

*Wichtige Informationen von Frau Merrins zur Klausur*

*Aufgabenstellung genau lesen.*

*Nur das beantworten bzw. lösen, was in der Aufgabenstellung gefragt ist, nicht mehr und nicht weniger*

*präzise antworten*

*Bei Diagrammen die Beschriftungen für die x-Achse und y-Achse nicht vergessen.*

*Im Streudiagramm die Regressionsgerade über den gesamten Bereich des Streudiagramms (bzw. über den letzten Wert auf der x-Achse bzw. y-Achse zeichnen.*

*Es muss ersichtlich sein, dass die Regressionsgerade eine lineare Funktion visualisiert und diese Funktion über die Beispielwerte aus der Aufgabenstellung hinaus geht (Die lineare Funktion ist unendlich)*

*Zum Zeichnen des Streudiagramms ein separates Blatt nehmen*

*Wenn **im Streudiagramm oberhalb der Regressionsgerade  $f(x)$**  geschrieben und die Gerade somit als lineare Funktion gekennzeichnet wird, gibt das einen Zusatzpunkt*

*Die Skala der Achsen im Streudiagramm muss nicht bei 0 beginnen, sondern kann auch mit dem kleinsten y- oder x-Wert beginnen (wenn sich das Diagramm dadurch besser zeichnen lässt.) Dann muss am Achsen-Schnittpunkt (x,y) ein Viertelkreis gezeichnet werden (siehe Beispiel)*

*Für die Regressionsgerade müssen mittels der Regressionsgleichung  $\hat{y} = a + b \cdot x$  nur 2 Punkte ermittelt werden. Für den ersten Punkt kann für  $x = 0$  eingesetzt werden. Dann entspricht der Startpunkt den Koordinaten (x-Wert = 0, y-Wert = Wert zum Regressionskoeffizienten a) ➔ Regressionskoeffizient a wird anhand der dazugehörigen Formel berechnet.*

*Für den 2. Punkt einen x-Wert nehmen, der einigen Abstand zum x-Wert des ersten Punktes hat.*

*Das kann, muss aber kein Wert aus der Lösungstabelle sein. Es sollte ein Wert sein, mit dem man gut die Regressionsfunktion berechnen kann.*

*Beide Punkte im Streudiagramm einzeichnen und mit der Geraden verbinden (Gerade aber den gesamten Bereich des Streudiagramms zeichnen. Die Gerade muss über alle in der Aufgabenstellung genannten x-Werte und über den gesamten Bereich des Streudiagramms gezeichnet werden.*

*Die Punkte im Streudiagramm als Punkte oder kleine Kreise, kleine Quadrate oder Rauten zeichnen, **WICHTIG** die Punkte müssen exakt den Koordinaten x, y entsprechen (Präzision ist ihr wichtig!)*

*programmierbare Taschenrechner sind erlaubt (da alle Schulrechner heute programmierbar sind)*

*Es dürfen auch die Programmierfunktionen genutzt werden*

**WICHTIG**

*Der Rechenweg muss ersichtlich sein ➔ Tabelle ausfüllen ➔ Funktion schreiben ➔ Werte aus Tabelle einsetzen ➔ Werte aus Zwischenrechnung ➔ Ergebnis*

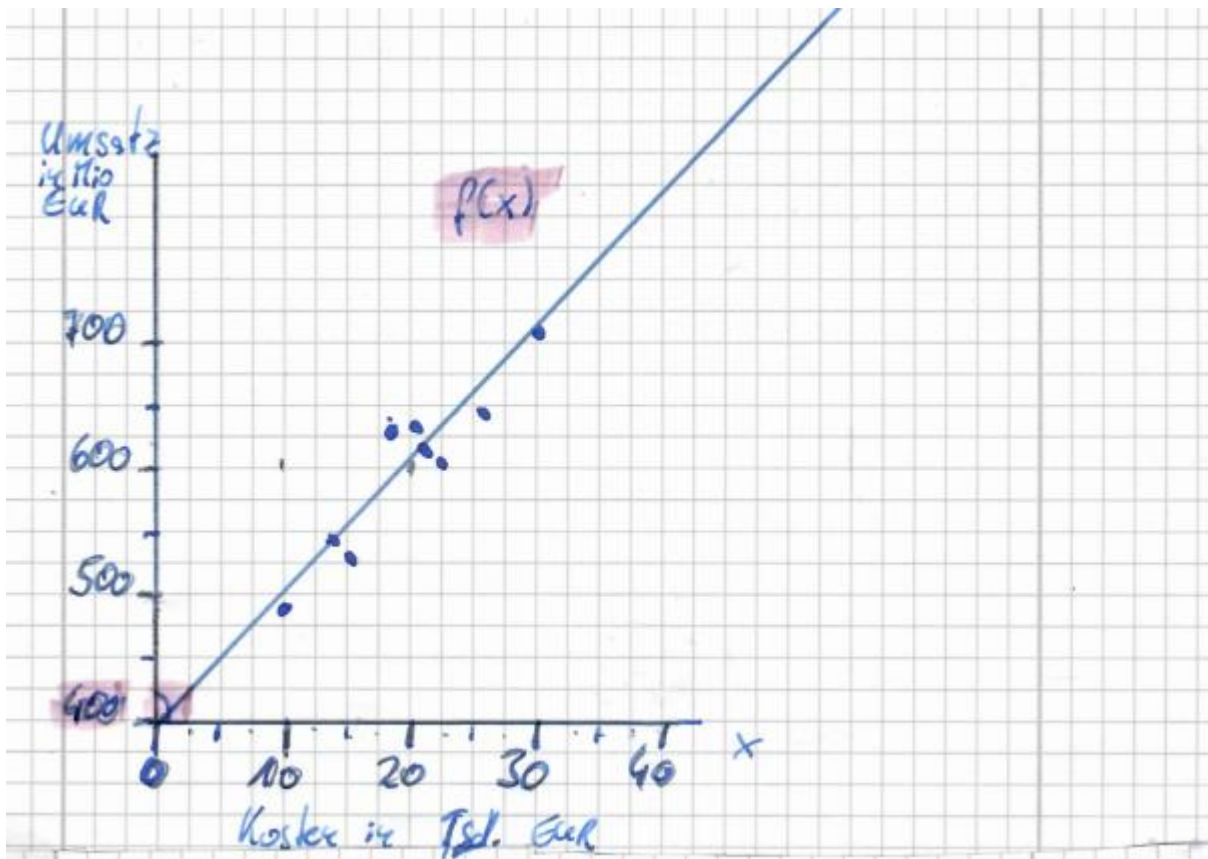
## Übungsaufgabe in der Präsenz

### Aufgabenstellung

Ein Filialleiter testet für neun bezüglich des Standortes, der Verkaufsfläche und des Zeitraums vergleichbare Filialen den Einfluss der Kosten auf den Umsatz.

- a. Zeichnen Sie das Streudiagramm und die Regressionsgerade zur Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!
- b. Ermitteln Sie die Korrelationskoeffizienten
- c. Berechnen Sie das Bestimmtheitsmaß
- d. Interpretieren Sie das Bestimmtheitsmaß
- e. Prognostizieren Sie die geschätzten Kosten bei einem Umsatz von 900 Mio EUR
- f. Prognostizieren Sie den geschätzten Umsatz bei Kosten von 30.000 EUR
  
- g. Zeichnen Sie das Streudiagramm und die Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!

Zeichnen Sie das Streudiagramm und die Regressionsgerade zur Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!



Regressionsgerade zur Regressionsfunktion

Es ist ausreichend, 2 Punkte f. die Regressionsgerade zu ermitteln.

Regressionsfunktion:  $f(x) = a + b \cdot x_i$

für  $a$  und  $b$  die zuvor ermittelten Werte einsetzen

$$a = 400 \quad b = 10$$

$$f(x) = 400 + 10x_i$$

Für  $x_i$  können beliebige Werte eingesetzt werden. Es müssen keine  $x_i$ -Werte aus der Tabelle sein.

Wir nehmen für  $x_1 = 0$

$$f(0) = 400 + 10 \cdot 0 \rightarrow f(0) = 400$$

1. Koordinatenpunkt f. d. Regressionsgerade (0, 400)

Wir nehmen für  $x_2 = 20$

$$f(20) = 400 + 10 \cdot 20 \rightarrow f(20) = 600$$

2. Koordinatenpunkt f. d. Regressionsgerade (20, 600)

Zum Zeichnen des Streudiagramms ein separates Blatt nehmen

Bei Diagrammen die Beschriftungen für die x-Achse und y-Achse nicht vergessen.

Im Streudiagramm die Punkte zu den x,y-Koordinaten als Punkte oder kleine Kreise, kleine Quadrate oder Rauten zeichnen.

WICHTIG die Punkte müssen exakt den Koordinaten x, y entsprechen (Präzision ist wichtig! Evtl. millimeter-Papier verwenden)

Im Streudiagramm die Regressionsgerade über den gesamten Bereich des Streudiagramms (bzw. über den letzten Wert auf der x-Achse bzw. y-Achse zeichnen.

Es muss ersichtlich sein, dass die Regressionsgerade eine lineare Funktion visualisiert und diese Funktion über die Beispielwerte aus der Aufgabenstellung hinaus geht (Die lineare Funktion ist unendlich)

Wenn im Streudiagramm oberhalb der Regressionsgerade  $f(x)$  geschrieben und die Gerade somit als lineare Funktion gekennzeichnet wird, gibt das einen Zusatzpunkt

Die Skala der Achsen im Streudiagramm muss nicht bei 0 beginnen, sondern kann auch mit dem kleinsten y- oder x-Wert beginnen (wenn sich das Diagramm dadurch besser zeichnen lässt.) Dann muss am Achsen-Schnittpunkt (x,y) ein Viertelkreis gezeichnet werden (siehe Beispiel)

Für die Regressionsgerade müssen mittels der Regressionsgleichung  $\hat{y}(x) = a + b \cdot x_i$  nur 2 Punkte ermittelt werden.

Für den ersten Punkt kann für  $x = 0$  eingesetzt werden.

Dann entspricht der Startpunkt den Koordinaten (x-Wert = 0, y-Wert = Wert zum Regressionskoeffizienten a) Regressionskoeffizient a wird anhand der dazugehörigen Formel berechnet.

Für den 2. Punkt einen x-Wert nehmen, der einigen Abstand zum x-Wert des ersten Punktes hat.

Das kann, muss aber kein Wert aus der Lösungstabelle sein.

Es sollte ein Wert sein, mit dem man gut die Regressionsfunktion berechnen kann.

Beide Punkte im Streudiagramm einzeichnen und mit der Geraden verbinden (Gerade aber den gesamten Bereich des Streudiagramms zeichnen. Die Gerade muss über alle in der Aufgabenstellung genannten x-Werte und über den gesamten Bereich des Streudiagramms gezeichnet werden.

In diesem Diagramm wurden für die Bestimmung des **ersten x.y-Koordinatenpunkt für x der Wert 0 eingesetzt.**

$$x = 0$$

$$\hat{y}(x) = a + b \cdot x_i$$

$$a = 400, b = 10$$

$$\hat{y}(0) = 400 + 10 \cdot 0$$

$$x = 0$$

$$\hat{y} = 400$$

Der **erste x.y-Koordinatenpunkt ist somit 0, 400**

$$(x = 0, y = 400)$$

In diesem Diagramm wurden für die Bestimmung des **zweiten x.y-Koordinatenpunkt** für x der Wert 20 **eingesetzt**.

$$x = 20$$

$$\hat{y}(x) = a + b \cdot x_i$$

$$a = 400, b = 10$$

$$\hat{y}(20) = 400 + 10 \cdot 20$$

$$x = 20$$

$$\hat{y} = 600$$

Der **zweite x.y-Koordinatenpunkt** ist somit **20, 600**

$$(x = 20, y = 600)$$

Beide Koordinatenpunkte (0, 400) und (20, 600) im Diagramm einzeichnen und mit der Geraden verbinden.  
Die Gerade jedoch über die gesamte Breite des Diagramms zeichnen“

i Filial-Nr.	x <sub>i</sub> Kosten in Tsd. €	y <sub>i</sub> Umsatz in Mio. €	x <sub>i</sub> * y <sub>i</sub>	x <sub>i</sub> <sup>2</sup>	y <sub>i</sub> <sup>2</sup>
1	10	480	4.800	100	230.400
2	18	632	11.376	324	399.424
3	30	702	21.060	900	492.804
4	22	630	13.860	484	396.900
5	26	645	16.770	676	416.025
6	14	545	7.630	196	297.025
7	24	606	14.544	576	367.236
8	21	630	13.230	441	396.900
9	15	530	7.950	225	280.900
<b>SUMMEN</b>	<b>180</b>	<b>5.400</b>	<b>111.220</b>	<b>3.922</b>	<b>3.277.614</b>

$$\bar{x} = 180 : 9 = 20$$

$$\bar{y} = 5.400 : 9 = 600$$

## Regressionsfunktion

$$\hat{y}(x) = a + b \cdot x_i$$

| a und b sind die Regressionskoeffizienten  
| und müssen berechnet werden

*Hinweis:*

*Warum schreibt man  $\hat{y}$  (y-Dach)? und nicht einfach y?*

*$\hat{y}$  kennzeichnet Schätzwerte (Das Ergebnis der Regressionsrechnung ist somit immer ein Schätzwert (Prognosewert)).*

### Berechnung der Regressionskoeffizienten a und b

*Hinweis:*

*Der Nenner in nachfolgenden Formeln (hier 2.898) muss nur einmal berechnet werden, da er in beiden Formeln identisch ist!*

*Formel zur Berechnung des Nenners zu beiden Regressionskoeffizienten  $(n \cdot \sum x_i^2) - (\sum x_i)^2$*

### Berechnung des Regressionskoeffizienten a

$$a = \frac{(\sum x_i^2 \cdot \sum y_i) - (\sum x_i \cdot \sum (x_i \cdot y_i))}{(n \cdot \sum x_i^2) - (\sum x_i)^2}$$

$$a = \frac{(3.922 \cdot 5.400) - (180 \cdot 111.220)}{(9 \cdot 3.922) - 180^2} = \frac{1.159.200}{2.898} = 400$$

$$a = 400$$

### Berechnung des Regressionskoeffizienten b

$$b = \frac{(n \cdot \sum (x_i \cdot y_i)) - (\sum x_i \cdot \sum y_i)}{(n \cdot \sum x_i^2) - (\sum x_i)^2}$$

$$= \frac{(9 \cdot 111.220) - (180 \cdot 5.400)}{(9 \cdot 3.922) - 180^2} = \frac{28.980}{2.898} = 10$$

$$b = 10$$

### Regressionsfunktion mit Regressionskoeffizienten a und b

$$\hat{y}(x) = 400 + 10x_i$$

*Hinweis:*

*Die Regressionskoeffizienten müssen in Worten interpretiert werden können. ➡ welche Abhängigkeit zwischen abhängiger Variable b und unabhängiger Variable a?*

*Was bedeuten in der Regressionsrechnung die Regressionskoeffizienten b und a?*

*b ist ein Faktor im veränderlichen Term der Regressionsrechnung, nur der Regressionskoeffizient a ist konstant.*

*Wenn z. B. x den Wert 0 hat (keine Kosten (unabhängige Variable x) entstehen bzw. keine Investitionen getätigt werden, wird immer noch ein Umsatz erwirtschaftet*

*$\hat{y}$  hat dann den Wert des Regressionskoeffizient a ➡  $\hat{y} = a + b \cdot x = a + b \cdot 0 \Rightarrow \hat{y} = a$*

## Korrelationskoeffizient $r_{x,y}$

Ermitteln Sie die Korrelationskoeffizienten

$$r_{x,y} = \frac{COV(x, y)}{s_x \cdot s_y}$$

$$r_{x,y} = \frac{\left(\frac{1}{n} * \sum(x_i * y_i)\right) - (\bar{x} * \bar{y})}{\sqrt{\left(\frac{1}{n} * \sum x_i^2\right) - \bar{x}^2} * \sqrt{\left(\frac{1}{n} * \sum y_i^2\right) - \bar{y}^2}}$$

$$r_{x,y} = \frac{\left(\frac{1}{9} * 111.220\right) - (20 * 600)}{\sqrt{\left(\frac{1}{9} * 3.922\right) - 20^2} * \sqrt{\left(\frac{1}{9} * 3.277.614\right) - 600^2}} = \frac{357,77}{\sqrt{435,77 - 400} * \sqrt{364.179,33 - 360.000}}$$

$$r_{x,y} = \frac{357,77}{\sqrt{35,77} * \sqrt{4.179,33}} = \frac{357,77}{5,98 * 64,64773} = \frac{357,77}{386,593} = 0,925$$

Korrelationskoeffizient  $r_{x,y} = 0,925$

## Bestimmtheitsmaß $R^2$

Berechnen Sie das Bestimmtheitsmaß  $R^2$

*Hinweis:*

*Bei einfacher linearer Regression (Regression mit nur 2 Variablen) entspricht das Bestimmtheitsmaß  $R^2$  dem Quadrat des Korrelationskoeffizienten  $r_{x,y}$ .*

Bestimmtheitsmaß  $R^2 = 0,925^2 = 0,856 = 85,6 \%$

*Hinweis:*

*Das Bestimmtheitsmaß in immer auch in % anzugeben.*



## Interpretation des Bestimmtheitsmaßes $R^2$ .

*Hinweis:*

*Des Bestimmtheitsmaß  $R^2$  muss in Worten interpretiert werden können*

*z. B.*

*Bestimmtheitsmaß ist ein Gütemaß des Modells, beschreibt wie gut das Modell der Realität entspricht.*

*Bestimmtheitsmaß ist Anteil der Varianz zur abhängigen Variable der sich durch den Anteil der Varianz der unabhängigen variable erklären lässt.*

*Z. B.*

*52% der Varianz der Umsätze (Unterschiede zu den Umsätzen (abhängiges Merkmal)) lassen sich durch die Varianz der Kosten (Kostenunterschiede (unabhängiges Merkmal)) erklären. Die übrigen 48% der Varianz zu den Kosten werden durch andere Einflussgrößen (Faktoren) erklärt.*

Interpretieren Sie das Bestimmtheitsmaß zur Aufgabenstellung.

85,6% der Varianz zum Umsatz (Umsatzunterschiede) lassen sich durch die Kostenvarianz erklären. Die übrigen 14,4% der Varianz zum Umsatz werden durch andere Einflussgrößen (Faktoren) erklärt.

Zu den übrigen Einflussgrößen können die Marktgröße, die Lage des Marktes, die Kaufkraft der Kunden u. a. sein.

Ein Bestimmtheitsmaß von 0,856 bzw. 85,6% kennzeichnet eine starke positive Korrelation zwischen dem abhängigen Merkmal (der abhängigen Variablen) „Umsatz“ und dem unabhängigen Merkmal (der unabhängigen Variablen) „Kosten“.

## Prognosewerte

Prognostizieren Sie die geschätzten Kosten bei einem Umsatz von 900 Mio EUR

$$y \text{ (Umsatz)} = 900 \text{ €}$$

$$\begin{array}{ll} \hat{y}(x) = a + b \cdot x_i & | \hat{y} = 900, a = 400, b = 10 \\ & | \text{Werte für } \hat{y} \text{ und } a \text{ und } b \text{ in die Gleichung einsetzen} \\ 900 = 400 + 10 \cdot x & | -400, : 10 \\ 500 : 10 = x & \\ x = 50 \text{ Tsd. €} & \end{array}$$

Hinweis:

Für den Wert der unabhängigen Variable immer auch die Einheit (hier „Tsd. €“ ) angeben.

Bei einem Umsatz von 900 Mio. € betragen die geschätzten Kosten voraussichtlich 50 Tsd. €.

Prognostizieren Sie den geschätzten Umsatz bei Kosten von 30.000 EUR.

$$x \text{ (Kosten)} = 30 \text{ Tsd. €}$$

$$\begin{array}{ll} \hat{y}(x) = a + b \cdot x_i & | x = 30, a = 400, b = 10 \\ & | \text{Werte für } a \text{ und } b \text{ und } x \text{ in die Gleichung einsetzen} \\ \hat{y}(30) = 400 + 10 \cdot 30 & \\ \hat{y}(30) = 700 \text{ Mio €} & \end{array}$$

Hinweis:

Für den Wert der abhängigen Variable immer auch die Einheit (hier „Mio. €“ ) angeben.

Bei Kosten von 30 Tsd. € beträgt der geschätzte Umsatz voraussichtlich 700 Mio. €.