Wirtschaftsstatistik
Übungsblatt mit Lösungen Modul 6
Korrelation und Regression

Aufgabe 1

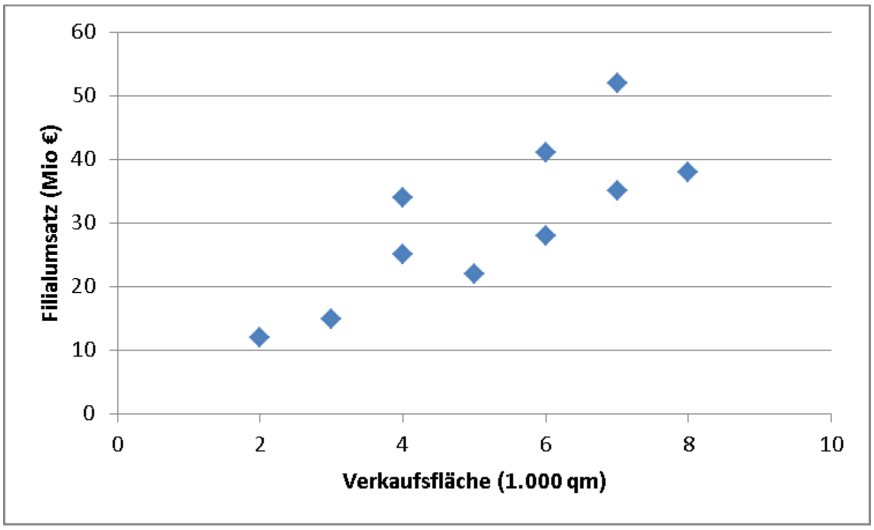
Filiale	A	B	C	D	E	F	G	H	I	J
Verkaufsfläche (1.000 qm)	7	5	6	3	8	2	4	6	4	7
Filialumsatz (Mio €)	35	22	41	15	38	12	34	28	25	52

a) Visualisieren Sie für das obige Datenmaterial den Zusammenhang zwischen den beiden Merkmalen Verkaufsfläche und Filialumsatz in einem geeigneten Diagramm. Wie wird diese grafische Darstellung genannt?

b) Welche Erkenntnisse liefert Ihnen eine erste qualitative Zusammenhangs- bzw. Abhängigkeitsanalyse auf der Basis des in a) erstellten Diagramms?

Lösung der Aufgabe 1

a) Streudiagramm



b) Es gibt einen positiven und „relativ“ starken Zusammenhang zwischen Verkaufsfläche und Umsatz der Filialen. Der Zusammenhang ist annähernd linear.

Aufgabe 2

Monate	Januar	Februar	März	April
Produzierte Menge (in Mio. Stück)	5	3	2	4
Produktionskosten (Mio. €)	8	4	6	5

a) Berechnen Sie eine lineare Regressionsfunktion $\hat{Y} = \hat{Y}(x) = a + b*x$, die den Zusammenhang zwischen produzierter Menge und Kosten möglichst gut charakterisiert
(x produzierte Menge in Mio. Stück und Y Kosten in Mio. €)
Tipp: transponiere und erweitere obere Tabelle, um Zwischenberechnungen einzufügen
(s. z. B. Folien 20 und 39 Modul 6)
b) Zeichnen Sie das Streudiagramm und die Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!
c) Interpretieren Sie die beiden Regressionskoeffizienten der berechneten Regressionsfunktion betriebswirtschaftlich.
Wie nennt man eine solche Funktion in der Betriebswirtschaftslehre?
d) Erstellen Sie auf der Basis der in a) ermittelten Regressionsfunktion eine Kostenprognose für den Monat Mai, in dem eine Produktionsmenge von 6 Mio. Stück geplant ist
e) Berechnen Sie den Korrelationskoeffizienten und das Bestimmtheitsmaß zur obigen Regressionsrechnung
f) Interpretieren Sie das Bestimmtheitsmaß
g) Berechnen Sie die Varianz der Produktionsmengen, der Produktionskosten und der Regressionswerte (Summen aus der Berechnungstabelle verwenden).
Zeigen Sie an diesem Beispiel, dass das Bestimmtheitsmaß angibt, welcher Anteil der Varianz der Produktionskosten erklärt wird durch die Regressionsfunktion bzw. die Varianz der Produktionsmengen.
Tipp: s. Varianzzerlegung Modul 6

Lösung der Aufgabe 2

a)

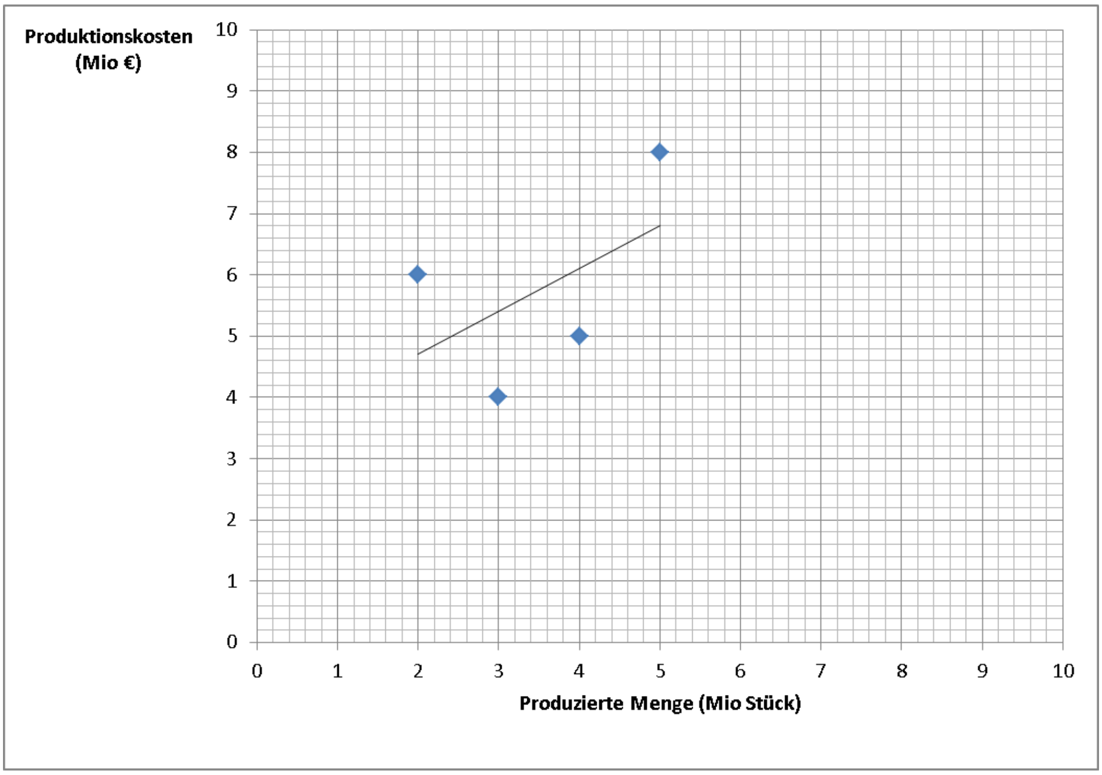
Monat	Produzierte Menge (Mio Stück) x_i	Produktionskosten (Mio €) Y_i	x^2_i	x_iY_i	Y^2_i
Januar	5	8	25	40	64
Februar	3	4	9	12	16
März	2	6	4	12	36
April	4	5	16	20	25
Σ	14	23	54	84	141

$$a = \frac{54 * 23 - 14 * 84}{4 * 54 - 14^2} = \frac{1242 - 1176}{216 - 196} = \frac{66}{20} = 3{,}3$$

$$b = \frac{4 * 84 - 14 * 23}{20} = \frac{336 - 322}{20} = \frac{14}{20} = 0{,}7$$

Regressionsfunktion Ŷ = Ŷ(x) = a + b*x = 3,3 + 0,7*x (Mio €)

b)



c) a: Fixkosten: 3,3 Mio € , b: variable Stückkosten: 0,70 €
Die Funktion nennt man in der Betriebswirtschaftslehre eine (lineare) Kostenfunktion

d) Kostenprognose auf der Basis der in a) ermittelten Regressionsfunktion für den Monat Mai, in dem eine Produktionsmenge von 6 Mio. Stück geplant ist:
Regressionsfunktion

Ŷ(6) = 3,3 + 0,7*6 = 3,3 + 4,2 = 7,5 (Mio €)

e)

$$r = \frac{\frac{1}{4} * 84 - 3{,}5 * 5{,}75}{\sqrt{\frac{1}{4} * 54 - 3{,}5^2} * \sqrt{\frac{1}{4} * 141 - 5{,}75^2}} = \frac{21 - 20{,}125}{\sqrt{1{,}25} - \sqrt{2{,}1875}} = \frac{0{,}875}{1{,}6536}$$
$$= 0{,}529$$

$$R^2 = 0{,}28$$

f) Interpretation des Bestimmtheitsmaßes: nur 28% der Varianz der Produktionskosten werden erklärt durch die Varianz der Produktionsmengen, d.h., nur 28% der monatlichen Kostenunterschiede werden dadurch erklärt, dass die monatlichen Produktionsmengen unterschiedlich waren. Die restlichen 72% lassen sich nur durch andere Einflussgrößen erklären.

g)

$$\bar{x} = \frac{1}{4} * 14 = 3{,}5$$

$${s_x}^2 = \frac{1}{4} * 54 - 3{,}5^2 = 13{,}5 - 12{,}25 = 1{,}25$$

$$\bar{Y} = \frac{1}{4} * 23 = 5{,}75$$

$${s_Y}^2 = \frac{1}{4} * 141 - 5{,}75^2 = 35{,}25 - 33{,}0625 = 2{,}1875$$

$$\bar{\hat{Y}} == \frac{1}{4} * (6{,}8 + 5{,}4 + 4{,}7 + 6{,}1) = 5{,}75$$

$${s_{\hat{Y}}}^2 = \frac{1}{4} * (6{,}8^2 + 5{,}4^2 + 4{,}7^2 + 6{,}1^2) - 5{,}75^2 = 33{,}765 - 33{,}0625 = 0{,}6125$$

$$\frac{{s_{\hat{Y}}}^2}{{s_Y}^2} = \frac{0{,}6125}{2{,}1875} = 0{,}28 = R^2$$

Aufgabe 3

Die Marktforschungsabteilung eines Unternehmens will für ein neues Produkt die Abhängigkeit der Absatzmenge x vom Preis Y empirisch untersuchen. Dazu wird in 4 vergleichbar großen Testmärkten unter sonst annähernd gleichen Rahmenbedingungen 8 Wochen lang jeweils ein unterschiedlicher Preis gefordert. Die Ergebnisse dieses Tests stehen in der folgenden Tabelle:

Testmärkte	**A**	**B**	**C**	**D**
x: Preis pro Stück (€)	1,90	2,10	1,50	2,50
Y: Absatzmenge in 8 Wochen (Stück)	5.000	4.000	8.000	3.000

a) Berechnen Sie eine lineare Regressionsfunktion $\hat{Y} = \hat{Y}(x) = a + b*x$, die die Abhängigkeit zwischen Preis und Absatzmenge möglichst gut charakterisiert. (x Preis in € und Y Absatzmenge in Stück)

b) Zeichnen Sie das Streudiagramm und die Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!
c) Interpretieren Sie die beiden Regressionskoeffizienten der in a) berechneten Regressionsfunktion betriebswirtschaftlich
d) Berechnen Sie den Korrelationskoeffizienten und das Bestimmtheitsmaß zur obigen Regressionsrechnung
e) Interpretieren Sie das Bestimmtheitsmaß für diesen Fall
f) Erstellen Sie auf der Basis der in a) ermittelten Regressionsfunktion eine Absatzmengen-Prognose für einen Testmarkt M, der den Testmärkten A, B, C und D entspricht. Welche Absatzmenge ist dort in 8 Wochen zu erwarten bei einem Preis von € 1,70
g) Berechnen Sie für die Daten die Varianz der empirisch ermittelten Absatzmengen und die Varianz der dazugehörigen Regressionswerte (= entsprechende Absatzmengen auf der Regressionsfunktion). Was ergibt der Quotient „Varianz der Regressionswerte zur Varianz der Absatzmengen“?

<u>Lösung der Aufgabe 3</u>

a)

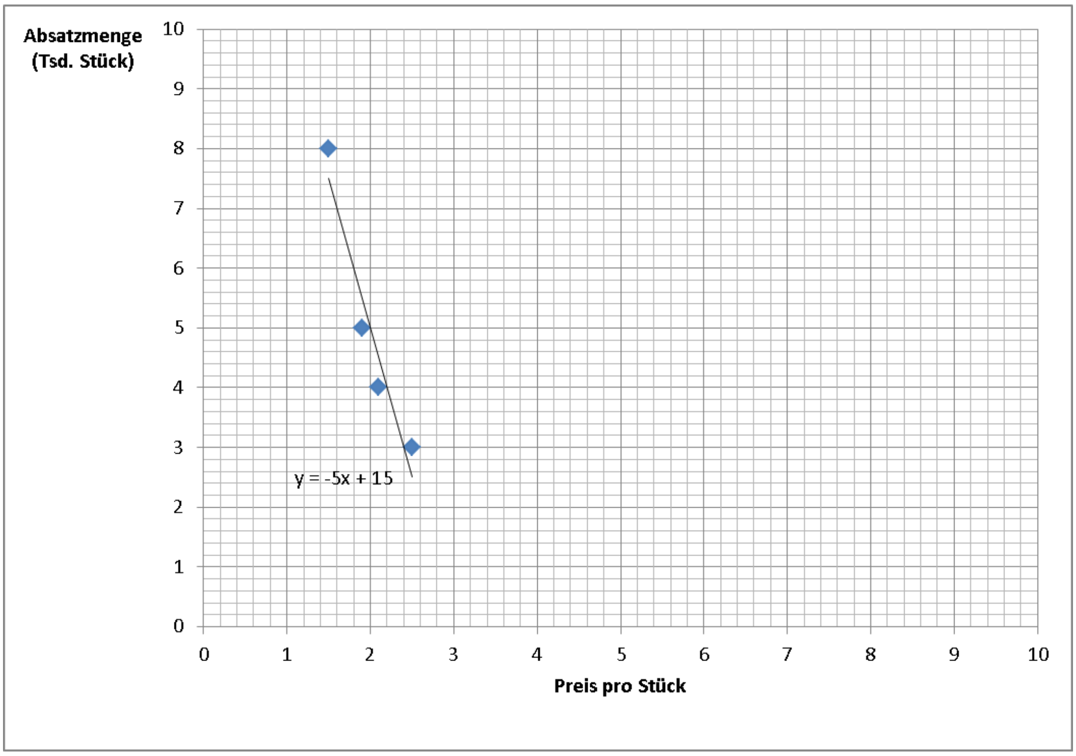
Testmarkt i	**Preis (€) x_i**	**Absatzmenge (Stück) Y_i**	**x^2_i**	**x_iY_i**	**Y^2_i**
A	1,90	5.000	3,61	9.500	25.000.000
B	2,10	4.000	4,41	8.400	16.000.000
C	1,50	8.000	2,25	12.000	64.000.000
D	2,50	3.000	6,25	7.500	9.000.000
Σ	8	20.000	16,52	37.400	114.000.000

$$a = \frac{16{,}52 * 20.000 - 8 * 37.400}{4 * 16{,}52 - 8^2} = \frac{330.400 - 299.200}{66{,}08 - 64} = \frac{31.200}{2{,}08} = = 15.000$$

$$b = \frac{4 * 37.400 - 8 * 20.000}{2{,}08} = \frac{149.600 - 160.000}{2{,}08} = \frac{-10.400}{2{,}08} = -5.000$$

Regressionsfunktion $\hat{Y} = \hat{Y}(x) = a + b*x = 15.000 - 5.000*x$ (Mio €)

b)



c) Interpretation der beiden Regressionskoeffizienten:

a: bei einem Preis von 0,-€ werden 15.000 Stück abgesetzt. 15.000 stellen eine Obergrenze für den Absatz in einem Testmarkt dar.

b: Steigt der Preis um € 1,00, dann sinkt die Absatzmenge um 5.000 Stück

d)

$$r = \frac{\frac{1}{4} * 37.400 - 2 * 5.000}{\sqrt{\frac{1}{4} * 16{,}52 - 2^2} * \sqrt{\frac{1}{4} * 114.000.000 - 5.000^2}} = \frac{9.350 - 10.000}{\sqrt{0{,}13} - \sqrt{3.500.000}}$$

$$= \frac{-650}{674{,}537} = -0{,}9636$$

$R^2 = 0{,}9286$

e) 92,86% der Absatzmengenunterschiede (der Varianz der Absatzmengen) lassen sich erklären durch die Unterschiede bei den Preisen in den verschiedenen Testmärkten (Varianz der Preise).

f) Prognose:

$\hat{Y}(1{,}7)$ = 15.000 – 5.000*1,7 = 15.000 – 8.500 = 6.500 (Stück)

g) $s_Y{}^2 = \frac{1}{4} * 114.000 - 5.000^2 = 3.500.000$

(Varianz der empirisch ermittelten Absatzmengen)

$\hat{y}_1 = \hat{y}(1{,}9) = 5.500$

$\hat{y}_2 = \hat{y}(2{,}1) = 4.500$

$\hat{y}_3 = \hat{y}(1{,}5) = 7.500$

$\hat{y}_4 = \hat{y}(2{,}5) = 2.500$

$$s_{\hat{Y}}{}^2 = \frac{1}{4} * (5.500^2 + 4.500^2 + 7.500^2 + 2.500^2) - 5.000^2$$

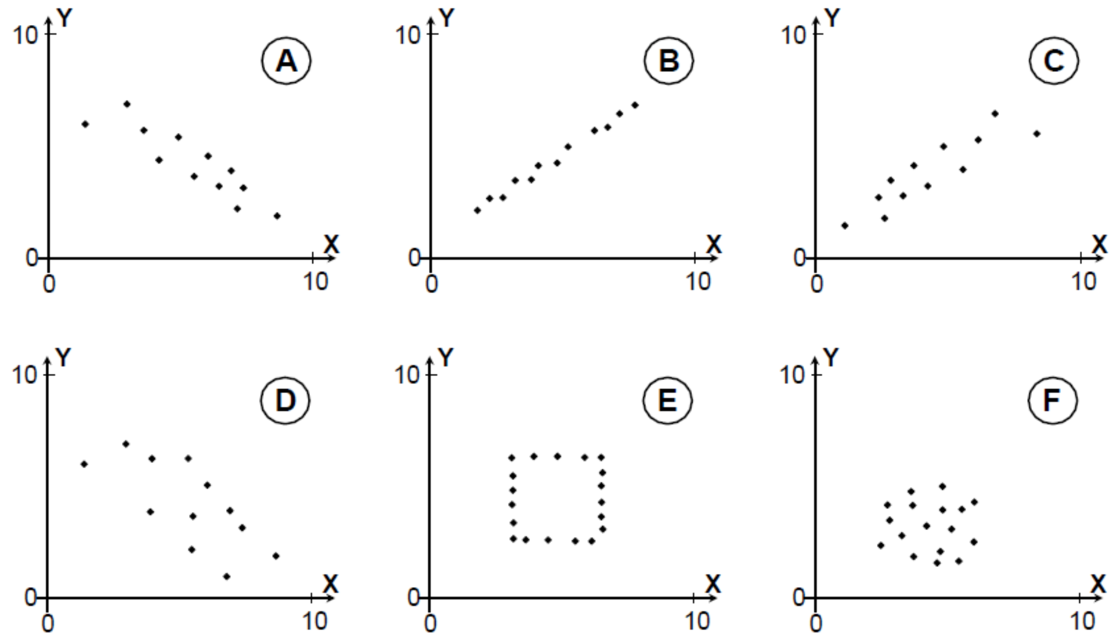
$$= \frac{1}{4} * 113.000.000 - 25.000.000 = 3.250.000$$

(Varianz der dazugehörigen Regressionswerte (= entsprechende Absatzmengen auf der Regressionsfunktion))

$$\frac{s_{\hat{Y}}{}^2}{s_Y{}^2} = \frac{3.250.000}{3.500.000} = 0{,}9286 = R^2$$

<u>Aufgabe 4</u>

In den folgenden Abbildungen sind 6 Streudiagramme dargestellt, die den Zusammenhang zwischen den zwei Merkmalen X und Y visualisieren. Bringen Sie die Korrelationskoeffizienten r_A, r_B, r_C, r_D, r_E, r_F nach ihren Werten in eine aufsteigende Rangfolge.
(r_A: Korrelationskoeffizient für das Streudiagramm A, usw.)



<u>Lösung der Aufgabe 4</u>

$-1 < r_A < r_D < r_F \approx 0 \approx r_E < r_C < r_B < +1$

Aufgabe 5

Es soll den Zusammenhang zwischen Preis P und Absatzmenge X empirisch untersucht werden. Dazu werden in 20 ausgewählten Testmärkten unterschiedliche Preise eingesetzt und Absatzmengen zu diesen Preisen dokumentiert. Die Regressionsrechnung liefert die folgende lineare Regressionsfunktion:

$$\hat{Y} = \hat{Y}(p) = a + b*p = 2.000 - 500*p \text{ (p in €, Y in Tsd. Stück)}$$

Der Korrelationskoeffizient beträgt r = - 0,75.
Nehmen Sie kurz Stellung zu den beiden folgenden Aussagen a) und b):

a) Aus der Regressionsfunktion kann man erkennen, dass der Zusammenhang zwischen Preis und Absatzmenge sehr stark ist.
b) Die Absatzmengenunterschiede in den Testmärkten lassen sich zu 75% durch die unterschiedliche Preispolitik erklären, d.h. durch die Preisunterschiede in den Testmärkten.

Lösung der Aufgabe 5

a) **Falsch!** Dies kann man nicht aus der Regressionsfunktion erkennen, sondern nur aus dem Korrelationskoeffizienten.
b) **Falsch!** Richtig: Die Absatzmengenunterschiede in den Testmärkten lassen sich zu **56,25%** durch die unterschiedliche Preissetzung erklären, d.h. durch die Preisunterschiede in den Testmärkten.
   $(0,75^2 = 0,5625)$

Aufgabe 6

Für ein neues Produkt wird untersucht, wie die Käufer auf unterschiedliche Preise reagieren. In 4 vergleichbar großen Testgeschäften in verschiedenen Gegenden wird das neue Produkt einen Monat lang zu jeweils unterschiedlichen Preisen angeboten. In der folgenden Tabelle sind für die Testphase die jeweils geforderten Preise (**p**) und die Reaktion der Käufer ausgedrückt in Absatzmengen des Produktes (**Y**) zusammengestellt.

Testgeschäft Nr.	**1**	**2**	**3**	**4**
x: Preis pro Stück (€)	3,00	1,50	1,00	2,50
Y: Absatzmenge (in Tausend Stück)	2	6	14	10

Nach der Methode der kleinsten Quadrate wurde mit Hilfe der Berechnungstabelle und den entsprechenden Formeln für die beiden Regressionskoeffizienten die folgende lineare Preis-Absatz-Funktion ermittelt:

$$\hat{Y} = \hat{Y}(x) = a + b*x = 16 - 4*x$$

a) Erstellen Sie eine Berechnungstabelle. Berechnen Sie die Regressionswerte und Residualwerte und tragen Sie diese in die Tabelle ein. Zeichnen Sie in das Koordinatensystem ein:
   a. die Beobachtungswertepaare durch ■
   b. die berechnete Regressionsfunktion,
   c. die Regressionswerte $\hat{Y}_i$ durch •,
   d. die Residualwerte (wenn farbiger Stift vorhanden, in Farbe)

b) Berechnen Sie den Korrelationskoeffizienten und das Bestimmtheitsmaß für die Daten in a) und interpretieren Sie die beiden Ergebnisse bezogen auf unser Beispiel

c) Prognostizieren Sie auf der Basis der Regressionsfunktion die monatliche Absatzmenge für einen den 4 Testgeschäften vergleichbaren Einzelhandelsbetrieb, wenn dort ein Preis von € 2,00 pro Stück verlangt wird

d) Was halten Sie von der Güte der obigen Prognose (unter c))?

e) Berechnen Sie mit Hilfe der Berechnungstabelle aus a) die Varianz ${s_Y}^2$ der Absatzmengen Y_i und die Varianz ${s_{\hat{Y}}}^2$ der entsprechenden Regressionswerte $\hat{Y}_i$. Zeigen Sie, dass hier die folgende Beziehung gilt:

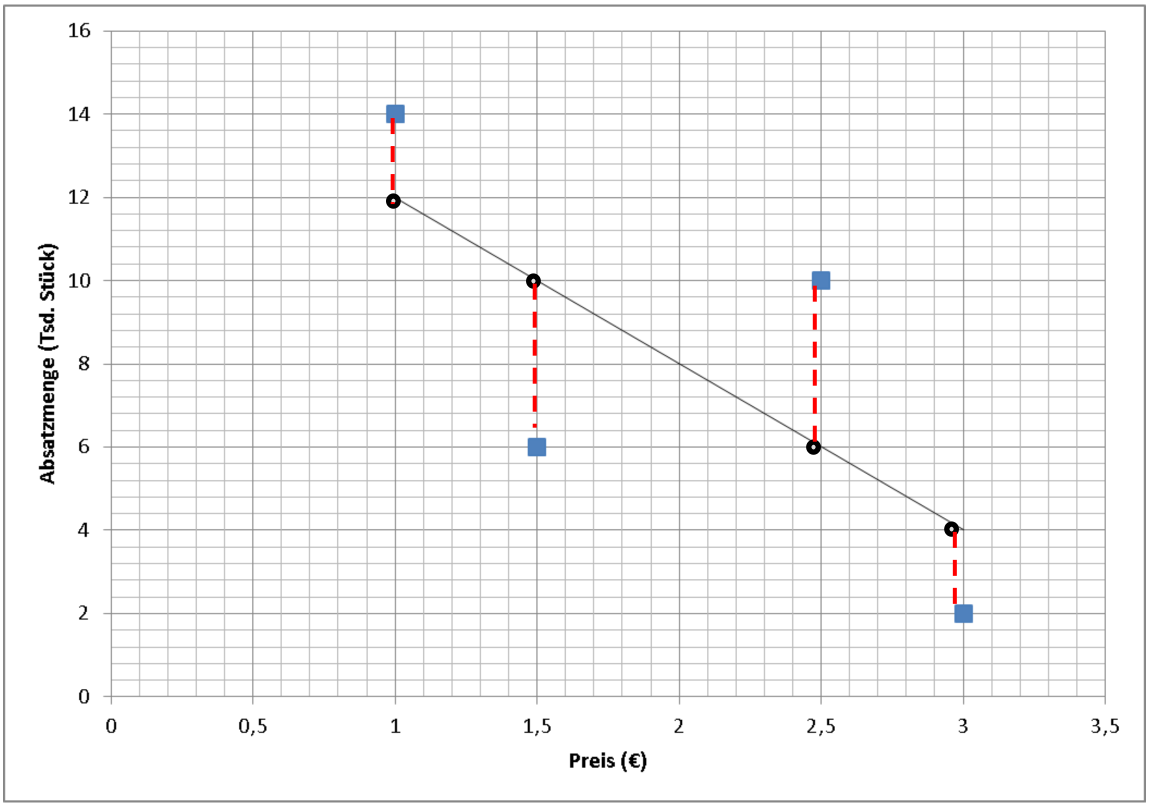
   Bestimmtheitsmaß = $R^2 = \frac{{s_{\hat{Y}}}^2}{{s_Y}^2}$

f) Berechnen Sie außerdem mit Hilfe der Berechnungstabelle aus a) die Varianz ${s_x}^2$ der Preise x_i. Zeigen Sie, dass die Varianz der Regressionswerte bestimmt wird durch die Varianz des unabhängigen Merkmals, d.h., dass hier folgende Beziehung gilt:
   ${s_{\hat{Y}}}^2 = b^2 * {s_x}^2$ (b → Regressionskoeffizient)

<u>Lösung der Aufgabe 6</u>

a)

Testgeschäft Nr.	**x_i Preis (€)**	**Y_i Absatzmenge (Tsd. Stück)**	**x^2_i**	**x_iY_i**	**Y^2_i**	**Regressionswerte $\hat{Y}_i$**	**Residualwerte (Y_i - $\hat{Y}_i$)**
1	3,00	2	9,00	6	4	4	-2
2	1,50	6	2,25	9	36	10	-4
3	1,00	14	1,00	14	196	12	+2
4	2,50	10	6,25	25	100	6	+4
Σ	8	32	18,50	54	336	32	0



b)

$$r = \frac{\frac{1}{4} * 54 - 2 * 8}{\sqrt{\frac{1}{4} * 18{,}5 - 2^2} * \sqrt{\frac{1}{4} * 336 - 8^2}} = \frac{13{,}5 - 16}{\sqrt{0{,}625} - \sqrt{20}} = \frac{-2{,}5}{12{,}5} = -0{,}707$$

$R^2 = 0{,}5$

50% der Varianz der Absatzmengen lassen sich erklären durch die Varianz der Preise bzw. durch die Regressionsfunktion. Die anderen 50% lassen sich nur durch andere Einflussfaktoren erklären.

c) $\hat{Y}(2) = 16 - 4*2 = 8$ (Tsd. Stück)

d) Die Prognosegüte ist nicht „überragend“, da sich nur 50% der Absatzmengenunterschiede durch die Preispolitik erklären lassen (s. R^2), die anderen 50% - durch andere Einflussgrößen, die bei dieser Untersuchung nicht berücksichtigt wurden.

e)

$$s_Y{}^2 = \frac{1}{4} * 336 - 8^2 = 20$$

$$s_{\hat{Y}}{}^2 = \frac{1}{4} * (4^2 + 10^2 + 12^2 + 6^2) - 8^2 = \frac{1}{4} * 296 - 64 = 74 - 64 = 10$$

$$\frac{s_{\hat{Y}}{}^2}{s_Y{}^2} = \frac{10}{20} = 0{,}5 = R^2$$

f)

$$s_x{}^2 = \frac{1}{4} * 18{,}5 - 2^2 = 0{,}625$$

$$b = -4$$

$$b^2 * s_x{}^2 = (-4)^2 * 0{,}625 = 10 = s_{\hat{Y}}{}^2$$

Aufgabe 7

Für 4 Monate liegen die Daten über den Hypothekenzinssatz **x** sowie über den saisonbereinigten monatlichen Auftragseingang **Y** im Bauhauptgewerbe vor, der auf den privaten Wohnungsbau entfällt.

Testgeschäft Nr.	**1**	**2**	**3**	**4**
x: Hypothekenzinssatz in %	6	7	5,5	7,5
Y: Auftragseingang (in Mrd. €)	30	40	50	20

a) Berechnen Sie nach der Methode der kleinsten Quadrate die lineare Regressionsfunktion, die die „mittlere" Abhängigkeit des Auftragseingangs im Bauhauptgewerbe vom Hypothekenzinssatz möglichst gut beschreibt
b) Interpretieren Sie für diesen Fall die beiden Regressionskoeffizienten
c) Zeichnen Sie in das Koordinatensystem die Wertepaare des Streudiagramms und die in a) berechnete Regressionsfunktion. Markieren Sie die Regressionswerte $\hat{Y}_i$
d) Berechnen Sie den Korrelationskoeffizienten und das Bestimmtheitsmaß zur obigen Regressionsrechnung
e) Interpretieren Sie das Bestimmtheitsmaß für diesen Fall
f) Prognostizieren Sie auf der Basis der Regressionsanalyse den (saisonbereinigten) monatlichen Auftragseingang im Bauhauptgewerbe, der zu erwarten ist bei einem Hypothekenzinssatz von 6,6% und von 7,2%
g) Welche Kennzahl kann man zur Beurteilung der Güte der obigen Prognose nutzen? Was halten Sie von der in f) erstellten Prognose?

Lösung der Aufgabe 7

a)

Monat i	Hypotheken-Zinssatz (in %) x_i	Auftragseingang (in Mrd. €) Y_i	x^2_i	x_iY_i	Y^2_i
1	6,0	30	36,00	180	900
2	7,0	40	49,00	280	1600
3	5,5	50	30,25	275	2500
4	7,5	20	56,25	150	400
Σ	26	140	171,5	885	5400

$$a = \frac{171{,}5 * 140 - 26 * 885}{4 * 171{,}5 - 26^2} = \frac{24010 - 23010}{686 - 676} = \frac{1000}{10} = 100$$

$$b = \frac{4 * 885 - 26 * 140}{10} = \frac{3540 - 3640}{10} = \frac{-100}{10} = -10$$

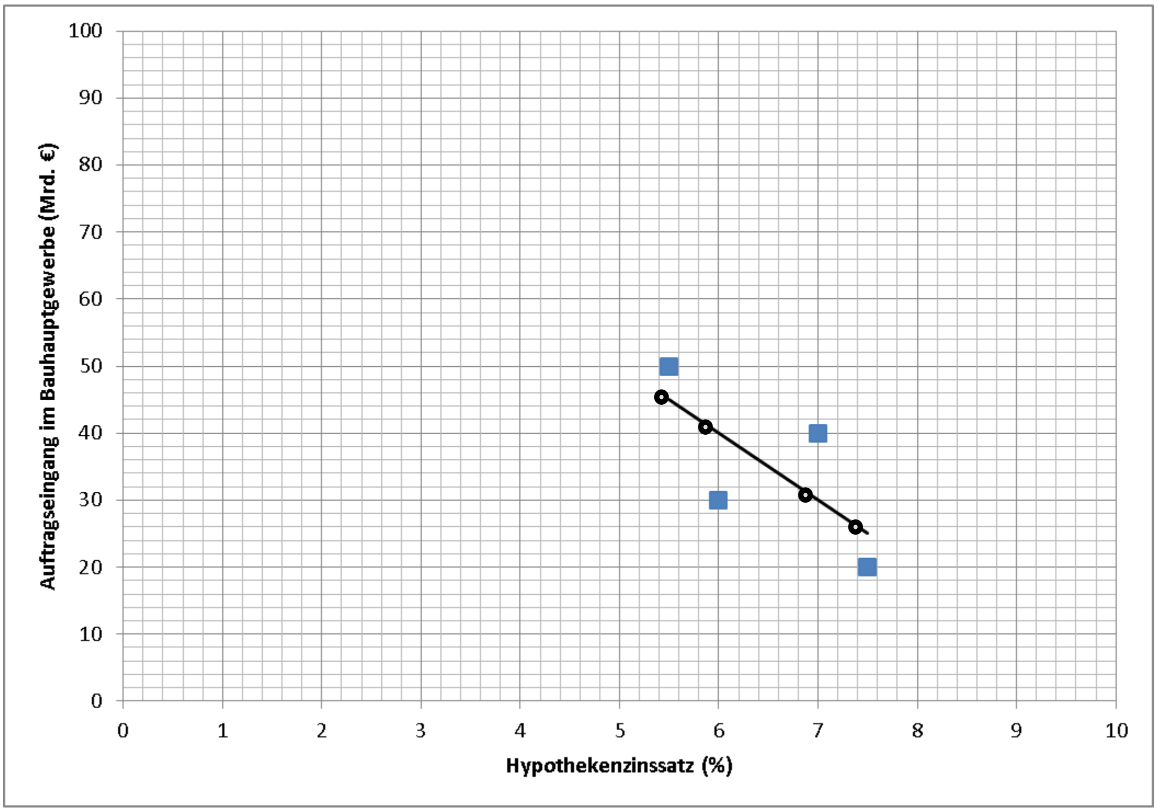
Regressionsfunktion Ŷ = Ŷ(x) = a + b*x = 100 - 10*x (Mrd. €)

b) Interpretation der beiden Regressionskoeffizienten:

a: 100 Mrd. €: Auftragseingang bei (theoretischen) Zinssatz von 0%.

b: Steigt der Zinssatz um 1%, dann sinkt der Auftragseingang um 10 Mrd. €.

c)



d)

$$r = \frac{\frac{1}{4} * 885 - 6{,}5 * 35}{\sqrt{\frac{1}{4} * 171{,}5 - 6{,}5^2} * \sqrt{\frac{1}{4} * 5400 - 35^2}} = \frac{221{,}25 - 227{,}5}{\sqrt{0{,}625} - \sqrt{125}}$$

$$= \frac{-6{,}25}{0{,}790 - 11{,}180} = -0{,}707$$

$R^2 = 0{,}5$

e) 50% der Varianz der Auftragseingänge im Bauhauptgewerbe lassen sich erklären durch die Varianz der Hypothekenzinssätze. D.h., 50% der Streuung bzw. Unterschiede bei den Auftragseingängen lassen sich erklären durch die Tatsache, dass die Hypothekenzinssätze in den betrachteten Monaten unterschiedlich hoch waren.

f) Prognose:
$\hat{Y}(6{,}6) = 100 - 10*6{,}6 = 100 - 66 = 34$ (Mrd. €)
$\hat{Y}(7{,}2) = 100 - 10*7{,}2 = 100 - 72 = 28$ (Mrd. €)

g) Das Bestimmtheitsmaß sowie der Korrelationskoeffizient können zur Beurteilung der Güte der obigen Prognose genutzt werden. Je näher der Wert des Korrelationskoeffizienten bei 1 liegt, umso „gehaltvoller" ist die Prognose. Entsprechendes gilt für das Bestimmtheitsmaß. In unserem Fall ist zu berücksichtigen, dass nur 50% der Unterschiede beim Volumen der Auftragseingänge durch den Hypothekenzinssatz erklärt werden. Für die übrigen 50% sind andere Einflussgrößen verantwortlich. D.h., die Prognose ist nur bedingt zu gebrauchen.