$

mit dem kleinsten AIC-Wert (338.2058) unter den "verschobenen" Modellen hat schlechtere Anpassung als das äquivalente Modell `mod.lm.R_19.5`

$$y_t = 3695.38 - 17.25 r_t + 2157.77\cos(t-0)$$

($AIC = 337.7801$). Der Unterschied im AIC-Wert ist aber sehr gering. Im Gegensatz zu den Modellen für den nichtverschobenen Datensatz sind die Residuen des Modells `mod.lm.R.V.5` normalverteilt, so dass hier viel eher der ANOVA-Test angewendet werden kann.

Unter den nichtlinearen Modellen ist das Modell `mod.nlm_19.R.1` $y_t = \beta + \gamma r_t + A\cos(\omega t) + B\sin(\omega t)$ dasjenige, das die beste Anpassung an den Datensatz RainM.ST_19 zeigt ($AIC = 337.1810$). Nach der Rücktransformation

```
> alpha <- sqrt((1588.519)^2+(-2493.789)^2)
> alpha
[1] 2956.751
> theta <- atan(-2493.789/1588.519)/pi
> theta
[1] -0.3194626
```

ergibt sich

$$y_t = 2990.4 + 3.21 r_t + 2956.751\cos(0.82t + 0.3194626\pi).$$

Die Analyse des Datensatzes hat gezeigt, dass das Modell

$$y_t = \beta + \gamma r_t + A\cos(t - 0)$$

das angebrachte Modell in beiden Abschnitten 5.3 und 5.4 war. Weil aber in Abschnitt 5.4 die Residuen des Modells normalverteilt waren, würde man sich für dieses Modell entscheiden. Das heißt, das Programm ILCYM könnte um das Modell `mod.lm.R.V.5` mit der Regressionsgleichung

$$y_t = 1811.35 + 16.94 r_{t-5} + 1974.53\cos(t - 0)$$

erweitert werden, um die Mottenanzahl der Symmetrischema Tangolias in Gebieten zu simulieren, in denen ähnliche Wetterbedingungen zu Huancayo (Peru) herrschen. Die Regionen, wo diese Gleichung eingesetzt werden könnte, sind in Abbildung 5.8 zu sehen. Die zweite Weltkarte (Abb.5.9) zeigt, wo Kartoffeln überhaupt angebaut werden.

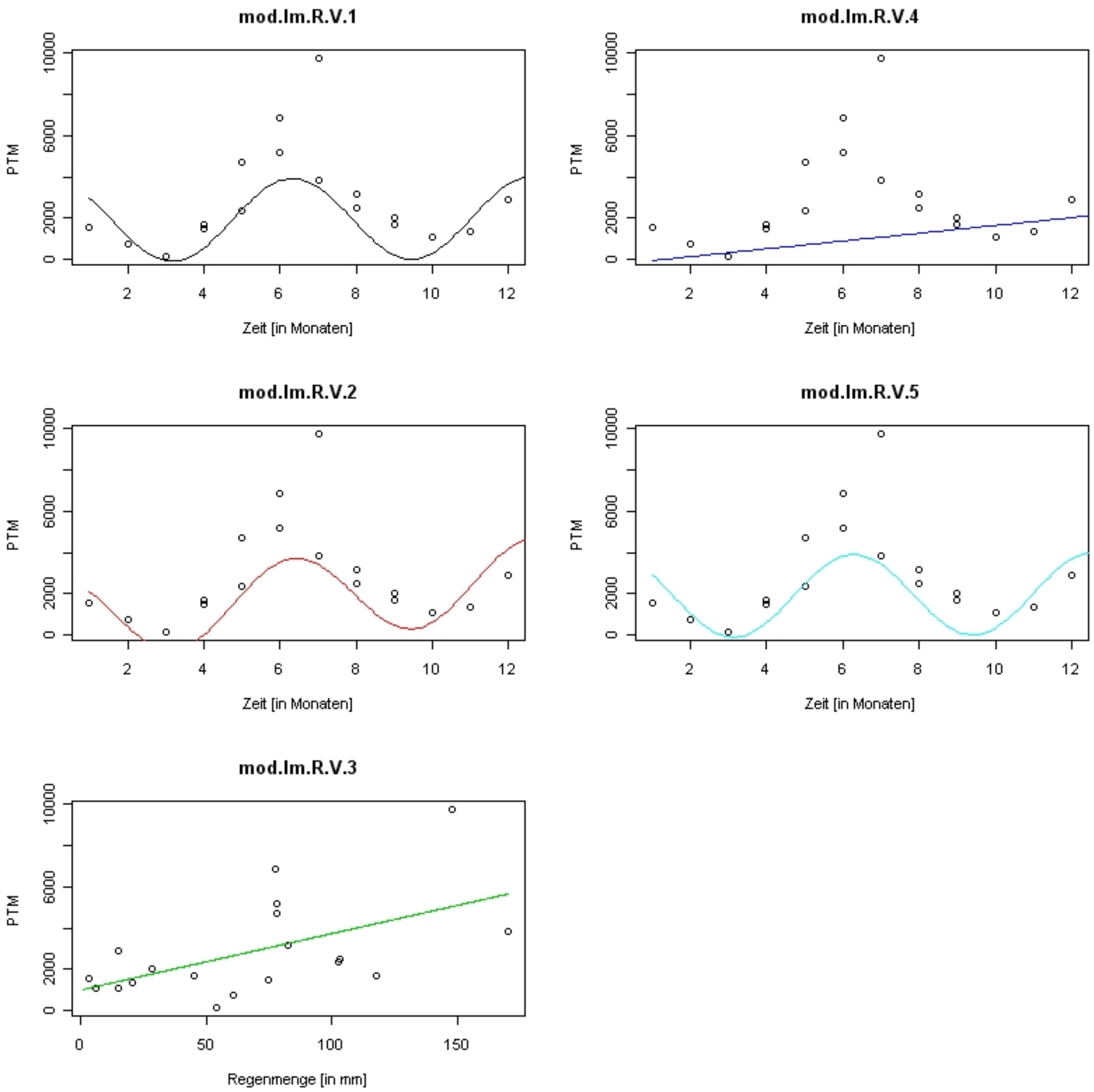

Abbildung 5.4: Vergleich der Modelle mit Regenmenge, mit Verschiebung, lineare Regression.

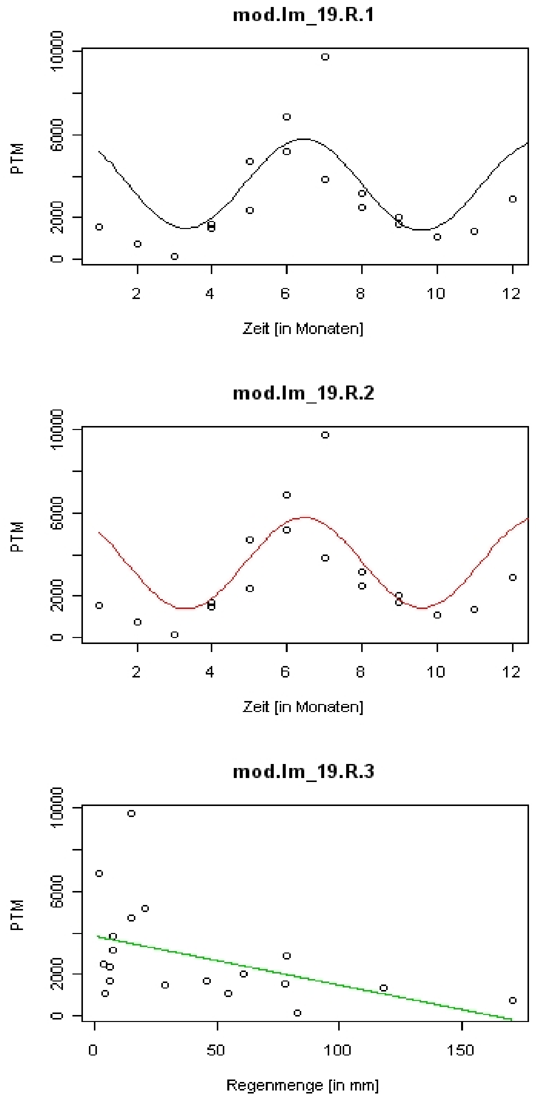

Abbildung 5.5: Vergleich der Modelle mit Regenmenge, Datensatz RainM.ST_19, lineare Regression.

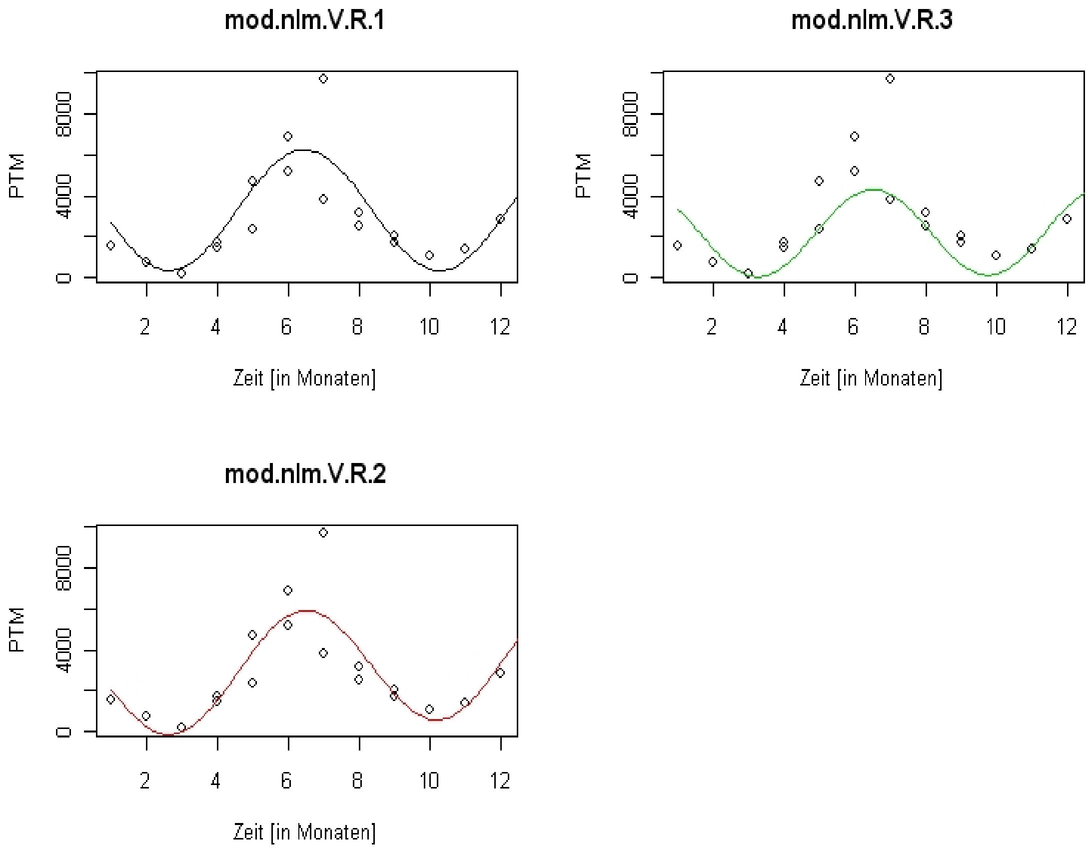

Abbildung 5.6: Vergleich der Modelle mit Regenmenge, nach der Verschiebung, nichtlineare Regression.

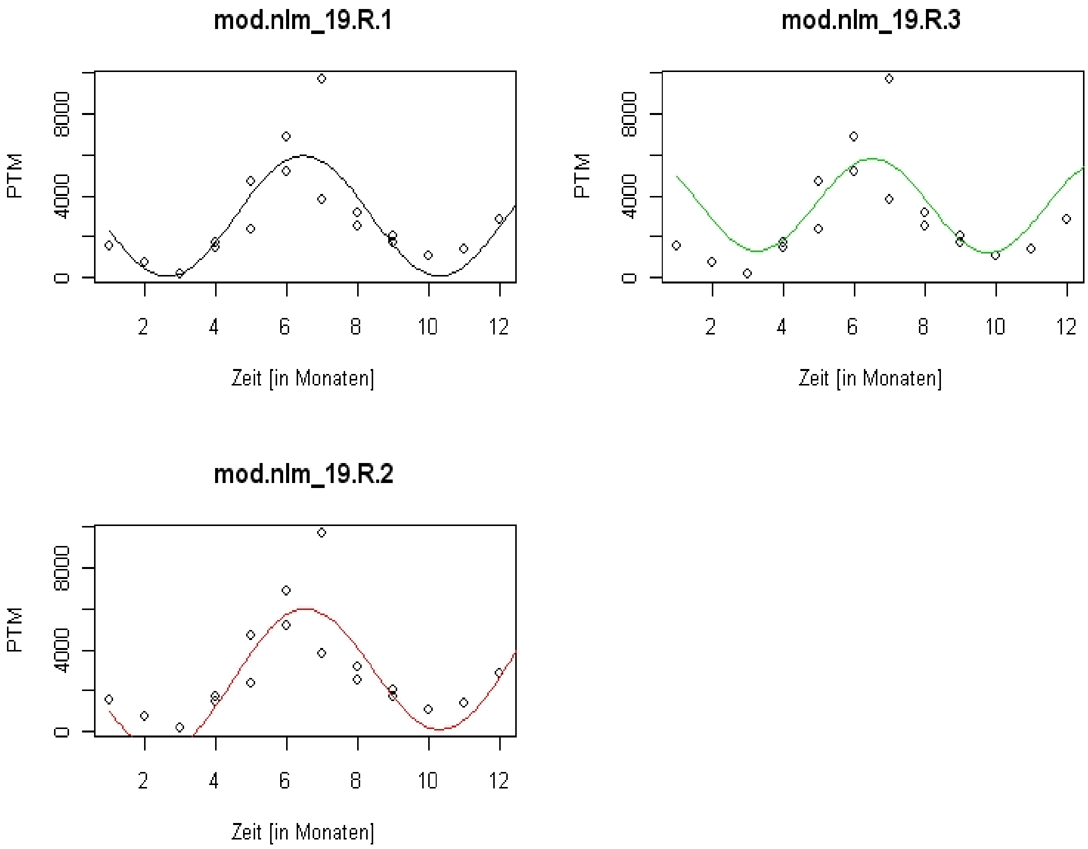

Abbildung 5.7: Vergleich der Modelle mit Regenmenge, Datensatz RainM.ST_19, nichtlineare Regression.

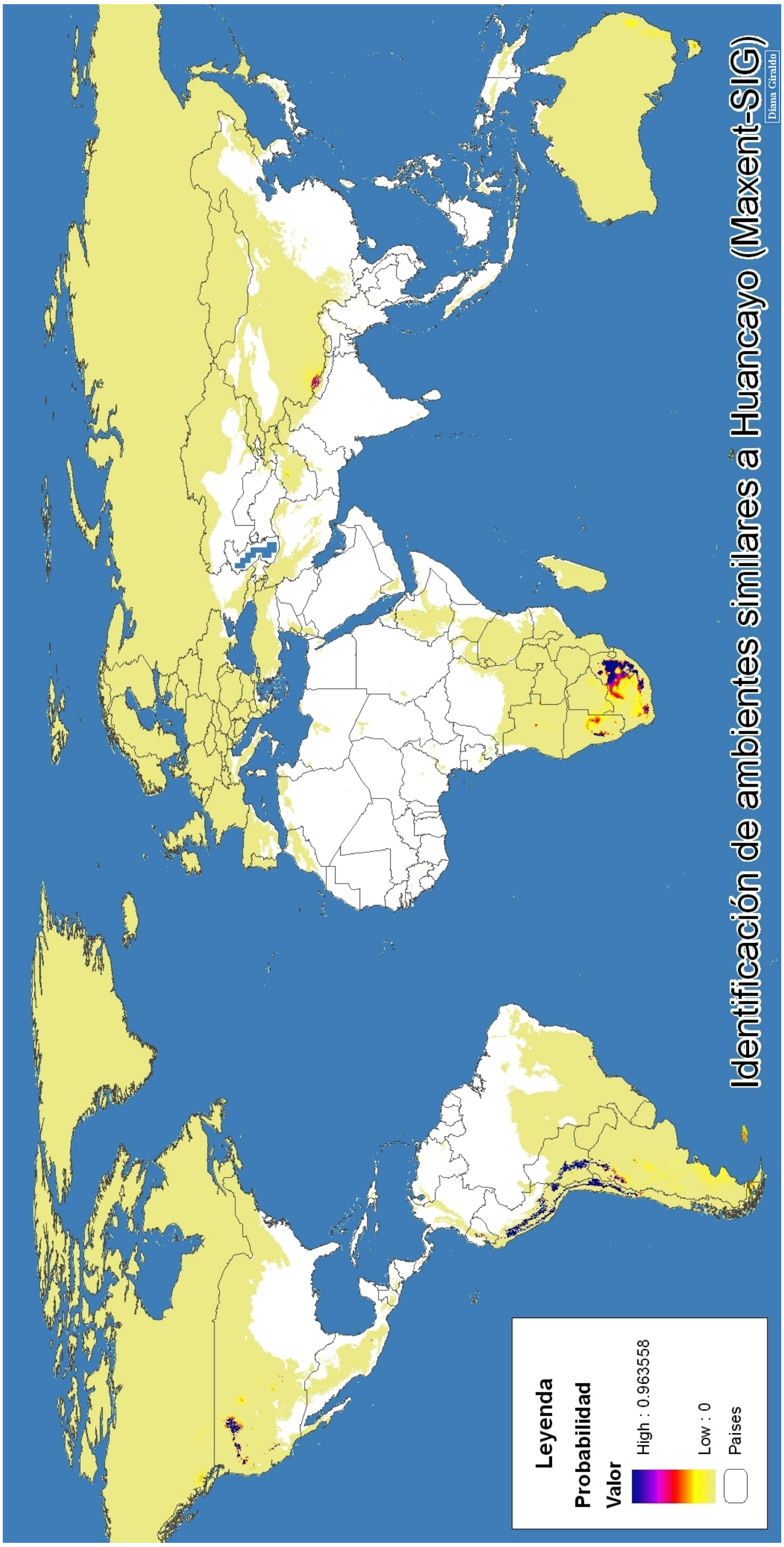

Abbildung 5.8: Weltkarte mit markierten Gebieten, wo ähliche Wetterbedingungen zu Huancayo (Peru) herrschen und Symmetrischema Tangolias vorkommen könnte.

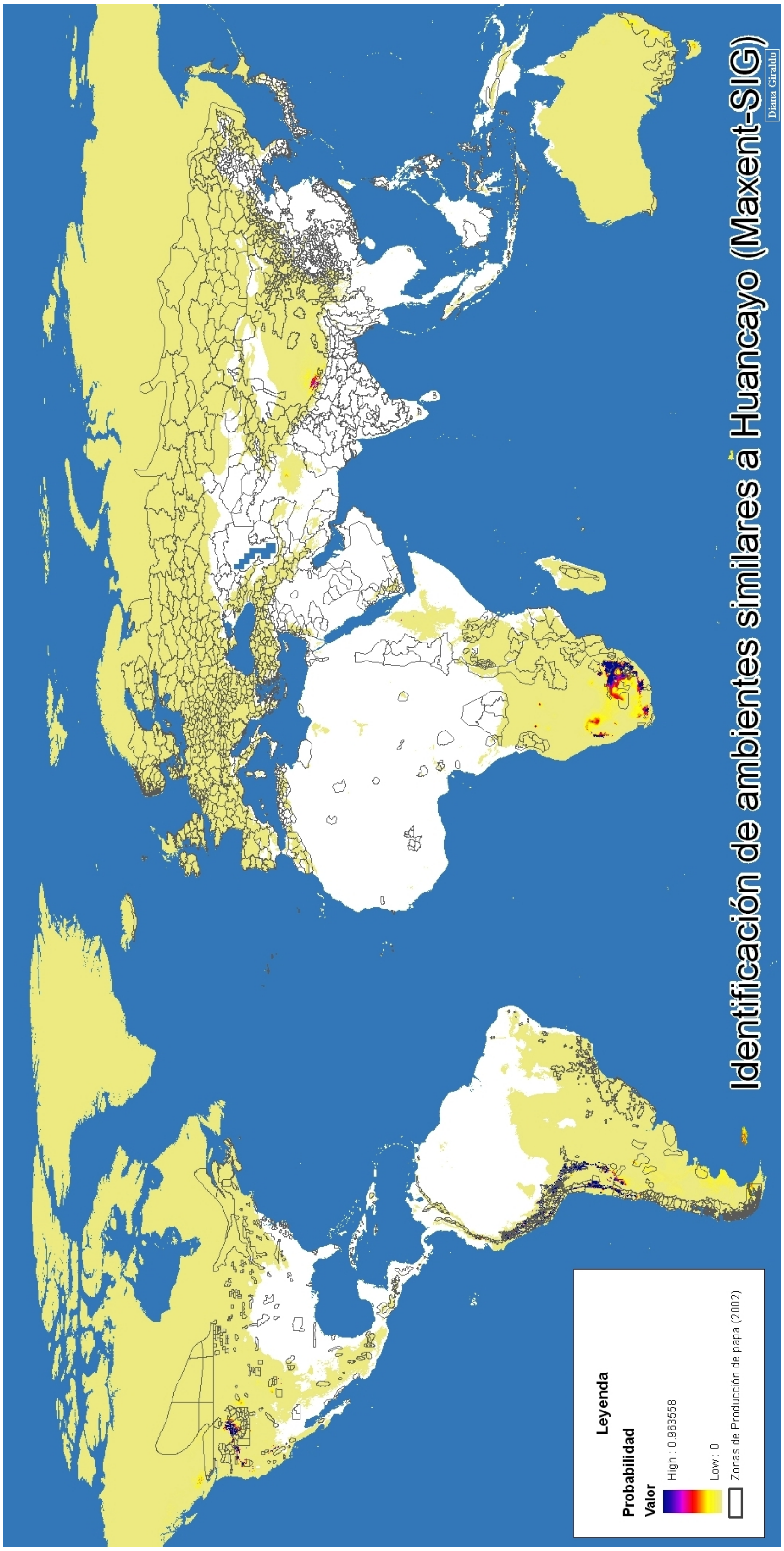

Abbildung 5.9: Weltkarte mit markierten Gebieten, wo Kartoffeln angebaut werden.

A Anhang

A.1 Anhang 1
Datensätze

A.1.1 Flugaktivität von Symmetrischema Tangolias an der CIP-Station 'La Victoria' - Rohdatensatz

Datensatz Wöchentliche Flugaktivität von Symmetrischema Tangolias an der CIP-Station 'La Victoria' in Huancayo, Peru, von November 1998 bis Oktober 2000 und der entsprechende Niederschlag in Milimeter.

Quelle [Keller 2003].

Analysis Kapitel 1

Tabelle A.1: Flugaktivität von Symmetrischema Tangolias an der CIP-Station 'La Victoria' in Huancayo von November 1998 bis Oktober 2000 und der entsprechende Niederschlag.

Date	Month	Rainfall	Ptm
11/03/1998	11	6.6	1036
11/10/1998	11	20.8	1003
11/17/1998	11	41.6	1212
11/24/1998	11	5.7	461
12/01/1998	12	19.3	455
12/08/1998	12	17.7	582
12/15/1998	12	24.6	700
12/22/1998	12	16.4	914
12/29/1998	12	24.5	976
01/05/1999	1	6.3	656
01/12/1999	1	20	298
01/19/1999	1	27.9	650
01/26/1999	1	24.2	858
02/02/1999	2	12.3	730
02/09/1999	2	33.9	315
02/16/1999	2	54.8	452

Fortsetzung auf der nächsten Seite

Flugaktivität von Symmetrischema Tangolias – Fortsetzung

Date	Month	Rainfall	Ptm
02/23/1999	2	47.1	1065
03/02/1999	3	39.1	425
03/09/1999	3	13.2	218
03/16/1999	3	14.3	240
03/23/1999	3	31.5	268
03/30/1999	3	5.2	256
04/06/1999	4	4.2	371
04/13/1999	4	11.5	331
04/20/1999	4	12.9	382
04/27/1999	4	0	363
05/04/1999	5	3.9	561
05/11/1999	5	2.3	347
05/18/1999	5	0	356
05/25/1999	5	0	1108
06/01/1999	6	0	1064
06/08/1999	6	10.9	838
06/15/1999	6	1	645
06/22/1999	6	8.5	690
06/29/1999	6	0	1932
07/06/1999	7	4.1	3465
07/13/1999	7	0	2874
07/20/1999	7	10.9	1252
07/27/1999	7	0	2150
08/03/1999	8	0	484
08/10/1999	8	0	635
08/17/1999	8	0	470
08/24/1999	8	0	381
08/31/1999	8	3.6	549
09/07/1999	9	6.6	438
09/14/1999	9	14	389

Fortsetzung auf der nächsten Seite

Flugaktivität von Symmetrischema Tangolias – Fortsetzung

Date	Month	Rainfall	Ptm
09/21/1999	9	11.3	590
09/28/1999	9	28.9	597
10/05/1999	10	50.2	216
10/12/1999	10	0	258
10/19/1999	10	1.4	252
10/26/1999	10	2.6	330
11/02/1999	11	13.2	226
11/09/1999	11	15.5	183
11/16/1999	11	24.6	232
11/23/1999	11	40.3	359
11/30/1999	11	24.2	371
12/07/1999	12	31	652
12/14/1999	12	24.3	1054
12/21/1999	12	11.5	547
12/28/1999	12	11.5	616
01/04/2000	1	17.9	591
01/11/2000	1	15.9	348
01/18/2000	1	13.3	349
01/25/2000	1	30.5	285
02/01/2000	2	40.5	193
02/08/2000	2	15.3	234
02/15/2000	2	29.3	150
02/22/2000	2	32.8	86
02/29/2000	2	52.6	105
03/07/2000	3	26.9	36
03/14/2000	3	32.4	25
03/21/2000	3	12.9	25
03/28/2000	3	10.6	38
04/04/2000	4	13	26
04/11/2000	4	5.2	22

Fortsetzung auf der nächsten Seite

Flugaktivität von Symmetrischema Tangolias – Fortsetzung

Date	Month	Rainfall	Ptm
04/18/2000	4	26.2	758
04/25/2000	4	1.1	862
05/02/2000	5	1.7	840
05/09/2000	5	4.5	768
05/16/2000	5	7.6	983
05/23/2000	5	1.2	1412
05/30/2000	5	0	699
06/06/2000	6	0	1315
06/13/2000	6	1.8	1940
06/20/2000	6	0	1620
06/27/2000	6	0	1974
07/04/2000	7	0.5	586
07/11/2000	7	0	1186
07/18/2000	7	0	1035
07/25/2000	7	6.8	1017
08/01/2000	8	0	362
08/08/2000	8	10.2	836
08/15/2000	8	1.4	986
08/22/2000	8	0	465
08/29/2000	8	9.8	496
09/05/2000	9	0	624
09/12/2000	9	0.3	480
09/19/2000	9	2.6	286
09/26/2000	9	3.4	280
10/03/2000	10	0	106
10/10/2000	10	2.1	476
10/17/2000	10	2	480

A.1.2 Flugaktivität von Symmetrischema Tangolias an der CIP-Station 'La Victoria' - die Wochenfänge des Rohdatensatzes A.1.1 werden zusammengefasst

Datensatz Monatliche Flugaktivität von Symmetrischema Tangolias an der CIP-Station 'La Victoria' in Huancayo, Peru, von November 1998 bis Oktober 2000 und der entsprechende Niederschlag in Milimeter.

Analysis Kapitel 1 und Abschnitt 5.3

Date	Month	Rainfall	PTM
1998	11	74.7	3712
1998	12	102.5	3627
1999	1	78.4	2462
1999	2	148.1	2562
1999	3	103.3	1407
1999	4	28.6	1447
1999	5	6.2	2372
1999	6	20.4	5169
1999	7	15	9741
1999	8	3.6	2519
1999	9	60.8	2014
1999	10	54.2	1056
1999	11	117.8	1371
1999	12	78.3	2869
2000	1	77.6	1573
2000	2	170.5	768
2000	3	82.8	124
2000	4	45.5	1668
2000	5	15	4702
2000	6	1.8	6849
2000	7	7.3	3824
2000	8	7.3	3145
2000	9	6.3	1670
2000	10	4.1	1062

A.1.3 Flugaktivität von Symmetrischema Tangolias an der CIP-Station 'La Victoria' - nach der Verschiebung

Datensatz Monatliche Flugaktivität von Symmetrischema Tangolias an der CIP-Station 'La Victoria' in Huancayo, Peru, von November 1998 bis Oktober 2000 und der entsprechende Niederschlag in Milimeter, nach der Verschiebung der Regenmenge gegen die Anzahl der Kartoffelmotte.

Analysis Kapitel 1 und 5.4

Tabelle A.3: Flugaktivität von Symmetrischema Tangolias nach der Verschiebung. Der Datensatz month.ptm.rain.

Year	Month	PTM	YearR	MonthR	Rainfall
1999	4	1447	1998	11	74.7
1999	5	2372	1998	12	102.5
1999	6	5169	1999	1	78.4
1999	7	9741	1999	2	148.1
1999	8	2519	1999	3	103.3
1999	9	2014	1999	4	28.6
1999	10	1056	1999	5	6.2
1999	11	1371	1999	6	20.4
1999	12	2869	1999	7	15.0
2000	1	1573	1999	8	3.6
2000	2	768	1999	9	60.8
2000	3	124	1999	10	54.2
2000	4	1668	1999	11	117.8
2000	5	4702	1999	12	78.3
2000	6	6849	2000	1	77.6
2000	7	3824	2000	2	170.5
2000	8	3145	2000	3	82.8
2000	9	1670	2000	4	45.5
2000	10	1062	2000	5	15.0

A.2 Anhang 2
R – Quellcodes

Im Folgenden werden die Quellcodes aufgeführt und erläutert, die in dieser Arbeit aufgerufen wurden. Es wurde die freie Software R Version 2.9.2 vom 24.08.2009 ([R Development Core Team 2009]) benutzt.

A.2.1 Plotten des Niederschlags gegen die Flugaktivität der Kartoffelmotte

```
PlotRain<-function(data, title, xla="Zeit [in Monaten]", yla="Anzahl der
  + Maennchen im Monat", j=4, k=3) # j - Spalte mit PTM,
   # k - Spalte mit Niederschlag
{
N <- length(data[,1])
set.seed(22)
matplot(data[,j], xaxt="n", type="n", main=title, xlab=xla, ylab=yla)
lines(data[,j], col="red", lty=1)
matpoints(data[,j], bg="red", pch=21)
atn <- seq(1,N,1) # monatlich
atn3 <- seq(1,N,3) # woechentlich
axis(1, at=atn, labels=c("N","D","J","F","M","A","M","J","J","A",
+ "S","O","N","D","J","F","M","A","M","J","J","A","S","O"))
axis(2,lty=1,lwd=0.5)
par(new=TRUE)
plot(data[,k], bg="limegreen", pch=21, axes=FALSE, main=title,
+ xlab=xla, ylab=yla)
grid(17, 17, lty=3)
lines(data[,k], col="green4", lty=1)
axis(4,col.axis = "green4", lty = 1, lwd = 0.5)
legend("topright", inset=.05, c("PTM","Niederschlag"), lty=c(1,1), box()
+ col=c("red","green4"), xjust=0.5, yjust=0.4)
}
```

A.2.2 Verschiebung des Niederschlags gegen die Flugaktivität der Kartoffelmotte

Einlesen der Daten.

```
> library(RODBC)
> library(MASS)
# Monatsfaenge
> channel <- odbcConnectExcel("LaVictoriaM.xls")
> RainM.ST<-sqlFetch(channel, "LV_ST_RainM")
```

Verschiebung des Datensatzes.

```
# Trennung von PTM und Rainfall
> month.ptm <- RainM.ST[,c(1,2,4)]
> month.rain <- RainM.ST[,c(1,2,3)]
# Verschiebung von PTM (um 5 Monate)
> month.ptm.versch <- month.ptm[6:24,]
> month.rain.versch <- month.rain[1:19,]
# Zusammensetzen von PTM und Rainfall
> month.ptm.rain <- cbind(month.ptm.versch,month.rain.versch)
# Umbenennung
> names(month.ptm.rain) <- c("Year", "Month", "PTM", "YearR",
+ "MonthR", "Rainfall")
```

Den verschobenen Datensatz auf Korrelation testen.

```
> cor(month.ptm.rain[,c(3,6)])
               PTM  Rainfall
PTM      1.0000000 0.5484252
Rainfall 0.5484252 1.0000000

# Korrelation mithilfe von cor.test
> cor.test(month.ptm.rain[,3],month.ptm.rain[,6],method = "spearman")
+ $estimate
```

```
rho
0.5853445
Warning message:
In cor.test.default(month.ptm.rain[, 3], month.ptm.rain[, 6], method
= "spearman") :
Cannot compute exact p-values with ties
```

A.2.3 Plotten circulärer Rohdaten

Es werden 100 generierte Beobachtungen von der Von-Mises-Verteilung mit der mittleren Richtung $\vartheta = 3$ und dem Konzentrationsparameter $\kappa = 0$ geplottet:

1 library(CircStats): Circular Data Plot - ermöglicht das Plotten von Kreisdaten

2 100 Beobachtungen werden generiert und unter `data.vm` gespeichert

3 `data.vm` wird geplottet

```
> library(CircStats) #1
> data.vm <- rvm(100, 0, 3) #2
> circ.plot(data.vm, stack=TRUE, bins=150) #3
```

A.2.4 Windrose-Diagramm zum Datensatz 1.1

Der Quellcode wurde von der Seite [web] entnommen.

```
clock.plot <- function (x, col = rainbow(n), ...)
{ if( min(x)<0 ) x <- x-min(x)
if( max(x)>1 ) x <- x/max(x)
n <- length(x)
if(is.null(names(x))) names(x) <- 0:(n-1)
m <- 1.05
plot(0, type = 'n', xlim = c(-m,m), ylim = c(-m,m),
axes = F, xlab = '', ylab = '', ...)
a <- pi/2 - 2*pi/200*0:200
polygon(cos(a),sin(a))
```

```
v <- .02
a <- pi/2 - 2*pi/n*0:n
segments((1+v)*cos(a), (1+v)*sin(a), (1-v)*cos(a), (1-v)*sin(a))
segments(cos(a), sin(a), 0, 0, col='light grey',lty = 3)
ca <- -2*pi/n*(0:50)/50
for (i in 1:n)
{ a <- pi/2 - 2*pi/n*(i-1)
b <- pi/2 - 2*pi/n*i
polygon( c(0, x[i]*cos(a+ca), 0), c(0, x[i]*sin(a+ca), 0),
col=col[i] )
v <- .1
text((1+v)*cos(a), (1+v)*sin(a), names(x)[i])}}
```

A.2.5 Veranschaulichung des Richtungsmedians

```
> library(CircStats)

# GERADE ZAHL VON BEOBACHTUNGEN
> data.gerade <- rvm(10, 1, 3)
> circ.plot(data.gerade, stack=TRUE, bins=150, dotsep = 23, main = "(a)",
+ shrink=1.1)
> data.gerade <- sort(data.gerade)
> median<-(data.gerade[5]+data.gerade[6])/2
> arrows(cos(median),sin(median),1.5*cos(median),1.5*sin(median),col=2)
> text(1.6*cos(median),1.6*sin(median),expression(paste(tilde(theta))),
+ cex = 1.3,col=2)

# UNGERADE ZAHL VON BEOBACHTUNGEN
> data.ungerade <- rvm(9, 1, 3)
> circ.plot(data.ungerade,stack=TRUE,bins=150,dotsep = 23,main ="(b)",
+ shrink=1.1)
> data.ungerade<-sort(data.ungerade)
> arrows(cos(data.ungerade[5]),sin(data.ungerade[5]),
+ 1.5*cos(data.ungerade[5]),1.5*sin(data.ungerade[5]),
```

```
+ col=2)
> text(1.6*cos(data.ungerade[5]),1.6*sin(data.ungerade[5]),
+ expression(paste(tilde(theta))),cex = 1.3, col=2)
```

A.2.6 Finden der Regressionskurve für den Datensatz RainM.ST

A.2.6.1 Lineares Modell

Der Datensatz `RainM.ST` wird aus einer Excel-Datei eingelesen. Die für das Programm notwendigen Pakete haben folgende Funktionen:

1 library (RODBC): ermöglicht das Laden der Datensätze aus einer Excel-Datei

2 library (stats): das AIC-Kriterium

3-4 der Datensatz `LV_ST_RainM` wird aus der Datei `LaVictoriaM.xls` eingelesen und unter `RainM.ST` gespeichert

```
> library(RODBC) #1
> library(stats) #2
> channel<-odbcConnectExcel("LaVictoriaM.xls") #3
> RainM.ST <- sqlFetch(channel, "LV_ST_RainM") #4
```

Die Modelle werden mit dem Shapiro-Wilk-Test auf die Normalverteilung getestet.

```
> shapiro.test(aov(PTM~cos(Month)+sin(Month),data=RainM.ST)$residuals)
+ $p.value
[1] 0.01537650
> shapiro.test(aov(PTM~Month+cos(Month)+sin(Month),data=RainM.ST)
+ $residuals)$p.value
[1] 0.0158398
> shapiro.test(aov(PTM~Month,data=RainM.ST)$residuals)$p.value
[1] 0.001152614
> shapiro.test(aov(PTM~cos(Month),data=RainM.ST)$residuals)$p.value
[1] 0.009226567
```

```
> shapiro.test(aov(PTM~Rainfall+cos(Month)+sin(Month),data=RainM.ST)
+ $residuals)$p.value
[1] 0.01120242
> shapiro.test(aov(PTM~Month+Rainfall+cos(Month)+sin(Month),data=RainM.ST)
+ $residuals)$p.value
[1] 0.01046258
> shapiro.test(aov(PTM~Rainfall,data=RainM.ST)$residuals)$p.value
[1] 0.01660329
> shapiro.test(aov(PTM~Month+Rainfall,data=RainM.ST)$residuals)$
+ p.value
[1] 0.01556918
> shapiro.test(aov(PTM~cos(Month)+Rainfall,data=RainM.ST)$residuals)$
+ p.value
[1] 0.005433883
```

ANOVA-Tests:

```
> anova(lm(PTM~cos(Month)+sin(Month)+Month,data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
cos(Month)  1 49718144 49718144 18.4963 0.0003483 ***
sin(Month)  1   242637   242637  0.0903 0.7669396
Month       1  1037753  1037753  0.3861 0.5413917
Residuals  20 53760144  2688007
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(lm(PTM~sin(Month)+cos(Month)+Month,data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
sin(Month)  1   107210   107210  0.0399 0.8437231
cos(Month)  1 49853571 49853571 18.5467 0.0003435 ***
Month       1  1037753  1037753  0.3861 0.5413917
```

```
Residuals  20 53760144  2688007
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

> anova(lm(PTM~Month+sin(Month)+cos(Month),data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value   Pr(>F)
Month       1  1940734  1940734  0.7220 0.405547
sin(Month)  1   965592   965592  0.3592 0.555668
cos(Month)  1 48092207 48092207 17.8914 0.000411 ***
Residuals  20 53760144  2688007
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

> anova(lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
Rainfall    1 16460380 16460380  6.8741 0.0167832 *
cos(Month)  1 41797266 41797266 17.4551 0.0005104 ***
sin(Month)  1   770855   770855  0.3219 0.5770992
Month       1   233653   233653  0.0976 0.7581594
Residuals  19 45496524  2394554
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

> anova(lm(PTM~Rainfall+sin(Month)+cos(Month)+Month,data=RainM.ST))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
Rainfall    1 16460380 16460380  6.8741 0.0167832 *
sin(Month)  1   753663   753663  0.3147 0.5813426
cos(Month)  1 41814458 41814458 17.4623 0.0005094 ***
Month       1   233653   233653  0.0976 0.7581594
```

```
Residuals  19 45496524  2394554
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Parameterschätzung der neun linearen Modelle:

```
> mod.lm.1<-lm(PTM~cos(Month)+sin(Month),data=RainM.ST)
> mod.lm.1
Call:
lm(formula = PTM ~ cos(Month) + sin(Month), data = RainM.ST)
Coefficients:
(Intercept)   cos(Month)   sin(Month)
     2922.2       2094.8        138.9

> mod.lm.2<-lm(PTM~cos(Month)+sin(Month)+Month,data=RainM.ST)
> mod.lm.2
Call:
lm(formula = PTM ~ cos(Month) + sin(Month) + Month, data = RainM.ST)
Coefficients:
(Intercept)   cos(Month)   sin(Month)        Month
    2491.29      2066.45       267.52        66.29

> mod.lm.3<-lm(PTM~Month,data=RainM.ST)
> mod.lm.3
Call:
lm(formula = PTM ~ Month, data = RainM.ST)
Cofficients:
(Intercept)        Month
    2285.93        82.38

> mod.lm.4<-lm(PTM~cos(Month),data=RainM.ST)
> mod.lm.4
Call:
lm(formula = PTM ~ cos(Month), data = RainM.ST)
Coefficients:
```

```
(Intercept)   cos(Month)
       2921         2091

> mod.lm.R.1<-lm(PTM~Rainfall+cos(Month)+sin(Month),data=RainM.ST)
> mod.lm.R.1
Call:
lm(formula = PTM ~ Rainfall + cos(Month) + sin(Month), data = RainM.ST)
Coefficients:
(Intercept)     Rainfall   cos(Month)   sin(Month)
    3627.44       -13.03      1945.17       249.75

> mod.lm.R.2<-lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,data=RainM.ST)
> mod.lm.R.2
Call:
lm(formula = PTM ~ Rainfall+cos(Month)+sin(Month)+Month,data=RainM.ST)
Coefficients:
(Intercept)     Rainfall   cos(Month)   sin(Month)        Month
    3398.80       -12.64      1935.91       308.51        31.98

> mod.lm.R.3<-lm(PTM~Rainfall,data=RainM.ST)
> mod.lm.R.3
Call:
lm(formula = PTM ~ Rainfall, data = RainM.ST)
Coefficients:
(Intercept)     Rainfall
    3758.22       -17.16

> mod.lm.R.4<-lm(PTM~Rainfall+Month,data=RainM.ST)
> mod.lm.R.4
Call:
lm(formula = PTM ~ Rainfall + Month, data = RainM.ST)
Coefficients:
(Intercept)     Rainfall        Month
    3542.35       -16.69        29.21
```

```
> mod.lm.R.5<-lm(PTM~Rainfall+cos(Month),data=RainM.ST)
> mod.lm.R.5
Call:
lm(formula = PTM ~ Rainfall + cos(Month), data = RainM.ST)
Coefficients:
(Intercept)     Rainfall   cos(Month)
    3597.98       -12.54      1944.77
```

Werte für das AIC:

```
> AIC(mod.lm.1,mod.lm.2,mod.lm.3,mod.lm.4)
         df      AIC
mod.lm.1  4 427.4957
mod.lm.2  5 429.0368
mod.lm.3  3 440.5990
mod.lm.4  3 425.6017

> AIC(mod.lm.R.1,mod.lm.R.2,mod.lm.R.3,mod.lm.R.4,mod.lm.R.5)
           df      AIC
mod.lm.R.1  5 425.1542
mod.lm.R.2  6 427.0313
mod.lm.R.3  3 436.9453
mod.lm.R.4  4 438.8823
mod.lm.R.5  4 423.5554
```

A.2.6.2 Nichtlineares Modell

Um nichtlineare Modelle bilden zu können, wird das Paket `library (nls2)` notwendig:

```
> library(nls2)
```

Es werden 4 Regressionsfunktionen definiert:

1 Funktionen `fct1.nlm` und `fct2.nlm` für die Modelle ohne den Niederschlag

2 Funktionen `fct1.nlm.R` und `fct1.nlm.R` werden für die Modelle mit aufgenommener Regenmenge benutzt.

```
> fct1.nlm<-function(beta, A, B, omega, t){
+ beta+A*cos(omega*t)+B*sin(omega*t)}

> fct2.nlm<-function(beta0, A, B, omega, beta1, t){
+ beta0+A*cos(omega*t)+B*sin(omega*t)+ beta1*t}

> fct3.nlm<-function(beta,A,omega,t){
+  beta+A*cos(omega*t)}

> fct1.nlm.R<-function(beta, gamma, A, B, omega, t, r){
+ beta+gamma*r+A*cos(omega*t)+B*sin(omega*t)}

> fct2.nlm.R<-function(beta0, gamma, A, B, omega, beta1, t, r){
+ beta0+gamma*r+A*cos(omega*t)+B*sin(omega*t) + beta1*t}

> fct3.nlm.R<-function(beta, gamma, A, omega, t, r){
+  beta+gamma*r+A*cos(omega*t)}
```

Danach werden 4 Gitter definiert, die als Startwerte der Funktion `nls2()` dienen werden [Ritz u. Streibig 2008]. Die Startwerte werden von den entsprechenden linearen Modellen übernommen.

1 Gitter für das Modell `mod.nlm.1`

2 Gitter für das Modell `mod.nlm.2`

2 Gitter für das Modell `mod.nlm.3`

4 Gitter für das Modell `mod.nlm.R.1`

5 Gitter für das Modell `mod.nlm.R.2`

6 Gitter für das Modell `mod.nlm.R.3`

```
> grid1.nlm<-expand.grid(list(beta=seq(2921,2923,by=0.1),A=seq(2093,2095,
+ by=0.1),B=seq(138,140,by=0.1),omega=seq(0.9,1.1,by=0.1))) #1
> grid2.nlm<-expand.grid(list(beta0=seq(2491,2492,by=0.1),A=seq(2066,2067,
+ by=0.1),B=seq(267,268,by=0.1),omega=seq(0.9,1.0,by=0.1),
+ beta1=seq(66,68,by=0.1))) #2
> grid3.nlm<-expand.grid(list(beta=seq(2921,2923,by=0.1),A=seq(2093,2095,
+ by=0.1),omega=seq(0.9,1.1,by=0.1))) #3
> grid1.nlm.R<-expand.grid(list(beta=seq(3627,3627.5,by=0.1),
+ gamma=seq(-13.5,-13,by=0.1),A=seq(1945,1946,by=0.1),
+ B=seq(249,259,by=0.1),omega=seq(0.9,1.0,by=0.1))) #4
> grid2.nlm.R<-expand.grid(list(beta0=seq(3398,3399,by=0.2),
+ gamma=seq(-13,-12,by=0.2),A=seq(1935.5,1936,by=0.2),B=seq(308,309,
+ by=0.2),omega=seq(0.9,1.0,by=0.1),beta1=seq(31.5,32,by=0.2))) #5
> grid3.nlm.R<-expand.grid(list(beta=seq(3787,3788,by=0.2),gamma=seq(-18,
+ -17,by=0.2),A=seq(1935.5,1936,by=0.2),omega=seq(0.9,1.0,by=0.1))) #6
```

Nun wird, zu jeder definierten Regressionsfunktion, mithilfe der Funktion `nls2()`, jeweils 1 Modell erzeugt, das als Startwerte die vorher definierten Gitter haben. Es wird der `Gauss-Newton` Algorithmus benutzt.

1 Modell `mod.nlm.1` zur Funktion `fct1.nlm`

2 Modell `mod.nlm.2` zur Funktion `fct2.nlm`

3 Modell `mod.nlm.3` zur Funktion `fct3.nlm`

4 Modell `mod.nlm.R.1` zur Funktion `fct1.nlm.R`

5 Modell `mod.nlm.R.2` zur Funktion `fct2.nlm.R`

6 Modell `mod.nlm.R.3` zur Funktion `fct3.nlm.R`

```
> mod.nlm.1<-nls2(PTM~fct1.nlm(beta,A,B,omega,Month),data=RainM.ST,
+ start=grid1.nlm) #1
> mod.nlm.2<-nls2(PTM~fct2.nlm(beta0,A,B,omega,beta1,Month),data=RainM.ST,
+ start=grid2.nlm) #2
> mod.nlm.3<-nls2(PTM~fct3.nlm(beta,A,omega,Month),data=RainM.ST,
+ start=grid3.nlm) #3
```

```
> mod.nlm.R.1<-nls2(PTM~fct1.nlm.R(beta,gamma,A,B,omega,Month,Rainfall),
+ data=RainM.ST,start=grid1.nlm.R) #4
> mod.nlm.R.2<-nls2(PTM~fct2.nlm.R(beta0,gamma,A,B,omega,beta1,Month,
+ Rainfall),data=RainM.ST,start=grid2.nlm.R) #5
> mod.nlm.R.3<-nls2(PTM~fct3.nlm.R(beta,gamma,A,omega,Month,Rainfall),
+ data=RainM.ST,start=grid3.nlm.R) #6
```

Parameterschätzung und `summary()`:

```
> mod.nlm.1
Nonlinear regression model
  model:  PTM ~ fct1.nlm(beta,A,B,omega,Month)
      beta          A          B      omega
 3275.8446  1725.7991 -1841.7000     0.8585
 residual sum-of-squares: 39575394
Number of iterations to convergence: 7
Achieved convergence tolerance: 2.626e-06

> mod.nlm.2
Nonlinear regression model
  model:  PTM ~ fct2.nlm(beta0,A,B,omega,beta1,Month)
     beta0          A          B      omega      beta1
 2819.4343  1757.7457 -1797.5942     0.8562    70.8235
 residual sum-of-squares: 38254036
Number of iterations to convergence: 4
Achieved convergence tolerance: 2.208e-06

> mod.nlm.3
Nonlinear regression model
  model:  PTM ~ fct3.nlm(beta, A, omega, Month)
     beta         A     omega
3020.3695 2242.3720    0.9613
 residual sum-of-squares: 50616024
Number of iterations to convergence: 3
Achieved convergence tolerance: 7.327e-07
```

```
> mod.nlm.R.1
Nonlinear regression model
  model:  PTM ~ fct1.nlm.R(beta,gamma,A,B,omega,Month,Rainfall)
      beta      gamma          A          B      omega
 3094.8408     4.5592  1604.7952 -2176.8922     0.8388
 residual sum-of-squares: 39196706
Number of iterations to convergence: 9
Achieved convergence tolerance: 6.007e-06

> mod.nlm.R.2
Nonlinear regression model
  model:  PTM ~ fct2.nlm.R(beta0,gamma,A,B,omega,beta1,Month,Rainfall)
     beta0      gamma          A          B      omega      beta1
 2116.0388    10.9112  1457.8060 -2570.0132     0.8124   111.3664
 residual sum-of-squares: 36525553
Number of iterations to convergence: 7
Achieved convergence tolerance: 3.947e-06

> mod.nlm.R.3
Nonlinear regression model
     beta     gamma         A     omega
3573.7407  -10.9230 2045.8453    0.9734
 residual sum-of-squares: 44739083
Number of iterations to convergence: 4
Achieved convergence tolerance: 2.026e-07

> summary(mod.nlm.1)
Formula: PTM ~ fct1.nlm(beta,A,B,omega,Month)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta   3.276e+03  3.068e+02  10.676 1.04e-09 ***
A      1.726e+03  6.738e+02   2.561   0.0186 *
B     -1.842e+03  6.265e+02  -2.939   0.0081 **
omega  8.585e-01  4.273e-02  20.092 9.90e-15 ***
```

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

Residual standard error: 1407 on 20 degrees of freedom
Number of iterations to convergence: 7
Achieved convergence tolerance: 2.626e-06

> summary(mod.nlm.2)
Formula: PTM ~ fct2.nlm(beta0, A, B, omega, beta1, Month)
Parameters:
Estimate Std. Error t value Pr(>|t|)
beta0  2.819e+03  6.438e+02   4.380 0.000322 ***
A      1.758e+03  6.734e+02   2.610 0.017211 *
B     -1.798e+03  6.417e+02  -2.801 0.011396 *
omega  8.562e-01  4.328e-02  19.782 3.89e-14 ***
beta1  7.082e+01  8.733e+01   0.811 0.427404
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1419 on 19 degrees of freedom
Number of iterations to convergence: 4
Achieved convergence tolerance: 2.208e-06

> summary(mod.nlm.3)
Formula: PTM ~ fct3.nlm(beta, A, omega, Month)
Parameters:
       Estimate Std. Error t value Pr(>|t|)
beta  3.020e+03  3.256e+02   9.278 7.08e-09 ***
A     2.242e+03  4.742e+02   4.729 0.000114 ***
omega 9.613e-01  2.608e-02  36.863  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1553 on 21 degrees of freedom
Number of iterations to convergence: 3
```

```
Achieved convergence tolerance: 7.327e-07

> summary(mod.nlm.R.1)
Formula: PTM ~ fct1.nlm.R(beta,gamma,A,B,omega,Month,Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta   3.095e+03  5.169e+02   5.987 9.23e-06 ***
gamma  4.559e+00  1.011e+01   0.451   0.6571
A      1.605e+03  7.828e+02   2.050   0.0544 .
B     -2.177e+03  9.098e+02  -2.393   0.0272 *
omega  8.388e-01  5.504e-02  15.241 4.16e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1436 on 19 degrees of freedom
Number of iterations to convergence: 9
Achieved convergence tolerance: 6.005e-06

> summary(mod.nlm.R.2)
Formula: PTM ~ fct2.nlm.R(beta0,gamma,A,B,omega,beta1,Month,Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta0  2116.0388   968.8305   2.184   0.0424 *
gamma    10.9112    11.2600   0.969   0.3454
A      1457.8060   842.1212   1.731   0.1005
B     -2570.0132   888.5385  -2.892   0.0097 **
omega     0.8124     0.0518  15.684  6.1e-12 ***
beta1   111.3664    94.7568   1.175   0.2552
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1424 on 18 degrees of freedom
Number of iterations to convergence: 7
Achieved convergence tolerance: 3.947e-06
```

```
> summary(mod.nlm.R.3)
Formula: PTM ~ fct3.nlm.R(beta,gamma,A,omega,Month,Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta  3573.74070  463.77172   7.706 2.07e-07 ***
gamma  -10.92295    6.72332  -1.625 0.119896
A     2045.84535  468.55056   4.366 0.000299 ***
omega    0.97336    0.02903  33.528  < 2e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual standard error: 1496 on 20 degrees of freedom
Number of iterations to convergence: 4
Achieved convergence tolerance: 2.026e-07
```

Werte für das AIC:

```
> AIC(mod.nlm.1,mod.nlm.2,mod.nlm.R.1,mod.nlm.R.2)
             df      AIC
mod.nlm.1     5 421.6850
mod.nlm.2     6 422.8700
mod.nlm.3     4 425.5904
mod.nlm.R.1   6 423.4542
mod.nlm.R.2   7 423.7603
mod.nlm.R.3   5 424.6283
```

A.2.7 Finden der Regressionskurve für den Datensatz month.ptm.rain

```
> shapiro.test(aov(PTM~Rainfall+cos(Month)+sin(Month),
+ data=month.ptm.rain)$residuals)$p.value
[1] 0.3783714
> shapiro.test(aov(PTM~Month+Rainfall+cos(Month)+sin(Month),
+ data=month.ptm.rain)$residuals)$p.value
```

```
[1] 0.6351482
> shapiro.test(aov(PTM~Rainfall,data=month.ptm.rain)$residuals)$p.value
[1] 0.1266351
> shapiro.test(aov(PTM~Rainfall+Month,data=month.ptm.rain)$residuals)$
+ p.value
[1] 0.02677987
> shapiro.test(aov(PTM~Rainfall+cos(Month),data=month.ptm.rain)$residuals)
+ $p.value
[1] 0.3435277

> bartlett.test(PTM~Month+Rainfall,data=month.ptm.rain)
Error in bartlett.test.default(c(1447, 2372, 5169, 9741, 2519, 2014, 1056,:
  there must be at least 2 observations in each group

> anova(lm(PTM~cos(Month)+Rainfall+sin(Month)+Month,data=month.ptm.rain))
Analysis of Variance Table

Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
cos(Month)  1 51645771 51645771 19.9503 0.0005322 ***
Rainfall    1 10203962 10203962  3.9417 0.0670501 .
sin(Month)  1    42964    42964  0.0166 0.8993259
Month       1  2978866  2978866  1.1507 0.3015434
Residuals  14 36242015  2588715
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,data=month.ptm.rain))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value   Pr(>F)
Rainfall    1 30411947 30411947 11.7479 0.004083 **
cos(Month)  1 31437786 31437786 12.1442 0.003643 **
sin(Month)  1    42964    42964  0.0166 0.899326
Month       1  2978866  2978866  1.1507 0.301543
```

```
Residuals  14 36242015  2588715
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

> summary(lm(PTM~cos(Month)+Rainfall,data=month.ptm.rain))
Call:
lm(formula = PTM ~ cos(Month) + Rainfall, data = month.ptm.rain)
Residuals:
    Min      1Q  Median      3Q     Max
-2364.4  -855.5  -339.2   841.6  3932.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1811.351    676.691   2.677  0.01654 *
cos(Month)  1974.527    551.662   3.579  0.00251 **
Rainfall      16.941      8.308   2.039  0.05831 .
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual standard error: 1567 on 16 degrees of freedom
Multiple R-squared: 0.6117,Adjusted R-squared: 0.5631
F-statistic:  12.6 on 2 and 16 DF,  p-value: 0.000517

> summary(lm(PTM~cos(Month),data=month.ptm.rain))
Call:
lm(formula = PTM ~ cos(Month), data = month.ptm.rain)
Residuals:
     Min       1Q   Median       3Q      Max
-2686.17 -1083.05    16.98   680.50  4975.68

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2978.9      392.7   7.585 7.48e-07 ***
cos(Month)    2369.6      562.5   4.213 0.000585 ***
---
```

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

Residual standard error: 1706 on 17 degrees of freedom
Multiple R-squared: 0.5108,Adjusted R-squared: 0.482
F-statistic: 17.75 on 1 and 17 DF,  p-value: 0.0005851

> summary(lm(PTM~Rainfall,data=month.ptm.rain))
Call:
lm(formula = PTM ~ Rainfall, data = month.ptm.rain)
Residuals:
    Min      1Q  Median      3Q     Max
-2547.4 -1506.2  -177.2   976.8  4695.9

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   989.55     828.68   1.194   0.2488
Rainfall       27.38      10.13   2.704   0.0150 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2039 on 17 degrees of freedom
Multiple R-squared: 0.3008,Adjusted R-squared: 0.2596
F-statistic: 7.312 on 1 and 17 DF,  p-value: 0.01504

# M O D E L L E
> mod.lm.R.V.1<-lm(PTM~Rainfall+cos(Month)+sin(Month),
+ data=month.ptm.rain)
> mod.lm.R.V.2<-lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,
+ data=month.ptm.rain)
> mod.lm.R.V.3<-lm(PTM~Rainfall,data=month.ptm.rain)
> mod.lm.R.V.4<-lm(PTM~Rainfall+Month,data=month.ptm.rain)
> mod.lm.R.V.5<-lm(PTM~cos(Month)+Rainfall,data=month.ptm.rain)

# S U M M A R Y
> mod.lm.R.V.1
```

```
Call:
lm(formula = PTM~Rainfall+cos(Month)+sin(Month),data=month.ptm.rain)
Coefficients:
(Intercept)     Rainfall   cos(Month)   sin(Month)
    1829.67        16.71      1978.42        68.03

> mod.lm.R.V.2
Call:
lm(formula = PTM~Rainfall+cos(Month)+sin(Month)+Month,data=month.ptm.rain)
Coefficients:
(Intercept)     Rainfall   cos(Month)   sin(Month)        Month
     759.87        18.92      1946.10       149.16       137.75

> mod.lm.R.V.3
Call:
lm(formula = PTM ~ Rainfall, data = month.ptm.rain)
Coefficients:
(Intercept)     Rainfall
     989.55        27.38

> mod.lm.R.V.4
Call:
lm(formula = PTM~Rainfall+Month,data=month.ptm.rain)
Coefficients:
(Intercept)     Rainfall        Month
    -250.76        30.04       158.75

# A I C
> AIC(mod.lm.R.V.1,mod.lm.R.V.2,mod.lm.R.V.3,mod.lm.R.V.4,mod.lm.R.V.5)
             df      AIC
mod.lm.R.V.1  5 340.1850
mod.lm.R.V.2  6 340.6842
mod.lm.R.V.3  3 347.3809
mod.lm.R.V.4  4 348.2596
mod.lm.R.V.5  4 338.2058
```

A.2.8 Finden der Regressionskurve für den Datensatz RainM.ST_19

```
> RainM.ST_19<-RainM.ST[6:24,]

> shapiro.test(aov(PTM~Rainfall+cos(Month)+sin(Month),data=RainM.ST_19)
+ $residuals)$p.value
[1] 0.02823879
> shapiro.test(aov(PTM~Month+Rainfall+cos(Month)+sin(Month),
+ data=RainM.ST_19)$residuals)$p.value
[1] 0.02687639
> shapiro.test(aov(PTM~Rainfall,data=RainM.ST_19)$residuals)$p.value
[1] 0.01940782
> shapiro.test(aov(PTM~Month+Rainfall,data=RainM.ST_19)$residuals)$p.value
[1] 0.01725636
> shapiro.test(aov(PTM~cos(Month)+Rainfall,data=RainM.ST_19)$residuals)
+ $p.value
[1] 0.007708397

> bartlett.test(PTM~Month+Rainfall,data=RainM.ST_19)
Error in bartlett.test.default(c(1447, 2372, 5169, 9741, 2519, 2014, 1056,:
  there must be at least 2 observations in each group

> anova(lm(PTM~cos(Month)+Rainfall+sin(Month)+Month,data=RainM.ST_19))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
cos(Month)  1 51645771 51645771 19.4862 0.0005884 ***
Rainfall    1 11073841 11073841  4.1782 0.0602318 .
sin(Month)  1  1192173  1192173  0.4498 0.5133370
Month       1    96523    96523  0.0364 0.8513935
Residuals  14 37105271  2650376
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> anova(lm(PTM~cos(Month)+sin(Month)+Rainfall+Month,data=RainM.ST_19))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value    Pr(>F)
cos(Month)  1 51645771 51645771 19.4862 0.0005884 ***
sin(Month)  1   760946   760946  0.2871 0.6004892
Rainfall    1 11505067 11505067  4.3409 0.0560167 .
Month       1    96523    96523  0.0364 0.8513935
Residuals  14 37105271  2650376
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

> anova(lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,data=RainM.ST_19))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value   Pr(>F)
Rainfall    1 21432778 21432778  8.0867 0.013010 *
cos(Month)  1 41286833 41286833 15.5777 0.001461 **
sin(Month)  1  1192173  1192173  0.4498 0.513337
Month       1    96523    96523  0.0364 0.851394
Residuals  14 37105271  2650376
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

> anova(lm(PTM~Rainfall+sin(Month)+cos(Month)+Month,data=RainM.ST_19))
Analysis of Variance Table
Response: PTM
           Df   Sum Sq  Mean Sq F value   Pr(>F)
Rainfall    1 21432778 21432778  8.0867 0.013010 *
sin(Month)  1  1731573  1731573  0.6533 0.432445
cos(Month)  1 40747433 40747433 15.3742 0.001537 **
Month       1    96523    96523  0.0364 0.851394
Residuals  14 37105271  2650376
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

```
> summary(lm(PTM~Rainfall+cos(Month),data=RainM.ST_19))
Call:
lm(formula = PTM ~ Rainfall + cos(Month), data = RainM.ST_19)
Residuals:
    Min      1Q  Median      3Q     Max
-1949.9  -776.2  -110.5   410.5  4677.6

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3695.385    488.312   7.568 1.13e-06 ***
Rainfall     -17.247      8.028  -2.148 0.047355 *
cos(Month)  2157.770    520.201   4.148 0.000757 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1549 on 16 degrees of freedom
Multiple R-squared: 0.6203,Adjusted R-squared: 0.5728
F-statistic: 13.07 on 2 and 16 DF,  p-value: 0.0004321

> summary(lm(PTM~cos(Month),data=RainM.ST_19))
Call:
lm(formula = PTM ~ cos(Month), data = RainM.ST_19)
Residuals:
     Min       1Q   Median       3Q      Max
-2686.17 -1083.05    16.98   680.50  4975.68

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2978.9      392.7   7.585 7.48e-07 ***
cos(Month)    2369.6      562.5   4.213 0.000585 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1706 on 17 degrees of freedom
```

```
Multiple R-squared: 0.5108,Adjusted R-squared: 0.482
F-statistic: 17.75 on 1 and 17 DF,  p-value: 0.0005851

> summary(lm(PTM~Rainfall,data=RainM.ST_19))
Call:
lm(formula = PTM ~ Rainfall, data = RainM.ST_19)
Residuals:
    Min      1Q  Median      3Q     Max
-2676.3 -1409.4  -433.7   914.3  6259.5

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3834.90     680.84   5.633 2.98e-05 ***
Rainfall      -23.56      11.02  -2.138   0.0473 *
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual standard error: 2165 on 17 degrees of freedom
Multiple R-squared: 0.212,Adjusted R-squared: 0.1656
F-statistic: 4.573 on 1 and 17 DF,  p-value: 0.0473

> mod.lm_19.R.1<-lm(PTM~Rainfall+cos(Month)+sin(Month),data=RainM.ST_19)
> mod.lm_19.R.2<-lm(PTM~Rainfall+cos(Month)+sin(Month)+Month,
+ data=RainM.ST_19)
> mod.lm_19.R.3<-lm(PTM~Rainfall,data=RainM.ST_19)
> mod.lm_19.R.4<-lm(PTM~Rainfall+Month,data=RainM.ST_19)
> mod.lm_19.R.5<-lm(PTM~Rainfall+cos(Month),data=RainM.ST_19)

> mod.lm_19.R.1
Call:
lm(formula = PTM~Rainfall+cos(Month)+sin(Month),data=RainM.ST_19)
Coefficients:
(Intercept)     Rainfall   cos(Month)   sin(Month)
    3721.47       -17.62      2144.94       351.12
```

```
> mod.lm_19.R.2
Call:
lm(formula = PTM~Rainfall+cos(Month)+sin(Month)+Month,
data=RainM.ST_19)
Coefficients:
(Intercept)     Rainfall   cos(Month)   sin(Month)        Month
    3544.36       -17.29      2152.28       369.33        24.61

> mod.lm_19.R.3
Call:
lm(formula = PTM ~ Rainfall, data = RainM.ST_19)
Coefficients:
(Intercept)     Rainfall
    3834.90       -23.56

> mod.lm_19.R.4
Call:
lm(formula = PTM ~ Rainfall + Month, data = RainM.ST_19)
Coefficients:
(Intercept)     Rainfall        Month
    4045.00       -23.93       -29.05

> mod.lm_19.R.5
Call:
lm(formula = PTM ~ Rainfall + cos(Month), data = RainM.ST_19)
Coefficients:
(Intercept)     Rainfall   cos(Month)
    3695.38       -17.25      2157.77

> AIC(mod.lm_19.R.1,mod.lm_19.R.2,mod.lm_19.R.3,mod.lm_19.R.4,
+ mod.lm_19.R.5)
              df      AIC
mod.lm_19.R.1  5 339.1808
mod.lm_19.R.2  6 341.1314
mod.lm_19.R.3  3 349.6526
```

```
mod.lm_19.R.4  4 351.6192
mod.lm_19.R.5  4 337.7801
```

A.2.9 Finden der Regressionskurve für den Datensatz month.ptm.rain - nichtlineare Modelle

```
# F U N K T I O N E N
> fct1.nlm.V.R<-function(beta, gamma, A, B, omega, t, r){
+ beta+gamma*r+A*cos(omega*t)+B*sin(omega*t)}

> fct2.nlm.V.R<-function(beta0, gamma, A, B, omega, beta1, t, r){
+ beta0+gamma*r+A*cos(omega*t)+B*sin(omega*t) + beta1*t}

> fct3.nlm.V.R<-function(beta, gamma, A, omega, t, r){
+  beta+gamma*r+A*cos(omega*t)}

# G I T T E R
# Die Startwerte werden von den aequivalenten linearen Modellen
# uebernommen.
> grid1.nlm.V.R<-expand.grid(list(beta=seq(1829.5,1830,by=0.2),
+ gamma=seq(16.5,17,by=0.2),A=seq(1978,1979,by=0.2),B=seq(68,68.5,
+ by=0.2),omega=seq(0.9,1.0,by=0.1)))
> grid2.nlm.V.R<-expand.grid(list(beta0=seq(759.5,760,by=0.2),
+ gamma=seq(18.5,19,by=0.2),A=seq(1946,1946.5,by=0.2),B=seq(149,149.5,
+ by=0.2),omega=seq(0.9,1.0,by=0.1),beta1=seq(137.5,138,by=0.2)))
> grid3.nlm.V.R<-expand.grid(list(beta=seq(1829.5,1830,by=0.2),gamma=
+ seq(16.5,17,by=0.2),A=seq(1978,1979,by=0.2),omega=seq(0.9,1.0,by=0.1)))

#  M O D E L L E
> mod.nlm.V.R.1<-nls2(PTM~fct1.nlm.V.R(beta,gamma,A,B,omega,Month,
+ Rainfall),data=month.ptm.rain,start=grid1.nlm.V.R)
> mod.nlm.V.R.2<-nls2(PTM~fct2.nlm.V.R(beta0,gamma,A,B,omega,beta1,
+ Month,Rainfall),data=month.ptm.rain,start=grid2.nlm.V.R)
> mod.nlm.V.R.3<-nls2(PTM~fct3.nlm.V.R(beta,gamma,A,omega,Month,Rainfall),
```

```
+ data=month.ptm.rain,start=grid3.nlm.V.R)

# S U M M A R Y
> mod.nlm.V.R.1
Nonlinear regression model
  model: PTM ~ fct1.nlm.V.R(beta,gamma,A,B,omega,Month,Rainfall)
     beta     gamma         A         B     omega
 3311.633    -2.709  1653.951 -2455.804     0.824
 residual sum-of-squares: 30163830
Number of iterations to convergence: 7
Achieved convergence tolerance: 4.455e-06

> mod.nlm.V.R.2
Nonlinear regression model
  model: PTM ~ fct2.nlm.V.R(beta0,gamma,A,B,omega,beta1,Month,Rainfall)
     beta0      gamma          A           B      omega      beta1
 2451.5557     0.5323  1714.2260 -2276.6486     0.8328    94.6593
 residual sum-of-squares: 28793374
Number of iterations to convergence: 6
Achieved convergence tolerance: 1.560e-06

> mod.nlm.V.R.3
Nonlinear regression model
     beta     gamma         A     omega
2102.1914   13.2941 2121.8253    0.9684
 residual sum-of-squares: 37977908
Number of iterations to convergence: 7
Achieved convergence tolerance: 2.358e-06

> summary(mod.nlm.V.R.1)
Formula: PTM ~ fct1.nlm.V.R(beta,gamma,A,B,omega,Month,Rainfall)
Parameters:
Estimate Std. Error t value Pr(>|t|)
beta   3.312e+03  9.019e+02   3.672  0.00251 **
gamma -2.709e+00  1.175e+01  -0.231  0.82094
```

```
A      1.654e+03  9.736e+02   1.699  0.11145
B     -2.456e+03  1.040e+03  -2.361  0.03325 *
omega  8.240e-01  5.812e-02  14.177 1.07e-09 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual standard error: 1468 on 14 degrees of freedom
Number of iterations to convergence: 7
Achieved convergence tolerance: 4.431e-06

> summary(mod.nlm.V.R.2)
Formula: PTM ~ fct2.nlm.V.R(beta0,gamma,A,B,omega,beta1,Month,Rainfall)
Parameters:
Estimate Std. Error t value Pr(>|t|)
beta0  2451.5557  1418.7615   1.728   0.1077
gamma     0.5323    12.5486   0.042   0.9668
A      1714.2260   951.9936   1.801   0.0950 .
B     -2276.6486  1107.3224  -2.056   0.0604 .
omega     0.8328     0.0624  13.346 5.78e-09 ***
beta1    94.6593   120.0893   0.788   0.4447
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual standard error: 1488 on 13 degrees of freedom
Number of iterations to convergence: 6
Achieved convergence tolerance: 1.560e-06

> summary(mod.nlm.V.R.3)
Formula: PTM ~ fct3.nlm.V.R(beta, gamma, A, omega, Month, Rainfall)
Parameters:
       Estimate Std. Error t value Pr(>|t|)
beta  2.102e+03  7.692e+02   2.733   0.0154 *
gamma 1.329e+01  9.491e+00   1.401   0.1816
A     2.122e+03  6.029e+02   3.519   0.0031 **
omega 9.684e-01  3.568e-02  27.139 3.64e-14 ***
```

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

Residual standard error: 1591 on 15 degrees of freedom
Number of iterations to convergence: 7
Achieved convergence tolerance: 2.358e-06

# A I C
> AIC(mod.nlm.V.R.1,mod.nlm.V.R.2,mod.nlm.V.R.3)
              df      AIC
mod.nlm.V.R.1  6 337.1963
mod.nlm.V.R.2  7 338.3128
mod.nlm.V.R.3  5 339.5731
```

A.2.10 Finden der Regressionskurve für den Datensatz RainM.ST_19, nichtlineare Regression

```
# F U N K T I O N E N
> fct1.nlm_19.R<-function(beta,gamma,A,B,omega,t,r){
beta+gamma*r+A*cos(omega*t)+B*sin(omega*t)}

> fct2.nlm_19.R<-function(beta0,gamma,A,B,omega,beta1,t,r){
beta0+gamma*r+A*cos(omega*t)+B*sin(omega*t)+beta1*t}

> fct3.nlm_19.R<-function(beta, gamma, A, omega, t, r){
+  beta+gamma*r+A*cos(omega*t)}

# G I T T E R
> grid1.nlm_19.R<-expand.grid(list(beta=seq(1829.5,1830,by=0.2),
+ gamma=seq(16.5,17,by=0.2),A=seq(1978,1979,by=0.2),B=seq(68,68.5,
+ by=0.2),omega=seq(0.9,1.0,by=0.1)))

> grid2.nlm_19.R<-expand.grid(list(beta0=seq(759.5,760,by=0.2),
+ gamma=seq(18.5,19,by=0.2),A=seq(1946,1946.5,by=0.2),B=seq(149,149.5,
```

```
+ by=0.2),omega=seq(0.9,1.0,by=0.1),beta1=seq(137.5,138,by=0.2)))

> grid3.nlm_19.R<-expand.grid(list(beta=seq(1829.5,1830,by=0.2),
+ gamma=seq(16.5,17,by=0.2),A=seq(1978,1979,by=0.2),omega=seq(0.9,1.0,
+ by=0.1)))

# M O D E L L E
> mod.nlm_19.R.1<-nls2(PTM~fct1.nlm_19.R(beta,gamma,A,B,omega,Month,
+ Rainfall),data=RainM.ST_19,start=grid1.nlm_19.R)

> mod.nlm_19.R.2<-nls2(PTM~fct2.nlm_19.R(beta0,gamma,A,B,omega,beta1,
+ Month,Rainfall),data=RainM.ST_19,start=grid2.nlm_19.R)

> mod.nlm_19.R.3<-nls2(PTM~fct3.nlm_19.R(beta,gamma,A,omega,Month,
+ Rainfall),data=RainM.ST_19,start=grid3.nlm_19.R)

# S U M M A R Y
> mod.nlm_19.R.1
Nonlinear regression model
 model: PTM ~ fct1.nlm_19.R(beta,gamma,A,B,omega,Month,Rainfall)
     beta     gamma         A         B     omega
 2990.401     3.213  1588.519 -2493.789     0.819
 residual sum-of-squares: 30139702

> mod.nlm_19.R.2
Nonlinear regression model
 model: PTM ~ fct2.nlm_19.R(beta0,gamma,A,B,omega,beta1,Month,Rainfall)
    beta0     gamma         A         B     omega     beta1
 1930.001     8.803  1368.832 -2903.356     0.800   125.310
 residual sum-of-squares: 27899317
Number of iterations to convergence: 6
Achieved convergence tolerance: 1.910e-06

> mod.nlm_19.R.3
Nonlinear regression model
```

```
  model:  PTM ~ fct3.nlm_19.R(beta, gamma, A, omega, Month, Rainfall)
    beta    gamma        A    omega
3630.442  -14.578 2278.150    0.966
 residual sum-of-squares: 35846256
Number of iterations to convergence: 4
Achieved convergence tolerance: 2.029e-06

> summary(mod.nlm_19.R.1)
Formula: PTM ~ fct1.nlm_19.R(beta,gamma,A,B,omega,Month,Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta   2.990e+03  5.993e+02   4.990 0.000198 ***
gamma  3.213e+00  1.228e+01   0.262 0.797427
A      1.589e+03  1.055e+03   1.506 0.154296
B     -2.494e+03  1.056e+03  -2.362 0.033181 *
omega  8.190e-01  6.589e-02  12.429 5.95e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1467 on 14 degrees of freedom

Number of iterations to convergence: 9
Achieved convergence tolerance: 3.925e-06

> summary(mod.nlm_19.R.2)
Formula: PTM ~ fct2.nlm_19.R(beta0,gamma,A,B,omega,beta1,Month,
    Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta0  1.930e+03  1.192e+03   1.620   0.1293
gamma  8.803e+00  1.323e+01   0.665   0.5175
A      1.369e+03  1.152e+03   1.188   0.2562
B     -2.903e+03  1.019e+03  -2.850   0.0137 *
omega  8.000e-01  6.176e-02  12.952 8.32e-09 ***
beta1  1.253e+02  1.225e+02   1.023   0.3249
```

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

Residual standard error: 1465 on 13 degrees of freedom

Number of iterations to convergence: 6
Achieved convergence tolerance: 1.910e-06

> summary(mod.nlm_19.R.3)
Formula: PTM ~ fct3.nlm_19.R(beta, gamma, A, omega, Month, Rainfall)
Parameters:
        Estimate Std. Error t value Pr(>|t|)
beta  3630.44192  492.71997   7.368 2.34e-06 ***
gamma  -14.57802    8.46796  -1.722 0.105703
A     2278.14952  534.93951   4.259 0.000687 ***
omega    0.96605    0.03115  31.016 5.08e-15 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual standard error: 1546 on 15 degrees of freedom

Number of iterations to convergence: 4
Achieved convergence tolerance: 2.029e-06

# A I C
> AIC(mod.nlm_19.R.1,mod.nlm_19.R.2,mod.nlm_19.R.3)
                 df      AIC
mod.nlm_19.R.1   6 337.1810
mod.nlm_19.R.2   7 337.7135
mod.nlm_19.R.3   5 338.4756
```

Literaturverzeichnis

[web] *From Data to Graphics.* `http://zoonek2.free.fr/UNIX/48_R/03.html`,

[Batschelet 1981] Batschelet, Edward: *Circular Statistics in Biology.* Academic Press, 1981

[Burnham u. Anderson 2002] Burnham, Kenneth P. ; Anderson, David R.: *Model Selection and Multimodel Inference.* Springer, 2002

[Fisher 1993] Fisher, N. I.: *Statistical Analysis of Circular Data.* Cambridge University Press, 1993

[Fisher 2000] Fisher, Nicholas I.: *Statistical analysis of circular data.* Cambridge Univ. Press, Cambridge, 2000

[Heuser 2008] Heuser, Harro: *Lehrbuch der Analysis Teil 2.* Viewweg+Teubner, 2008

[J.A.Greenwood u. a. 1953] J.A.Greenwood ; E.J.Gumbel ; D.Durand: The circular normal distribution: Theory and tables. In: *Journal of the American Statistical Association* 15.3 (1953), Nr. 48, S. 131–152

[Keller 2003] Keller, Stefan: *Advances in Crop Research*, Universität Hohenheim, Diss., 2003

[Kohn 2005] Kohn, Wolfgang: *Statistik: Datenanalysis und Wahrscheinlichkeitsrechnung.* Springer, Berlin, Heidelberg, New York, 2005

[Krengel 2005] Krengel, Ulrich: *Einführung in die Wahrscheinlichkeitstheorie und Statistik.* Vieweg, Wiesbaden, 2005

[Lee 2010] Lee, Alan: Circular data. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (2010), Nr. 4, S. 477–486

[Mardia 1972] Mardia, Kanti V.: *Statistics of Directional Data.* Academic Press, London, 1972

[Mardia u. Jupp 2000] MARDIA, Kanti V. ; JUPP, Peter E.: *Directional Statistics.* Wiley, Chichester, 2000

[R Development Core Team 2009] R DEVELOPMENT CORE TEAM: *R: A language and environment for statistical computing.* Version: 2009. `http://cran.r-project.org/`

[Ritz u. Streibig 2008] RITZ, Christian ; STREIBIG, Jens C.: *Nonlinear Regression with R.* Springer, 2008

[Sachs u. Hedderich 2006] SACHS, Lothar ; HEDDERICH, Jürgen: *Angewandte Statistik: Methodensammlung mit R.* Springer, Berlin, Heidelberg, New York, 2006

[Siudak 2007] SIUDAK, Anna Z.: *Robuste Parameterschätzung,Modelldiskriminierung und optimale Versuchsplanung am Beispiel von In-vitro-Datensätzen zur Benzaldehydlyase*, Carl von Ossietzky Universität Oldenburg, Diss., 2007

[Small 1996] SMALL, Christopher G.: *The statistical theory of shape.* Springer, New York, NY [u.a.], 1996

[Sporleder u. a. 2009] SPORLEDER, Marc ; SIMON, Reinhard ; GONZALES, Juan ; CHAVEZ, Daniel ; MENDIBURU, Felipe ; KROSCHEL, Jürgen: *Manual for ILCYM Version 1.1 Insect Life Cycle Modeling.* Version: 2009. `http://research.cip.cgiar.org/confluence/display/ilcym/Home`

[Weber u. Ulrich 2007] WEBER, Hubert ; ULRICH, Helmut: *Laplace-Transformation, Grundlagen, Fourierreihen und Fourierintegral, Anwendungen.* B.G. Teubner Verlag, 2007

[Zucchini 2009] ZUCCHINI, Walter: *Statistik für Bachelor- und Masterstudenten: eine Einführung für Wirtschafts- und Sozialwissenschaftler.* Springer, Berlin [u.a.], 2009

Erklärung

Ich versichere hiermit, dass ich meine Diplomarbeit mit dem Thema

> *Zyklische Modellierung von Mottenanzahlen Statistische Methoden für circuläre Daten*

selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Kassel, den 28. Dezember 2010

Agnieszka Wenska