Michael Buhlmann

Mathematikaufgaben
> Daten, Statistik
> Korrelation

Aufgabe: Für eine Baumart wird der Zusammenhang zwischen Alter (Jahre) und Stammdurchmesser (Zentimeter) untersucht. Dabei stehen die folgenden Daten tabellarisch zur Verfügung:

Alter (Jahre)	2	5	8	10	14	16	20	25	31	35	40	46	50
Durchmesser (Zentimeter)	6	8	8	12	18	21	22	30	28	31	32	35	36

Berechne den Korrelationskoeffizienten zwischen den Merkmalen. Welche Schlüsse können hinsichtlich des Zusammenhangs zwischen Baumalter und -durchmesser gezogen werden?

1. 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 ($-1 \le r \le 1$) Stärke und Richtung einer linearen Beziehung zwischen X und Y charakterisiert. Der Korrelationskoeffizient errechnet sich als:

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

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 < |r| ≤ 0,3
schwach	0,3 < |r| ≤ 0,5
mittel	0,5 < |r| ≤ 0,8
stark	0,8 < |r| < 1
funktional	|r| = 1

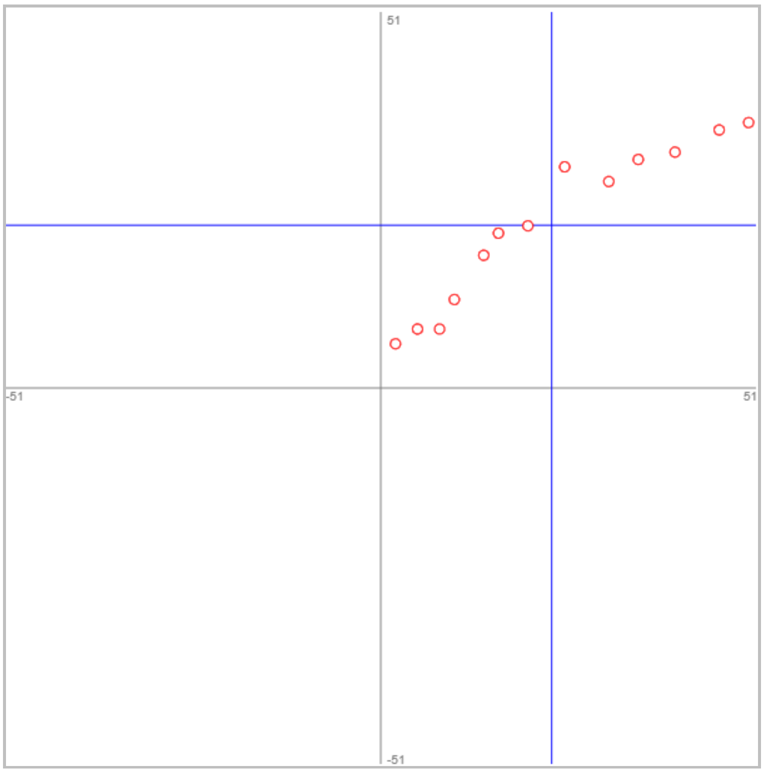
II. Der Korrelationskoeffizient in Zusammenhang mit den Merkmalen Baumalter und -durchmesser berechnet sich über die folgende Tabelle:

Merkmal Alter [x], Merkmal Durchmesser [y]; Tabelle:

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	2	6	-21.2308	450.7456	-16.0769	258.4675	341.3254
2	5	8	-18.2308	332.3609	-14.0769	198.1598	256.6331
3	8	8	-15.2308	231.9763	-14.0769	198.1598	214.4024
4	10	12	-13.2308	175.0533	-10.0769	101.5444	133.3254

5	14	18	-9.2308	85.2071	-4.0769	16.6213	37.6331
6	16	21	-7.2308	52.284	-1.0769	1.1598	7.787
7	20	22	-3.2308	10.4379	-0.0769	0.0059	0.2485
8	25	30	1.7692	3.1302	7.9231	62.7751	14.0178
9	31	28	7.7692	60.3609	5.9231	35.0828	46.0178
10	35	31	11.7692	138.5148	8.9231	79.6213	105.0178
11	40	32	16.7692	281.2071	9.9231	98.4675	166.4024
12	46	35	22.7692	518.4379	12.9231	167.0059	294.2485
13	50	36	26.7692	716.5917	13.9231	193.8521	372.7101
Summe:	302	287	-	3056.3077	-	1410.9231	1989.7692
Mittelwert:	23.2308	22.0769	-	-	-	-	-
Standardabweichung:			-	15.333	-	10.4179	-
Korrelationskoeffizient r =							0.9582

III. Mit r = 0,9582 als Korrelationskoeffizienten liegt damit ein starker, schon durch die Biologie des Wachstums von Bäumen begründeter Zusammenhang zwischen Baumalter und -durchmesser vor.



2. 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:

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

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	$0 < \|r\| \leq 0,3$
schwach	$0,3 < \|r\| \leq 0,5$
mittel	$0,5 < \|r\| \leq 0,8$
stark	$0,8 < \|r\| < 1$
funktional	$\|r\| = 1$

II. Der Korrelationskoeffizient in Zusammenhang mit den Merkmalen Baumalter und -durchmesser berechnet sich über die folgende Tabelle:

Merkmal Alter [x], Merkmal Durchmesser [y]; Tabelle:

i	**Wert x_i**	**Quadrat x_i^2**	**Wert y_i**	**Quadrat y_i^2**	**Produkt x_i*y_i**
1	2	4	6	36	12
2	5	25	8	64	40
3	8	64	8	64	64
4	10	100	12	144	120
5	14	196	18	324	252
6	16	256	21	441	336
7	20	400	22	484	440
8	25	625	30	900	750
9	31	961	28	784	868
10	35	1225	31	961	1085
11	40	1600	32	1024	1280
12	46	2116	35	1225	1610
13	50	2500	36	1296	1800
Summe:	302	10072	287	7747	8657
Mittelwert:	23.2308	-	22.0769	-	-
Korrelationskoeffizient r =					0.9582

III. Mit r = 0,9582 als Korrelationskoeffizienten liegt damit ein starker, schon durch die Biologie des Wachstums von Bäumen begründeter <u>Zusammenhang</u> zwischen Baumalter und -durchmesser vor.

www.michael-buhlmann.de / 06.2023 / Aufgabe 1877