

Grundlagen und Anwendung der Wahrscheinlichkeitstheorie

Wintersemester 2022/2023

L02 Projekt Datenanalyse

Gruppe 06

Björn Kämpfert

Felix Luis Fritsche

Benedict Casper Satter

Inhalt

Datensatz 2

2.1 Struktur und Inhalt

2.2 Skalenvariante

2.3 Verwendete Software und genutzte Funktionen

2.5 – 2.7 Urliste, Rangliste und Excel-Datei (Anhang)

2.8 Modus, arithmetischer Mittelwert, Median

2.9 Spannweite

2.10 mittlere Abweichung vom Median

2.11 Stichprobenvarianz

2.12 Variationskoeffizient

2.13 Box-Whisker-Plot

2.14 Scatterplot

2.16 Zusammenfassung

2.17 / 2.18 Quartile und Dezile, Quartilabstand

2.19 Kovarianz

2.20 Korrelationskoeffizient

2.1 / 2.3 Struktur und Inhalt / Maßnahmen zur Datenbereinigung

Nach Absprache mit Dr.-Ing. Christian De Schryver haben wir für Datensatz 2 und 3 uns die Datensätze der Gruppe 05 besorgt, da unsere unrealistische Werte hatten.

Dieser Datensatz befasst sich mit Exportquoten der Jahre 2015 bis 2021, Werte für die Jahre 2016 und 2020 fehlen jedoch, weshalb wir uns dazu entschieden haben, die Werte aus dem Datensatz zu entfernen. Die vorliegende Datei ist nach Jahr sortiert, die Exportquote wird in Prozent angegeben. Die Zeilenanzahl liegt bei 14 Zeilen, die Quelle und der Stand des Datensatzes sind in den letzten Zeilen angegeben. Die Datei stammt vom Statistischen Bundesamt (Destatis), der angegebene Stand des Datensatzes ist der 12. September 2022.

Es liegt eine Datei mit Datensatz in UTF 8-Kodierung vor. Der Datensatz ist in einer CSV-Datei mit einer Größe von 170 Bytes vorhanden. Die Trennung in den Urdateien verwendet als Dezimalzeichen ein Komma und als Trennzeichen ein Semikolon.

2.4 Verwendete Software und genutzte Funktionen

Zum Bearbeiten des Datensatzes wird Visual Studio Code benutzt, mit der Programmiersprache Python, dem Jupyter Notebook und den extra Bibliotheken pandas, numpy, matplotlib.

2.8 Modus, arithmetischer Mittelwert, Median

Modus: n.A., da es jeden Wert nur einmal gibt

Arithmetischer Mittelwert

| | |
|-------------|--------|
| Jahr | 2018 |
| Exportquote | 47.02% |

Median

| | |
|-------------|--------|
| Jahr | 2018 |
| Exportquote | 47.00% |

2.9 Spannweite

| | |
|-------------|------|
| Jahr | 6.0 |
| Exportquote | 0.6% |

2.10 mittlere Abweichung vom Median

| | |
|-------------|--------|
| Jahr | 1.6 |
| Exportquote | 0.184% |

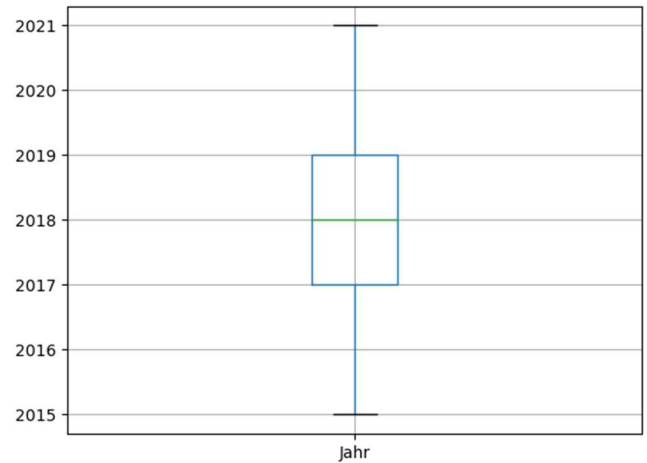
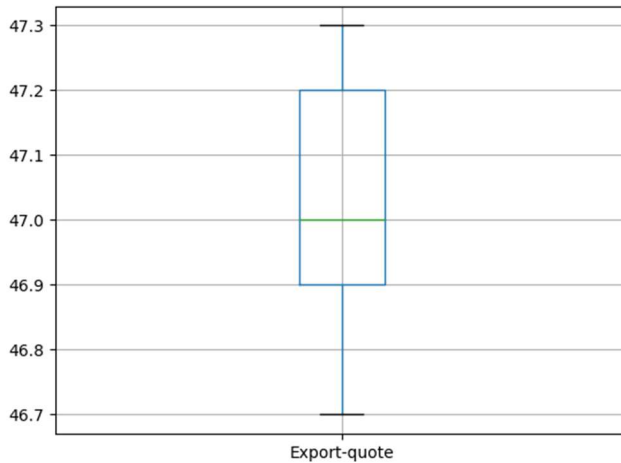
2.11 Stichprobenvarianz

| | |
|-------------|--------|
| Jahr | 5000 |
| Exportquote | 0.057% |

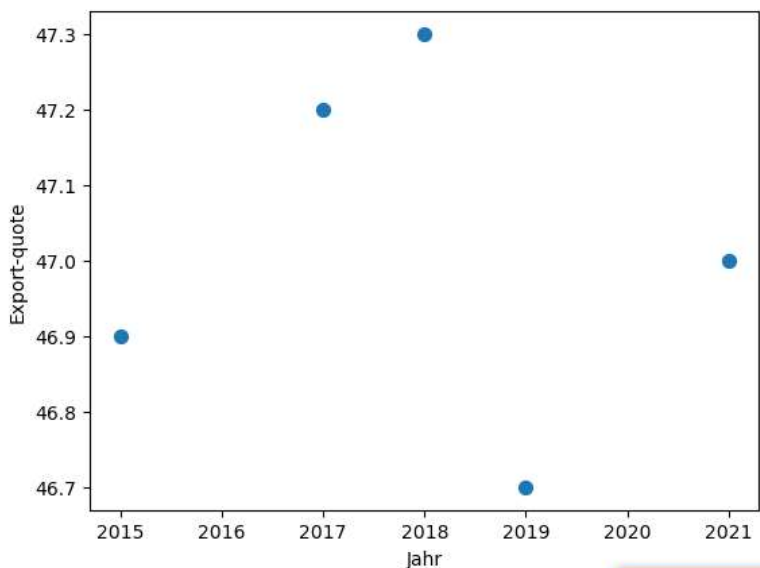
2.12 Variationskoeffizient

| | |
|-------------|-----------|
| Jahr | 0.111 |
| Exportquote | 0.507756% |

2.13 Box-Whisker-Plot



2.14 Scatterplot



2.16 Zusammenfassung

Der Datensatz erfasst mit 7 Jahren einen eher kleinen Zeitraum, besonders wenn man beachtet, dass zwei der Jahre keine Angaben haben. Der arithmetische Mittelwert für die Exportquote liegt bei 47.02% und somit nur .02% vom Median entfernt, die mittlere Abweichung vom Median von nur 0.184% zeigt, dass sich die Exportquote über diesen Zeitraum nicht groß verändert hat, beziehungsweise dass die Werte alle nahe beieinander liegen. Die Spannweite der Exportquote beträgt aber auch nur 0.6 Prozent, wodurch diese Erkenntnis bestätigt wird. Der Variationskoeffizient fällt mit 0.507

aufgrund der geringen Distanz der Werte eher gering aus. Die Varianz von 0.057 zeigt einen geringen (quadrierten) Abstand der Werte vom arithmetischen Mittel.

2.17 / 2.18 Quartile und Dezile, Quartilabstand

Quartilsabstand: $Q3 - Q1 = 0.3\%$

| | Jahreszahl | Exportquote |
|------------|------------|-------------|
| <i>min</i> | 2015 | 46.7 % |
| 25% | 2017 | 46.9 % |
| 50% | 2018 | 47.0 % |
| 75% | 2019 | 47.2 % |
| <i>max</i> | 2021 | 47.3 % |

| | Jahreszahl | Exportquote |
|-----|------------|-------------|
| 0.1 | 2015.8 | 46.78 % |
| 0.2 | 2016.6 | 46.86 % |
| 0.3 | 2017.2 | 46.92 % |
| 0.4 | 2017.6 | 46.96 % |
| 0.5 | 2018.0 | 47.00 % |
| 0.6 | 2018.4 | 47.08 % |
| 0.7 | 2018.8 | 47.16 % |
| 0.8 | 2019.4 | 47.22 % |
| 0.9 | 2020.2 | 47.26 % |
| 1.0 | 2021.0 | 47.30 % |

2.19 Kovarianz

-0.05

2.20 Korrelationskoeffizient

-0.094

Quellenverzeichnis

Folien aus Vorlesung

<https://pandas.pydata.org/>

<https://studyflix.de/statistik/empirische-varianz-2016>

<https://www.geeksforgeeks.org/>

<https://matheguru.com/stochastik/kovarianz.html>