

## Statistik

Dr. Thomas Zehrt

Zeitreihen

### Motivation

Typische Beispiele für Zeitreihen sind

- die täglichen Schlusskurse des SMI

| Nummer      | 1       | 2       | 3       | 4       | 5       |
|-------------|---------|---------|---------|---------|---------|
| Datum       | 03.08.  | 04.08.  | 05.08.  | 06.08.  | 07.08.  |
| Schlusskurs | 5954.66 | 5964.43 | 5927.70 | 5959.29 | 5933.70 |

Wir könnten und werden zur Darstellung dieser Zeitreihe einfacher  $y_1 = 5954.66$ ,  $y_2 = 5964.43$ ,  $y_3 = 5927.70$ ,  $y_4 = 5959.29$  und  $y_5 = 5933.70$  schreiben und natürlich erinnert diese Darstellung an eine (mathematische) Zahlenfolge.

- die täglichen Schlusskurse eines börsengehandelten Wertpapiers,
- die monatlichen Arbeitslosenzahlen,
- die Quartalsumsätze eines Unternehmens.

Bei allen diesen Zeitreihen wird ein Merkmalsträger über eine gewisse Zeit hin beobachtet und die Ausprägungen des Merkmals werden zeitlich hintereinander, meist in regelmässigen Abständen, gemessen bzw. bestimmt. In der Zeitreihenanalyse geht es darum, eine Zeitreihe auf Gesetzmässigkeiten zu untersuchen und daraus eventuell Vorhersagen für den weiteren Verlauf der Zeitreihe zu machen.

### Einführung in Geogebra und R

Sicher ist Ihnen in den letzten Wochen aufgefallen, wie mühsam, langwierig und fehleranfällig statistische Berechnungen schon für relativ kleine Datensätze sein können. In der Praxis müssen solche Rechnungen aber für deutlich grössere Datensätze durchgeführt werden. Dabei ist der Einsatz von Computern und spezieller Software unumgänglich.

Trotzdem würde ich Ihnen stets raten, zu jedem Thema mindestens ein (kleineres) Beispiel per Hand zu bearbeiten. Alleine aus der Nutzung einer Software kann kein Verständnis des Problems bzw. des zugehörigen Lösungsalgorithmus erwachsen und dieses Verständnis ist (oft, nicht immer) unerlässlich. Insbesondere wenn Sie wissenschaftlich arbeiten (Seminararbeiten, Bachelorarbeiten, Masterarbeiten) wollen!

- Geogebra ist eine kostenlose dynamische Mathematiksoftware für Schüler und Studenten. Geogebra unterstützt innovatives Lernen im Fach Mathematik und kann von der Homepage

*www.geogebra.org*

herunter geladen und (meist problemlos) auf dem privaten Rechner installiert werden. Tun Sie das bitte! Mittlerweile gibt es auch Versionen für Tablets und Smartphones.

- In diesem Kurs möchte ich Ihnen auch eine (kurze) Einführung in R geben, ein mächtiges und **frei verfügbares** Programm zur Bearbeitung statistischer Probleme. Nach (fast garantierten) Startschwierigkeiten werden Sie (hoffentlich) einsehen, dass es sich sehr gut mit R arbeiten lässt. Sie finden das Programm unter der Adresse

*stat.ethz.ch/CRAN*

zum download und ich würde Ihnen empfehlen, das Programm auf Ihrem privaten Rechner zu installieren.

## Benötigtes Schulwissen

- Summenzeichen,
- Lösen von linearen Gleichungssystemen (2 Gleichungen und 2 Variablen),
- ein wenig Gefühl im Umgang mit Zahlenfolgen ist hilfreich,
- Eigenschaften der Logarithmen, insbesondere  $\log_a(u \cdot v) = \log_a(u) + \log_a(v)$ . Falls also

$$y = G \cdot S \cdot R$$

für reelle Zahlen  $y, G, S, R > 0$  gilt, so gilt auch

$$\log_a(y) = \log_a(G \cdot S \cdot R) = \log_a(G) + \log_a(S) + \log_a(R).$$

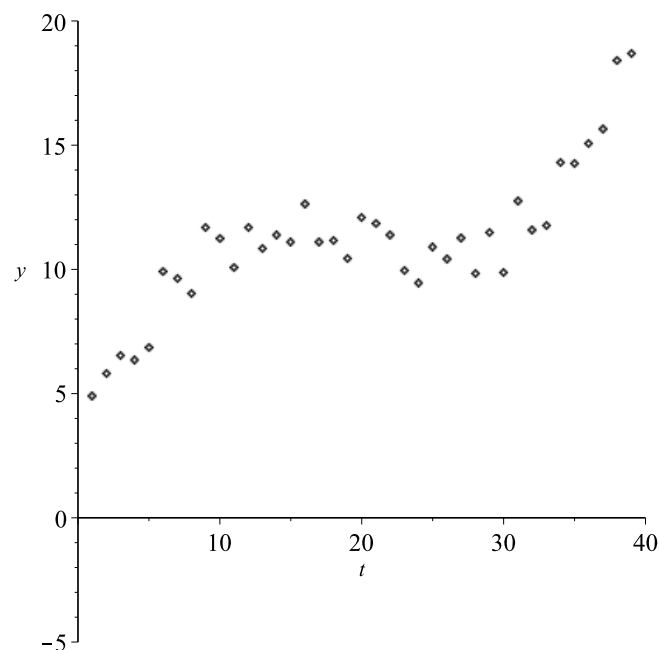
**Merke:** Logarithmen machen Produkte zu Summen!

# 1 Problemstellung

Unter einer Zeitreihe

$$y_i = y(t_i) \quad \text{für } i = 1, 2, 3, \dots$$

verstehen wir eine zeitlich geordnete Folge von Beobachtungen eines quantitativen Merkmals  $Y$ , d.h.  $y_i$  ist die Ausprägung des beobachteten Merkmals zum Zeitpunkt  $t_i$ .

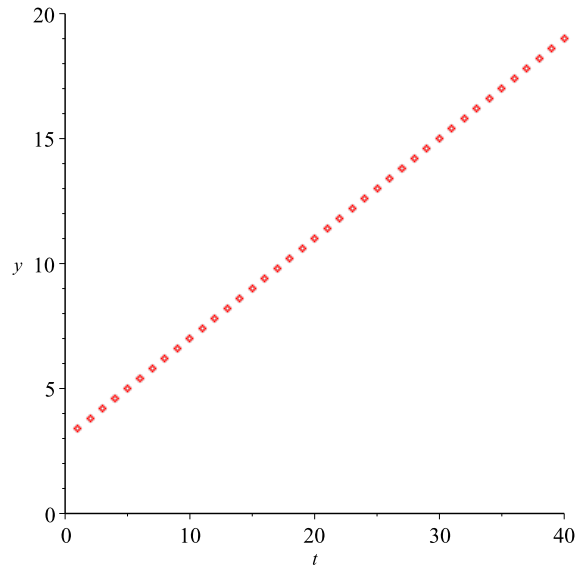


Dass wir  $t_i$  anstelle von  $t$  schreiben bedeutet, dass wir unser Merkmal nur an einigen Zeitpunkten bestimmen (können) und eigentlich nichts darüber wissen, welche Grösse unser Merkmal zwischen diesen (Mess)zeitpunkten hat. Zeitreihen kommen in der Praxis recht häufig vor (Bruttoinlandprodukt, Umsätze einer Firma, mein Guthaben oder mein Gewicht der letzten 10 Jahre) und man versucht zunächst, die Zeitreihe in Komponenten zu zerlegen, die dann einzeln untersucht werden können. Für die Zeitreihenanalyse verwendet man häufig ein additives Zeitreihenmodell

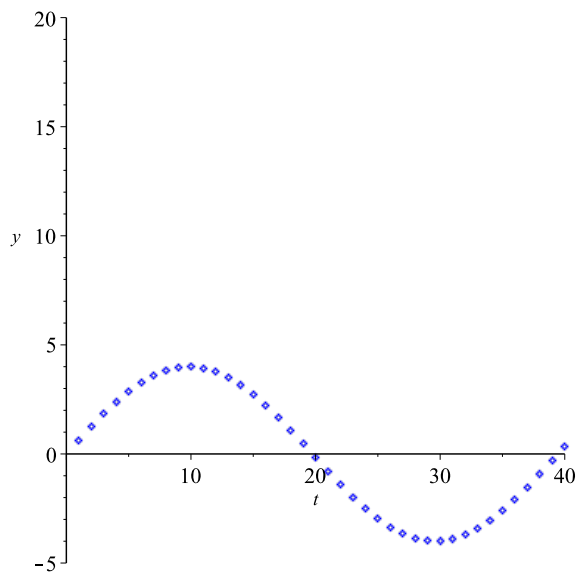
$$y(t_i) = G(t_i) + S(t_i) + R(t_i) \quad \text{für } i = 1, 2, 3, \dots$$

wobei die einzelnen Summanden die folgenden Bedeutungen haben:

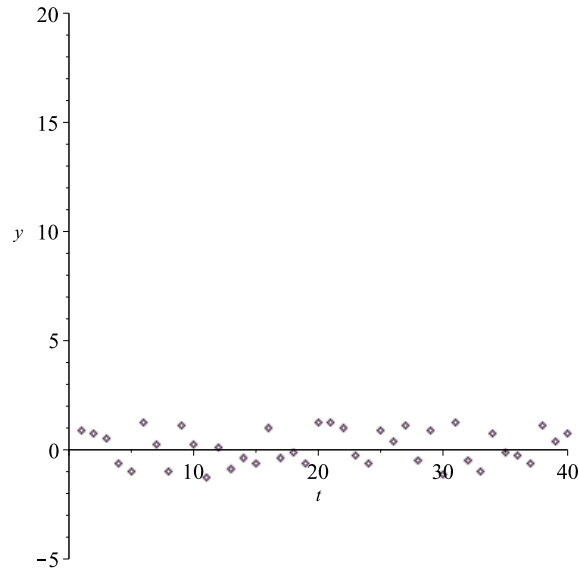
1.  $G$  die glatte Komponente (langfristig wirkend; meist monotone Funktion, die z.B. das kontinuierliche Wirtschaftswachstum modelliert),



2.  $S$  die Saisonkomponente (beschreibt im allgemeinen regelmässige periodische Einfüsse beispielsweise die Frühjahrsbelegung auf dem Arbeitsmarkt) und

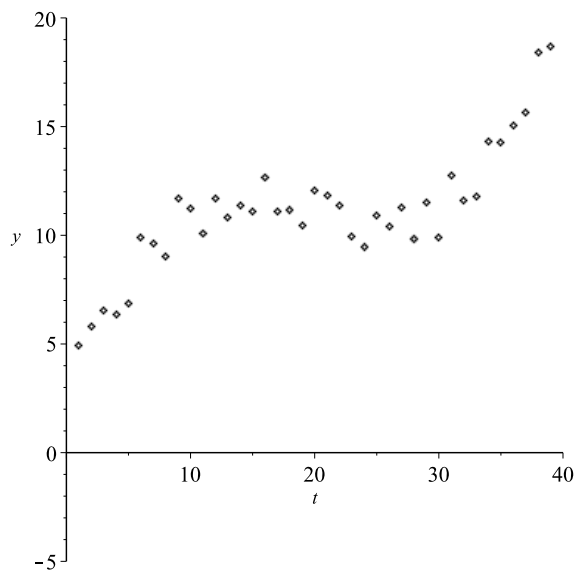
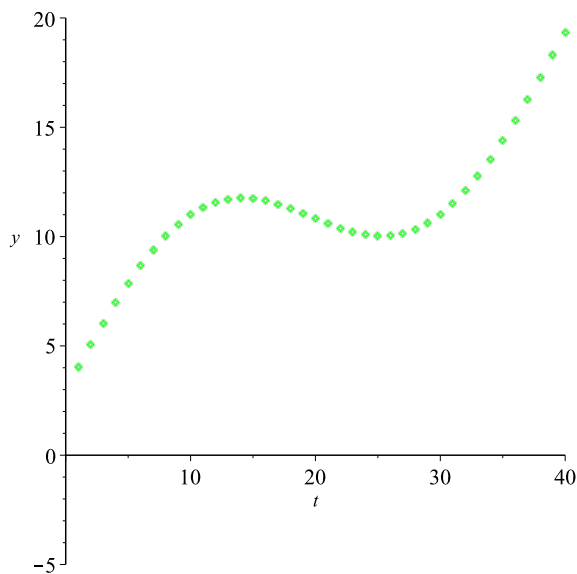


3.  $R$  die reguläre Komponente (modelliert im allgemeinen regellose und zufällige Einflüsse).



Diese drei Komponenten einer Zeitreihe sind **nicht** direkt zu **beobachten**, diese Aufteilung ist also willkürlich, scheint aber sinnvoll zu sein.

Im ersten Bild sehen Sie nun die Zeitreihe  $G(t_i) + S(t_i)$  und im zweiten Bild die gesamte Zeitreihe  $G(t_i) + S(t_i) + R(t_i)$ .



**Aufgabe 1.1** *Ein multiplikatives Zeitreihenmodell*

$$y(t_i) = G(t_i) \cdot S(t_i) \cdot R(t_i),$$

kann leicht auf ein additives Modell zurückgeführt werden. Überlegen Sie wie.

**Lösung:**

Die Aufgabe der Zeitreihenanalyse ist nun die Schätzung von glatter Komponente und Saisonkomponenten aus einer gegebenen endlichen Zeitreihe (d.h. aus einer zeitlich geordneten Reihe von  $n$  Messungen des Merkmals).

$$y_1 = y(t_1), y_2 = y(t_2), \dots, y_n = y(t_n).$$

Die Notation der Zeitreihe mit Indizes (oder sogar Multiindizes) ist sicher etwas gewöhnungsbedürftig, insbesondere auch dann, wenn die Zeitreihenmessungen nicht in regelmässigen Abständen vollzogen werden.

**Beispiel 1.1** Für die drei SMI Schlusskurse 5954.66 (am 03.08.2009), 5964.43 (am 04.08.2009) und 5933.70 (am 07.08.2009) bedeutet das  $t_1 = 03.08.2009$ ,  $t_2 = 04.08.2009$  und  $t_3 = 07.08.2009$  sowie z.B.  $y_3 = y(07.08.2009) = 5933.70$ .

**Ergänzung** Der Quotient aus zwei aufeinander folgenden Zeitreihenwerten ist der Wachstumsfaktor dieser Periode:

$$q_t = \frac{y_t}{y_{t-1}}$$

für  $t = 1, 2, \dots$ . Das geometrische Mittel

$$\hat{q} = \left( \prod_{t=1}^T q_t \right)^{1/T} = \left( \frac{y_T}{y_0} \right)^{1/T}$$

liefert den durchschnittlichen Wachstumsfaktor in den  $T$  Perioden.

Bezieht man den Wert eines Merkmals zum Zeitpunkt  $t$  auf den Wert des Merkmals zum Zeitpunkt  $t = 0$  so liegt eine so genannte Messziffer vor. So erhält man für die  $T$  Perioden die Messziffernfolge  $(y_t/y_0)$  für  $t = 1, 2, \dots, T$ . Mit Hilfe dieser Messziffernfolge lässt sich die zeitliche Entwicklung eines Merkmals recht gut erkennen.

## 2 Bestimmung der glatten Komponente

Bei der Zeitreihenanalyse beginnt man zunächst mit der glatten Komponente, indem man für  $G$  einen bestimmten Funktionstyp unterstellt. Wichtige glatte Komponenten sind

1. die lineare glatte Komponente  $G(t_i) = at_i + b$ ,
2. die quadratische oder parabolische glatte Komponente  $G(t_i) = at_i^2 + bt_i + c$  und
3. die exponentielle glatte Komponente  $G(t_i) = a \cdot e^{bt_i}$

mit den (zu bestimmenden Konstanten)  $a, b, c$ . Ein wichtiges Verfahren zur Schätzung der Trendkomponenten ist die aus dem Kapitel über Regressionsrechnung bekannte Methode der kleinsten Quadrate. Wir wollen den Fall der lineare glatte Komponente genau besprechen.

Für die lineare glatte Komponente  $G(t_i) = at_i + b$  erhält man zunächst die modifizierte Modellgleichung der Zeitreihe:

$$y(t_i) = G(t_i) + S(t_i) + R(t_i) = at_i + b + \underbrace{S(t_i) + R(t_i)}_{\text{Störung}} \approx at_i + b$$

Diese Gleichung entspricht dem linearen Regressionsmodell, wenn man den Teil  $S(t_i) + R(t_i)$  als die Störung interpretiert, die verhindert, dass die Punktwolke genau auf einer Modellkurve (Gerade) liegt. Damit ergeben sich die optimalen Parameter  $\hat{a}$  und  $\hat{b}$  für die glatte Komponente als Lösung der beiden Gleichungen

$$\sum_{i=1}^n t_i y_i = a \sum_{i=1}^n t_i^2 + b \sum_{i=1}^n t_i \quad \text{und} \quad \sum_{i=1}^n y_i = a \sum_{i=1}^n t_i + b \cdot n$$

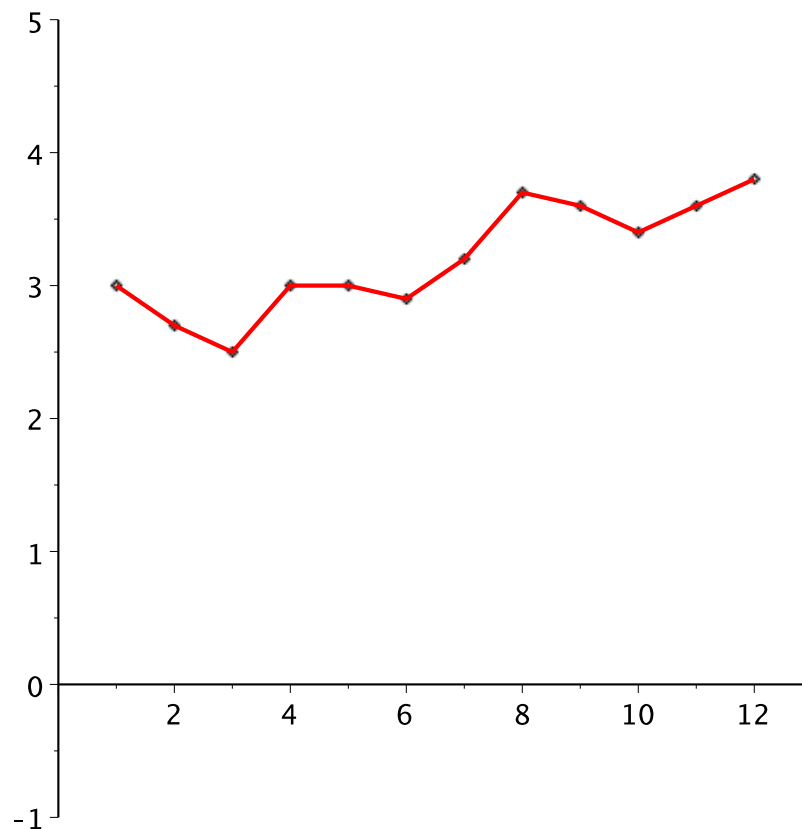
und wir schreiben  $\hat{G}(t_i) = \hat{a}t_i + \hat{b}$  für unsere, mittels der Methode der kleinsten Quadrate bestimmten, **geschätzte** glatte Komponente.

Auch an anderer Stelle werden wir den „Hut  $\hat{\cdot}$ “, nutzen um anzudeuten, dass es sich um eine Schätzung handelt.

**Aufgabe 2.1** Der Umsatz  $y_i = y(t_i)$  ( $i = 1, 2, \dots, 12$ ) einer Firma (angegeben in Mio. CHF) wird drei Jahre lang in jedem Quartal bestimmt:

| Quartal | 1-01 | 2-01 | 3-01 | 4-01 | 1-02 | 2-02 | 3-02 | 4-02 | 1-03 | 2-03 | 3-03 | 4-03 |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|
| $t_i$   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
| $y_i$   | 3.0  | 2.7  | 2.5  | 3.0  | 3.0  | 2.9  | 3.2  | 3.7  | 3.6  | 3.4  | 3.6  | 3.8  |

Skizze:



1. Bestimmen Sie mit Hilfe der Methode der kleinsten Quadrate die lineare glatte Komponente der Zeitreihe.

**Lösung:**

Es gilt  $n = 12$  und:

$$\sum_{i=1}^{12} t_i = 78, \quad \sum_{i=1}^{12} t_i^2 = 650, \quad \sum_{i=1}^{12} y_i = 38.4, \quad \sum_{i=1}^{12} t_i y_i = 263.9$$

Das lineare Gleichungssystem ist:

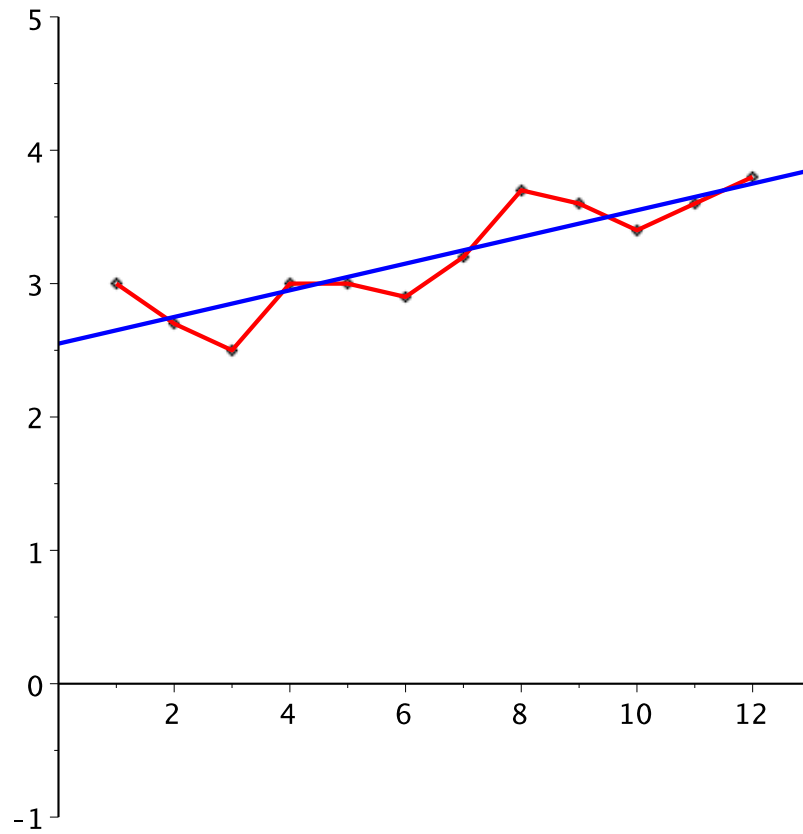
$$263.9 = a \cdot 650 + b \cdot 78$$

$$38.4 = a \cdot 78 + b \cdot 12$$



mit den Lösungen  $\hat{a} = 0.1$  und  $\hat{b} = 2.55$ . Die glatte Komponente (Regressionsgerade) ist somit  $\hat{G}(t) = 0.1 \cdot t + 2.55$ .

Skizze:



**Definition 2.1** Ist nun  $y_1 = y(t_1), y_2 = y(t_2), \dots, y_n = y(t_n)$  eine Zeitreihe, so bezeichnet man (im Fall des additiven Zeitreihenmodells) die neue Zeitreihe  $y_1^* = y^*(t_1), y_2^* = y^*(t_2), \dots, y_n^* = y^*(t_n)$  mit

$$y^*(t_i) := y(t_i) - \hat{G}(t_i)$$

als die zugehörige trendbereinigte Zeitreihe.

**Aufgabe 2.2** Der Umsatz  $y_i = y(t_i)$  ( $i = 1, 2, \dots, 12$ ) einer Firma (angegeben in Mio. CHF) wird drei Jahre lang in jedem Quartal bestimmt:

| Quartal | 1-01 | 2-01 | 3-01 | 4-01 | 1-02 | 2-02 | 3-02 | 4-02 | 1-03 | 2-03 | 3-03 | 4-03 |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|
| $t_i$   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
| $y_i$   | 3.0  | 2.7  | 2.5  | 3.0  | 3.0  | 2.9  | 3.2  | 3.7  | 3.6  | 3.4  | 3.6  | 3.8  |

- Bestimmen Sie mit Hilfe der Methode der kleinsten Quadrate die lineare glatte Komponente der Zeitreihe.

**Lösung:**  $\hat{G}(t) = 0.1 \cdot t + 2.55$

- Bestimmen Sie die trendbereinigte Zeitreihe.

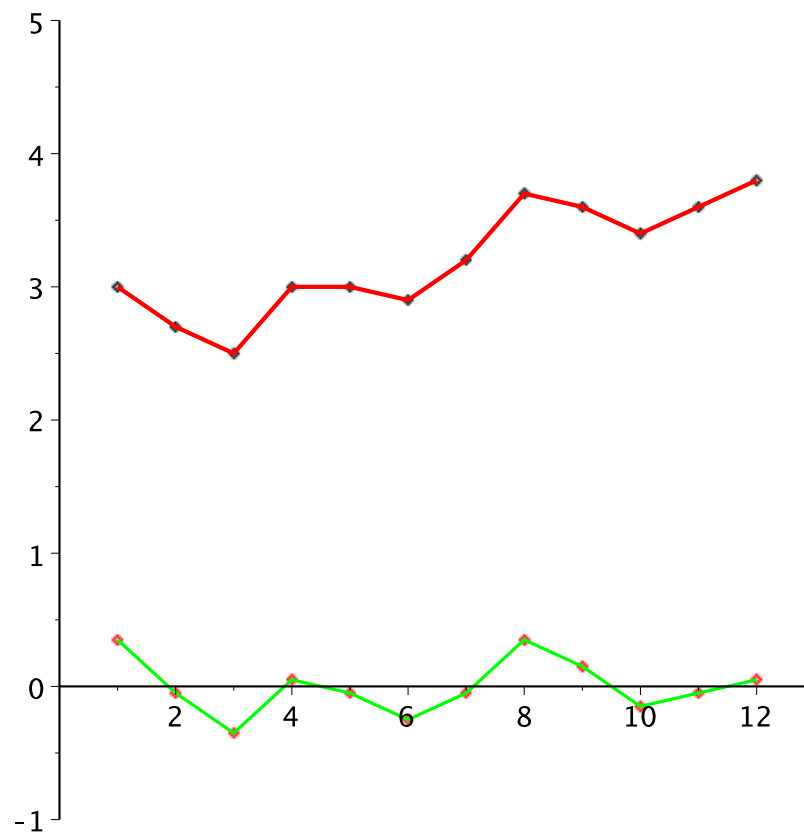
**Lösung:** Wir berechnen für alle  $i = 1, 2, \dots, 12$  den Wert

$$y_i^* = y_i - \hat{G}(t_i) = y_i - 0.1 \cdot i - 2.55$$

und das ergibt:

$$\begin{aligned} y_1^* &= 3.0 - 0.1 \cdot 1 - 2.55 = 0.35 \\ y_2^* &= 2.7 - 0.1 \cdot 2 - 2.55 = -0.05 \\ y_3^* &= 2.5 - 0.1 \cdot 3 - 2.55 = -0.35 \\ y_4^* &= 3.0 - 0.1 \cdot 4 - 2.55 = 0.05 \\ y_5^* &= 3.0 - 0.1 \cdot 5 - 2.55 = -0.05 \\ y_6^* &= 2.9 - 0.1 \cdot 6 - 2.55 = -0.25 \\ y_7^* &= 3.2 - 0.1 \cdot 7 - 2.55 = -0.05 \\ y_8^* &= 3.7 - 0.1 \cdot 8 - 2.55 = 0.35 \\ y_9^* &= 3.6 - 0.1 \cdot 9 - 2.55 = 0.15 \\ y_{10}^* &= 3.4 - 0.1 \cdot 10 - 2.55 = -0.15 \\ y_{11}^* &= 3.6 - 0.1 \cdot 11 - 2.55 = -0.05 \\ y_{12}^* &= 3.8 - 0.1 \cdot 12 - 2.55 = 0.05 \end{aligned}$$

Skizze der trendbereinigten Zeitreihe (grün):



### 3 Bestimmung der Saisonkomponente

Der Einfachheit halber wollen wir in diesem Kapitel stets annehmen, dass die Zeitreihenwerte in regelmässigen Abständen erhoben werden, d.h. dass  $t_{i+1} - t_i$  für alle  $i$  stets den selben Wert (meist 1) hat.

Eine Zeitreihe kann sehr unterschiedliche zyklische Schwankungen aufweisen und bei ökonomischen Zeitreihen muss man insbesondere jahreszeitliche und konjunkturbedingte unterscheiden. Entsprechend ist die Zykluslänge  $k$  festzulegen, d.h. die Anzahl der Zeitperioden, aus denen ein Zyklus besteht. Wir wollen in Zukunft stets einen konstanten Zyklus annehmen, was bedeutet, dass sich die zyklische Komponente nach  $k$  (Mess)Perioden stets exakt wiederholt. Mathematisch würden wir das wie folgt ausdrücken: Die Funktion  $S$  hat die Eigenschaft  $S(t_i) = S(t_{i+k})$  für alle  $i$ .

Bei saisonalen Zyklen ergibt sich die Zykluslänge unmittelbar aus der Anzahl der Perioden pro Jahr (z.B.  $k = 4$  bei Quartalswerte oder  $k = 12$  bei Monatswerten). Ist die Zykluslänge nicht bekannt, wie z.B. bei Konjunkturzyklen, muss man meist mit verschiedenen Werten für  $k$  experimentieren.

Der Schätzzeitraum sollte so gewählt werden, dass alle Zyklusphasen gleichmässig repräsentiert werden. Mathematisch bedeutet das, dass  $n$  (Anzahl der Beobachtungen) ein ganzzahliges Vielfaches von  $k$  (Zykluslänge) sein muss. Der Schätzzeitraum umfasst dann  $m = \frac{n}{k}$  Zyklen.

Um nun die zyklische Komponente zu schätzen, liegt es nahe, sie als **durchschnittlichen Wert der trendbereinigten Zeitreihe in entsprechenden Zyklusphasen** zu bestimmen:

$$\begin{aligned}
 \hat{S}(t_i) &= \hat{S}(t_{i+k}) = \hat{S}(t_{i+2 \cdot k}) = \dots = \hat{S}(t_{i+(m-1) \cdot k}) \\
 &:= \frac{1}{m} \sum_{j=0}^{m-1} y^*(t_{i+j \cdot k}) \\
 &= \frac{1}{m} \sum_{j=0}^{m-1} \left( y(t_{i+j \cdot k}) - \hat{G}(t_{i+j \cdot k}) \right)
 \end{aligned} \tag{1}$$

**Aufgabe 3.1** Der Umsatz  $y_i = y(t_i)$  ( $i = 1, 2, \dots, 12$ ) einer Firma (angegeben in Mio. CHF) wird drei Jahre lang in jedem Quartal bestimmt:

| Quartal | 1-01 | 2-01 | 3-01 | 4-01 | 1-02 | 2-02 | 3-02 | 4-02 | 1-03 | 2-03 | 3-03 | 4-03 |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|
| $t_i$   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
| $y_i$   | 3.0  | 2.7  | 2.5  | 3.0  | 3.0  | 2.9  | 3.2  | 3.7  | 3.6  | 3.4  | 3.6  | 3.8  |

1. Bestimmen Sie mit Hilfe der Methode der kleinsten Quadrate die lineare glatte Komponente der Zeitreihe.

**Lösung:**  $\hat{G}(t) = 0.1 \cdot t + 2.55$

2. Bestimmen Sie die trendbereinigte Zeitreihe.

**Lösung:** 0.35, -0.05, -0.35, 0.05, -0.05, -0.25, -0.05, 0.35, 0.15, -0.15, -0.05, 0.05

3. Bestimmen Sie dann gemäss Formel (1) die Werte  $\hat{S}_1, \dots, \hat{S}_{12}$  der zyklischen Komponente.

**Lösung:** Es gilt sicher  $n = 12$ ,  $k = 4$  und

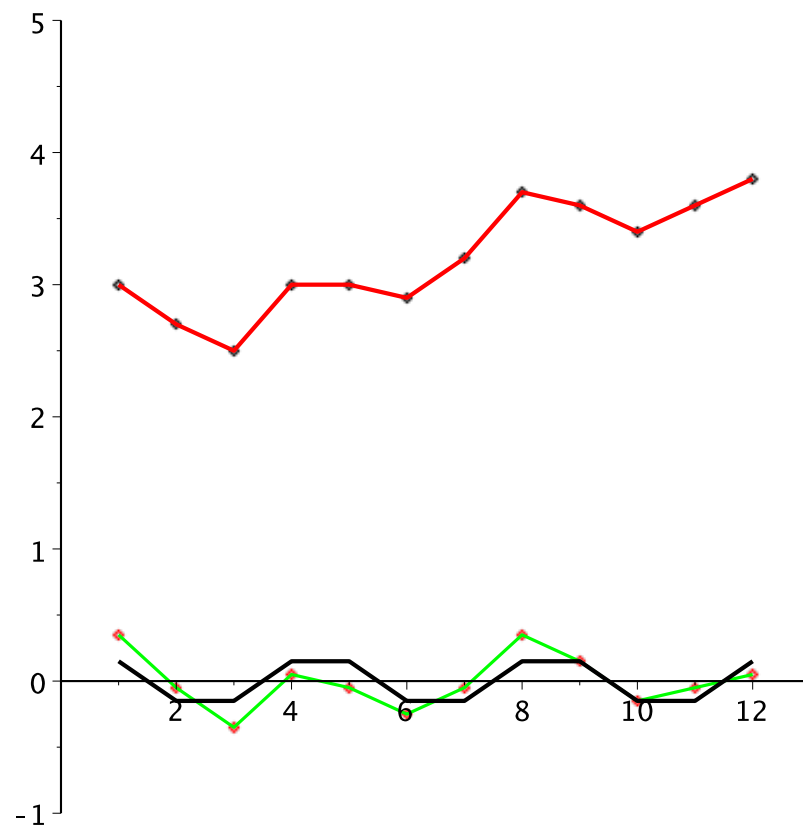
$$\begin{aligned} \hat{S}_1 = \hat{S}_{1+4} = \hat{S}_{1+2 \cdot 4} = \hat{S}_1 = \hat{S}_5 = \hat{S}_9 &= \frac{1}{3}(y_1^* + y_5^* + y_9^*) \\ &= \frac{1}{3}(0.35 - 0.05 + 0.15) = 0.15 \end{aligned}$$

$$\begin{aligned} \hat{S}_2 = \hat{S}_{2+4} = \hat{S}_{2+2 \cdot 4} = \hat{S}_2 = \hat{S}_6 = \hat{S}_{10} &= \frac{1}{3}(y_2^* + y_6^* + y_{10}^*) \\ &= \frac{1}{3}(-0.05 - 0.25 - 0.15) = -0.15 \end{aligned}$$

$$\begin{aligned} \hat{S}_3 = \hat{S}_{3+4} = \hat{S}_{3+2 \cdot 4} = \hat{S}_3 = \hat{S}_7 = \hat{S}_{11} &= \frac{1}{3}(y_3^* + y_7^* + y_{11}^*) \\ &= \frac{1}{3}(-0.35 - 0.05 - 0.05) = -0.15 \end{aligned}$$

$$\begin{aligned} \hat{S}_4 = \hat{S}_{4+4} = \hat{S}_{4+2 \cdot 4} = \hat{S}_4 = \hat{S}_8 = \hat{S}_{12} &= \frac{1}{3}(y_4^* + y_8^* + y_{12}^*) \\ &= \frac{1}{3}(0.05 + 0.35 + 0.05) = 0.15 \end{aligned}$$

Skizze der Saisonkomponente (schwarz):



4. Bestimmen Sie dann die Schätzwerte  $\hat{R}_1, \dots, \hat{R}_{12}$  der regulären Komponente.

**Lösung:** Für alle  $i = 1, 2, \dots, 12$  gilt

$$\hat{R}_i = y_i - \hat{G}(t_i) - \hat{S}_i = y_i^* - \hat{S}_i$$

also

$$\begin{aligned}\hat{R}_1 &= y_1^* - \hat{S}_1 = 0.35 - 0.15 = 0.2 \\ \hat{R}_2 &= y_2^* - \hat{S}_2 = -0.05 + 0.15 = 0.1 \\ \hat{R}_3 &= y_3^* - \hat{S}_3 = -0.35 + 0.15 = -0.2 \\ \hat{R}_4 &= y_4^* - \hat{S}_4 = 0.05 - 0.15 = -0.1 \\ \hat{R}_5 &= y_5^* - \hat{S}_5 = -0.05 - 0.15 = -0.2 \\ \hat{R}_6 &= y_6^* - \hat{S}_6 = -0.25 + 0.15 = -0.1 \\ \hat{R}_7 &= y_7^* - \hat{S}_7 = -0.05 + 0.15 = 0.1 \\ \hat{R}_8 &= y_8^* - \hat{S}_8 = 0.35 - 0.15 = 0.2 \\ \hat{R}_9 &= y_9^* - \hat{S}_9 = 0.15 - 0.15 = 0 \\ \hat{R}_{10} &= y_{10}^* - \hat{S}_{10} = -0.15 + 0.15 = 0 \\ \hat{R}_{11} &= y_{11}^* - \hat{S}_{11} = -0.05 + 0.15 = 0.1 \\ \hat{R}_{12} &= y_{12}^* - \hat{S}_{12} = 0.05 - 0.15 = -0.1\end{aligned}$$

## 4 Einige R-Befehle

### Hilfeseite

?quantile öffnet Hilfeseite zum Befehl „quantile“

### Grundrechenarten

|     |                            |
|-----|----------------------------|
| +   | Addition                   |
| -   | Subtraktion                |
| *   | Multiplikation             |
| /   | Division                   |
| 2^3 | Potenzieren 2 <sup>3</sup> |

### Zuweisungen (Werte an Variablen)

|  |  |
|--|--|
| $x \leftarrow 2$ oder $2 \rightarrow x$        | der Variablen $x$ wird der Wert 2 zugewiesen             |
| $x \leftarrow c(1, 2, 3, 4)$                   | $x$ wird der Vektor (1, 2, 3, 4) zugewiesen              |
| $x \leftarrow 1 : 4$                           | $x$ wird der Vektor (1, 2, 3, 4) zugewiesen              |
| $x \leftarrow \text{seq}(1, 4, \text{by} = 1)$ | $x$ wird der Vektor(Zahlenfolge) (1, 2, 3, 4) zugewiesen |

### Statistische Funktionen (t Vektor oder Folge)

|                         |   |
|-------------------------|---|
| min(t)                  | Minimum   |
| max(t)                  | Maximum   |
| mean(t)                 | arithmetisches Mittel                           |
| median(t)               | Median  |
| quantile(t, probs=0.75) | 0.75-Quantil                                    |
| summary(t)              |   |
| var(t)                  | Varianz (mit Vorfaktor $1/(n - 1)$ )            |
| sd(t)                   | Standardabweichung (mit Vorfaktor $1/(n - 1)$ ) |
| boxplot(t)              | Boxplot von t                                   |
| sort(t)                 | Rangwertreihe (sortiert)                        |
| table(t)                | Häufigkeitstabelle                              |

### Lineare Regression (x und y gleichlange Vektoren)

lm(y ~ x) Lineares Modell  $\hat{y}_i = \hat{a}x_i + \hat{b}$

### Grafiken

|  |  |
|--|--|
| plot(NULL, xlim=c(-5,5), ylim=c(0,10)) | zeichnet leeres Koordinatensystem                        |
| abline(a=0, b=1)                       | zeichnet die Gerade $a + bx$                             |
| curve(x <sup>2</sup> , from -5, to =5) | zeichnet Graphen von $f(x) = x^2$ für $-5 \leq x \leq 5$ |

### Pseudozufallszahlen (PZZ) (Simulation)

|             |   |
|-------------|---|
| set.seed(1) | setzt Startwert (=1) des PZZ-Generators           |
| runif(5)    | erzeugt 5 PZZ aus der Gleichverteilung auf (0, 1) |
| rnorm(5)    | erzeugt 5 PZZ aus der Standardnormalverteilung    |

### Schleifen

|   |                |
|---|----------------|
| for( i in 1:100 ) { Anweisungen }           | for-Schleife   |
| while( logischer Ausdruck ) { Anweisungen } | while-Schleife |



## 5 Übungsaufgabe (darf und sollte mit R gelöst werden)

Ein Lebensmittelhersteller möchte die Menge eines hergestellten Produktes zeitlich besser der Nachfrage anpassen, um das Verderben produzierter Überkapazitäten zu vermindern. Dazu wird eine Aufstellung der pro Monat verkauften Mengen (in Tonnen) des Produktes im Zeitraum 2003-2005 vorgenommen:

| Monat     | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  |
|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Jahr 2003 | 1.2 | 0.7 | 2.0 | 2.2 | 3.2 | 2.7 | 3.8 | 4.3 | 3.6 | 2.9 | 2.9 | 7.7 |
| Jahr 2004 | 1.4 | 0.7 | 2.2 | 2.3 | 3.5 | 2.4 | 4.0 | 4.1 | 3.3 | 2.7 | 3.1 | 6.9 |
| Jahr 2005 | 1.3 | 0.8 | 1.9 | 2.1 | 2.9 | 2.5 | 4.1 | 4.0 | 3.2 | 2.6 | 3.0 | 7.6 |

1. Skizzieren Sie die Zeitreihe.
2. Bestimmen Sie mit Hilfe der Methode der kleinsten Quadrate die lineare glatte Komponente der Zeitreihe.
3. Bestimmen Sie die trendbereinigte Zeitreihe.
4. Bestimmen Sie dann gemäss Formel (1) die Werte  $\hat{S}_1, \dots, \hat{S}_{36}$  der zyklischen Komponente.

Ich kann nur davon abraten, diese Aufgabe per Hand zu rechnen. Falls Sie es doch versuchen wollen (Respekt!), hier einige Zwischenergebnisse:

$$\sum_{i=1}^{36} t_i = 666,$$

$$\sum_{i=1}^{36} y_i = 109.8,$$

$$\sum_{i=1}^{36} t_i^2 = 16'206,$$

$$\sum_{i=1}^{36} t_i y_i = 2170.6 \text{ und}$$

$$\hat{G}(t) = 0.0358 \cdot t + 2.3867$$

Öffnen Sie R und geben Sie Zeile für Zeile die folgenden Befehle ein:

```
t <- seq(1 : 36)
t
y <- c(1.2, 0.7, 2.0, 2.2, 3.2, 2.7, 3.8, 4.3, 3.6, 2.9, 2.9, 7.7,
      1.4, 0.7, 2.2, 2.3, 3.5, 2.4, 4.0, 4.1, 3.3, 2.7, 3.1, 6.9,
      1.3, 0.8, 1.9, 2.1, 2.9, 2.5, 4.1, 4.0, 3.2, 2.6, 3.0, 7.6)
y
summary(y)
plot(NULL, xlim = c(-1, 37), ylim = c(-3, 10), xlab = "Zeit", ylab = "")
points(t, y)
lines(t, y)
Trend <- lm(y ~ t)
G <- fitted.values(Trend)
G
abline(Trend, col = 134)
Trendb <- c(y - G)
Trendb
points(t, Trendb, col = 50)
lines(t, Trendb, col = 50)
Ss <- 1 : 12
for(i in 1 : 12) {Ss[i] <- 1/3*(Trendb[i+12*0]+Trendb[i+12*1]+Trendb[i+12*2])}
Ss
Null <- rep(0, 36)
Null
S <- Ss + Null
S
points(t, S, col = 25)
lines(t, S, col = 25)
R <- Trendb - S
R
windows()
plot(t, R)
```