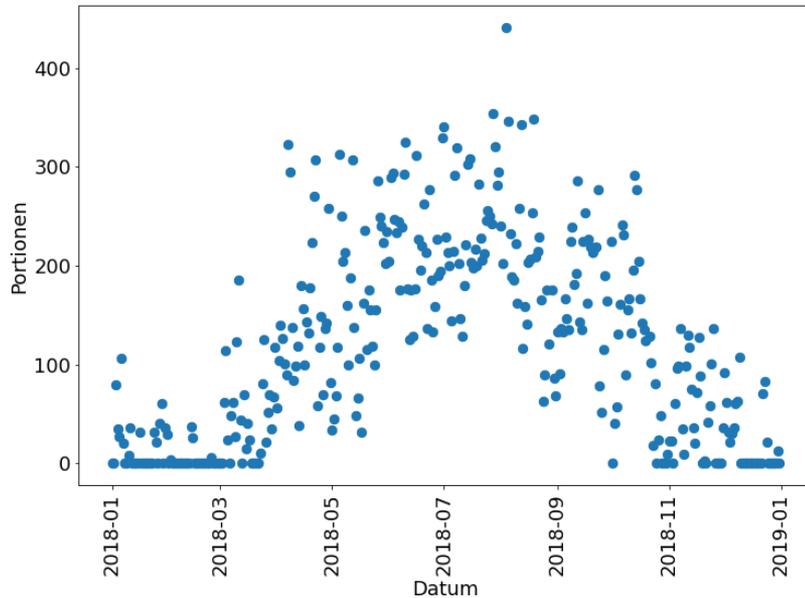


# Lineare Regression

Modelle für lineare Zusammenhänge

## Eisverkäufe 2018



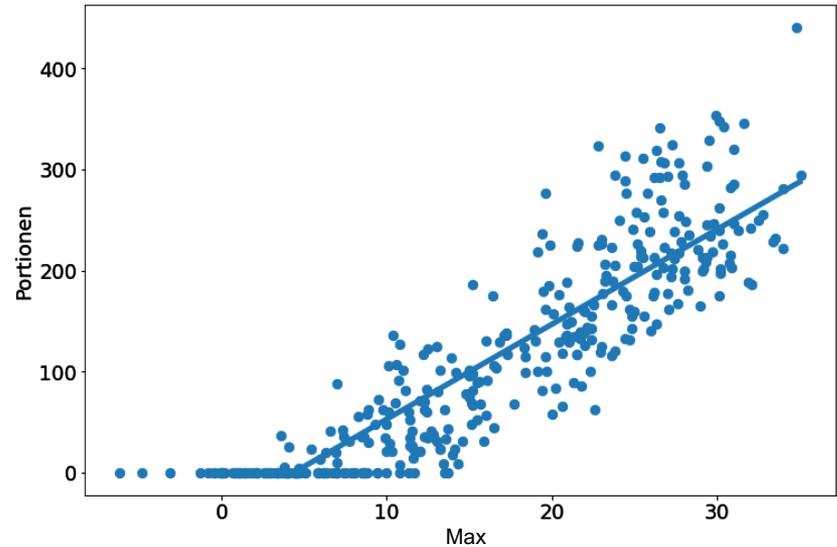
- Offensichtlich: Zusammenhang zwischen Jahreszeit und Eisverkauf
- Aber nicht linear!
- Korrelationskoeffizient daher nicht sinnvoll
- Vermutung: Zusammenhang aufgrund des Wetters
  - Im Winter kalt: wenig Verkäufe
  - Im Sommer warm: viel Verkäufe

## Zusammenhang zwischen Eis und Höchsttemperatur

- Schon eher linear
  - Durch Linie beschreibbar
  - Linie als "Modell" für den Zusammenhang zwischen Temperatur und Verkäufen
  - Eisverkäufe sind abhängig von der Temperatur
  - **Abhängige** Variable: Portionen
  - **Unabhängige** Variable: Temperatur

Erklärte Variable

Erklärende Variable



**Ein Modell beschreibt den Zusammenhang zwischen einer *abhängigen* Variablen und einer oder mehreren *unabhängigen* Variablen.**

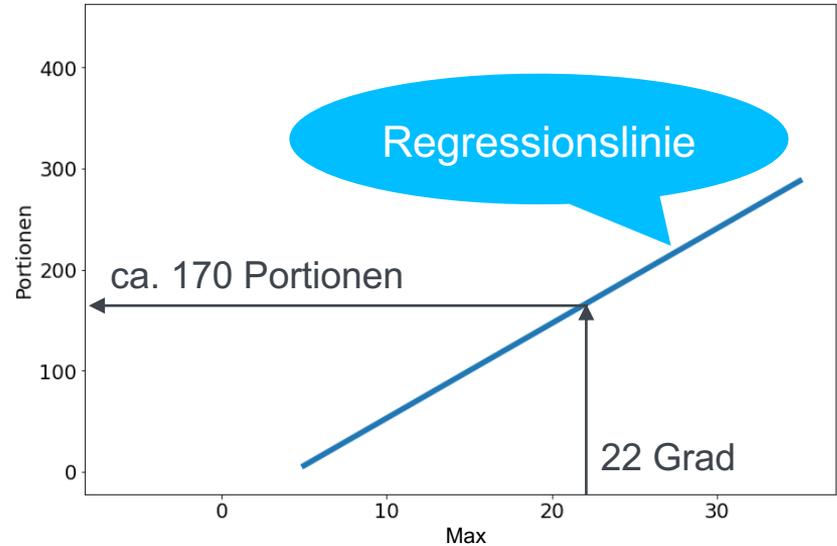
**Die abhängige Variable nennt man auch *erklärte* Variable; die unabhängigen Variablen nennt man auch *erklärende* Variablen.**

# Lineares Modell

- Linie als abstraktes Modell für die Daten
- Starke Vereinfachung: Linie statt Datenwolke
- Kann für Prognosen eingesetzt werden

Lineare  
Regression!

- Unabhängige Variable = Einflussgröße
- Abhängige Variable = Zielgröße



**Bei der linearen Regression bezeichnet man die abhängige Variable auch als Zielgröße, und die unabhängigen Variablen als Einflussgrößen.**



**Quiz - Drag and Drop:  
Abhängige/Unabhängige  
Variable**



**Quiz - Single Choice:  
Abhängige/Unabhängige  
Variable**

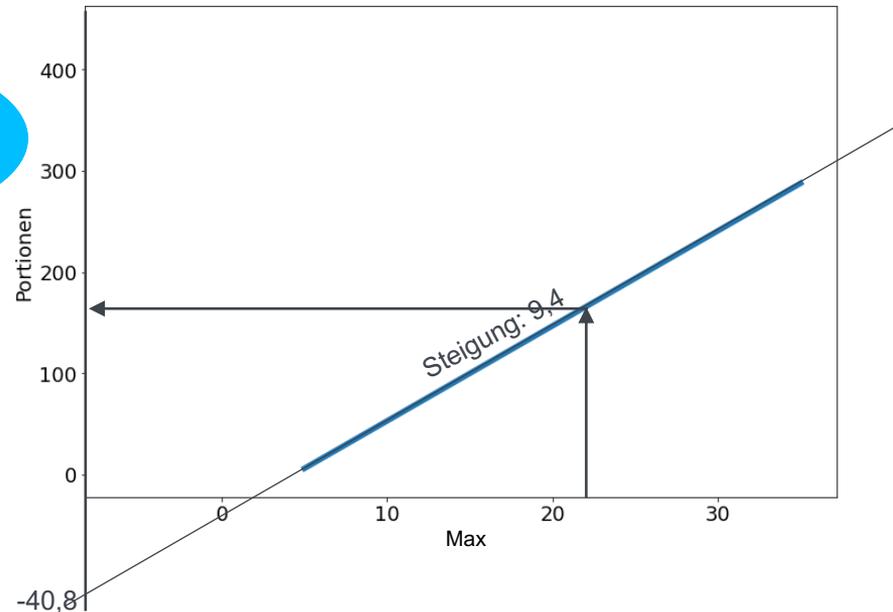
## Prognosen mit diesem Modell

- Geradengleichung für diese Linie:  
 $y = 9,4 x - 40,8$
- x-Achse: Temperatur
- y-Achse: Portionen
- also:  
 $\text{Portionen} = 9,4 * \text{Temperatur} - 40,8$

Nur zwei  
Parameter statt  
365 Punkte!

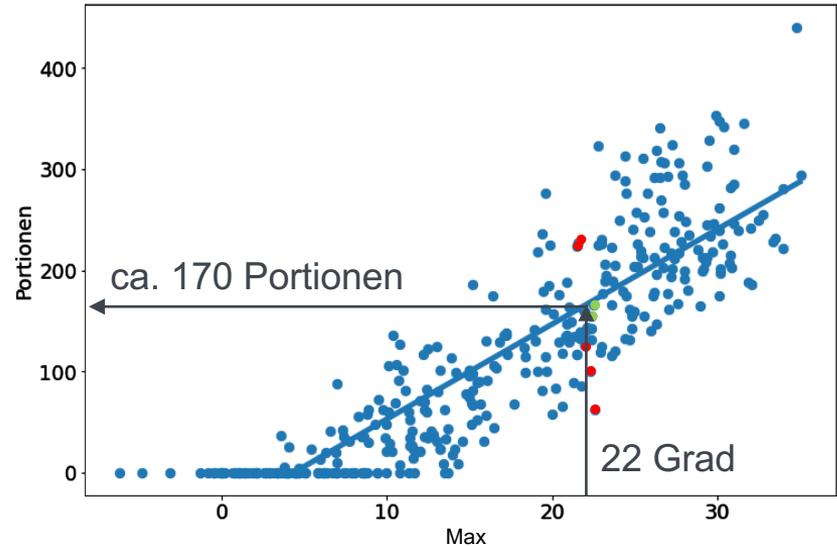
Beispiel: 22 Grad

$$\begin{aligned} \text{Portionen} &= 9,4 * 22 - 40,8 \\ &= 166 \end{aligned}$$



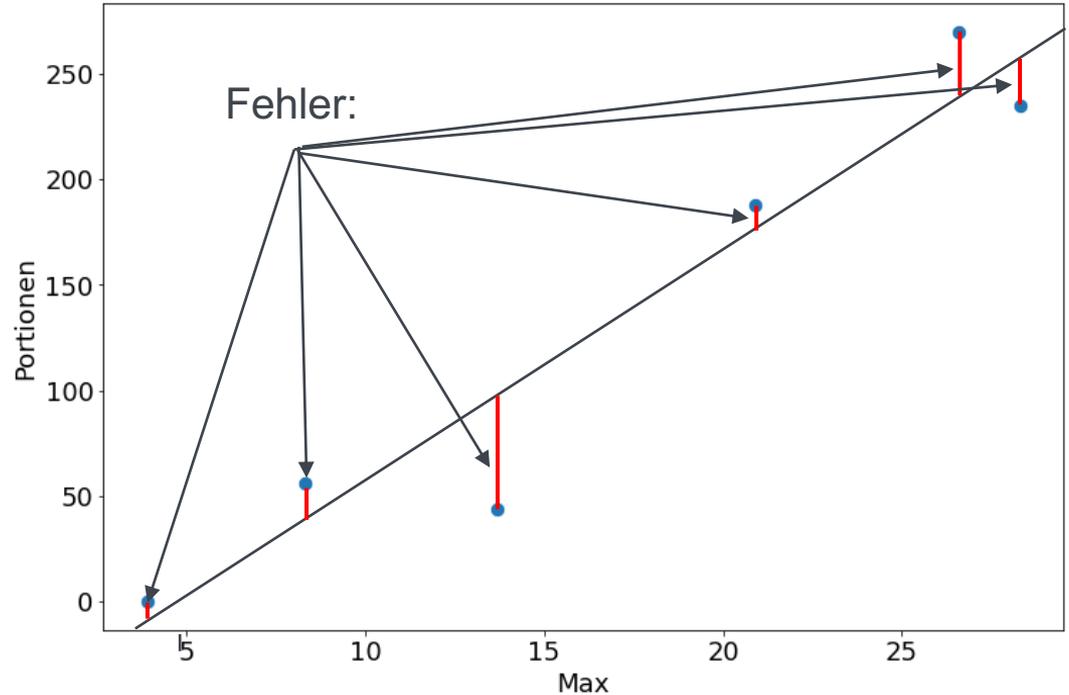
## Fehler mit diesem Modell

- In einigen Fällen Prognose nicht korrekt
- Ursache: die Daten bilden keine exakte Linie
  - Korrelation: 0.89  $\Rightarrow$  stark!
- Maß für die Güte von linearen Regressionsmodellen:
  - der Determinationskoeffizient (oder auch: das Bestimmtheitsmaß)
  - verwandt mit Korrelationskoeffizienten

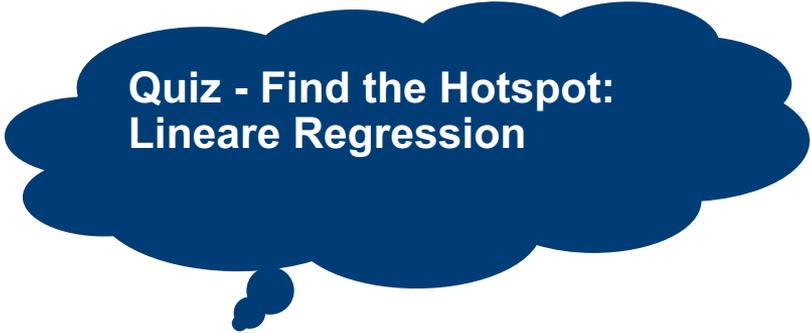


## Wie kommt man zu diesem Modell?

Linie wird so gewählt,  
dass die Summe der  
Fehlerquadrate minimal ist.



**Die Regressionslinie wird so gewählt, dass die Summe der Fehlerquadrate auf den zu modellierenden Daten minimiert wird.**

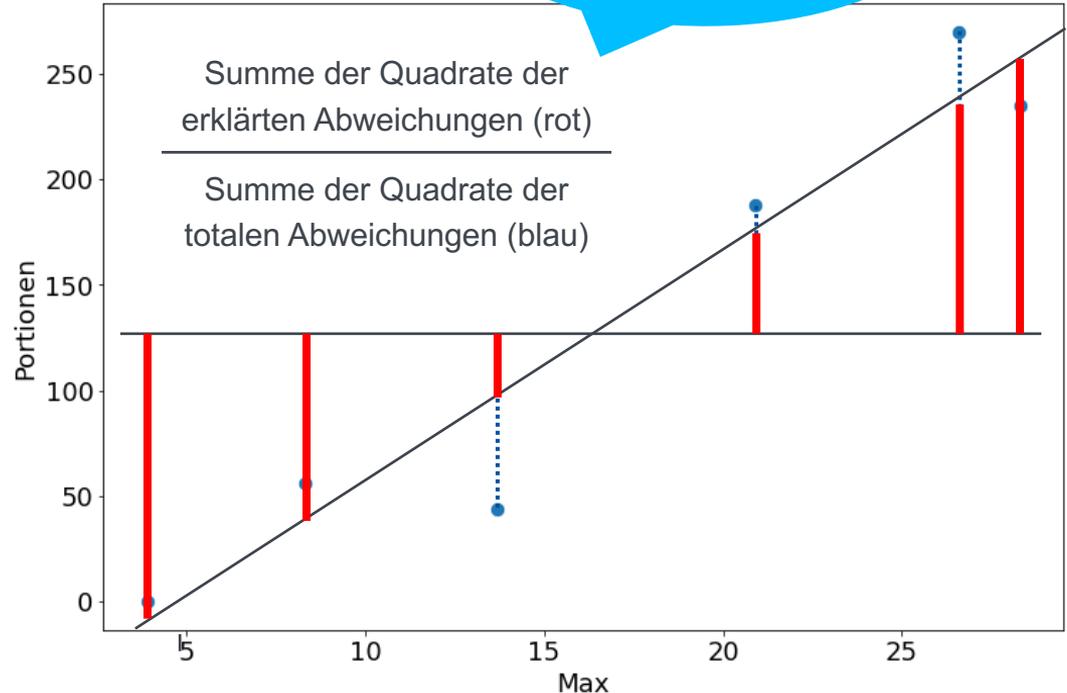


**Quiz - Find the Hotspot:  
Lineare Regression**

## Wie gut ist das Modell?

- Wir suchen eine Erklärung für die Veränderungen bei den verkauften Portionen
- Abweichungen vom Mittelwert erklären/vorhersagen (blau gestrichelt)
- Bewertung: Vergleich mit Abweichung der Vorhersagen vom Mittelwert (rot)
- Gut: die Vorhersagen erklären einen großen Anteil der Abweichungen

Determinationskoeffizient



**Die Güte von linearen  
Regressionsmodellen wird durch  
den Determinationskoeffizienten  
gemessen.**

**Er berechnet sich aus der Summe  
der Quadrate der erklärten  
Abweichungen geteilt durch die  
Summe der Quadrate der totalen  
Abweichungen.\***

**\*Das entspricht der erklärten Varianz geteilt durch die Gesamt-Varianz.**

## Determinationskoeffizient und Korrelationskoeffizient

- Der Determinationskoeffizient ist das Quadrat des Korrelationskoeffizienten
- Er wird daher auch oft als  $R^2$  bezeichnet.

## Dr. Antje Schweitzer

Universität Stuttgart  
Institut für Maschinelle Sprachverarbeitung



**Universität Stuttgart**

Institut für Maschinelle Sprachverarbeitung  
Institut für Software Engineering



**IHK** Industrie- und Handelskammer  
Reutlingen

Reutlingen | Tübingen | Zollernalb



**IHK** Region Stuttgart



**IHK** Industrie- und Handelskammer  
Karlsruhe



# Lizenzbestimmungen

“Lineare Regression” von Antje Schweitzer, KI B<sup>3</sup> / Uni Stuttgart

Das Werk - mit Ausnahme der folgenden Elemente:

- Logos der Verbundpartner und des Förderprogramms
- im Quellenverzeichnis aufgeführte Medien

ist lizenziert unter:

 [CC BY 4.0 \(https://creativecommons.org/licenses/by/4.0/deed.de\)](https://creativecommons.org/licenses/by/4.0/deed.de)

(Namensnennung 4.0 International)