

RPTU Kaiserslautern

# DATENPROJEKT

Datensatz 1

verfasst von

Jens Bleymehl

427741 (Matrikelnummer)

Isaac Lee Zen Xue

427008 (Matrikelnummer)



Grundlagen und Anwendungen der Wahrscheinlichkeitstheorie

Wintersemester 2023 / 2024

Dozent: Christian De Schryver

# Quellenverzeichnis

## Videos

- <https://studyflix.de/statistik/korrelationskoeffizient-22>
-  Boxplot erstellen, Median, unteres/oberes Quartil, Minimum, Maximum | Mathe ...
-  Quantile, Quartile berechnen - Statistik

## Webseiten

- <https://rechneronline.de/korrelation/varianz-standardabweichung.php>

## R1.1 - Übersicht

Dieser Datensatz stammt von der Webseite "Climate Reanalyzer", einer Ressource des Climate Change Institute der University of Maine, und konzentriert sich darauf, die durchschnittliche Meerestemperatur weltweit im Breitengradbereich von 60°N bis 60°S zu analysieren. Diese Temperaturwerte beziehen sich hier jeweils auf den 25. August des betrachteten Jahres, werden von 1981 bis 2023 zusammengefasst und bilden so die Grundlage des Datensatzes. Das Ziel besteht darin, Einblicke in die Temperaturentwicklung der Erde zu erhalten und zu ermitteln, ob diese Entwicklung jahresabhängig sein könnte. Der Datensatz liegt als einzelne Datei in tabellarischer Form vor. Die Datenquelle befindet sich unter [https://climatoreanalyzer.org/clim/sst\\_daily/](https://climatoreanalyzer.org/clim/sst_daily/).

Bemerkung: Durch das Fehlen eines Temperaturwertes bei dem ersten Datenpunkt 1981 wird dieses Jahr aus der Analyse ausgeschlossen.

## R1.2 - Skalenvariante

Skalenvariante für die Betrachtung der Jahre/Temperatur : Intervallskala.

## R1.3 - Benutzte Softwares und Funktionen

Die Analyse dieses Datensatzes wurde unter Verwendung der Software GitHub durchgeführt, zusammen mit der Programmiersprache Python und -umgebung Visual Studio Code, sowie verschiedenen Tools von Google (wie Google Sheets und Google Docs).

## R1.4 - Modus, Mittelwert und Median

Modus der Temperaturen : 20.14°C aus den Jahren 1983, 1988, 1989 und 1990

Daraus folgen die Mittelwerte: 2002.5 (Jahreszahlen) bzw.  $855.59^{\circ}\text{C} / 43 \approx 19.897^{\circ}\text{C}$  (Temperaturen). Median: 2002.5 (Jahreszahl),  $20.35^{\circ}\text{C}$  (Temperatur)

### R1.8 - Spannweite

Der Zeitraum, in dem die Temperaturen aufgezeichnet wurden, ist zwischen 1982 und 2023. Die Temperaturen der Meeresoberfläche variieren zwischen einem Minimum von  $19.91^{\circ}\text{C}$  Celsius und einem Maximum von  $21.09^{\circ}\text{C}$ .

### R1.9 - Mittlere Abweichung vom Median

Die mittlere Abweichung vom Median der Jahreszahlen beträgt 10.50.

Die mittlere Abweichung vom Median der Temperaturen beträgt  $0.21^{\circ}\text{C}$ .

### R1.10 - Stichprobenvarianz

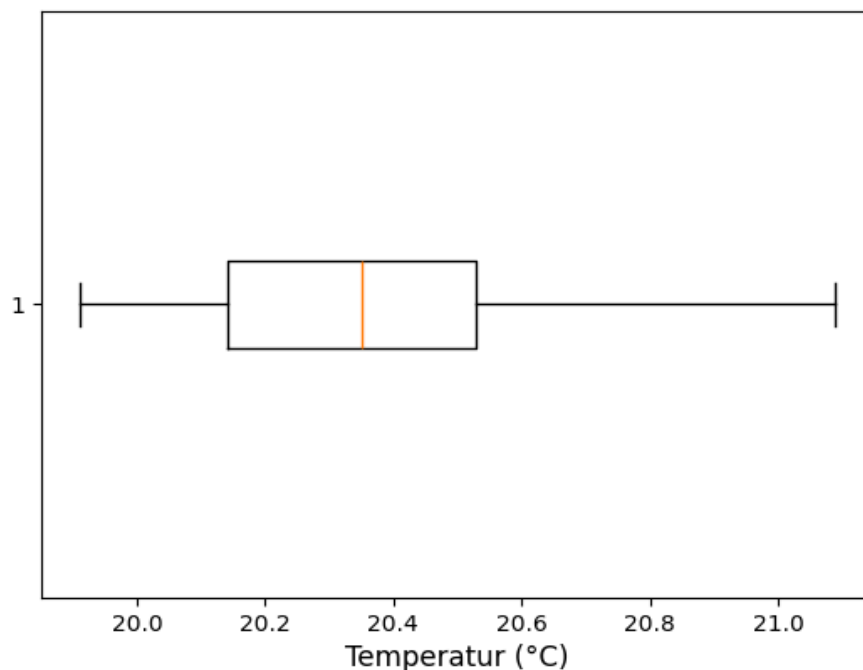
Die Stichprobenvarianz der Jahreszahlen beträgt  $s^2 = 150.5 \Rightarrow$  Standardabweichung  $s \approx 12.27$ .

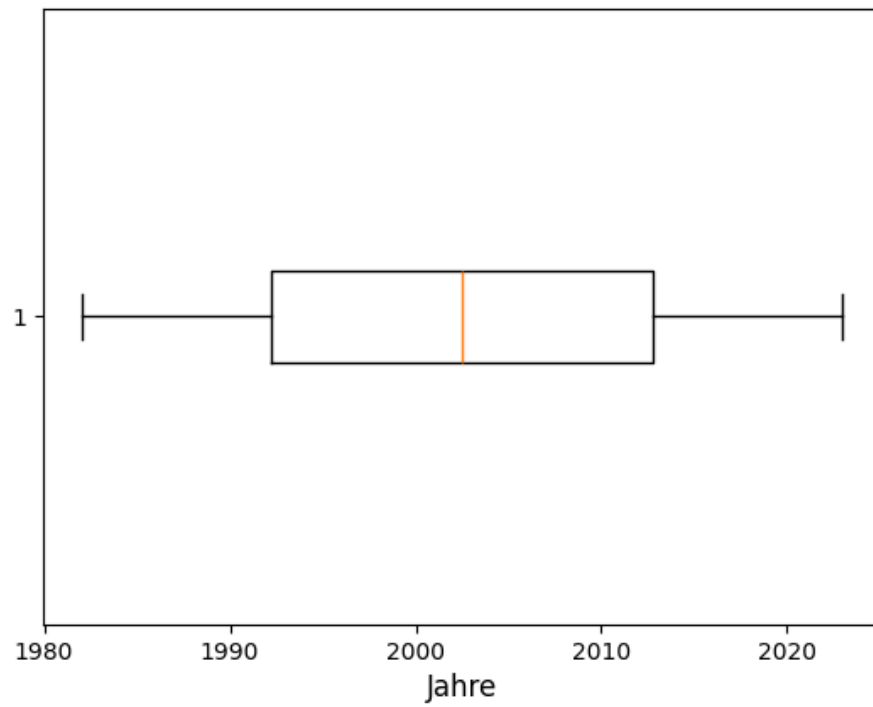
Die Stichprobenvarianz der Temperaturen beträgt  $s^2 = 0.0693 \Rightarrow$  Standardabweichung  $s \approx 0.2632$

### R1.11 - Variationskoeffizient

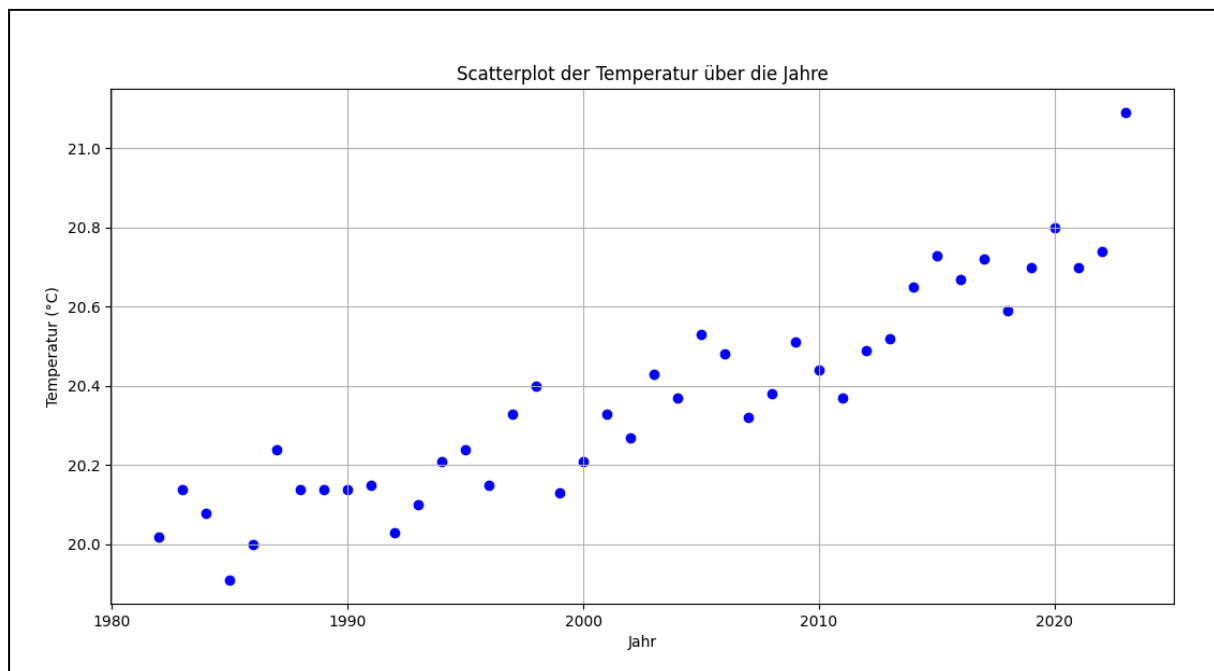
Die Standardabweichung wird durch den Mittelwert geteilt. Der Variationskoeffizient der Jahreszahlen beträgt somit  $\approx 0.006127$  ( $\approx 0.61\%$ ) und der der Temperaturen  $0.0132$  ( $\approx 1.32\%$ ).

### R1.12 - Box-Whisker-Plot der Jahreszahlen und Temperaturen





### R1.13 - Scatterplot



## **R1.15 - Quartile und Dezile (mit linearer Interpolation)**

### Jahreszahlen:

$$Q1 (00.25) = 1991.5$$

$$Q2 (00.50) = 2002.5 \text{ (Median)}$$

$$Q3 (00.75) = 2013.5$$

$$D1 (00.10) = 1985.5$$

$$D2 (00.20) = 1989.5$$

$$D3 (00.30) = 1993.5$$

$$D4 (00.40) = 1998.5$$

$$D5 (00.50) = 2002.5$$

$$D6 (00.60) = 2006.5$$

$$D7 (00.70) = 2011.5$$

$$D8 (00.80) = 2015.5$$

$$D9 (00.90) = 2019.5$$

### Temperaturen:

$$Q1 (00.25) = 20.14^{\circ}\text{C}$$

$$Q2 (00.50) = 20.35^{\circ}\text{C (Median)}$$

$$Q3 (00.75) = 20.56^{\circ}\text{C}$$

$$D1 (00.10) = 20.055^{\circ}\text{C}$$

$$D2 (00.20) = 20.14^{\circ}\text{C}$$

$$D3 (00.30) = 20.15^{\circ}\text{C}$$

$$D4 (00.40) = 20.255^{\circ}\text{C}$$

$$D5 (00.50) = 20.35^{\circ}\text{C (Median)}$$

$$D6 (00.60) = 20.415^{\circ}\text{C}$$

$$D7 (00.70) = 20.505^{\circ}\text{C}$$

$$D8 (00.80) = 20.66^{\circ}\text{C}$$

$$D9 (00.90) = 20.725^{\circ}\text{C}$$

## **R1.16 - Quartilsabstand $R_{Q0.5}$**

Es gilt:  $R_{Q0.5} = Q3 - Q1$ . Somit beträgt der Quartilsabstand der Jahreszahlen :  $R_{Q0.5}$ : 22 und der der Temperaturen  $R_{Q0.5}$ :  $0.42^{\circ}\text{C}$ .

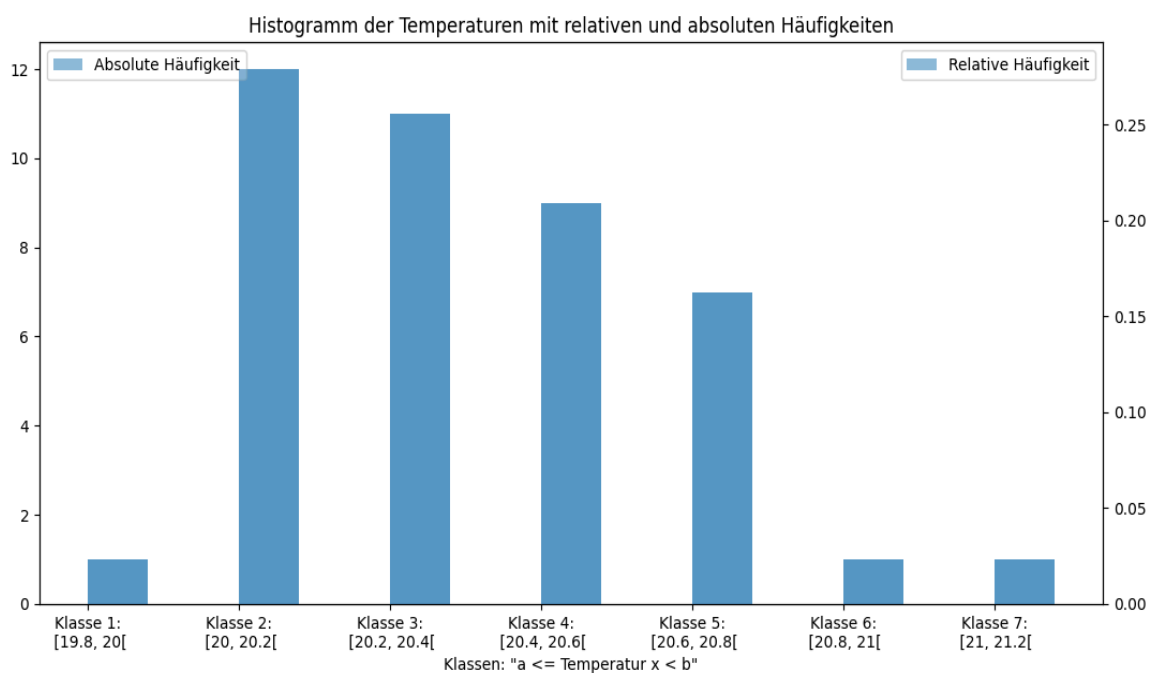
## R1.17 - Kovarianz

Die Kovarianz zwischen Temperaturdaten und Jahreszahlen beträgt: 2.9872.

## R1.18 - Korrelationskoeffizient

Der Korrelationskoeffizient zwischen Temperaturdaten und Jahreszahlen beträgt: 0.9472.

## R1.19 - Histogramm



Der Rangkorrelationskoeffizient nach Spearman zwischen Temperaturdaten und Jahreszahlen beträgt: 0.9332.

## R1.14 - Fazit

Der Datensatz deutet wegen des hohen Korrelationskoeffizienten darauf hin, dass Temperatur und Jahreszahl stark miteinander korrelieren.

Die Oberflächentemperaturen steigen im Laufe der Zeit allmählich, was auch durch den Korrelationskoeffizienten im Zusammenhang mit den steigenden Jahreszahlen zu begründen ist. Es ist also wahrscheinlich, dass die Temperaturentwicklung (anhand dieser Daten) von der Zeitentwicklung abhängt.