Statistik im Bachelor-Studium der BWL und VWL

Methoden, Anwendung, Interpretation

3., aktualisierte Auflage

Max C. Wewel

Mit herausnehmbarer Formelsammlung

Pearson

$$\hat{s}_2 = \hat{s}_6 = \hat{s}_{10} = \frac{1}{3}\left[(y_2 - \hat{g}_2) + (y_6 - \hat{g}_6) + (y_{10} - \hat{g}_{10}) \right]$$
$$= \frac{1}{3}(-0{,}05 - 0{,}25 - 0{,}15) = -0{,}15 \quad [10^6\,€]$$

$$\hat{s}_3 = \hat{s}_7 = \hat{s}_{11} = \frac{1}{3}\left[(y_3 - \hat{g}_3) + (y_7 - \hat{g}_7) + (y_{11} - \hat{g}_{11}) \right]$$
$$= \frac{1}{3}(-0{,}35 - 0{,}05 - 0{,}05) = -0{,}15 \quad [10^6\,€]$$

$$\hat{s}_4 = \hat{s}_8 = \hat{s}_{12} = \frac{1}{3}\left[(y_4 - \hat{g}_4) + (y_8 - \hat{g}_8) + (y_{12} - \hat{g}_{12}) \right]$$
$$= \frac{1}{3}(0{,}05 + 0{,}35 + 0{,}05) = 0{,}15 \quad [10^6\,€] \,.$$

Dies bedeutet, dass die Umsätze saisonbedingt **im ersten und vierten Quartal** eines Jahres jeweils um **150 000 € über dem Trendwert** und **im zweiten und dritten Quartal** jeweils um **150 000 € unter dem Trendwert** liegen. Deshalb sollten die Ex-ante-Trendprognosen aus Tabelle 3.6 um diese Saisoneinflüsse korrigiert werden (vgl. Tabelle 3.8 und Abbildung 3.9).

Quartal	t	$\hat{g}_t$	$\hat{s}_t$	$\hat{y}_t$
2014 / 1	13	3,85	0,15	4,0
2014 / 2	14	3,95	–0,15	3,8
2014 / 3	15	4,05	–0,15	3,9
2014 / 4	16	4,15	0,15	4,3

Tabelle 3.8: Ex-ante-Prognosen unter Berücksichtigung der Saisoneinflüsse

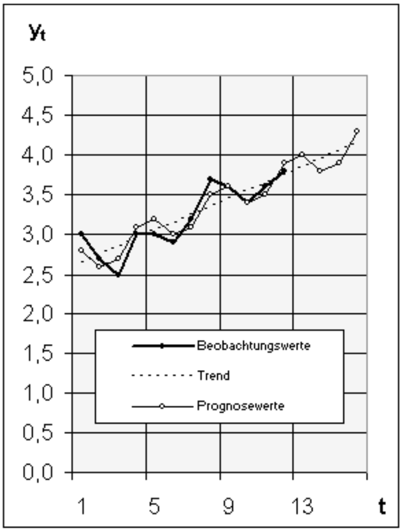

Abbildung 3.9: Prognosewerte der Umsatzreihe ohne und mit Saisonfigur

3.3.4 Beurteilung der Zeitreihenzerlegung

Im hier betrachteten additiven Zeitreihenmodell gilt – analog zum linearen Regressionsmodell – die **Streuungszerlegung**

$$\sigma_y^2 = \sigma_{\hat{g}}^2 + \sigma_{\hat{s}}^2 + \sigma_{\hat{r}}^2 , \tag{3.22}$$

wobei nun

- die Varianz der beiden systematischen Komponenten $\sigma_{\hat{g}}^2$ und $\sigma_{\hat{s}}^2$ als **erklärte Streuung** und
- die Varianz der Restkomponente $\sigma_{\hat{r}}^2$ als **Reststreuung**

bezeichnet werden.

Die Güte des Zeitreihenmodells und der aus ihm abgeleiteten Prognosen kann daher wieder anhand des Anteils der erklärten Streuung an der Gesamtstreuung beurteilt werden. Je nach Einbeziehung der zyklischen Komponente gibt es zwei **Bestimmtheitsmaße**:

$$r_{\hat{g}}^2 = \frac{\sigma_{\hat{g}}^2}{\sigma_y^2} = \frac{\sigma_{ty}^2}{\sigma_t^2 \cdot \sigma_y^2} \tag{3.23}$$

und

$$r_{\hat{g}+\hat{s}}^2 = \frac{\sigma_{\hat{g}}^2 + \sigma_{\hat{s}}^2}{\sigma_y^2} . \tag{3.24}$$

Während das Bestimmtheitsmaß (3.23) **nur** den Erklärungsanteil der **Trendfunktion** angibt, enthält (3.24) **zusätzlich** noch den Erklärungsanteil der **zyklischen Komponente**.

Beispiel

Umsatzprognose

Im Fall der additiven Zerlegung der Umsatz-Zeitreihe erhält man die folgende **Streuungszerlegung** (vgl. Tabelle 3.9):

$$\sigma_y^2 = \frac{1{,}92}{12} = 0{,}16\ [10^{12}€^2] \qquad \sigma_{\hat{g}}^2 = \frac{1{,}43}{12} = 0{,}11917\ [10^{12}€^2]$$

$$\sigma_{\hat{s}}^2 = \frac{0{,}27}{12} = 0{,}0225\ [10^{12}€^2] \qquad \sigma_{\hat{r}}^2 = \frac{0{,}22}{12} = 0{,}01833\ [10^{12}€^2]$$

und damit die **Bestimmtheitsmaße**

$$r_{\hat{g}}^2 = \frac{0{,}11917}{0{,}16} = 0{,}7448 \qquad \text{und} \qquad r_{\hat{g}+\hat{s}}^2 = \frac{0{,}11917 + 0{,}0225}{0{,}16} = 0{,}8854 \cdot$$

Somit wird die **Streuung des Umsatzes zu 74,48 % durch den linearen Trend und zu weiteren 14,06 % durch die konstante Saisonfigur erklärt**. Mit einem Bestimmtheitsmaß von insgesamt 88,54 % liefert die additive Zeitreihenzerlegung mithin eine gute bis sehr gute Erklärung des Zeitreihenverlaufs (was auch aus Abbildung 3.9 hervorgeht).

t	y_t	$(y_t-\bar{y})^2$	$\hat{g}_t$	$(\hat{g}_t-\bar{\hat{g}})^2$	$\hat{s}_t$	$\hat{s}_t^2$	$\hat{r}_t$	$\hat{r}_t^2$
1	3,0	0,04	2,65	0,3025	0,15	0,0225	0,2	0,04
2	2,7	0,25	2,75	0,2025	–0,15	0,0225	0,1	0,01
3	2,5	0,49	2,85	0,1225	–0,15	0,0225	–0,2	0,04
4	3,0	0,04	2,95	0,0625	0,15	0,0225	–0,1	0,01
5	3,0	0,04	3,05	0,0225	0,15	0,0225	–0,2	0,04
6	2,9	0,09	3,15	0,0025	–0,15	0,0225	–0,1	0,01
7	3,2	0	3,25	0,0025	–0,15	0,0225	0,1	0,01
8	3,7	0,25	3,35	0,0225	0,15	0,0225	0,2	0,04
9	3,6	0,16	3,45	0,0625	0,15	0,0225	0	0
10	3,4	0,04	3,55	0,1225	–0,15	0,0225	0	0
11	3,6	0,16	3,65	0,2025	–0,15	0,0225	0,1	0,01
12	3,8	0,36	3,75	0,3025	0,15	0,0225	–0,1	0,01
---	38,4	1,92	38,4	1,43	0	0,27	0	0,22

Tabelle 3.9: Streuungszerlegung im additiven Zeitreihenmodell

3.3.5 Saisonbereinigung

In der empirischen Wirtschaftsforschung wird die Zeitreihenanalyse oft zur Saisonbereinigung verwendet. Zahlreiche ökonomische Zeitreihen wie z.B. die Arbeitslosenzahlen oder die Nachfrage nach bestimmten Konsumgütern weisen nämlich ausgeprägte saisonale Schwankungen auf. Dadurch wird die Beurteilung der Veränderungen besonders am „aktuellen Rand“ erschwert. So ist z.B. fraglich, inwieweit ein Anstieg der Beschäftigung im Frühling oder eine Umsatzsteigerung in der Vorweihnachtszeit auf konjunkturelle Belebungen hinweisen oder nur Ausdruck jährlich wiederkehrender Saisoneinflüsse sind.

Zur Analyse der längerfristigen Entwicklungstendenzen empfiehlt es sich daher, den Saisoneinfluss aus der „Ursprungsreihe“ y_t $(t=1,\ldots,n)$ zu eliminieren. Diesen Vorgang nennt man **Saisonbereinigung**. Allgemein bezeichnet man die Elimination kurzfristiger Schwankungen aus Zeitreihen auch als **Glättung** von Zeitreihen.

Die Saisonbereinigung setzt eine **Schätzung der saisonalen Einflüsse** $\hat{s}_t$ $(t=1,\ldots,n)$ voraus, z.B. mit Hilfe des oben dargestellten Modells der additiven Zeitreihenzerlegung. Die **saisonbereinigte Zeitreihe** $\hat{y}_t^s$ $(t=1,\ldots,n)$ erhält man dann durch Subtraktion der geschätzten Saisoneinflüsse von den Ursprungswerten:

$$\hat{y}_t^s \;=\; y_t - \hat{s}_t \qquad (t=1,\ldots,n)\,. \tag{3.25}$$

Beispiel

Saisonbereinigter Umsatz

Die in diesem Kapitel verwendete Umsatz-Zeitreihe weist, wie die Analysen in den vorangehenden Abschnitten gezeigt haben, deutliche jahreszeitliche Schwankungen auf. Zur Beurteilung des Unternehmenserfolgs erscheint es daher sinnvoll, die Saisoneinflüsse aus der Umsatzreihe „herauszurechnen“.

Die nachfolgende Tabelle 3.10 enthält sowohl die **Ursprungswerte y_t** sowie die **saisonbereinigten Umsätze $\hat{y}_t^s$**. Aus dem dazugehörigen Zeitreihendiagramm (Abbildung 3.10) erkennt man, dass die saisonbereinigte Reihe deutlich „glatter“ verläuft als die Ursprungsreihe; sie enthält keine systematischen Schwankungen mehr.

Quartal	t	y_t	$\hat{s}_t$	$\hat{y}_t^s$
2011 / 1	1	3,0	0,15	2,85
2011 / 2	2	2,7	–0,15	2,85
2011 / 3	3	2,5	–0,15	2,65
2011 / 4	4	3,0	0,15	2,85
2012 / 1	5	3,0	0,15	2,85
2012 / 2	6	2,9	–0,15	3,05
2012 / 3	7	3,2	–0,15	3,35
2012 / 4	8	3,7	0,15	3,55
2013 / 1	9	3,6	0,15	3,45
2013 / 2	10	3,4	–0,15	3,55
2013 / 3	11	3,6	–0,15	3,75
2013 / 4	12	3,8	0,15	3,65
---	---	38,4	0	38,4

Tabelle 3.10: Saisonbereinigung

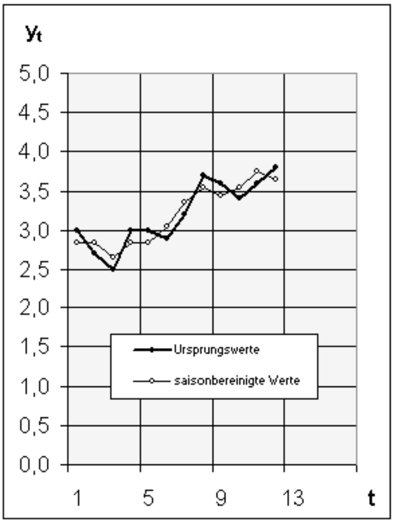

Abbildung 3.10: Saisonbereinigung

3.3.6 Glättung durch gleitende Durchschnitte

Eine einfache Methode der Glättung von historischen Zeitreihen besteht darin, kurzfristige zufällige, saisonale oder auch konjunkturelle Schwankungen zu eliminieren, indem man die Zeitreihenwerte über mehrere, aufeinander folgende Perioden mittelt. Die Ursprungswerte y_t werden dabei ersetzt durch **gleitende Durchschnitte** $\bar{y}_t^{(k)}$, wobei die Zahl k deren „Länge", d.h. die Anzahl der Perioden bei der Durchschnittsbildung angibt. Sollen durch die Glättung systematische Zyklusschwankungen ausgeglichen werden, so muss k der Zykluslänge entsprechen (siehe Abschnitt 3.3.3). Generell nimmt die glättende Wirkung mit steigendem k zu.

Je nachdem, aus welchen Perioden die Zeitreihenwerte bei der Mittelung stammen, können gleitende Durchschnitte **nachlaufend** (d.h. nur unter Einbeziehung von früheren Perioden), **zentriert** (d.h. unter Einbeziehung gleich vieler früherer wie späterer Perioden) oder auch **vorlaufend** (d.h. nur unter Einbeziehung späterer Perioden) definiert werden. Für analytische Zwecke wie die Bereinigung einer Zeitreihe von zyklischen Schwankungen werden zumeist zentrierte Durchschnitte verwendet, gelegentlich aber auch nachlaufende Durchschnitte, etwa bei der Analyse von Aktienkursen (z.B. 30-Tage- bzw. 200-Tage-Durchschnitte). Vorlaufende Durchschnitte sind jedoch in der Praxis selten.

- **Nachlaufende gleitende Durchschnitte** der Länge k ($\leq n$) sind allgemein definiert durch:

$$\overleftarrow{y}_t^{(k)} = \frac{1}{k}\sum_{j=0}^{k-1} y_{t-j} \quad (t = k,\dots,n) \;, \tag{3.26}$$

so dass beispielsweise nachlaufende gleitende Dreier-Durchschnitte berechnet werden nach der Formel:

$$\overleftarrow{y}_t^{(3)} = \frac{1}{3}\sum_{j=0}^{2} y_{t-j} = \frac{y_t + y_{t-1} + y_{t-2}}{3} \quad (t = 3,\dots,n)\,.$$

- Bei **zentrierten gleitenden Durchschnitten** ist zu unterscheiden, ob ihre Länge k ($\leq n$) ungeradzahlig oder geradzahlig ist. Die allgemeine Definition lautet:

$$\overleftrightarrow{y}_t^{(k)} = \begin{cases} \dfrac{1}{k}\displaystyle\sum_{j=-\frac{k-1}{2}}^{\frac{k-1}{2}} y_{t+j} & \textit{falls k ungerade} \\ \dfrac{1}{k}\left(\tfrac{1}{2}y_{t-\frac{k}{2}} + \displaystyle\sum_{j=-\frac{k-2}{2}}^{\frac{k-2}{2}} y_{t+j} + \tfrac{1}{2}y_{t+\frac{k}{2}}\right) & \textit{falls k gerade}\,. \end{cases} \tag{3.27}$$

Zentrierte gleitende Dreier- bzw. Vierer-Durchschnitte werden z.B. gebildet durch:

$$\overleftrightarrow{y}_t^{(3)} = \frac{1}{3}\sum_{j=-1}^{1} y_{t+j} = \frac{y_{t-1} + y_t + y_{t+1}}{3} \quad (t = 2,\dots,n-1)$$

bzw.

$$\overleftrightarrow{y}_t^{(4)} = \frac{1}{4}\left(\tfrac{1}{2}y_{t-2} + \sum_{j=-1}^{1} y_{t+j} + \tfrac{1}{2}y_{t+2}\right) = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4} \quad (t = 3,\dots,n-2)\,.$$

Beispiel

Glättung der Umsatz-Zeitreihe

Berechnet man für die vierteljährliche Umsatz-Zeitreihe **zentrierte gleitende Vier-Quartale-Durchschnitte**, so erhält man auf sehr einfache Weise eine **saisonbereinigte Reihe** der Quartalsumsätze.

Quartal	t	y_t	$\ddot{y}_t^{(4)}$
2011 / 1	1	3,0	---
2011 / 2	2	2,7	---
2011 / 3	3	2,5	2,8000
2011 / 4	4	3,0	2,8250
2012 / 1	5	3,0	2,9375
2012 / 2	6	2,9	3,1125
2012 / 3	7	3,2	3,2750
2012 / 4	8	3,7	3,4125
2013 / 1	9	3,6	3,5250
2013 / 2	10	3,4	3,5875
2013 / 3	11	3,6	---
2013 / 4	12	3,8	---

Tabelle 3.11: Gleitende Vier-Quartale-Durchschnitte

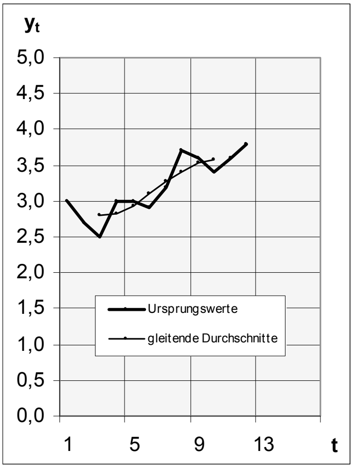

Abbildung 3.11: Gleitende Vier-Quartale-Durchschnitte

Wie das Beispiel zeigt, sind zentrierte gleitende Durchschnitte wegen der fehlenden Werte am Ende des Beobachtungszeitraums nicht geeignet, Entwicklungen am „aktuellen Rand" zu beurteilen. Im Übrigen sind gleitende Durchschnitte generell **rein deskriptiver Natur**; sie besitzen – anders als eine Trendfunktion – keinen darüber hinausgehenden Hypothesen-Charakter und bieten insofern keinen Ansatzpunkt zur Prognose der künftigen Entwicklung.

3.4 Aufgaben

Aufgabe 3.1

Der Personalchef eines Unternehmens vermutet einen Zusammenhang zwischen der Lage auf dem regionalen Arbeitsmarkt und der Zahl der Krankmeldungen in seinem Betrieb. Die durchschnittlichen Arbeitslosenzahlen in der Region (X) und die Zahl der Krankmeldungen im Betrieb (Y) betrugen in den letzten zwölf Jahren:

Arbeitslose [1000]	5	6	6	7	17	15	15	13	12	9	8	7
Krankmeldungen [100]	18	16	14	13	2	3	4	6	7	10	7	8

a) Berechnen Sie für die beiden Zeitreihen jeweils die Standardabweichung, die Kovarianz und den Korrelationskoeffizienten! Beurteilen Sie die Korrelation der beiden Merkmale!

b) Bestimmen Sie nach der Methode der kleinsten Quadrate die Koeffizienten der Regressionsgerade und zeichnen Sie die Gerade in das Streuungsdiagramm ein!

c) Berechnen und interpretieren Sie das Bestimmtheitsmaß!

Aufgabe 3.2

Ein Professor meint, dass bei den Studierenden des Studiengangs BWL ein starker Zusammenhang zwischen der Mathematiknote im ersten Semester (X) und der Examensnote (Y) besteht. Bei 20 Absolventen ergab sich folgende zweidimensionale Verteilung dieser beiden (als intervallskaliert betrachteten) Noten:

x_i \ y_j	**1**	**2**	**3**	**4**
1	1	1	0	0
2	2	1	1	0
3	0	3	2	1
4	0	5	2	1

a) Bestimmen Sie die Durchschnittsnoten, die Varianzen von X und Y sowie die Kovarianz!

b) Schätzen Sie die Regressionsgerade $y = a + b\ x$ nach der Methode der kleinsten Quadrate!

c) Wie beurteilen Sie das Ergebnis der Regressionsanalyse? Was besagt hier das Bestimmtheitsmaß?

d) Bestimmen Sie die Varianzen der Ex-post-Prognosewerte und der Residuen, $\sigma^2_{\hat{y}}$ und $\sigma^2_{\hat{u}}$!

e) Prognostizieren sie die Examensnote eines Studierenden, der in der Mathematik-Klausur eine 1 bzw. eine 4 hatte.

Aufgabe 3.3

Andreas interessiert sich für den Kauf eines Gebrauchtwagens vom Typ „Smart". In der Samstagsausgabe der Regionalzeitung findet er acht passende Angebote:

Tachostand [1000 km]	14	9	30	37	17	6	20	27
Preis [1000 €]	11,7	12,0	9,6	9,3	10,8	12,3	10,2	10,5

a) Andreas nimmt an, dass der Preis (Y) linear vom Tachostand (X) abhängt. Schätzen Sie die Regressionsgerade nach der Methode der kleinsten Quadrate!

b) Beurteilen Sie die Güte der Regression anhand des Bestimmtheitsmaßes!

c) Andreas erfährt von seinem Freund Simon, dass dieser ihm seinen gebrauchten „Smart" mit Tachostand 40000 km zum Freundschaftspreis von 9000 € verkaufen würde. Wie beurteilen Sie das Angebot?

Aufgabe 3.4

Für ein neues Produkt soll die Preis-Absatz-Funktion geschätzt werden. Dazu wurde das Produkt in neun vergleichbaren Supermärkten zu Stückpreisen (p_t) zwischen 60 und 100 Euro angeboten und jeweils die Absatzmenge (q_t) innerhalb einer Woche ermittelt.

Preis [€/Stück]	60	65	70	75	80	85	90	95	100
Absatz [Stück]	23	19	12	13	16	9	11	9	5

a) Berechnen Sie die Standardabweichungen der Absatzpreise und -mengen sowie die Kovarianz und den Korrelationskoeffizienten!

b) Schätzen Sie die Preis-Absatz-Funktion $q = a + b\,p$ und geben Sie an, wie viel Prozent der Mengenstreuung nicht durch die Preisstreuung erklärt werden kann! Wie beurteilen Sie demnach die Güte der Regressionsgeraden?

c) Prognostizieren Sie die wöchentliche Absatzmenge sowie den Umsatz in den insgesamt 150 (gleich großen) Supermärkten, wenn das Produkt in einer Aktionswoche zum Stückpreis von 70 € angeboten wird!

Aufgabe 3.5

Die Absatzzahlen eines europäischen Nutzfahrzeug-Herstellers auf dem nordamerikanischen Markt entwickelten sich in den Jahren 2006 bis 2010 wie folgt:

Halbjahr	I/06	II/06	I/07	II/07	I/08	II/08	I/09	II/09	I/10	II/10
Absatz [1000 Stück]	33	33	31	29	25	23	21	20	18	17

a) Berechnen Sie die Trendgerade für die Absatzentwicklung und beurteilen Sie die Güte der Anpassung mit Hilfe des Bestimmtheitsmaßes!
b) Ermitteln Sie die Halbjahres-Saisonfigur! Wie beurteilen Sie deren Beitrag zur Erklärung des Zeitreihenverlaufs?
c) Geben Sie Prognosewerte für das erste und zweite Halbjahr 2011 ohne und mit Berücksichtigung des Saisoneinflusses an!
d) Halten Sie eine Saisonbereinigung der Absatzzahlen für sinnvoll? (Begründung?)

Aufgabe 3.6
Die Beschäftigtenzahl (Quartalsdurchschnitte) einer Straßenbaufirma entwickelte sich in den Jahren 2011 bis 2013 wie folgt:

Quartal	I/11	II/11	III/11	IV/11	I/12	II/12	III/12	IV/12	I/13	II/13	III/13	IV/13
Beschäftigte	124	172	202	186	144	189	233	206	173	195	252	204

a) Wie viele Mitarbeiter beschäftigte die Firma im Durchschnitt der Jahre 2011 bis 2013, und wie hoch war die Standardabweichung der Quartalswerte?
b) Berechnen Sie die Trendgerade und beurteilen Sie die Güte der Anpassung mit dem Bestimmtheitsmaß!
c) Ermitteln Sie die Saisonfigur und beurteilen Sie deren Beitrag zur Erklärung der Beschäftigungsschwankungen!
d) Erstellen Sie Ex-ante-Prognosen für die Quartale des Jahres 2014!

Aufgabe 3.7
Ein Spielzeughersteller hatte in den Jahren 2011 und 2012 die folgenden vierteljährlichen Umsatzzahlen zu verzeichnen:

Quartal	I/11	II/11	III/11	IV/11	I/12	II/12	III/12	IV/12
Umsatz [1000 €]	143,5	145,0	155,0	194,5	150,5	150,0	161,0	200,5

a) Bestimmen Sie die Trendgerade für die Umsatzentwicklung und erklären Sie, was der geschätzte Koeffizient $\hat{b}$ hier aussagt!
b) Berechnen Sie die konstante Saisonfigur bei additiver Zeitreihenzerlegung!
c) Beurteilen Sie die Erklärungsgüte des additiven Modells der Zeitreihenzerlegung!
d) Um wie viel Prozent ist der saisonbereinigte Umsatz im letzten Quartal 2012 gegenüber dem Vorquartal gestiegen?
e) Berechnen Sie – soweit möglich – die zentrierten gleitenden Vier-Quartale-Durchschnitte der Umsatzzeitreihe!

Aufgabe 3.8

Ein Unternehmen führt für seine Mitarbeiter eine freiwillige Fortbildung durch, die jeweils am Montag-, Mittwoch- und Freitagabend stattfindet. In den ersten vier Wochen ergaben sich folgende Teilnehmerzahlen:

Woche	1	1	1	2	2	2	3	3	3	4	4	4
Tag	Mo	Mi	Fr	Mo	Mi	Fr	Mo	Mi	Fr	Mo	Mi	Fr
Teilnehmerzahl	95	99	55	79	87	51	63	75	39	59	67	35

a) Berechnen Sie die Trendgerade für die Teilnehmerzahl und das dazugehörige Bestimmtheitsmaß!

b) Bestimmen Sie mit Hilfe der geschätzten Trendwerte die „Saisoneinflüsse“ der Wochentage $\hat{s}_{Mo}$, $\hat{s}_{Mi}$ und $\hat{s}_{Fr}$!

c) Zu wie viel Prozent wird die Fluktuation der Teilnehmerzahl durch Trend und Saisonfigur erklärt? Wie beurteilen Sie die Erklärungsgüte des Zeitreihenmodells insgesamt?

d) Bestimmen Sie für die Wochen zwei bis vier die nachlaufenden gleitenden Dreier-Durchschnitte der Teilnehmerzahlen!

e) Prognostizieren Sie mit dem Zerlegungsmodell die Teilnehmerzahlen am Montag, Mittwoch und Freitag der fünften Woche!

Indexzahlen

4

ÜBERBLICK