

Mathematikaufgaben

> Daten, Statistik

> Korrelation, lineare Regression

Aufgabe: Ein mittelständisches Unternehmen stellt für die letzten fünf Jahre die Kosten für das Produktmarketing und den Umsatz gegenüber:

	2020	2021	2022	2023	2024
Marketingkosten (Millionen €)	5	4	6	7	3
Umsatz (Millionen €)	41	39	49	55	32

Berechne Korrelationskoeffizient und Bestimmtheitsmaß. Ermittle die Gleichung der Geraden, die als lineare Regression den Zusammenhang zwischen Marketingkosten und Umsatz wiedergibt. Beeinflusst das Marketing den Produktumsatz des Unternehmens?

Lösung: I. Allgemein gilt: Mit Daten und Statistik verbunden sind die zahlenmäßige Erfassung von Sachverhalten und Beobachtungen und deren grafische Darstellung. Daten ergeben damit ein mathematisches Modell, das mit Hilfe mathematischer Methoden analysiert und bewertet werden kann. Sind zwei Zufallsvariablen X, Y gegeben, so lässt sich ein stochastischer Zusammenhang (Korrelation) zwischen diesen Zufallsgrößen nachweisen mittels des sog. (Bravais-Pearson-) Korrelationskoeffizienten r , der als Maßzahl $(-1 \leq r \leq 1)$ Stärke und Richtung einer linearen Beziehung zwischen X und Y charakterisiert. Der Korrelationskoeffizient errechnet sich als:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{bzw.} \quad r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \cdot \sqrt{\sum_{i=1}^n y_i^2 - n \bar{y}^2}},$$

wobei die Wertepaare (x_i, y_i) Stichprobenpaare von Ausprägungen x_i, y_i der Zufallsvariablen X und Y sind, $i = 1, 2, \dots, n$. Ausprägungen sind Werte eines Merkmals; Merkmale sind geeignete, charakteristische Eigenschaften eines Untersuchungsobjekt innerhalb der Statistik und können als Zufallsvariablen aufgefasst werden.

Mit dem Korrelationskoeffizienten lässt sich die Stärke des stochastischen Zusammenhangs zwischen den Zufallsvariablen wie folgt einschätzen:

Zusammenhang	Korrelationskoeffizient
kein	$r = 0$
gering	$0 < r \leq 0,3$
schwach	$0,3 < r \leq 0,5$
mittel	$0,5 < r \leq 0,8$
stark	$0,8 < r < 1$
funktional	$ r = 1$

Das Bestimmtheitsmaß R^2 errechnet sich als Quadrat des Korrelationskoeffizienten r mit: $R^2 = r^2$ ($0 \leq R^2 \leq 1$) und charakterisiert damit die Größe der Abhängigkeit zwischen den Zufallsvariablen X und Y.

II. Unter der Voraussetzung einer z.B. starken Korrelation r zwischen zwei Merkmalen X und Y mit

deren jeweiligen Ausprägungen $x_i, y_i, i = 1, 2, \dots, n$ gibt die lineare Regression einen linearen Zusammenhang zwischen den vorgegebenen Zufallsvariablen X als Einfluss- und Y als Zielgröße wieder. Es entsteht eine Regressionsgerade $y = a + bx$ durch die Wertepaare $(x_i, y_i), i = 1, 2, \dots, n$, auch auf der Grundlage der Mittelwerte $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ vermöge der Berechnungen:

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{bzw.} \quad b = \frac{n \cdot \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \cdot \left(\sum_{i=1}^n y_i \right)}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (\text{Geradensteigung})$$

$$a = \bar{y} - b \cdot \bar{x} \quad \text{bzw.} \quad a = \frac{\left(\sum_{i=1}^n x_i^2 \right) \cdot \left(\sum_{i=1}^n y_i \right) - \left(\sum_{i=1}^n x_i \right) \cdot \left(\sum_{i=1}^n x_i y_i \right)}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (\text{y-Achsenabschnitt der Geraden}).$$

III. Korrelationskoeffizient und Regressionsgerade in Zusammenhang mit den Merkmalen Marketingkosten und Umsatz berechnen sich (wahlweise) über die folgenden Tabellen:

Merkmal Marketingkosten (€) [x], Merkmal Umsatz (€) [y]; Tabelle (I):

i	Wert x_i	Wert y_i	Differenz $x_i - \bar{x}$	Quadrat $(x_i - \bar{x})^2$	Differenz $y_i - \bar{y}$	Quadrat $(y_i - \bar{y})^2$	Produkt $(x_i - \bar{x})(y_i - \bar{y})$
1	5	41	0	0	-2.2	4.84	0
2	4	39	-1	1	-4.2	17.64	4.2
3	6	49	1	1	5.8	33.64	5.8
4	7	55	2	4	11.8	139.24	23.6
5	3	32	-2	4	-11.2	125.44	22.4
Summe:	25	216	-	10	-	320.8	56
Mittelwert:	5	43.2	-	-	-	-	-
Standardabweichung:			-	1.4142	-	8.01	-
Regressionskoeffizient $r =$							0.9887
Steigung/Regressionsgerade $b =$							5.6
Achsenabschnitt/Regressionsgerade $a =$							15.2
Bestimmtheitsmaß $R^2 =$							0.9776

Merkmal Marketingkosten (€) [x], Merkmal Umsatz (€) [y]; Tabelle (II):

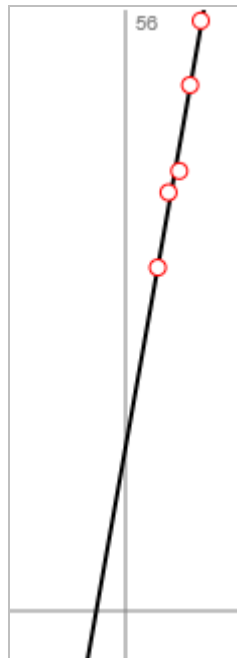
i	Wert x_i	Quadrat x_i^2	Wert y_i	Quadrat y_i^2	Produkt $x_i \cdot y_i$
1	5	25	41	1681	205
2	4	16	39	1521	156
3	6	36	49	2401	294
4	7	49	55	3025	385
5	3	9	32	1024	96
Summe:	25	135	216	9652	1136
Mittelwert:	5	-	43.2	-	-
Regressionskoeffizient $r =$					0.9887
Steigung/Regressionsgerade $b =$					5.6
Achsenabschnitt/Regressionsgerade $a =$					15.2
Bestimmtheitsmaß $R^2 =$					0.9776

Mit $r = 0,9887$ als Korrelationskoeffizienten liegt damit ein starker Zusammenhang zwischen Marketingkosten und Umsatz vor. Zu vermuten ist: Der Einsatz von Marketing beeinflusst den Unternehmensumsatz.

Die Regressionsgerade lautet:

$$y = 15,2 + 5,6x \quad (x \text{ [€]}, y \text{ [€]}).$$

IV. Graph von Regressionsgerade und Punktwolke ergeben nun folgendes Bild:



www.michael-buhlmann.de / 07.2025 / Aufgabe 2477