



# Statistik im Bachelor-Studium der BWL und VWL

Methoden, Anwendung, Interpretation

5., aktualisierte Auflage

Max C. Wewel  
Anja Blatter



Pearson

**EXTRAS**  
**ONLINE**



### 3.1.3 Beispiel: Umsatzprognose

Um die Gemeinsamkeiten bzw. Unterschiede der Regressionsanalyse und Zeitreihenzerlegung zu verdeutlichen, sollen beide Ansätze am gleichen Beispiel einer betrieblichen Umsatzprognose aufgezeigt werden.

#### Beispiel

##### Umsatzprognose

In einem Betrieb soll zu Beginn des Jahres 2024 eine **Ex-ante-Umsatzprognose** für die vier Quartale dieses Jahres erstellt werden. Dabei stehen dem Unternehmen die Quartalsdaten der Umsätze in den zurückliegenden Jahren 2021 bis 2023 zur Verfügung.

Während diese Informationen für eine Zeitreihenanalyse ausreichen, benötigt man für eine Regressionsanalyse weitere Daten, da man den **Umsatz** (= abhängige Variable) durch mindestens ein anderes Merkmal erklären will. Als Hypothese mag dabei die Vorstellung dienen, dass der Umsatz von der Höhe der **Werbeausgaben** (= unabhängige Variable) abhängt. Für diese beiden Merkmale liegen folgende Quartalsdaten vor.

Quartal	Umsatz [Mio. €]	Werbeausgaben [1000 €]
Schäzbereich	beobachtete Werte	
2021 / 1	3,0	8
2021 / 2	2,7	9
2021 / 3	2,5	9
2021 / 4	3,0	10
2022 / 1	3,0	13
2022 / 2	2,9	9
2022 / 3	3,2	11
2022 / 4	3,7	11
2023 / 1	3,6	13
2023 / 2	3,4	12
2023 / 3	3,6	12
2023 / 4	3,8	15
Prognosebereich	Prognose	geplante Werte
2024 / 1	?	19
2024 / 2	?	13
2024 / 3	?	17
2024 / 4	?	21

Tabelle 3.1: Daten zur Umsatzprognose

## 3.2 Regressionsanalyse

### 3.2.1 Problemstellung

Die Regressionsanalyse stellt eine Weiterentwicklung der Korrelationsanalyse dar (vgl. Abschnitt 2.5), wobei hier grundsätzlich vorausgesetzt wird, dass die beiden betrachteten Merkmale  $X$  und  $Y$  **quantitativ** sind. Während die Merkmale in der Korrelationsanalyse völlig gleich behandelt werden, wird in der Regressionsanalyse unterstellt, dass das erste Merkmal ( $X$ ) unabhängig ist und das zweite Merkmal ( $Y$ ) vom ersten Merkmal abhängig ist. Diesen Zusammenhang beschreibt die **Regressionsfunktion**

$$y = f(x).$$

Welches Merkmal unabhängig und welches abhängig ist, richtet sich i.a. nach der vermuteten Kausalität. Sehr oft ist die im Modell **erklärte (abhängige) Variable Y** eine Zielgröße (Gewinn, Umsatz etc.) und die **erklärende (unabhängige) Variable X** eine Instrumentgröße (Absatzpreis, Werbeausgaben etc.). Als Regressionsfunktion wird meist eine lineare Funktion gewählt, weil dies die einfachste Form der Abhängigkeit ist und alle differenzierbaren Funktionen lokal durch eine lineare Funktion angenähert werden können.

Ohnehin haben Regressionsfunktionen immer nur approximative Charakter, weil sich Zusammenhänge zwischen ökonomischen Merkmalen, die i.d.R. das Verhalten von Wirtschaftssubjekten (Produzenten, Konsumenten etc.) wiederspiegeln, naturgemäß nicht durch einfache Funktionen exakt beschreiben lassen. Dies bedeutet, dass beim Einsetzen von tatsächlichen zweidimensionalen Beobachtungswerten

$$(x_t, y_t) \quad (t = 1, 2, \dots) \tag{3.1}$$

– seien sie aus dem Schätz- oder Prognosebereich – fast immer eine **Abweichung** (bzw. ein **Fehler** oder eine **Störung**)  $u_t$  zwischen dem Beobachtungswert der abhängigen Variablen  $y_t$  und dem Funktionswert  $f(x_t)$  auftritt. Unter Berücksichtigung dieser empirischen Abweichung lautet das **allgemeine Regressionsmodell**

$$y_t = f(x_t) + u_t \quad (t = 1, 2, \dots) \tag{3.2}$$

bzw. das **lineare Regressionsmodell**

$$y_t = a + b x_t + u_t \quad (t = 1, 2, \dots). \tag{3.3}$$

Im Rahmen einer Regressionsanalyse sind nun im wesentlichen zwei Probleme zu behandeln:

- **Schätzproblem:** Wie kann die Regressionsfunktion bzw. – im linearen Fall – wie können die beiden Regressionskoeffizienten  $a$  und  $b$  optimal passend zu den vorliegenden Beobachtungswerten des Schätzbereichs numerisch bestimmt werden?
- **Beurteilungsproblem:** Wie lässt sich die Aussagefähigkeit der so geschätzten Regressionsfunktion und damit die Güte der aus ihr abgeleiteten Prognosen beurteilen?

### 3.2.2 Bestimmung der Regressionskoeffizienten

Empirischer Ausgangspunkt des Schätzproblems der Regressionsanalyse sind die zweidimensionalen Beobachtungswerte des Schätzbereichs

$$(x_t, y_t) \quad (t = 1, \dots, n),$$

welche in einem Streudiagramm veranschaulicht werden können. Die Regressionsfunktion soll den unterstellten linearen Zusammenhang zwischen diesen Beobachtungswerten der beiden Merkmale  $X$  und  $Y$  „optimal“ beschreiben.

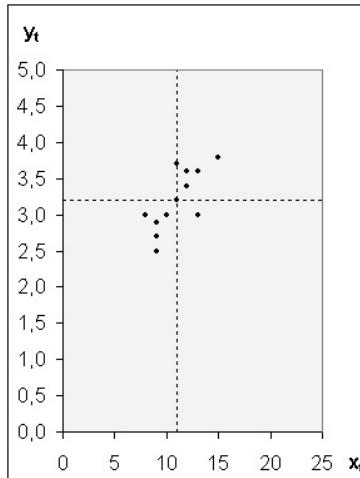


Abbildung 3.2: Streudiagramm zur Umsatzprognose (vgl. Tabelle 3.1)

Als Optimalitätskriterium wird das sog. **Kleinste-Quadrat-Prinzip** (oder kurz: KQ-Prinzip) angewandt, nach dem die Summe der quadrierten Fehler im Schätzbereich minimiert wird. Für das lineare Regressionsmodell bedeutet das:

$$\sum_{t=1}^n u_t^2 = \sum_{t=1}^n (y_t - f(x_t))^2 =: q(a, b) \rightarrow \min! \quad (3.4)$$

Im Streudiagramm wird also diejenige Gerade gesucht, für die die Summe der quadrierten senkrechten Abstände zu den zweidimensionalen Beobachtungswerten am kleinsten ist.

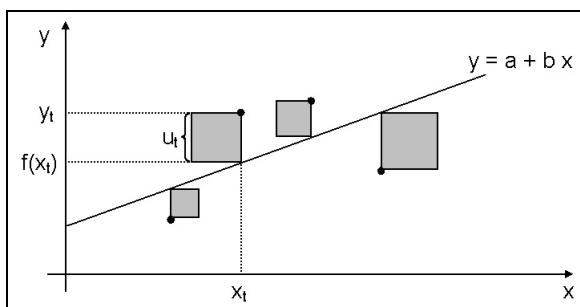


Abbildung 3.3: Kleinste-Quadrat-Prinzip

### Exkurs

## Ableitung der Kleinst-Quadrat-Schätzungen

Gesucht wird das lokale (und zugleich globale) Minimum der Funktion

$$q(a, b) = \sum_{t=1}^n (a + b x_t - y_t)^2.$$

Die **notwendigen Bedingungen** für ein lokales Extremum sind die **Normalgleichungen**:

$$q_a(\hat{a}, \hat{b}) = 2 \sum_{t=1}^n (\hat{a} + \hat{b} x_t - y_t) = 2 \left( n \hat{a} + \hat{b} \sum_{t=1}^n x_t - \sum_{t=1}^n y_t \right) \stackrel{!}{=} 0$$

$$\Leftrightarrow \hat{a} + \hat{b} \bar{x} - \bar{y} = 0 \Leftrightarrow \hat{a} = \bar{y} - \hat{b} \bar{x}$$

$$q_b(\hat{a}, \hat{b}) = 2 \sum_{t=1}^n (\hat{a} + \hat{b} x_t - y_t) x_t = 2 \left( \hat{a} \sum_{t=1}^n x_t + \hat{b} \sum_{t=1}^n x_t^2 - \sum_{t=1}^n x_t y_t \right) \stackrel{!}{=} 0$$

$$\Leftrightarrow \underbrace{(\bar{y} - \hat{b} \bar{x})}_{\hat{a}} \bar{x} + \hat{b} \frac{1}{n} \sum_{t=1}^n x_t^2 - \frac{1}{n} \sum_{t=1}^n x_t y_t = 0$$

$$\Leftrightarrow \hat{b} \left( \frac{1}{n} \sum_{t=1}^n x_t^2 - \bar{x}^2 \right) = \frac{1}{n} \sum_{t=1}^n x_t y_t - \bar{x} \bar{y} \Leftrightarrow \hat{b} = \frac{\sigma_{xy}}{\sigma_x^2}.$$

Die **hinreichenden Zusatzbedingungen** für ein lokales Minimum sind erfüllt:

$$q_{aa}(\hat{a}, \hat{b}) = 2n > 0 \quad \text{und} \quad q_{bb}(\hat{a}, \hat{b}) = 2 \sum_{t=1}^n x_t^2 > 0 \quad \text{sowie}$$

$$q_{aa}(\hat{a}, \hat{b}) q_{bb}(\hat{a}, \hat{b}) - (q_{ab}(\hat{a}, \hat{b}))^2 = 4n \sum_{t=1}^n x_t^2 - \left( 2 \sum_{t=1}^n x_t \right)^2 = 4n^2 \left( \frac{1}{n} \sum_{t=1}^n x_t^2 - \bar{x}^2 \right) \\ = 4n^2 \sigma_x^2 > 0.$$

Nach dem Kleinst-Quadrat-Prinzip lautet die **geschätzte Regressionsgerade** also

$$\hat{y}_t = \hat{a} + \hat{b} x_t \tag{3.5}$$

mit den Regressionskoeffizienten

$$\hat{b} = \frac{\sigma_{xy}}{\sigma_x^2} \tag{3.6}$$

und

$$\hat{a} = \bar{y} - \hat{b} \bar{x}. \tag{3.7}$$

Die Schätzformel (3.6) für den Anstieg der Regressionsgerade zeigt folgenden Zusammenhang zur Korrelation der Merkmale  $X$  und  $Y$ :

<b>steigende</b> Regressionsgerade	$\Leftrightarrow$	<b>positive</b> Korrelation
<b>fallende</b> Regressionsgerade	$\Leftrightarrow$	<b>negative</b> Korrelation
<b>horizontale</b> Regressionsgerade	$\Leftrightarrow$	<b>keine</b> Korrelation

Ferner erkennt man durch Auflösen der Schätzformel (3.7) nach  $\bar{y}$ , dass die geschätzte Regressionsgerade durch den **Schwerpunkt**  $(\bar{x}, \bar{y})$  geht.

Setzt man nun in die geschätzte Regressionsgerade (3.5) bekannte bzw. angenommene Werte  $x_t$  der unabhängigen Variablen ein, so erhält man **bedingte Prognosewerte**  $\hat{y}_t$  für die abhängige Variable. Soweit es sich um Ex-post-Prognosen handelt, können diese mit den (bei der Schätzung benutzten) Beobachtungswerten verglichen werden. Die dabei auftretenden Differenzen heißen **Ex-post-Prognosefehler** oder kurz **Residuen**:

$$\hat{u}_t = y_t - \hat{y}_t \quad (t = 1, \dots, n). \quad (3.8)$$

Für die **Ex-post-Prognosewerte**  $\hat{y}_t$  ( $t = 1, \dots, n$ ) gilt im linearen Regressionsmodell aufgrund der o.g. Schwerpunkt-Eigenschaft:

$$\bar{\hat{y}} = \frac{1}{n} \sum_{t=1}^n \hat{y}_t = \frac{1}{n} \sum_{t=1}^n (\hat{a} + \hat{b} x_t) = \hat{a} + \hat{b} \bar{x} \stackrel{(3.7)}{=} \bar{y},$$

d.h. die Ex-post-Prognosewerte  $\hat{y}_t$  ( $t = 1, \dots, n$ ) entsprechen im Mittel den Beobachtungswerten  $y_t$  ( $t = 1, \dots, n$ ). Es treten also **keine systematischen Prognosefehler** auf. Für die Residuen gilt:

$$\bar{\hat{u}} = \bar{y} - \bar{\hat{y}} = 0.$$

### Beispiel

#### Umsatzprognose

Aus den Daten in Tabelle 3.1 soll die künftige Umsatzentwicklung des Unternehmens mit Hilfe eines Regressionsmodells prognostiziert werden, bei dem unterstellt wird, dass der **Umsatz des Unternehmens (Y)** näherungsweise eine lineare Funktion der im gleichen Zeitraum getätigten **Werbeausgaben (X)** ist. Zur Schätzung der Regressionskoeffizienten dient die folgende Arbeitstabelle.

Quartal	$t$	$x_t$	$x_t - \bar{x}$	$(x_t - \bar{x})^2$	$y_t$	$y_t - \bar{y}$	$(x_t - \bar{x})(y_t - \bar{y})$
2021 / 1	1	8	-3	9	3,0	-0,2	0,6
2021 / 2	2	9	-2	4	2,7	-0,5	1,0
2021 / 3	3	9	-2	4	2,5	-0,7	1,4
2021 / 4	4	10	-1	1	3,0	-0,2	0,2
2022 / 1	5	13	2	4	3,0	-0,2	-0,4
2022 / 2	6	9	-2	4	2,9	-0,3	0,6
2022 / 3	7	11	0	0	3,2	0	0
2022 / 4	8	11	0	0	3,7	0,5	0
2023 / 1	9	13	2	4	3,6	0,4	0,8
2023 / 2	10	12	1	1	3,4	0,2	0,2
2023 / 3	11	12	1	1	3,6	0,4	0,4
2023 / 4	12	15	4	16	3,8	0,6	2,4
---	---	132	0	48	38,4	0	7,2

Tabelle 3.2: Schätzung der Regressionsgerade nach dem Kleinst-Quadrate-Prinzip

$$\text{Man erhält: } \bar{x} = \frac{132}{12} = 11 [10^3] \quad \sigma_x^2 = \frac{48}{12} = 4 [10^6]$$

$$\bar{y} = \frac{38,4}{12} = 3,2 [10^6] \quad \sigma_{xy} = \frac{7,2}{12} = 0,6 [10^9]$$

$$\text{und somit: } \hat{b} = \frac{0,6}{4} = 0,15 [10^3] \quad \hat{a} = 3,2 - 0,15 \cdot 11 = 1,55 [10^6]$$

Die **geschätzte Regressionsgerade** lautet also:  $\hat{y}_t = 1,55 + 0,15 x_t$ .

Die Regressionskoeffizienten  $\hat{b}$  und  $\hat{a}$  lassen sich hier so interpretieren, dass ...

- jeder zusätzlich für Werbung ausgegebene Euro den Umsatz um ca. 150 erhöht und
- ohne Werbeausgaben nur mit einem Quartalsumsatz von 1,55 Mio. zu rechnen ist.

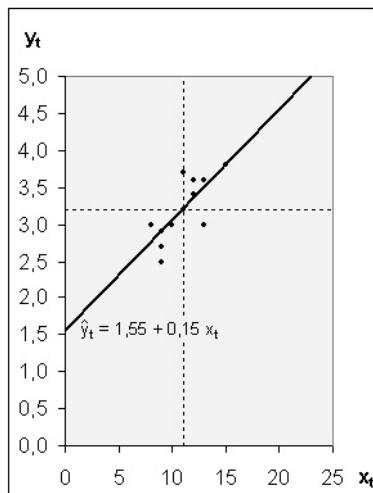


Abbildung 3.4: Regressionsgerade im Streudiagramm

Die geschätzte Regressionsgerade kann nun zu **bedingten Prognosen** genutzt werden. Für die ersten vier Quartale im Prognosebereich (vgl. Tabelle 3.1) ergeben sich die folgenden Ex-ante-Prognosewerte.

Quartal	$t$	$x_t$	$\hat{y}_t$
2024 / 1	13	19	4,4
2024 / 2	14	13	3,5
2024 / 3	15	17	4,1
2024 / 4	16	21	4,7

Tabelle 3.3: Ex-ante-Prognosen mit dem Regressionsmodell

### 3.2.3 Beurteilung des Regressionsmodells

Nach dem Kleinst-Quadrat-Prinzip kann gemäß den Formeln (3.6) und (3.7) immer eine optimale lineare Regressionsfunktion bestimmt werden, selbst wenn aus dem Streudiagramm klar hervorgeht, dass zwischen den Beobachtungswerten der beiden Merkmale,  $x_t$  und  $y_t$ , überhaupt kein Zusammenhang besteht. Daher ist es notwendig, das Schätzergebnis einer kritischen Prüfung zu unterziehen.

Die geschätzte Regressionsgerade und die daraus abgeleiteten Prognosen sind offenbar **umso zuverlässiger, je kleiner die Residuen** betragsmäßig sind. Ein Maß für die absolute Größe der Residuen ist deren Varianz:

$$\sigma_{\hat{u}}^2 = \frac{1}{n} \sum_{t=1}^n \hat{u}_t^2 .$$

Sie hängt allerdings von der Dimension der abhängigen Variablen  $Y$  ab und muss noch geeignet normiert werden. Hierzu bietet sich die im linearen Regressionsmodell allgemein gültige **Streuungszerlegung** an:

$$\sigma_y^2 = \sigma_{\hat{y}}^2 + \sigma_{\hat{u}}^2 . \quad (3.9)$$

Diese Beziehung besagt, dass die **Varianz der Beobachtungswerte**  $\sigma_y^2$  additiv zerlegt werden kann in

- die **Varianz der Ex-post-Prognosewerte**  $\sigma_{\hat{y}}^2$  (= erklärte Streuung) und
- die **Varianz der Residuen**  $\sigma_{\hat{u}}^2$  (= Reststreuung).

Die Varianz der Ex-post-Prognosewerte  $\sigma_{\hat{y}}^2$  wird als **erklärte Streuung** bezeichnet, weil sie derjenige Teil der Streuung der  $y_t$ -Werte ist, der aufgrund der Regressionsgerade aus der Streuung der  $x_t$ -Werte resultiert oder – anders ausgedrückt – mit der Streuung der  $x_t$ -Werte erklärt werden kann. Es gilt der folgende in Abbildung 3.5 verdeutlichte Zusammenhang:

$$\sigma_{\hat{y}}^2 = \hat{b}^2 \sigma_x^2 \stackrel{(3.6)}{=} \left( \frac{\sigma_{xy}}{\sigma_x^2} \right)^2 \sigma_x^2 = \frac{\sigma_{xy}^2}{\sigma_x^2} .$$

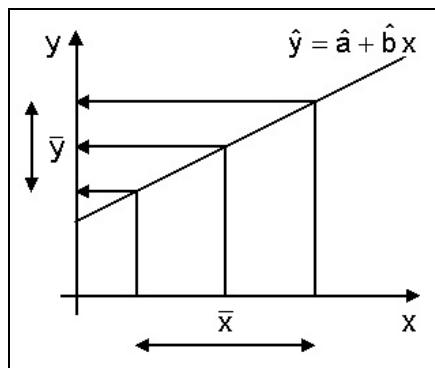


Abbildung 3.5: Streuungserklärung im Regressionsmodell

Im günstigsten Fall liegt ein **perfekter linearer Zusammenhang** zwischen den  $x_t$ - und  $y_t$ -Werten vor (vgl. Abbildung 3.6). Dann liegen alle zweidimensionalen Beobachtungswerte auf der geschätzten Regressionsgerade. Die Ex-post-Prognosewerte  $\hat{y}_t$  sind mit den Beobachtungswerten  $y_t$  identisch. Für die Varianzen in der Streuungszerlegung gilt somit:

$$\sigma_{\hat{y}}^2 = \sigma_y^2 \quad \text{und} \quad \sigma_{\hat{u}}^2 = 0.$$

In diesem Fall wird die Streuung der  $y_t$ -Werte mit dem Regressionsmodell also **vollständig** durch die Streuung der  $x_t$ -Werte erklärt.

Umgekehrt ist die Situation, wenn überhaupt **kein linearer Zusammenhang** erkennbar ist (vgl. Abbildung 3.7). Im ungünstigsten Fall völlig unkorrelierter  $x_t$ - und  $y_t$ -Werte verläuft die Regressionsgerade horizontal. Alle Ex-post-Prognosewerte  $\hat{y}_t$  entsprechen dann dem arithmetischen Mittel  $\bar{y}$  und es gilt für die Varianzen in der Streuungszerlegung:

$$\sigma_{\hat{y}}^2 = 0 \quad \text{und} \quad \sigma_{\hat{u}}^2 = \sigma_y^2,$$

d.h. die Streuung der  $y_t$ -Werte bleibt **vollständig unerklärt**, weil sie nicht auf die Streuung der  $x_t$ -Werte zurückgeführt werden kann.

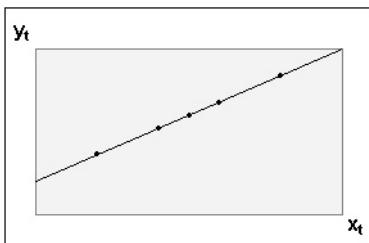


Abbildung 3.6: Perfekter linearer Zusammenhang

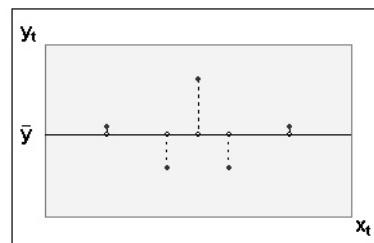


Abbildung 3.7: Kein linearer Zusammenhang

Als Ergebnis dieser Überlegungen ist festzuhalten, dass ein Regressionsmodell umso besser ist, je größer der Anteil der erklärten Streuung an der Gesamtstreuung ist. Dieser Anteil heißt **Bestimmtheitsmaß** und entspricht dem Quadrat des Korrelationskoeffizienten nach Bravais/Pearson:

$$r^2 = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2} = \frac{\sigma_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2} = 1 - \frac{\sigma_{\hat{u}}^2}{\sigma_y^2}. \quad (3.10)$$

Das Bestimmtheitsmaß ist auf das Einheitsintervall normiert:

$$0 \leq r^2 \leq 1. \quad (3.11)$$

Als sehr grobe Orientierung für die Interpretation mag die folgende **Faustregel** dienen:

$0 \leq r^2 < \frac{1}{3}$	$\rightarrow$ kein linearer Zusammenhang zwischen X und Y
$\frac{1}{3} \leq r^2 \leq \frac{2}{3}$	$\rightarrow$ schwach ausgeprägter linearer Zusammenhang zwischen X und Y
$\frac{2}{3} < r^2 \leq 1$	$\rightarrow$ stark ausgeprägter linearer Zusammenhang zwischen X und Y.

### Beispiel

#### Umsatzprognose

In Fortführung des Beispiels aus Tabelle 3.1 werden zur Beurteilung des Regressionsmodells die Ex-post-Prognosewerte und Residuen sowie die Komponenten der **Streuungszerlegung** ermittelt (vgl. Tabelle 3.4). Dabei ergibt sich wegen

$$\sigma_y^2 = \frac{1,92}{12} = 0,16 [10^{12-2}] \quad \sigma_{\hat{y}}^2 = \frac{1,08}{12} = 0,09 [10^{12-2}] \quad \sigma_{\hat{u}}^2 = \frac{0,84}{12} = 0,07 [10^{12-2}]$$

$$\text{das Bestimmtheitsmaß} \quad r^2 = \frac{0,09}{0,16} = 0,5625,$$

d.h. **56,25 % der Streuung des Umsatzes können** mit dem Regressionsmodell durch die Streuung der Werbeausgaben **erklärt werden**. Somit ist der unterstellte lineare Zusammenhang nur mäßig stark ausgeprägt.

Man sollte daher versuchen, das Regressionsmodell z.B. durch Verwendung eines anderen Funktionstyps oder durch zusätzliche erklärende Variablen zu verbessern (nicht-lineares bzw. multiples Regressionsmodell).

Quartal	t	y <sub>t</sub>	ŷ <sub>t</sub>	û <sub>t</sub>	(y <sub>t</sub> - ŷ <sub>t</sub> ) <sup>2</sup>	(ŷ <sub>t</sub> - ŷ̄) <sup>2</sup>	û <sub>t</sub> <sup>2</sup>
2021 / 1	1	3,0	2,75	0,25	0,04	0,2025	0,0625
2021 / 2	2	2,7	2,9	-0,2	0,25	0,09	0,04
2021 / 3	3	2,5	2,9	-0,4	0,49	0,09	0,16
2021 / 4	4	3,0	3,05	-0,05	0,04	0,0225	0,0025
2022 / 1	5	3,0	3,5	-0,5	0,04	0,09	0,25
2022 / 2	6	2,9	2,9	0	0,09	0,09	0
2022 / 3	7	3,2	3,2	0	0	0	0
2022 / 4	8	3,7	3,2	0,5	0,25	0	0,25
2023 / 1	9	3,6	3,5	0,1	0,16	0,09	0,01
2023 / 2	10	3,4	3,35	0,05	0,04	0,0225	0,0025
2023 / 3	11	3,6	3,35	0,25	0,16	0,0225	0,0625
2023 / 4	12	3,8	3,8	0	0,36	0,36	0
---	---	38,4	38,4	0	1,92	1,08	0,84

Tabelle 3.4: Ex-post-Prognosewerte, Residuen und Streuungszerlegung

## 3.3 Zeitreihenanalyse

### 3.3.1 Problemstellung

Eine **Zeitreihe** ist eine Folge von zeitlich hintereinander, meist in regelmäßigen Abständen bei demselben Merkmalsträger erhobenen Beobachtungswerten eines Merkmals:

$$y_t \quad (t = 1, \dots, n, n+1, n+2, \dots). \quad (3.12)$$

Typische Beispiele für Zeitreihen sind

- die täglichen Schlusskurse eines börsengehandelten Wertpapiers,
- die monatlich festgestellten Arbeitslosenzahlen in Deutschland,
- die Quartalsumsätze eines Unternehmens oder
- die jährlichen Produktionsmengen eines Stahlwerks.

Üblicherweise werden solche Zeitreihen in einem **Zeitreihendiagramm** grafisch dargestellt, wobei die lineare Verbindung aufeinander folgender Werte nur der besseren Veranschaulichung dient.

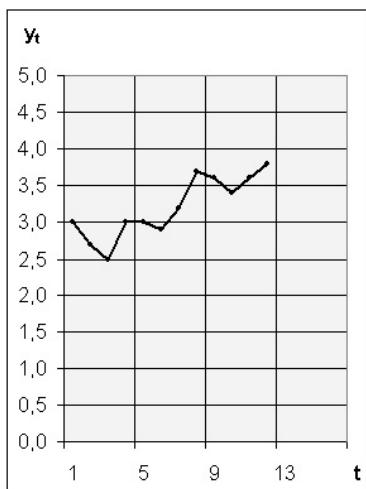


Abbildung 3.8: Zeitreihendiagramm der Umsatzreihe (Daten aus Tabelle 3.1)

In der **Zeitreihenanalyse** werden Zeitreihen auf Gesetzmäßigkeiten untersucht, die sich aus der zeitlichen Abfolge der Beobachtungswerte ergeben. Die einfachsten Verfahren zielen darauf ab, den Zeitreihenverlauf – so gut es geht – auf **systematische Komponenten** wie Trend und Saisoneinflüsse zurückzuführen. Neben den systematischen Komponenten wird noch eine **Restkomponente** berücksichtigt, in der alle nicht-systematischen („zufälligen“) Einflüsse auf die Zeitreihe zusammengefasst werden.

Das **additive Modell der Zeitreihenzerlegung** lautet:

$$y_t = g_t + s_t + r_t \quad (t = 1, 2, \dots), \quad (3.13)$$

wobei die Zeitreihe  $y_t$  in

- eine **glatte Komponente (Trend)**  $g_t$ ,
- eine **zyklische Komponente (z.B. Saisonfigur)**  $s_t$  und
- eine **Restkomponente**  $r_t$

zerlegt wird.

Wie in der Regressionsanalyse sind auch bei der Zeitreihenanalyse zwei Probleme zu behandeln:

- **Schätzproblem:** Wie lassen sich die systematischen Komponenten  $g_t$  und  $s_t$  aus den vorliegenden Zeitreihenwerten schätzen?
- **Beurteilungsproblem:** Wie lässt sich die Güte bzw. Aussagefähigkeit der Zeitreihenzerlegung beurteilen?

### 3.3.2 Bestimmung der glatten Komponente

Bei der Zeitreihenzerlegung beginnt man mit der glatten Komponente, indem man einen bestimmten Funktionstyp für die Trendfunktion unterstellt. Die einfachste Annahme ist die einer **linearen Trendfunktion**

$$g_t = a + bt .$$

Durch Einsetzen in die Modellgleichung (3.13) erhält man die Gleichung

$$y_t = a + bt + (s_t + r_t) ,$$

die dem linearen Regressionsmodell entspricht, wenn man die Werte der unabhängigen Variablen  $x_t$  durch die Zeitperioden  $t$  und die Störungen  $u_t$  durch die kombinierte Saison- und Restkomponente  $s_t + r_t$  ersetzt. Demnach können die Koeffizienten  $a$  und  $b$  der linearen Trendfunktion wieder nach dem **Kleinst-Quadrat-Prinzip** geschätzt werden:

$$\sum_{t=1}^n (s_t + r_t)^2 = \sum_{t=1}^n (y_t - a - bt)^2 =: q(a, b) \rightarrow \min!$$

Das Ergebnis ist die **geschätzte Trendgerade**

$$\hat{g}_t = \hat{a} + \hat{b} t \quad (3.14)$$

mit den analog zu (3.6) und (3.7) bestimmten Koeffizienten

$$\hat{b} = \frac{\sigma_{ty}}{\sigma_t^2} \quad (3.15)$$

und  $\hat{a} = \bar{y} - \hat{b} \frac{n+1}{2} . \quad (3.16)$

Da die Werte von  $t$  die Zahlen  $1, \dots, n$  sind, gilt bei der Berechnung:

$$\bar{t} = \frac{1}{n} \sum_{t=1}^n t = \frac{n+1}{2}$$

# Copyright

Daten, Texte, Design und Grafiken dieses eBooks, sowie die eventuell angebotenen eBook-Zusatzdaten sind urheberrechtlich geschützt. Dieses eBook stellen wir lediglich als **persönliche Einzelplatz-Lizenz** zur Verfügung!

Jede andere Verwendung dieses eBooks oder zugehöriger Materialien und Informationen, einschließlich

- der Reproduktion,
- der Weitergabe,
- des Weitervertriebs,
- der Platzierung im Internet, in Intranets, in Extranets,
- der Veränderung,
- des Weiterverkaufs und
- der Veröffentlichung

bedarf der **schriftlichen Genehmigung** des Verlags. Insbesondere ist die Entfernung oder Änderung des vom Verlag vergebenen Passwort- und DRM-Schutzes ausdrücklich untersagt!

Bei Fragen zu diesem Thema wenden Sie sich bitte an: **info@pearson.de**

## Zusatzdaten

Möglicherweise liegt dem gedruckten Buch eine CD-ROM mit Zusatzdaten oder ein Zugangscode zu einer eLearning Plattform bei. Die Zurverfügungstellung dieser Daten auf unseren Websites ist eine freiwillige Leistung des Verlags. **Der Rechtsweg ist ausgeschlossen.** Zugangscodes können Sie darüberhinaus auf unserer Website käuflich erwerben.

## Hinweis

Dieses und viele weitere eBooks können Sie rund um die Uhr und legal auf unserer Website herunterladen:

**<https://www.pearson-studium.de>**



Pearson