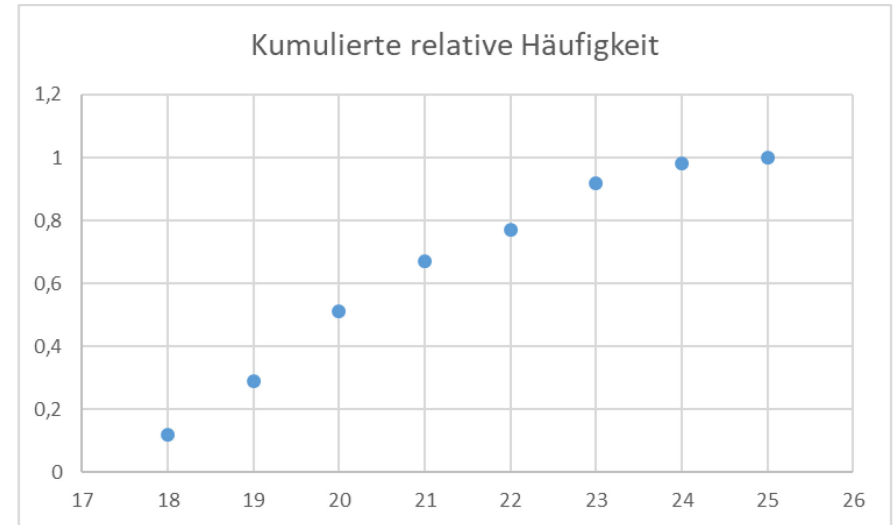


Übungsaufgaben

Alter	Anzahl	Rel Häufigkeit	Kumuliert
18	12	0,12	0,12
19	17	0,17	0,29
20	22	0,22	0,51
21	16	0,16	0,67
22	10	0,1	0,77
23	15	0,15	0,92
24	6	0,06	0,98
25	2	0,02	1



- a) jünger als 20 Jahre: 29%
- b) älter als 22 Jahre: 23%
- c) bis zu 23 Jahren: 92%
- d) genau 23 Jahre: 15%

Beschreiben Sie in Worten die Bedeutung bzw. Berechnung der folgenden Kennwerte:

Varianz: mittlere quadrierte Abweichung der Werte vom arithmetischen Mittel

Standardabweichung: Wurzel aus der Varianz

Variationskoeffizient: normierte Standardabweichung
(Standardabweichung / Mittelwert)

Ein mögliches Maß für den Zusammenhang zwischen zwei Variablen stellt der Korrelationskoeffizient nach Person dar.

Welche beiden Voraussetzungen müssen gegeben sein, damit dieser ein geeignetes Maß für den Zusammenhang zwischen zwei Variablen darstellt?

Skalenniveau metrisch

Zusammenhang linear

Welche beiden Extremwerte kann der Betrag des Korrelationskoeffizienten annehmen, und wie sind diese zu interpretieren?

0 → kein (linearer) Zusammenhang

1 → perfekter linearer Zusammenhang, alle Werte liegen auf einer Geraden

(- 1 → s.o. – Gerade hat negative Steigung)

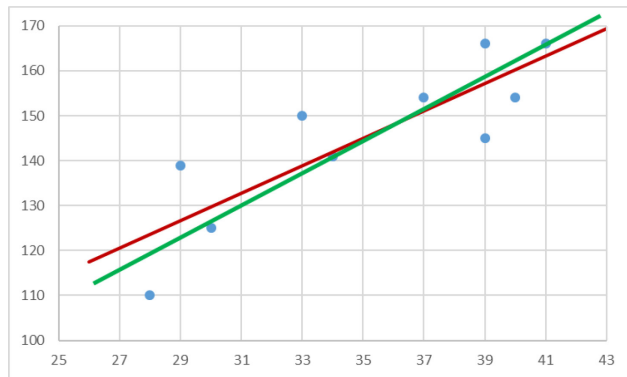
Antwort: Einfache lineare Regression



Skizzieren Sie Zielsetzung, Datengrundlage, Vorgehen und Einsatz der einfachen linearen Regression.

- Zielsetzung: Ein *abhängiges* Merkmal Y soll mit Hilfe des *unabhängigen* Merkmals X erklärt werden.
- Grundlage: Datensatz mit bekannten Werten für X und Y
(beide metrisch, linearer Zusammenhang)
- Vorgehen: *Ableiten eines Modells*, welches Y aus X erklärt
(Methode „kleinste Quadrate“)
- Einsatz: *Anwendung des Modells* auf neue Datensätze mit Werten für X und unbekanntem Y zur Vorhersage von Y

„Methode der kleinsten Quadrate“



Wie legt man die Gerade am „besten“ in die Punktwolke?

Ziel: Fehler möglichst minimal → Abstände der Punkte zur Geraden möglichst gering

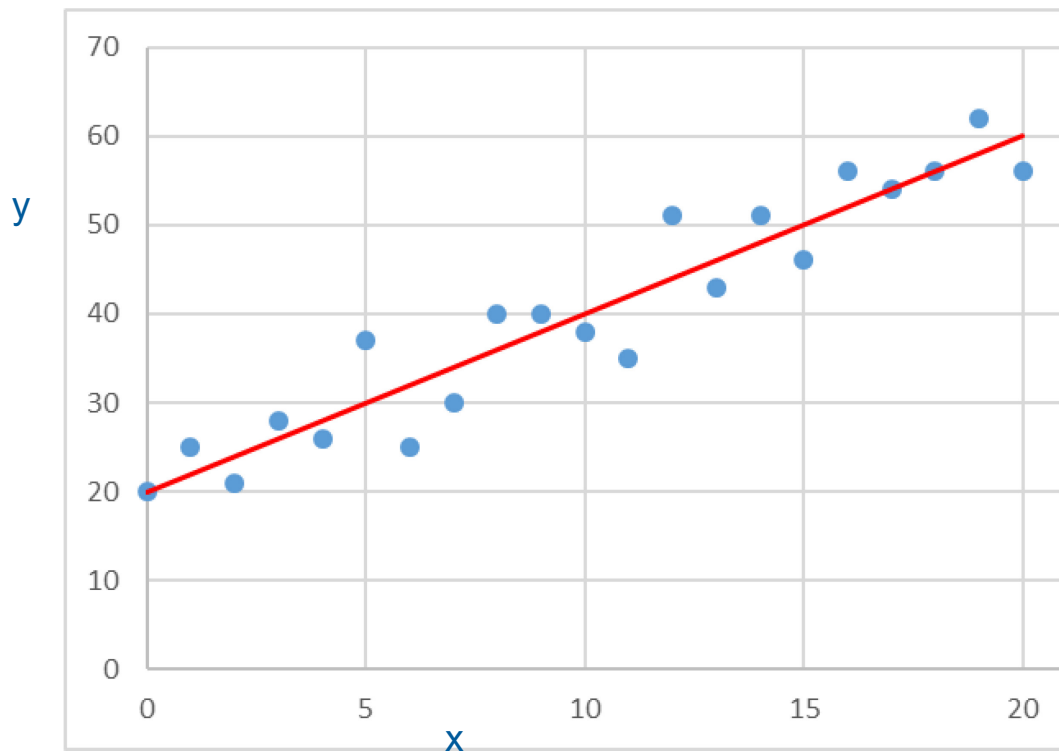
Positive und negative Abweichungen sollen sich nicht zu Null addieren, höhere Abweichungen stärker berücksichtigt werden.

→ Abweichungen werden quadriert und die *Summe der quadrierten Abweichungen minimiert*, um die Regressionsgerade zu ermitteln.

Lösung: Einfache lineare Regression



Für gegebene Werte von x und y wurde eine lineare Regression durchgeführt. Ermitteln Sie anhand der eingezeichneten Regressionsgeraden die Regressionsfunktion.



Regressionsfunktion:

$$y = 2x + 20$$

Welche Vorhersage trifft das Modell für $x = 25$?

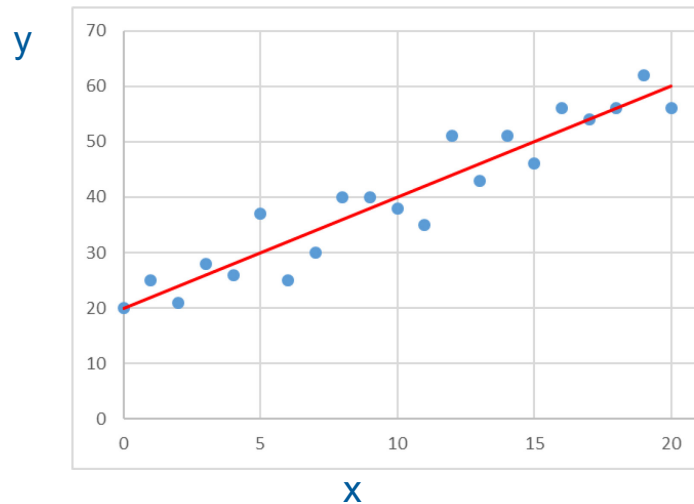
$$y = 70$$

Lösung: Bestimmtheitsmaß



Für die gegebenen Daten haben Sie die untenstehenden Werte ermittelt.

Berechnen Sie daraus das Bestimmtheitsmaß der Regression und interpretieren Sie dieses.



$$\text{Varianz}(y) = 163,0$$

$$\text{Varianz}(x) = 36,7$$

$$\text{Mittleres Fehlerquadrat der Regression} = 16,3$$

Bestimmtheitsmaß:

$$R^2 = 1 - (16,3 / 163) = 1 - 0,1 = 0,9$$

Interpretation:

90% der Varianz von y werden durch das Modell erklärt.