

Übungen zur Statistik für Wirtschaftswissenschaftler

Bearbeitet von
Von Prof. Dr. Josef Bleymüller, Prof. Dr. Rafael Weißbach, und Dr. Achim Dörre

1. Auflage 2018. Buch. IX, 178 S. Kartoniert
ISBN 978 3 8006 5873 2
Format (B x L): 16,0 x 24,0 cm
Gewicht: 354 g

[Wirtschaft > Volkswirtschaft > Wirtschaftsstatistik, Demographie](#)

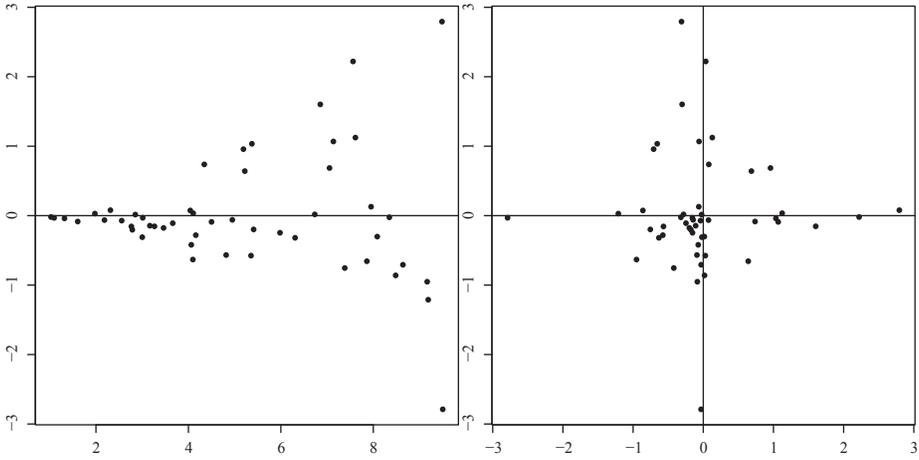
Zu [Inhaltsverzeichnis](#)

schnell und portofrei erhältlich bei

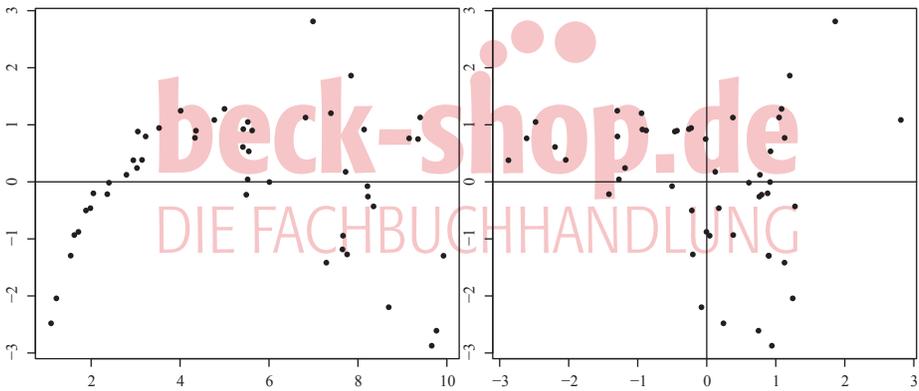
**beck-shop.de**
DIE FACHBUCHHANDLUNG

Die Online-Fachbuchhandlung beck-shop.de ist spezialisiert auf Fachbücher, insbesondere Recht, Steuern und Wirtschaft. Im Sortiment finden Sie alle Medien (Bücher, Zeitschriften, CDs, eBooks, etc.) aller Verlage. Ergänzt wird das Programm durch Services wie Neuerscheinungsdienst oder Zusammenstellungen von Büchern zu Sonderpreisen. Der Shop führt mehr als 8 Millionen Produkte.

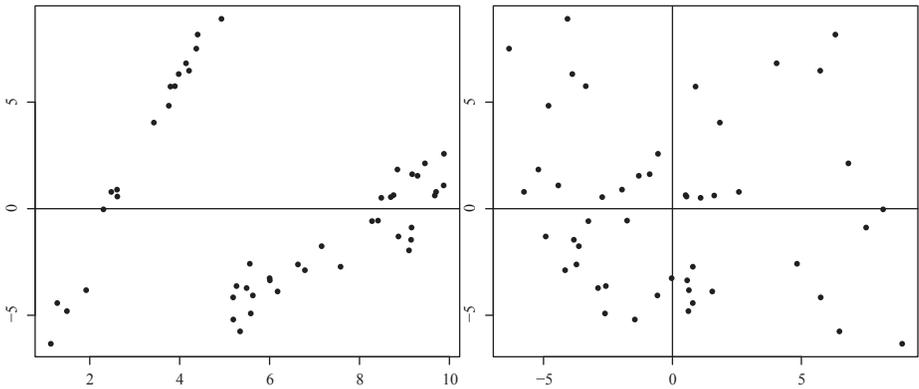
Situation 1:



Situation 2:



Situation 3:



Beurteilen Sie für jede der drei Regressionsanalysen grafisch, ob eine Annahmeverletzung (Heteroskedastizität, Fehlspezifikation, Strukturbruch) zu erkennen ist.

Aufgabe 22.5

Die folgende Tabelle gibt die installierte Leistung der Photovoltaikanlagen in Niedersachsen im Verlauf von 2004 bis 2015 wieder.

Jahr	Leistung (MWp)
2004	43
2005	109
2006	174
2007	256
2008	352
2009	709
2010	1512
2011	2286
2012	3042
2013	3316
2014	3461
2015	3560

- Bestimmen Sie die lineare Regressionsfunktion für die Entwicklung der installierten Leistung im Zeitverlauf. Nummerieren Sie zur Vereinfachung der Berechnungen die angegebenen Jahre x von 1 bis 12.
- Berechnen Sie die Residuen und zeichnen Sie ein Streuungsdiagramm der Wertepaare (x_i, e_i) . Erkennen Sie Hinweise auf einen oder mehrere Strukturbrüche?
- Zerlegen Sie den Datensatz in drei Teildatensätze gemäß der Jahreseinteilung 2004–2008, 2009–2012 und 2013–2015, und bestimmen Sie für jeden Teildatensatz eine lineare Regressionsfunktion. Zeichnen Sie das Streuungsdiagramm für den gesamten Datensatz und ergänzen Sie die drei geschätzten Regressionsfunktionen.
- Bestimmen Sie eine Punktprognose für die installierte Leistung der Photovoltaikanlagen für das Jahr 2018.

Lösungen

Lösung zu Aufgabe 22.1

- Da der monatliche Umsatz für *eine* weitere Stadt von Interesse ist, wird eine Prognose des individuellen Wertes y_0 bestimmt. Die Punktprognose lautet

$$\hat{y}_0 = 13 - x_0 = 13 - 5 = 8.$$

Der Erwartungswert des monatlichen Umsatzes in der zusätzlichen Stadt wird auf 8 Tsd. Euro geschätzt.

Das symmetrische Prognoseintervall wird darüber hinaus mittels

$$\hat{y}_0 - ts_F \leq y_0 \leq \hat{y}_0 + ts_F,$$

gebildet, wobei t das passende Quantil der Studentverteilung mit $\nu = n - 2 = 3$ Freiheitsgraden und s_F den Standardfehler des individuellen Prognosewerts bezeichnet. Es wird darüber hinaus die geschätzte Standardabweichung der Residuen $s_E = \sqrt{2}$ benötigt (siehe Aufgabe 21.2 (a)). Es gilt $t = t_{0,975;3} = 3,182$ und

$$s_F = s_E \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = \sqrt{2} \cdot \sqrt{1 + \frac{1}{5} + \frac{1}{40}} = \sqrt{2} \cdot \sqrt{1,225} \approx 1,57$$

Daher lautet das Prognoseintervall für den individuellen Prognosewert y_0 :

$$\begin{aligned} 8 - 3,182 \cdot 1,57 &\leq y_0 \leq 8 + 3,182 \cdot 1,57 \\ 3,02 &\leq y_0 \leq 12,98 \end{aligned}$$

Für den monatlichen Umsatz in einer Stadt bei einem Preis von $x_0 = 5$ wird der Bereich 3,02 bis 12,98 zum Konfidenzniveau von 95 % prognostiziert.

- (b) Da sich die Prognose in dieser Teilaufgabe auf *einige weitere Städte* bezieht, ist hier die Prognose für den durchschnittlichen monatlichen Umsatz pro Stadt von Interesse. Das Konfidenzintervall wird allgemein gebildet als

$$\hat{y}_0 - ts_{\hat{y}_0} \leq E(Y_0) \leq \hat{y}_0 + ts_{\hat{y}_0},$$

wobei $s_{\hat{y}_0}$ den geschätzten Standardfehler des durchschnittlichen Prognosewerts bezeichnet. Der Wert t ist wie in der vorigen Aufgabe das passende Quantil der Studentverteilung mit $\nu = n - 2 = 3$ Freiheitsgraden, d. h. $t = 3,182$. Der Standardfehler wird geschätzt mittels

$$s_{\hat{y}_0} = \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \approx 0,67.$$

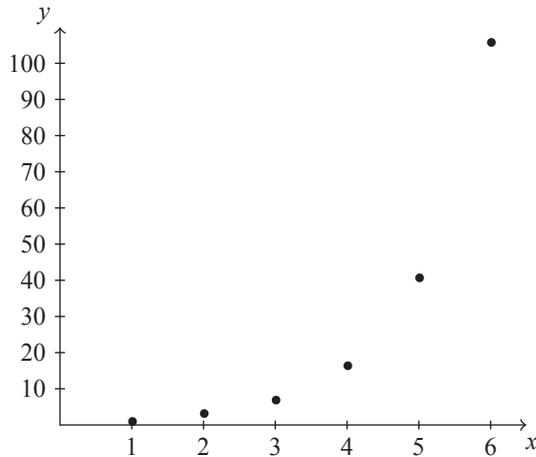
Das Konfidenzintervall lautet somit:

$$\begin{aligned} 8 - 3,182 \cdot 0,67 &\leq E(Y_0) \leq 8 + 3,182 \cdot 0,67 \\ 5,87 &\leq E(Y_0) \leq 10,13 \end{aligned}$$

Für den durchschnittlichen monatlichen Umsatz je Stadt wird zum Konfidenzniveau 95 % ein Bereich von 5,87 bis 10,13 prognostiziert.

Lösung zu Aufgabe 22.2

(a) Streuungsdiagramm:



Bestimmung der Regressionskoeffizienten:

$$b_2 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{153,58 - 3,5 \cdot 28,5}{15,167 - 12,25} \approx 18,457$$

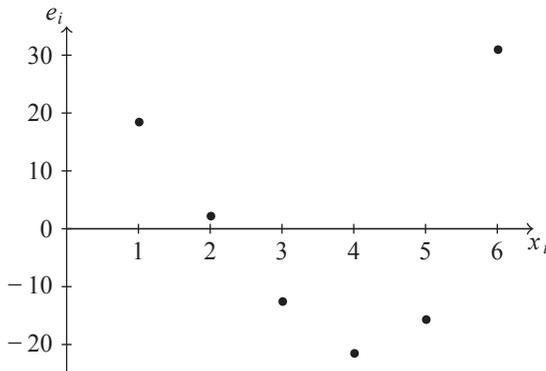
$$b_1 = \bar{y} - b_2 \bar{x} = 28,5 - 18,457 \cdot 3,5 \approx -36,1$$

(b) Bestimmung der Residuen $e_i = y_i - \hat{y}_i$:

Monat x_i	abgesetzte Menge y_i (in Tsd. Stück)	\hat{y}_i	e_i
1	0,5	-17,64	18,14
2	2,7	0,81	1,89
3	6,4	19,27	-12,87
4	15,9	37,73	-21,83
5	40,2	56,19	-15,99
6	105,3	74,64	30,66

Die Residuen besitzen offenbar systematisches Verhalten, was u.a. an der geringen Anzahl von Vorzeichenwechseln erkennbar ist. Ideal wäre dagegen ein zufälliges Streuen der Residuen um den Wert Null.

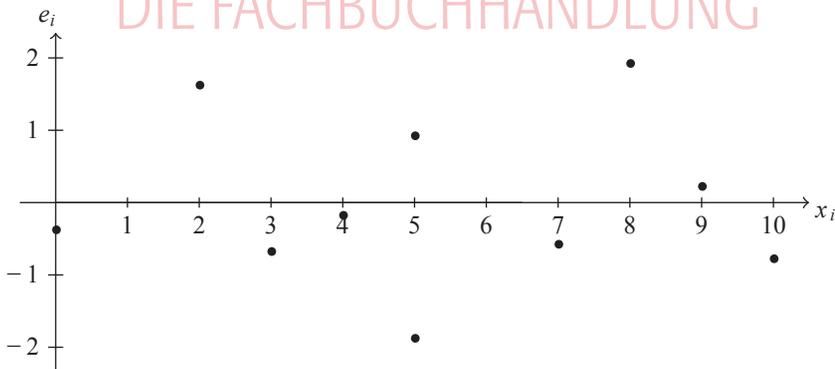
Grafisch bestätigt sich dieser Eindruck:



Abgesehen vom relativ geringen Stichprobenumfang in diesem Beispiel kann die Verwendung einer linearen Regressionsfunktion nicht als sinnvoll eingestuft werden. Das Streuungsdiagramm zeigt, dass stattdessen eher ein exponentieller Zusammenhang zwischen x und y plausibel ist. Ein Vorteil einer solchen Betrachtung bestünde darin, dass die geschätzten y -Werte der Regressionsfunktion dann allesamt nichtnegativ wären, was im beschriebenen Kontext, da es sich bei den Werten von y um abgesetzte Mengen handelt, inhaltlich erforderlich ist.

Lösung zu Aufgabe 22.3

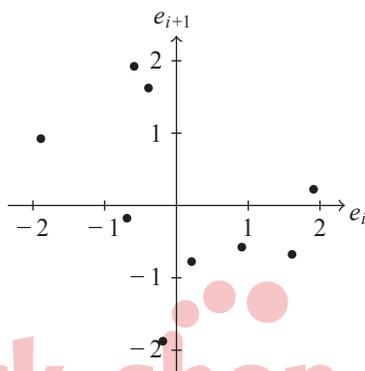
(a) Streuungsdiagramm der Residuen:



Für die zufälligen Störterme wird grundsätzlich angenommen, dass diese unverzerrt sind, keine Autokorrelation aufweisen und dieselbe Varianz besitzen (Homoskedastizität).

In der vorliegenden Situation streuen die Residuen ohne erkennbare Systematik um den Wert Null. Da die Intensität der Streuung bzw. die Distanz der Residuen zum Wert Null weder steigend noch fallend ist, gibt es kein direktes Indiz für Heteroskedastizität. Die Annahme der unverzerrten Störterme, d. h. dass diese im Erwartungswert Null sind, kann im Rahmen einer Regressionsanalyse nicht unmittelbar untersucht werden, da der Mittelwert der Residuen bei der Methode der kleinsten Quadrate generell bei Null liegt. Es gibt keine unmittelbar erkennbaren Verstöße gegen die Annahmen der Regressionsanalyse.

- (b) Um die Annahme der Autokorrelation näher zu untersuchen, wird ein Streuungsdiagramm der paarweise zugeordneten Residuen (e_i, e_{i+1}) erstellt:



Besonderes Augenmerk wird auf das gleichzeitige Vorliegen identischer Vorzeichen der Residuenpaare gelegt. So liegt z. B. das 8. Residuenpaar $(e_8, e_9) = (1,9; 0,2)$ im ersten Quadranten des Diagramms, d. h. auf das positive Residuum e_8 folgt ein ebenso positives Residuum e_9 . Dies spricht tendenziell für positive Autokorrelation. Die Residuenpaare $(e_2, e_3) = (1,6; -0,7)$, $(e_5, e_6) = (0,9; -0,6)$ und $(e_9, e_{10}) = (0,2; -0,8)$ liegen im vierten Quadranten des Koordinatensystems und sprechen dagegen tendenziell für negative Autokorrelation, da jeweils ein negativer Wert auf einen positiven folgt. Insgesamt ist dann von fehlender Autokorrelation auszugehen, wenn sich die Residuenpaare ausgewogen auf die einzelnen Quadranten verteilen.

Aufgrund des relativ geringen Stichprobenumfangs kann in diesem Beispiel keine definitive Aussage getroffen werden; ein deutliches Indiz für das Vorliegen von Autokorrelation wird nicht festgestellt.

Lösung zu Aufgabe 22.4

Situation 1: Es ist zu erkennen, dass die Streuung der Residuen mit zunehmenden x -Werten ansteigt. Dies ist ein klarer Hinweis auf Heteroskedastizität, da die Störterme augenscheinlich nicht dieselbe Varianz besitzen. Obwohl die Schätzer der Regressionskoeffizienten in dieser Situation unverzerrt sind, können deren Verteilungen nicht ohne Weiteres bestimmt werden. Infolgedessen sind Hypothesentests und Konfidenzintervalle mit den üblichen Formeln nicht zulässig. Im Streuungsdiagramm der gepaarten Residuen sind auffällig viele Werte im dritten Quadranten, was für positive Autokorrelation spricht. Jedoch liegen diese oft relativ nahe am Koordinatenursprung, weshalb eine genauere Untersuchung nötig wäre,

um eine verlässliche Aussage in Bezug auf die Autokorrelation zu erhalten. Insbesondere ist somit kein klares Indiz für eine Fehlspezifikation des Modells vorhanden.

Situation 2: Auffällig ist der systematische Verlauf der Residuen im Bereich $1 \leq x \leq 4$, da dort die Residuen tendenziell monoton wachsend verlaufen. Dies ist ein Indiz für ein fehlspezifiziertes Modell, d. h. die Annahme des linearen Zusammenhangs steht grundsätzlich in Frage. Darüber hinaus kann am Streudiagramm der gepaarten Residuen festgestellt werden, dass die Verteilung der Residuenpaare auf die Quadranten nicht ausgewogen ist. Deshalb liegt zusätzlich zur festgestellten Fehlspezifikation die Vermutung auf Autokorrelation nahe.

Situation 3: Die Residuen verlaufen im Bereich $1 \leq x \leq 5$ zunächst positiv wachsend, woraufhin ab $x = 5$ plötzlich die Höhe der Residuen sinkt. Dies ist ein klares Anzeichen für einen Strukturbruch. Dieser könnte eventuell auch den systematischen Anstieg der Residuen erklären. Bei korrekter Berücksichtigung des vermuteten Strukturbruchs könnten die resultierenden Residuen insofern ein reguläres Verhalten besitzen.

Lösung zu Aufgabe 22.5

(a) Zwischenwerte zur Bestimmung der Regressionskoeffizienten:

	Jahr x_i	Leistung y_i (MWp)	$x_i y_i$	x_i^2
	1	43	43	1
	2	109	218	4
	3	174	522	9
	4	256	1024	16
	5	352	1760	25
	6	709	4254	36
	7	1512	10584	49
	8	2286	18288	64
	9	3042	27378	81
	10	3316	33160	100
	11	3461	38071	121
	12	3560	42720	144
Summe	78	18820	178022	650
Mittelwert	6,5	1568,333	14835,17	54,167

Berechnung der Regressionskoeffizienten:

$$b_2 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \approx \frac{14835,17 - 6,5 \cdot 1568,333}{54,167 - (6,5)^2} \approx 389,5$$

$$b_1 = \bar{y} - b_2 \bar{x} \approx -963,1$$

Die Regressionsgerade lautet somit

$$\hat{y} = -963,1 + 389,5x.$$

(b) Berechnung der Residuen:

x_i	y_i	\hat{y}_i	e_i
1	43	-573,67	616,67
2	109	-184,21	293,21
3	174	205,24	-31,24
4	256	594,70	-338,70
5	352	984,15	-632,15
6	709	1373,61	-664,61
7	1512	1763,06	-251,06
8	2286	2152,52	133,48
9	3042	2541,97	500,03
10	3316	2931,42	384,58
11	3461	3320,88	140,12
12	3560	3710,33	-150,33

Streudiagramm der Wertepaare (x_i, e_i) :



Das Diagramm zeigt einen systematischen Verlauf der Residuen. Diese sind zunächst bis $x = 6$ monoton fallend, dann steigend bis $x = 9$ und dann wiederum monoton fallend. Dieser Verlauf deutet erkennbar einen oder mehrere Strukturbrüche an.

(c) Die folgende Tabelle enthält die Zwischenrechnungen zur Bestimmung der Regressionsfunktionen für die drei angegebenen Zeiträume:

Zeitraum	x_i	y_i	Zwischenwerte	Koeffizienten
2004–2008	1	43	$\overline{xy} = 713,4$	$b_1 = -42,7$ $b_2 = 76,5$
	2	109	$\overline{x} = 3; \overline{y} = 186,8$	
	3	174	$x^2 = 11$	
	4	256		
	5	352		
2009–2012	6	709	$\overline{xy} = 15126$	$b_1 = -3942,5$ $b_2 = 777,3$
	7	1512	$\overline{x} = 7,5; \overline{y} = 1887,25$	
	8	2286	$x^2 = 57,5$	
	9	3042		