

L02 Projekt Datenanalyse

In Grundlagen und Anwendung der
Wahrscheinlichkeitstheorie WS22/23

Gruppe 01

Lukas Annawald, Elias Motzel, Christian Käufer

Inhalt

- Datensatz 1 (ab Seite 2)
- Datensatz 2 (ab Seite 10)
- Datensatz 3 (ab Seite 17)
- Datensatz 4 (ab Seite 24)
- Quellenverzeichnis (Seite 26)

Datensatz 1

- 1.1 Struktur und Inhalt des Datensatzes
- 1.2 Skalenvariante
- 1.3 verwendete Software und genutzte Funktionen
- 1.4 Urliste
- 1.5 Rangliste
- 1.6 Excel-Datei
- 1.7 Modus, arithmetischer Mittelwert, Median
- 1.8 Spannweite
- 1.9 mittlere Abweichung vom Median
- 1.10 Stichprobenvarianz
- 1.11 Variationskoeffizient
- 1.12 Box-Whisker-Plot
- 1.13 Scatterplot
- 1.14 Zusammenfassung

1.1 Struktur und Inhalt des Datensatzes

Der Datensatz gibt Auskunft über die Exportquoten einzelner Jahre von 2015 bis 2021. Der Datensatz ist semistrukturiert.

Der Datensatz besteht aus 14 Zeilen. Zunächst wird in Zeile 3 definiert, dass jeweils Jahr und eine Exportquote durch ein „ ; “ getrennt werden. Die Exportquote wird in Prozent angegeben, wie man in Zeile 4 erkennen kann. In den folgenden Zeilen werden nun jeweils eine Jahreszahl und eine Dezimalzahl gegeben. Die Jahreszahlen sind absteigend von 2021 bis 2015 in Schritten von einem Jahr gegeben. Den Jahreszahlen wird jeweils eine Dezimalzahl zugeordnet. Es lässt sich keine Sortierung der Dezimalzahlen erkennen. Die Daten sind daher nach Jahreszahlen absteigend sortiert. Die Exportquoten (Dezimalzahlen) schwanken zwischen 43,0% und 47,3%.

Zuletzt sind in Zeile 13 und 14 Stand und Quelle der Daten angegeben.

Die Daten stammen vom Statistischen Bundesamt (Destatis) und sind auf dem Stand vom 12. September 2022. Die Daten liegen in einer .csv-Datei vor.

Da die meisten Berechnungen bei einer Intervallskala für die Variable “Jahreszahl” nicht zielführend sind, finden sich Berechnungen für “Jahreszahl” nur unter den entsprechend sinnvollen Unterpunkten.

1.2 Skalenvariante

Bei Jahren handelt es sich um eine Intervallskala. Die Intervallskala zählt zu den metrischen Skalen, die quantitative Werte wiedergeben. Eine Intervallskala unterteilt sich immer in gleich große Skalenabschnitte. Zwischen 1994 und 1999 liegt die gleiche Zeit wie zwischen 2001 und 2006.

Bei der Exportquote handelt es sich um eine Verhältnisskala. Die Exportquote ist eine Kennzahl, die das Verhältnis der Exporte zum

Bruttoinlandsprodukt einer Volkswirtschaft wiedergibt.

$$Exportquote = \frac{Exporte}{BIP} \times 100$$

1.3 Verwendete Software und genutzte Funktionen

VS Code (nur als Editor)

Excel: MODALWERT, MITTELWERT, MEDIAN, VARIANZ, Diagrammfunktionen, QUARTILE(ArrayA;1...), QUANTIL(Array;10/100...), KOVARIANZ.P, KORREL, STABW

1.4 Urliste

Siehe Data 1

1.5 Rangliste

Siehe Data 1

1.6 Excel Datei

Siehe Data 1

1.7 Modus, arithmetischer Mittelwert, Median

“Exportquote”:

Modus: nicht vorhanden, da jeder Wert genau einmal existiert

Arithmetischer Mittelwert: $46,31428571 \approx 46,31$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Median: 46,9

$$x_{med} = \begin{cases} x_{(\frac{n+1}{2})} & , \text{ falls } n \text{ ungerade} \\ \frac{1}{2} \cdot (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & , \text{ falls } n \text{ gerade} \end{cases}$$

“Jahreszahl”:

Modus: nicht vorhanden, da jeder Wert genau einmal existiert

arithmetischer Mittelwert: 2018

Median: 2018

1.8 Spannweite

“Exportquote”:

$$R = x_{\max} - x_{\min}$$

$$R = 47,3 - 43 = 4,3$$

“Jahreszahl”

$$R = 2021 - 2015 = 6$$

1.9 mittlere Abweichung vom Median

“Exportquote”:

$$MA_x = \frac{1}{n} \cdot \sum_{i=1}^n |x_i - x_{med}|$$

$$MA_x = 1/7 \cdot \sum_{i=1}^n |x_i - 4|$$

$$= 0,814285714$$

$$\approx 0,81$$

“Jahreszahl”:

$$MA_x = 1/7 \cdot \sum_{i=1}^n |x_i - 2018|$$

$$= 1,71428571$$

$$\approx 1,71$$

1.10 Stichprobenvarianz

“Exportquote”:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (\bar{x} - x_i)^2 \quad s^2 = \frac{1}{n-1} \cdot \left[\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right]$$

$$\text{Varianz} = 2,291428571$$

$$\approx 2,29$$

1.11 Variationskoeffizient

“Exportquote”

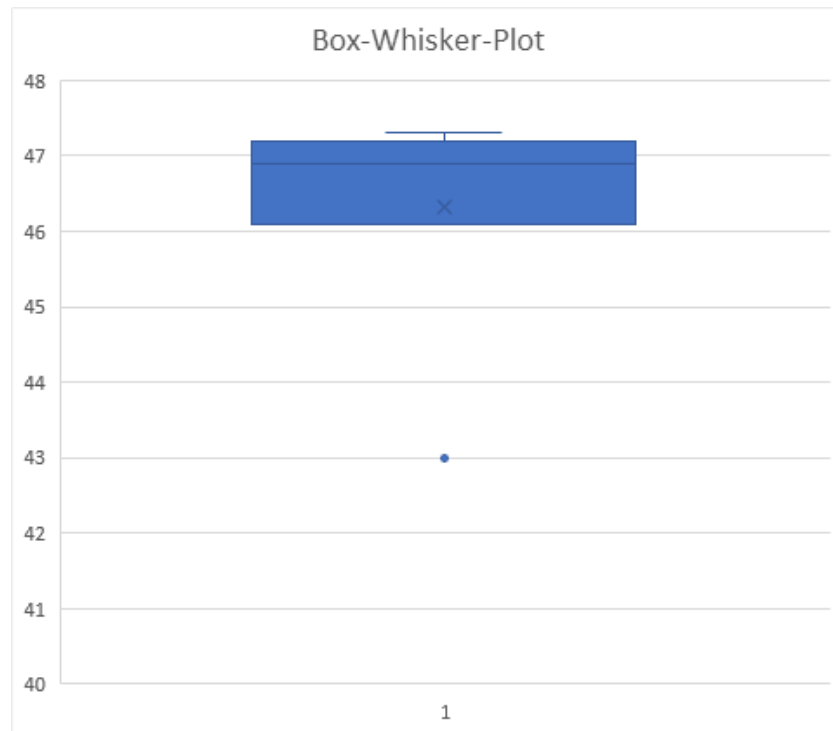
$$v = \frac{s}{\bar{x}}, \quad \bar{x} \neq 0$$

$$V = 0,032684225$$

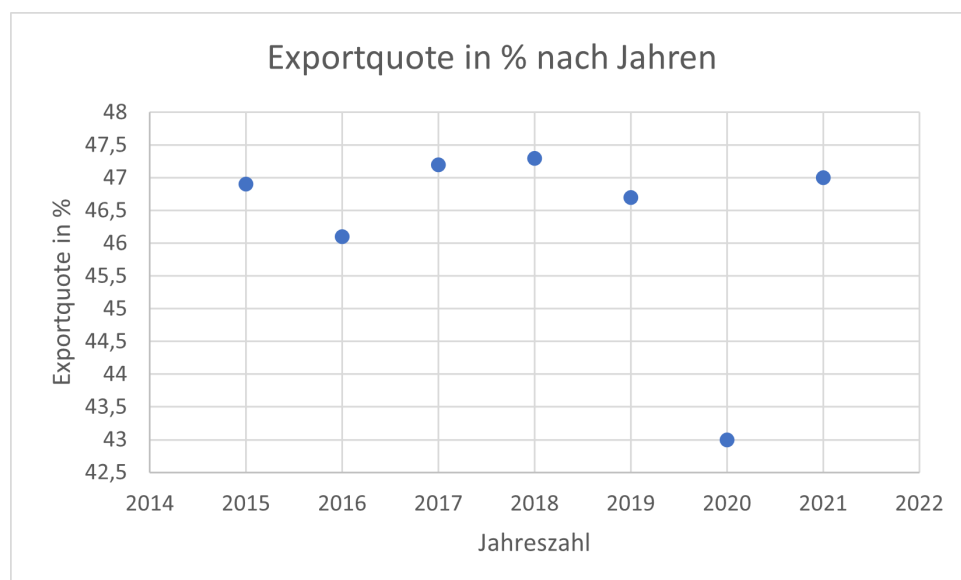
$$\approx 0,03$$

1.12 Box-Whisker-Plot

“Exportquote”:



1.13 Scatterplot



1.14 Zusammenfassung

Der gegebene Datensatz enthält die Exportquote von sieben Jahren, beginnend im Jahr 2015 bis 2021. Die Exportquoten schwanken zwischen 43% (2020) und 47,3% (2018) und haben insgesamt eine Spannweite von 4,3 Prozentpunkten, denn das Jahr 2020 macht eine Ausnahme und weicht um 3,9 Prozentpunkte vom Median ab. Dies wird auch im Box-Whisker-Plot sehr deutlich, da nur der Wert des Jahres 2020 unterhalb der Box liegt. Dies ist vermutlich auf die Corona-Pandemie und den damit einhergehenden Wirtschaftseinbruch zurückzuführen. Der arithmetische Mittelwert der Exportquote liegt bei 46,3%, während der Median bei 46,9% liegt. Es gibt keinen Modus für die Exportquote, da jeder Wert nur einmal auftritt. Die Jahreszahlen haben ebenfalls keinen Modus, da jeder Wert nur einmal auftritt. Die mittleren Abweichungen vom Median betragen 1,71 für die Jahreszahlen und 0,81 für die Exportquoten. Die Stichprobenvarianz beträgt ungefähr 4,7 für die Jahreszahlen und 2,29 für die Exportquoten. Der Variationskoeffizient ist 0,1 für die Jahreszahlen und 0,03 für die Exportquoten. Die Kovarianz zwischen den beiden Variablen beträgt ungefähr -0,91 und der Korrelationskoeffizient beträgt ungefähr -0,33, was auf eine mäßig starke negative Korrelation zwischen den Variablen hinweist. Der Rangkorrelationskoeffizient beträgt ungefähr -0,07 und zeigt ebenfalls eine schwache, negative Korrelation.

1.15 Quartile und Dezile

“Exportquote”:

1.Quartil		46,4	1.Dezil		44,86
2.Quartil/Median		46,9	2.Dezil		46,22
3.Quartil		47,1	3.Dezil		46,58
			4.Dezil		46,78
Quartilsabstand(Q3-Q1)		0,7	5.Dezil		46,9
			6.Dezil		46,96
			7.Dezil		47,04
			8.Dezil		47,16
			9.Dezil		47,24
			10.Dezil/Maximum		47,3

1.16 Quartilsabstand

“Exportquote”:

$$R_{Q0.5} = Q3 - Q1 = 0,7$$

1.17 Kovarianz

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

s = -0,914285714

1.18 Korrelationskoeffizient

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

r = -0,32619115

1.19 Klassen

“Exportquote”:

Klasse	Klassengrenze	Klassenhäufigkeit
1 unterhalb Mittelwert, außerhalb Standardabweichung	[0;44,8)	1
2 oberhalb Mittelwert, außerhalb Standardabweichung	(47,81;100]	0
3 Mittelwert + Standardabweichung	[46,31;47,81]	6
4 Mittelwert - Standardabweichung	[44,8;46.31)	0

“Jahreszahlen”:

Klasse	Klassengrenze	Klassenhäufigkeit
1 Jahreszahlen	[2015;2021]	7

1.20 Kontingenztafel (klassifizierte Daten)

Exportquote Jahreszahlen	1	2	3	4	Summe
1	1/7	0	6/7	0	1
Summe	1/7	0	6/7	0	1

1.21 Rangkorrelationskoeffizient

$$r_{sp} = \frac{s_{R_{xy}}}{s_{R_x} \cdot s_{R_y}} = \frac{\sum_{i=1}^n (R_{xi} - \overline{R_X})(R_{yi} - \overline{R_Y})}{\sqrt{\sum_{i=1}^n (R_{xi} - \overline{R_X})^2} \cdot \sqrt{\sum_{i=1}^n (R_{yi} - \overline{R_Y})^2}}$$

$$r = -0,0714287$$

Datensatz 2

- 2.1 Struktur und Inhalt des Datensatzes
- 2.2 bereinigte Daten
- 2.3 Maßnahmen zur Datenbereinigung
- 2.4 Verwendete Software und genutzte Funktionen
- 2.5 Excel-Datei
- 2.6 Urliste
- 2.7 Rangliste
- 2.8 Modus, arithmetischer Mittelwert und Median
- 2.9 Spannweite
- 2.10 mittlere Abweichung vom Median
- 2.11 Stichprobenvarianz
- 2.12 Variationskoeffizient
- 2.13 Box-Whisker-Plot
- 2.14 Scatterplot
- 2.15 weitere grafische Darstellung
- 2.16 textuelle Zusammenfassung
- 2.17 Quartile und Dezile
- 2.18 Quartilsabstand $R_Q0.5$
- 2.19 Kovarianz
- 2.20 Korrelationskoeffizient

2.1 Struktur und Inhalt des Datensatzes

Der semistrukturierte Datensatz enthält Informationen über die Exportquoten von 2015 bis 2021 und besteht aus 14 Zeilen. In der dritten Zeile wird definiert, dass Jahr und Exportquote durch das Trennzeichen " ; " getrennt werden. Die Exportquote wird als Prozentwert in der vierten Zeile angegeben. In den folgenden Zeilen werden nun jeweils eine Jahreszahl und eine Dezimalzahl gegeben. Die Jahreszahlen sind absteigend von 2021 bis 2015 in Schritten von einem Jahr gegeben. Den Jahreszahlen wird jeweils eine Dezimalzahl zugeordnet. Es lässt sich keine Sortierung der Dezimalzahlen erkennen. Die Daten sind daher nach Jahreszahlen absteigend sortiert. Die Exportquoten (Dezimalzahlen) schwanken zwischen 43,0% und 47,3%. Die Exportquote für das Jahr 2020 ist nicht verfügbar (NaN). Ebenso verhält es sich im Jahr 2016 (Haha).

Zuletzt werden in den Zeilen 13 und 14 Stand und Quelle der Daten angegeben.

Die Daten stammen vom Statistischen Bundesamt (Destatis) und sind auf dem Stand vom 12. September 2022. Die Daten liegen in einer .csv-Datei vor.

Da einige Berechnungen bei einer Intervallskala für die Variable "Jahreszahl" nicht zielführend sind, finden sich Berechnungen für "Jahreszahl" nur unter den entsprechend sinnvollen Unterpunkten.

2.2 bereinigte Daten

siehe Data 2

2.3 Maßnahmen zur Datenbereinigung

Um den Datensatz zu bereinigen, gibt es mehrere Möglichkeiten. Zum Beispiel kann man die fehlenden Werte auslassen oder auch durch einen Mittelwert ersetzen.

Eine weitere Möglichkeit wäre die Verwendung einer Interpolationsmethode, um die fehlenden Werte auf der Grundlage der vorhandenen Daten zu schätzen. Hierbei muss jedoch beachtet werden, dass Interpolation eine Annahme über die Kontinuität der Daten zwischen den verfügbaren Werten beinhaltet und somit auch Unsicherheiten in der Schätzung beinhalten kann.

Letztendlich hängt die Wahl der Methode davon ab, wie genau die Ergebnisse sein sollen und welche Annahmen bei der Analyse getroffen werden können.

Da die meisten Exportquoten im Datensatz jedoch nur geringfügig variieren und die Abweichungen zwischen den Jahren gering sind, ist es durchaus vertretbar, den Mittelwert der vorhandenen Werte für das fehlende Jahr zu verwenden. Allerdings sollte beachtet werden, dass dies eine Annahme ist und dass der Mittelwert möglicherweise nicht genau den tatsächlichen Wert für das fehlende Jahr widerspiegelt.

2.4 Verwendete Software und genutzte Funktionen

VS Code (nur als Editor)

Excel: MODALWERT, MITTELWERT, MEDIAN, VARIANZ, Diagrammfunktionen,
QUARTILE(ArrayA;1...), QUANTIL(Array;10/100...), KOVARIANZ.P, KORREL, STABW

2.5 Excel-Datei

siehe Data 2

2.6 Urliste

siehe Data 2

2.7 Rangliste

siehe Data 2

2.8 Modus, arithmetischer Mittelwert, Median

“Exportquote”:

Modus: 47,02

Arithmetischer Mittelwert: 47,02

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Median: 47,02

$$x_{med} = \begin{cases} x_{(\frac{n+1}{2})} & , \text{ falls } n \text{ ungerade} \\ \frac{1}{2} \cdot (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & , \text{ falls } n \text{ gerade} \end{cases}$$

“Jahreszahl”:

Modus: nicht vorhanden, da jeder Wert genau einmal existiert

arithmetischer Mittelwert: 2018

Median: 2018

2.9 Spannweite

"Exportquote":

$$R = x_{\max} - x_{\min}$$

$$R = 47,3 - 46,7 = 0,6$$

"Jahreszahl"

$$R = 2021 - 2015 = 6$$

2.10 mittlere Abweichung vom Median

"Exportquote":

$$MA_x = \frac{1}{n} \cdot \sum_{i=1}^n |x_i - x_{med}|$$

$$MA_x = 1/7 \cdot \sum_{i=1}^n |x_i - 4|$$

$$= 0,13142857$$

$$\approx 0,13$$

"Jahreszahl":

$$MA_x = 1/7 \cdot \sum_{i=1}^n |x_i - 2018|$$

$$= 1,71428571$$

$$\approx 1,71$$

2.11 Stichprobenvarianz

"Exportquote":

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (\bar{x} - x_i)^2 \quad s^2 = \frac{1}{n-1} \cdot \left[\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right]$$

$$\text{Varianz} = 0,038$$

2.12 Variationskoeffizient

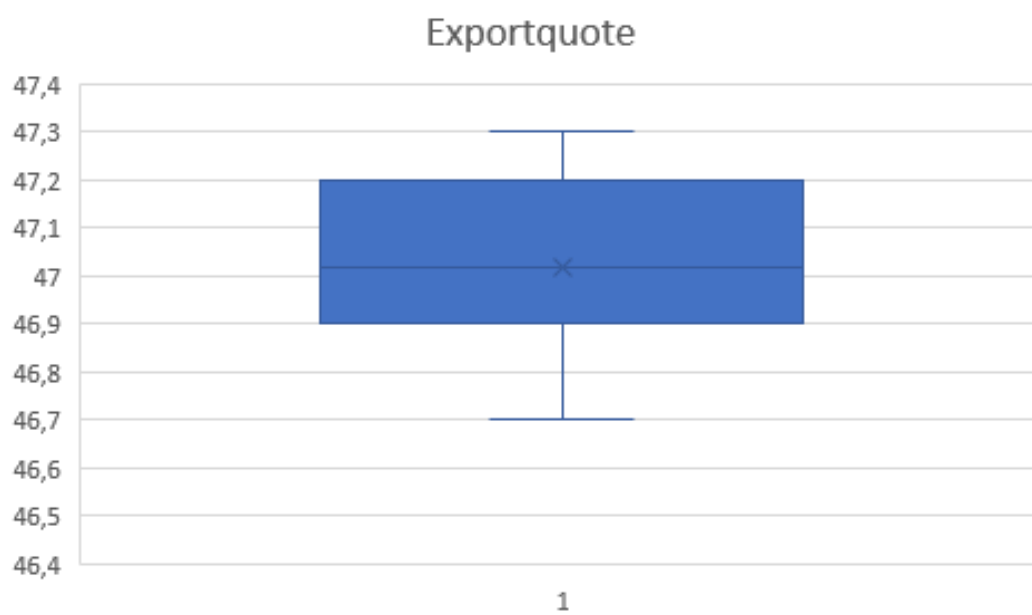
“Exportquote”

$$v = \frac{s}{\bar{x}}, \quad \bar{x} \neq 0$$

V = 0,00414581

≈ 0,004

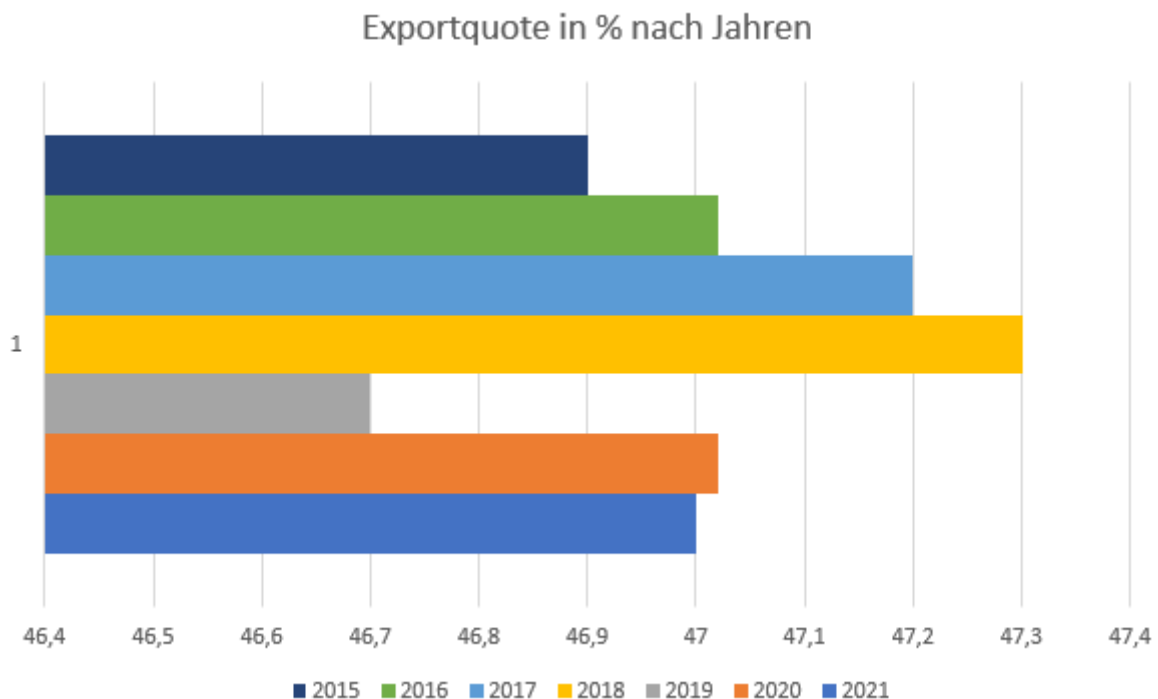
2.13 Box-Whisker-Plot



2.14 Scatterplot



2.15 weitere grafische Zusammenfassung



2.16 textuelle Zusammenfassung

Der gegebene Datensatz enthält die Exportquote von sieben Jahren, beginnend im Jahr 2015 bis 2021. Die Exportquoten schwanken zwischen 46,7% (2019) und 47,3% (2018) und haben insgesamt eine Spannweite von 0,6%. Der arithmetische Mittelwert und Median der Exportquote liegen beide bei 47,02%. Der Modus für die Exportquote beträgt durch die Bereinigung 47,02%(2016,2020). Die Jahreszahlen haben keinen Modus, da jeder Wert nur einmal auftritt. Der arithmetische Mittelwert und Median der Jahreszahlen sind beide 2018. Die Stichprobenvarianz beträgt ungefähr 4,7 für die Jahreszahlen und 0,038 für die Exportquoten. Der Variationskoeffizient ist 0,1 für die Jahreszahlen und 0,004 für die Exportquoten. Die Kovarianz zwischen den beiden Variablen beträgt ungefähr -0,029 und der Korrelationskoeffizient beträgt ungefähr -0,079, was auf eine schwache, negative Korrelation zwischen den Variablen hinweist. Der Rangkorrelationskoeffizient beträgt ungefähr -0,072 und zeigt ebenfalls eine schwache, negative Korrelation.

2.17 Quartile und Dezile

“Exportquote”:

1.Quartil		46,95	1.Dezil		46,82
2.Quartil/Median		47,02	2.Dezil		46,92
3.Quartil		47,11	3.Dezil		46,98
			4.Dezil		47,008
Quartilsabstand(Q3-Q1)		0,16	5.Dezil		47,02
			6.Dezil		47,02
			7.Dezil		47,056
			8.Dezil		47,164
			9.Dezil		47,24
			10.Dezil/Maximum		47,3

2.18 Quartilsabstand

“Exportquote”:

$$R_{Q0.5} = Q3 - Q1 = 0,16$$

2.19 Kovarianz

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$s = -0,0285714$$

2.20 Korrelationskoeffizient

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$r = -0,072075$$

Datensatz 3

- 3.1 Struktur und Inhalt des Datensatzes
- 3.2 zusammengeführte Datei
- 3.3 bereinigte Daten
- 3.4 Maßnahmen zur Datenbereinigung
- 3.5 Excel-Datei
- 3.6 Verwendete Software und genutzte Funktionen
- 3.7 Urliste
- 3.8 Rangliste
- 3.9 Modus, arithmetischer Mittelwert und Median
- 3.10 Spannweite
- 3.11 mittlere Abweichung vom Median
- 3.12 Stichprobenvarianz
- 3.13 Variationskoeffizient
- 3.14 Box-Whisker-Plot
- 3.15 Scatterplot
- 3.16 Curvefitting
- 3.17 Funktion des Curvefitting
- 3.18 Legende der Daten im Scatterplot
- 3.19 textuelle Zusammenfassung
- 3.20 Quartile und Dezile
- 3.21 Quartilsabstand $R_{Q0.5}$
- 3.22 Kovarianz
- 3.23 Korrelationskoeffizient

3.1 Struktur und Inhalt des Datensatzes

Die Daten liegen in zwei semistrukturierten Datensätzen vor und 3b enthält die Jahreszahlen von 2015 bis 2021, während 3a die dazu passenden Exportquoten enthält. Beide Datensätze bestehen aus 14 Zeilen. In Datensatz 3a wird in der dritten Zeile definiert, dass Jahr und Exportquote durch das Trennzeichen " ; " getrennt werden. Die Exportquote wird als Prozentwert in der vierten Zeile angegeben. In den folgenden Zeilen werden nun jeweils eine Nummer und eine Exportquote gegeben. Die Exportquoten schwanken zwischen 46,9% und 47,3%. Die Exportquote für Nummer 2 ist nicht verfügbar (NaN). Ebenso verhält es sich mit Nr.6 (Haha).

In 3b sind die Jahreszahlen absteigend von 2021 bis 2015 in Schritten von einem Jahr angegeben. Den Jahreszahlen wird jeweils ein Key zugeordnet. Die Jahreszahlen sind absteigend sortiert.

Zuletzt werden bei 3a und 3b in den Zeilen 13 und 14 Stand und Quelle der Daten angegeben. Die Daten stammen vom Statistischen Bundesamt (Destatis) und sind auf dem Stand vom 12. September 2022. Die Daten liegen in einer .csv-Datei vor.

Da einige Berechnungen bei einer Intervallskala für die Variable "Jahreszahl" nicht zielführend sind, finden sich Berechnungen für "Jahreszahl" nur unter den entsprechend sinnvollen Unterpunkten.

3.2 zusammengeführte Datei

siehe Data-3

3.3 bereinigte Daten

siehe Data-3

3.4 Maßnahmen zur Datenbereinigung

Um den Datensatz zu bereinigen, gibt es mehrere Möglichkeiten. Zum Beispiel kann man die fehlenden Werte auslassen oder auch durch einen Mittelwert ersetzen.

Eine weitere Möglichkeit wäre die Verwendung einer Interpolationsmethode, um die fehlenden Werte auf der Grundlage der vorhandenen Daten zu schätzen. Hierbei muss jedoch beachtet werden, dass Interpolation eine Annahme über die Kontinuität der Daten zwischen den verfügbaren Werten beinhaltet und somit auch Unsicherheiten in der Schätzung beinhalten kann.

Letztendlich hängt die Wahl der Methode davon ab, wie genau die Ergebnisse sein sollen und welche Annahmen bei der Analyse getroffen werden können.

Da die meisten Exportquoten im Datensatz jedoch nur geringfügig variieren und die Abweichungen zwischen den Jahren gering sind, ist es durchaus vertretbar, den Mittelwert der vorhandenen Werte für das fehlende Jahr zu verwenden. Allerdings sollte beachtet

werden, dass dies eine Annahme ist und dass der Mittelwert möglicherweise nicht genau den tatsächlichen Wert für das fehlende Jahr widerspiegelt.

3.5 Excel-Datei

siehe Data-3

3.6 Verwendete Software und genutzte Funktionen

VS Code (nur als Editor)

Excel: MODALWERT, MITTELWERT, MEDIAN, VARIANZ, Diagrammfunktionen, QUARTILE(ArrayA;1...), QUANTIL(Array;10/100...), KOVARIANZ.P, KORREL, STABW

3.7 Urliste

siehe Data 3

3.8 Rangliste

siehe Data 3

3.9 Modus, arithmetischer Mittelwert, Median

“Exportquote”:

Modus: 47,02

Arithmetischer Mittelwert: 47,02

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Median: 47,02

$$x_{med} = \begin{cases} x_{(\frac{n+1}{2})} & , \text{ falls } n \text{ ungerade} \\ \frac{1}{2} \cdot (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & , \text{ falls } n \text{ gerade} \end{cases}$$

“Jahreszahl”:

Modus: nicht vorhanden, da jeder Wert genau einmal existiert

arithmetischer Mittelwert: 2018

Median: 2018

3.10 Spannweite

"Exportquote":

$$R = x_{\max} - x_{\min}$$

$$R = 47,3 - 46,7 = 0,6$$

"Jahreszahl"

$$R = 2021 - 2015 = 6$$

3.11 mittlere Abweichung vom Median

"Exportquote":

$$MA_x = \frac{1}{n} \cdot \sum_{i=1}^n |x_i - x_{med}|$$

$$MA_x = 1/7 \cdot \sum_{i=1}^n |x_i - 4|$$

$$= 0,13142857$$

$$\approx 0,13$$

"Jahreszahl":

$$MA_x = 1/7 \cdot \sum_{i=1}^n |x_i - 2018|$$

$$= 1,71428571$$

$$\approx 1,71$$

3.12 Stichprobenvarianz

"Exportquote":

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (\bar{x} - x_i)^2 \quad s^2 = \frac{1}{n-1} \cdot \left[\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right]$$

$$\text{Varianz} = 0,038$$

3.13 Variationskoeffizient

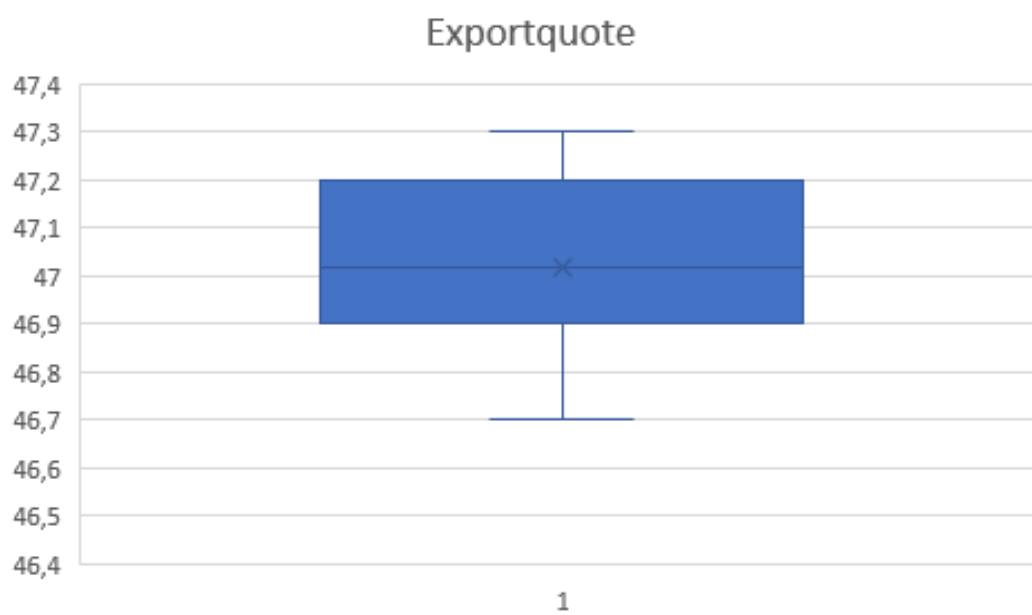
“Exportquote”

$$v = \frac{s}{\bar{x}}, \quad \bar{x} \neq 0$$

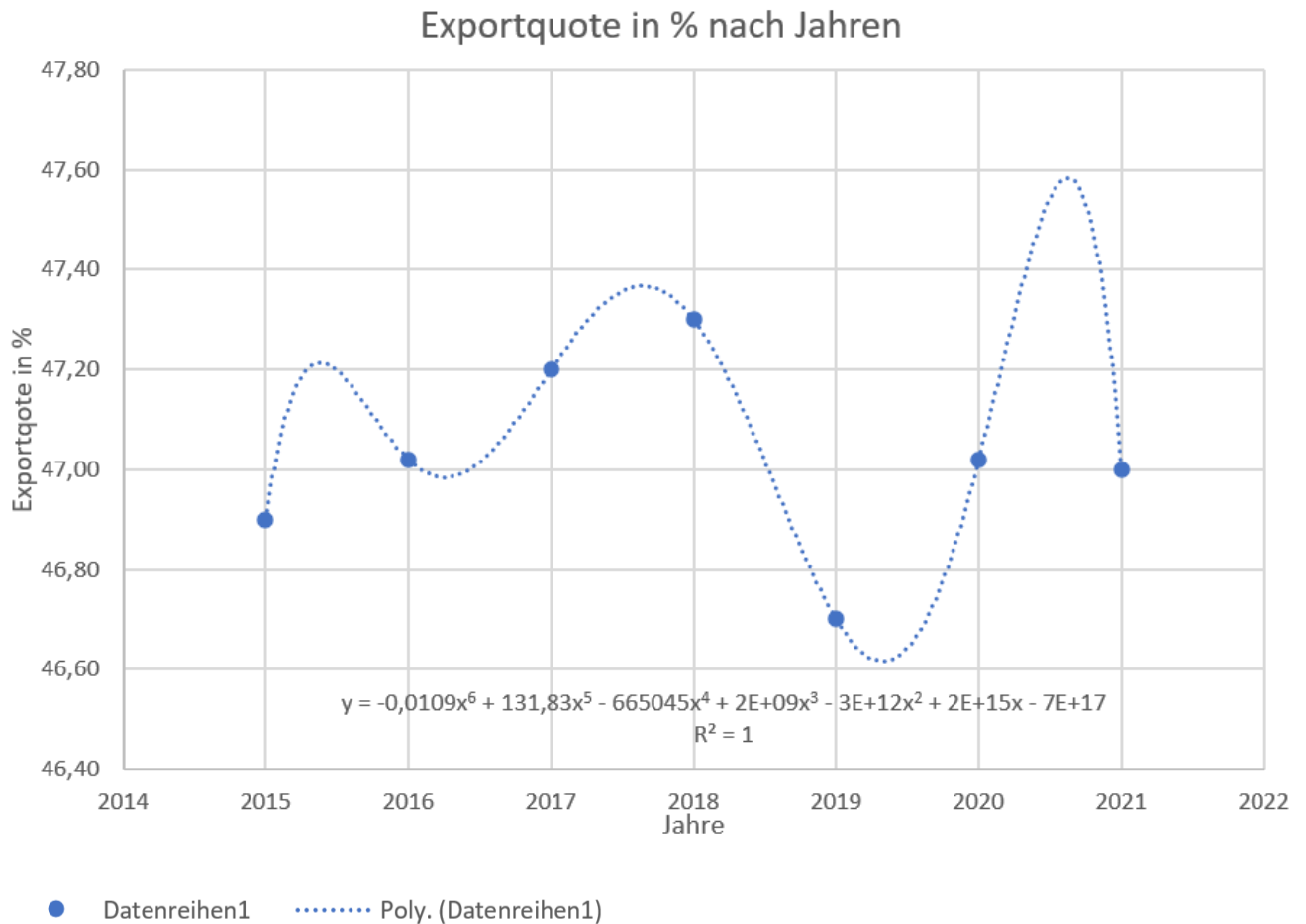
V = 0,00414581

≈ 0,004

3.14 Box-Whisker-Plot



3.15 Scatterplot + 3.16 Curvefitting+3.18 Legende



3.17 Funktion des Curvefitting

$$y = -0,0109x^6 + 131,83x^5 - 665045x^4 + 2E+09x^3 - 3E+12x^2 + 2E+15x - 7E+17$$

3.19 textuelle Zusammenfassung

Der gegebene Datensatz enthält die Exportquote von sieben Jahren, beginnend im Jahr 2015 bis 2021. Die Exportquoten schwanken zwischen 46,7% (2019) und 47,3% (2018) und haben insgesamt eine Spannweite von 0,6%. Der arithmetische Mittelwert und Median der Exportquote liegen beide bei 47,02%. Der Modus für die Exportquote beträgt durch die Bereinigung 47,02%(2016,2020). Die Jahreszahlen haben keinen Modus, da jeder Wert nur einmal auftritt. Der arithmetische Mittelwert und Median der Jahreszahlen sind beide 2018. Die Stichprobenvarianz beträgt ungefähr 4,7 für die Jahreszahlen und 0,038 für die Exportquoten. Der Variationskoeffizient ist 0,1 für die Jahreszahlen und 0,004 für die Exportquoten. Die Kovarianz zwischen den beiden Variablen beträgt ungefähr -0,029 und der Korrelationskoeffizient beträgt ungefähr -0,079, was auf eine schwache, negative Korrelation

zwischen den Variablen hinweist. Der Rangkorrelationskoeffizient beträgt ungefähr -0,072 und zeigt ebenfalls eine schwache, negative Korrelation.

3.20 Quartile und Dezile

“Exportquote”:

1.Quartil		46,95	1.Dezil		46,82
2.Quartil/Median		47,02	2.Dezil		46,92
3.Quartil		47,11	3.Dezil		46,98
			4.Dezil		47,008
Quartilsabstand(Q3-Q1)		0,16	5.Dezil		47,02
			6.Dezil		47,02
			7.Dezil		47,056
			8.Dezil		47,164
			9.Dezil		47,24
			10.Dezil/Maximum		47,3

3.21 Quartilsabstand

“Exportquote”:

$$R_{Q0.5} = Q3 - Q1 = 0,16$$

3.22 Kovarianz

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$s = -0,0285714$$

3.23 Korrelationskoeffizient

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$r = -0,072075$$

Datensatz 4

- 4.1 generierte Daten
- 4.2 bereinigte eingelesene Daten
- 4.3 Maßnahmen zur Datenbereinigung
- 4.4 verwendete Software und genutzte Funktionen
- 4.5 Modus, arithmetischer Mittelwert und Varianz
- 4.6 Stichprobenvarianz
- 4.7 Box-Whisker-Plot
- 4.8 textuelle Zusammenfassung

4.1 generierte Daten

siehe Data 4

4.2 bereinigte eingelesene Daten

siehe Data 4

4.3 Maßnahmen zur Datenbereinigung

keine Datenbereinigung notwendig

4.4 verwendete Software und genutzte Funktionen

Excel: MODALWERT, MITTELWERT, MEDIAN, VARIANZ, Diagrammfunktionen, KOVARIANZ.P

C-Programm zum Generieren der Daten

C-Programm zur Berechnung von Modus, arithmetischem Mittel, Median, Stichprobenvarianz (mit Excel Funktionen kontrolliert)

4.5 Modus, arithmetischer Mittelwert und Median

Modus: 15,40

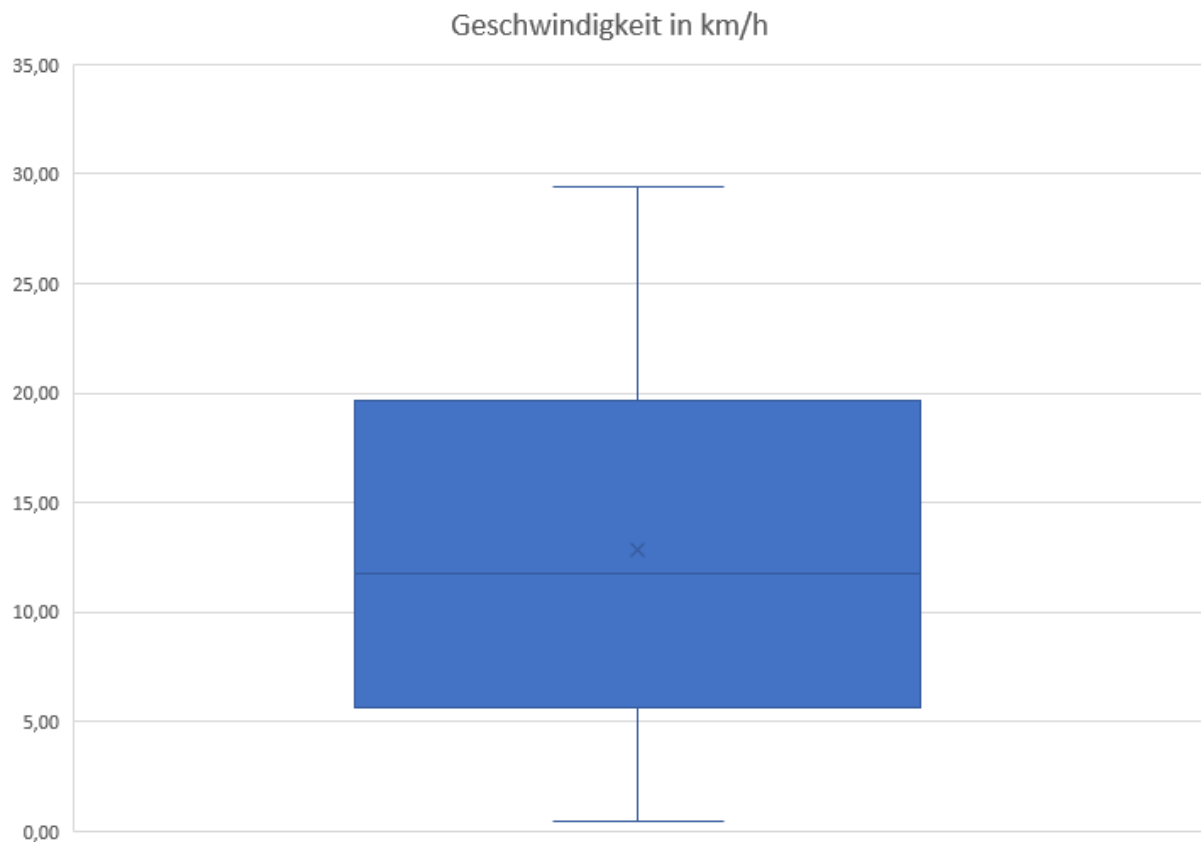
arithmetischer Mittelwert: 12,85

Median: 11,79

4.6 Stichprobenvarianz

Stichprobenvarianz: 74,71

4.7 Box-Whisker-Plot



4.8 textuelle Zusammenfassung

Die vorliegenden Daten sind Geschwindigkeiten in Kilometern pro Stunde. Die Datenreihe besteht aus 100 Beobachtungen und reicht von 0,42 km/h bis 29,43 km/h. Der Mittelwert der Daten beträgt 12,85 km/h und der Median liegt bei 11,79 km/h. Der Modus, also der am häufigsten auftretende Wert in der Datenreihe, ist 15,4 km/h. Die Stichprobenvarianz beträgt 74,71 (km/h)².

Quellenverzeichnis

Vorlesungsmaterial (Folien in Olat)

Buch "Wahrscheinlichkeitstheorie für Ingenieure" von Lothar Litz

Buch "Keine Panik vor Statistik!" von Markus Oestreich, Oliver Romberg