Michael Buhlmann

Mathematikaufgaben
> Daten, Statistik
> Korrelation, lineare Regression

Aufgabe: Für Restaurantbetriebe lässt sich ein Zusammenhang zwischen der für die Gäste zur Verfügung stehenden Fläche (Quadratmeter) und dem Jahresumsatz (Tausend €) vermuten. Dabei stehen die folgenden Daten tabellarisch zur Verfügung:

Fläche (Quadratmeter)	25	32	60	85	104	125	130	135	143	152	166	170	210
Umsatz (Tausend €)	42	45	60	110	122	175	220	190	184	205	248	251	286

Berechne Korrelationskoeffizient und Bestimmtheitsmaß. Ermittle die Gleichung der Geraden, die als lineare Regression den Zusammenhang zwischen Fläche und Umsatz wiedergibt.

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:

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

wobei die Wertepaare (x_i, y_i) Stichprobenpaare von Ausprägungen x_i, y_i der Zufallsvariablen X und Y sind, i = 1, 2, … 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	$0 < \lvert r\rvert \leq 0{,}3$
schwach	$0{,}3 < \lvert r\rvert \leq 0{,}5$
mittel	$0{,}5 < \lvert r\rvert \leq 0{,}8$
stark	$0{,}8 < \lvert r\rvert < 1$
funktional	$\lvert r\rvert = 1$

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, … 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, … 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:

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

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

III. Der Korrelationskoeffizient in Zusammenhang mit den Merkmalen Fläche und Umsatz berechnet sich über die folgende Tabelle:

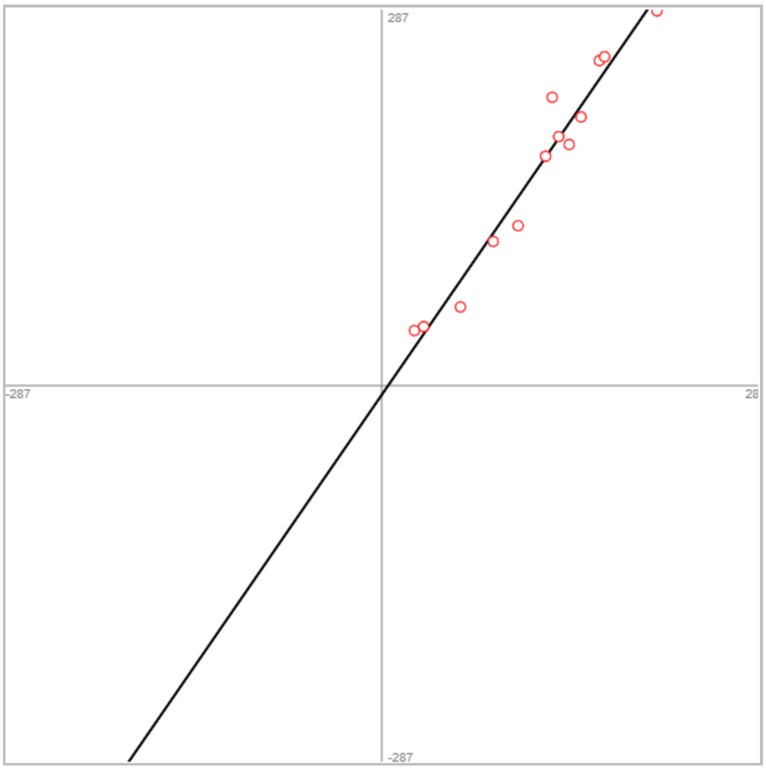
Merkmal Fläche [x], Merkmal Umsatz [y]; Tabelle:

i	**Wert x_i**	**Quadrat x_i^2**	**Wert y_i**	**Quadrat y_i^2**	**Produkt x_i*y_i**
1	25	625	42	1764	1050
2	32	1024	45	2025	1440
3	60	3600	60	3600	3600
4	85	7225	110	12100	9350
5	104	10816	122	14884	12688
6	125	15625	175	30625	21875
7	130	16900	220	48400	28600
8	135	18225	190	36100	25650
9	143	20449	184	33856	26312
10	152	23104	205	42025	31160
11	166	27556	248	61504	41168
12	170	28900	251	63001	42670
13	210	44100	286	81796	60060
Summe:	1537	218149	2138	431680	305623
Mittelwert:	118.2308	-	164.4615	-	-
Regressionskoeffizient r =					0.9785
Steigung/Regressionsgerade m =					1.4507
Achsenabschnitt/Regressionsgerade b =					-7.0528
Bestimmtheitsmaß R^2 =					0.9575

Mit r = 0, 9785 als Korrelationskoeffizienten liegt damit ein starker, schon durch wirtschaftliche Überlegungen begründeter Zusammenhang zwischen Fläche und Umsatz vor. Das Bestimmtheitsmaß beträgt R^2 = 0,9575. Die Regressionsgerade lautet:

y = -7,0528 + 1.4507x.

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



www.michael-buhlmann.de / 06.2023 / Aufgabe 1879