

Grundlagen und Anwendung der Wahrscheinlichkeitstheorie

Projekt Datenanalyse

GAWT

WS 2024/2025

bei

Dr. -Ing. Christian de Schryver

Gruppe 6

Datensätze 1

Projektersteller:

David Blanck

Sebastian Heß

Philipp Rosenbaum

Inhaltsverzeichnis

Datensatz 1	3
Beschreibung.....	3
Verwendete Python Anweisungen.....	4
Grafische Darstellungen	5
Auswertung der Daten	7
Datensatz 2.....	10
Beschreibung.....	10
Verwendete Python Anweisungen.....	11
Grafische Darstellungen	12
Auswertung der Daten	14
Datensatz 3.....	17
Beschreibung.....	17
Verwendete Python Anweisungen.....	18
Grafische Darstellungen	19
Auswertung der Datensätze	21
Datensatz 4.....	24
Beschreibung.....	25
Verwendete Python Anweisungen.....	26
Grafische Darstellungen	27
Auswertung der Daten	28

Datensatz 1

Beschreibung

Struktur und Inhalt

Der Datensatz 1 gibt die Elektrizitätserzeugung aus Steinkohle in Deutschland von 2002 bis 2023 an.

Die Daten werden in Jahr, Monat, Energieträger (Steinkohle) und Energieerzeugung (netto) getrennt. Die Energieerzeugung wird hierbei durch gemessene Megawattstunden (MWh) angegeben. Die Daten im Datensatz sind in unbereinigter Form vorhanden (also vollständig aber unsortiert). Bezogen auf die Quelle liegt uns nur ein Ausschnitt der Rohdaten vor. Außerdem sind die Spalten durch Semikolons getrennt. Die Spalten werden dadurch in Jahr (Spalte A), Monat (Spalte B), Energieträger Steinkohle (Spalte C) und Energieerzeugung (netto) (Spalte D) getrennt. Die Energieerzeugung wird hierbei durch gemessene Megawattstunden (MWh) angegeben. Die aufgenommenen Werte erstrecken sich vom Jahr 2002 bis 2023. Außerdem liegen zu jedem Jahr die erhobene Werte aus den Monaten April, August und Dezember vor.

Datenursprung

Die Daten stammen von dem statistischen Bundesamt (Destatis, Genisis-Tabelle: 43311-0002 (Steinkohle)) und sind vom Stand 09.04.2024.

Format

Die uns zur Verfügung gestellten Daten wurden auf drei separate Dateien aufgeteilt (Datensatz, Rohdaten, Quelle).

Datensatz:

- Name: data-1.csv
- Dateiformat: csv-Datei
- Kodierung: UTF-8
- Trennzeichen: Semikolon
- Anzahl Zeilen ohne Überschrift: 66
- Anzahl Spalten ohne Überschrift: 4

Rohdaten:

- Name: 43311-0002-Steinkohle.csv
- Dateiformat: csv-Datei
- Kodierung: UTF-8
- Trennzeichen: Semikolon

- Anzahl Zeilen: 528
- Anzahl Spalten: 4

Quelle:

- Name: Source.txt
- Dateiformat: txt-Datei

Verwendete Python Anweisungen

Für die Auswertung der Datensätze wurde Python 3.13.0 verwendet.

Anweisungen:

aus der Bibliothek statistics:

- median() - Bestimmen des Medians
- mean() - Mittelwert bestimmen
- mode() - Modus bestimmen
- multimode() - Modus bestimmen

aus der Bibliothek matplotlib:

- scatter() - zum Erstellen eines Scatterplots
- plot() - zum Erstellen eines Line Plots
- diverse Funktionen zum individualisieren der Plots wie z.B. anpassen des Titels

aus der Bibliothek numpy:

- std() - Standardabweichung
- cov() - Kovarianz
- var() - Stichprobenvarianz

aus der Bibliothek csv:

- open() - Öffnen von Dateien
- write() - Schreiben in Dateien
- close() - schließen von Dateien

Python built in Funktionen:

- abs() - Absolutwert
- len() - zum Bestimmen der Länge der Daten
- sort() - zum Sortieren der Daten

Grafische Darstellungen

Skalenvarianten

Elektrizitätserzeugung: Originalaufgelöst (in TWh), Verhältnisskaliert

Jahr: Originalaufgelöst, Intervallskaliert

Monat: Originalaufgelöst (nach Umwandlung in numerischen Wert), Intervallskaliert

Abbildung 1: Box-Plot-Elektrizitätserzeugung

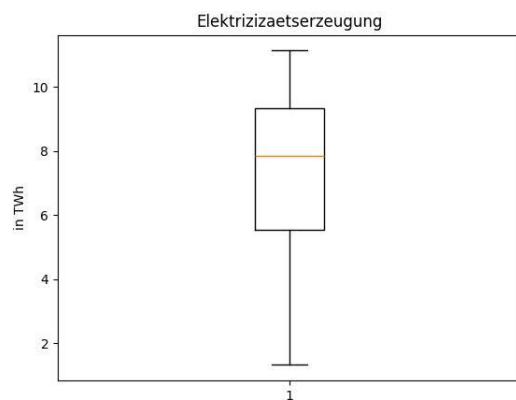


Abbildung 2: Box-Plot-Jahr

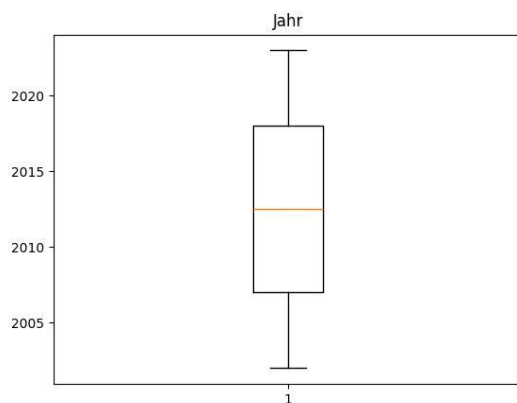


Abbildung 3: Box-Plot-Monat

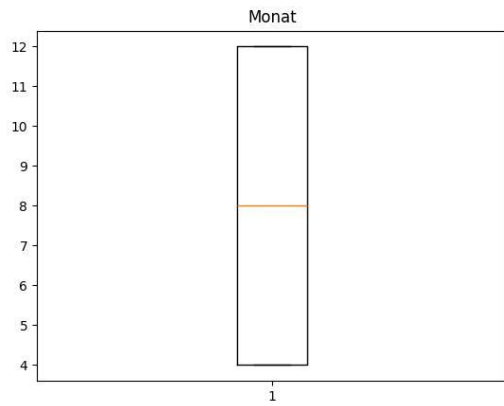


Abbildung 4: Scatterplot

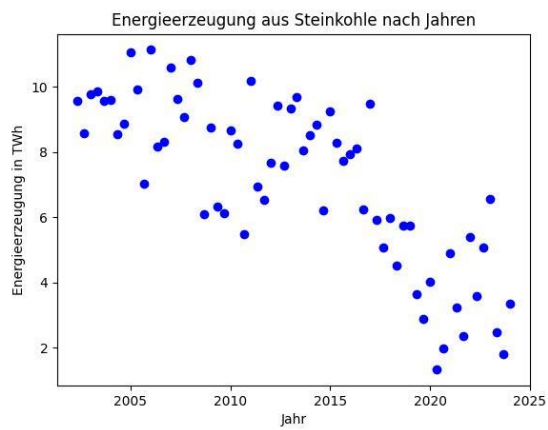
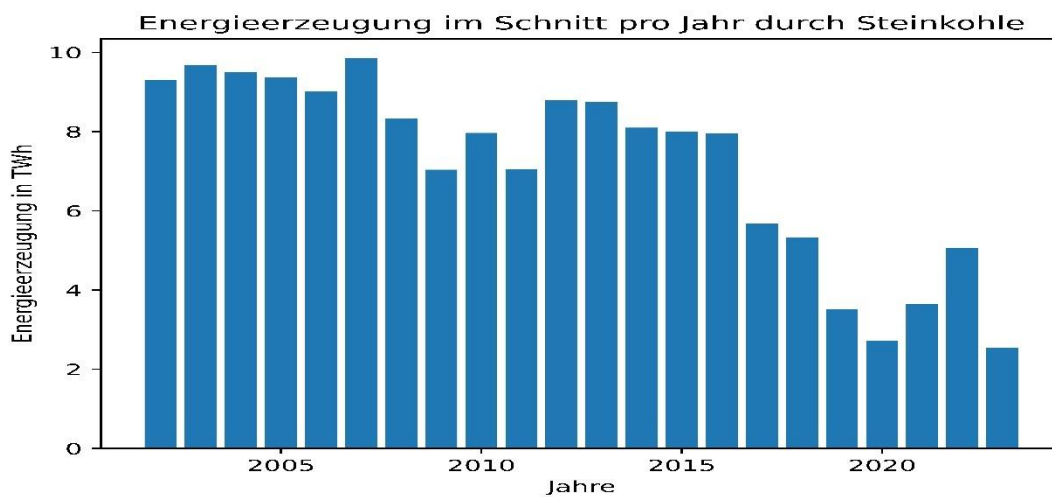


Abbildung 5: Histogramm Elektrizitätserzeugung



Auswertung der Daten

Lage und Streumaße

Alle Werte sind auf drei Nachkommastellen gerundet.

Variableneigenschaften

Elektrizitätserzeugung:

- Modus: 9567448
- Arithmetischer Mittelwert: 7145764.364
- Median: 7844663.000
- Spannweite: 9821158
- Mittlere Abweichung vom Median: 2129084.939
- Stichprobenvarianz: 6528443217279.838
- Variationskoeffizient: 0.358
- Kovarianz: 6628880805237.989
- Quartile:
 - q1: 5469250
 - q2: 7951870
 - q3: 9354832
 - q4: 11142512
- Dezile:
 - d1: 3220259
 - d2: 5062261
 - d3: 5926543
 - d4: 6526260
 - d5: 7951870
 - d6: 8315228
 - d7: 8884229
 - d8: 9567448
 - d9: 9921218
 - d10: 11142512

Jahr:

- Modus: 2002
- Arithmetischer Mittelwert: 2012.500
- Median: 2012.500
- Spannweite: 21
- Mittlere Abweichung vom Median: 5.500
- Stichprobenvarianz: 40.250
- Variationskoeffizient: 0.003

- Kovarianz: 40.869
- Quartile:
 - q1: 2007
 - q2: 2013
 - q3: 2018
 - q4: 2023
- Dezile:
 - d1: 2004
 - d2: 2006
 - d3: 2008
 - d4: 2010
 - d5: 2013
 - d6: 2015
 - d7: 2017
 - d8: 2019
 - d9: 2021
 - d10: 2023

Monat:

- Modus: 4
- Arithmetischer Mittelwert: 8.000
- Median: 8.000
- Spannweite: 8
- Mittlere Abweichung vom Median: 2.667
- Stichprobenvarianz: 10.667
- Variationskoeffizient: 0.408
- Kovarianz: 10.831
- Quartile:
 - q1: 4
 - q2: 8
 - q3: 12
 - q4: 12
- Dezile:
 - d1: 4
 - d2: 4
 - d3: 4
 - d4: 8
 - d5: 8
 - d6: 8
 - d7: 12
 - d8: 12

- d9: 12
- d10: 12

Zusammenfassung

Insgesamt lässt sich aus dem Datensatz erkennen, dass die Elektrizitätserzeugung durch Steinkohle vom Jahre 2002 bis 2023 immer weiter abgenommen hat.

Von April 2002 bis August 2004 lag die durchschnittliche Elektrizitätserzeugung bei 9.298.492 Megawattstunden (im folgenden MWh). Im Dezember 2004 stieg die Erzeugung sprunghaft an und erreichte dadurch Ihr Maximum von 11.066.104 MWh.

In den drei Jahren 2005, 2006 und 2007 erreichte die Erzeugung im Monat Dezember jeweils Ihr Jahresmaximum (über 10 Millionen MWh). Im restlichen Jahr betrug die durchschnittliche Erzeugung pro Monat 8.692.294,167 MWh.

Dieser sprunghafte Anstieg der Energieerzeugung (über 10 Millionen MWh) trat anschließend erneut im April 2008 und Dezember 2010 auf. Zwischen diesen zwei Ereignissen sank die durchschnittliche Elektrizitätserzeugung auf 7.099.546,429 MWh.

Nach 2010 lässt sich insgesamt ein negativer Trend der Elektrizitätserzeugung erkennen.

Betrachtet man die Werte der angegebenen Monate eines Jahres genauer, fällt auf, dass sich hier der Trend zu einer "V"-Form erkennen lässt. Das bedeutet, im Schnitt ist die erzeugte Energie im Monat April und Dezember höher als im August, der Wert liegt also unter dem Niveau von April und Dezember.

Am wenigsten Elektrizität durch Steinkohle wurde im April 2020 erzeugt (1.321.354 MWh) und am meisten im Dezember 2004 (11.066.104 MWh).

Datensatz 2

Beschreibung

Struktur und Inhalt

Im Datensatz 2 werden die Beschäftigten des Monats eines Jahres prozentualen Anteil zu dem Jahr 2015 angegeben.

Die Daten werden in Jahr, Monat, und Beschäftigte (2015=100) getrennt. Die Beschäftigten werden hierbei in den prozentualen Anteil zu dem Referenzwert zu 2015. Die Daten liegen in unbereinigter Form vor (sortiert, aber unvollständig, also fehlende Werte und in unterschiedlicher Darstellung (Numerisch und Strings). Außerdem liegt uns bezogen auf die Quelle nur ein Ausschnitt der Rohdaten vor. Die Spalten der Daten sind hierbei durch Semikolons in Jahr (Spalte A), Monat (Spalte B) und Beschäftigte (Spalte C) getrennt. Insgesamt liegen uns in diesem Datensatz 61 Werte vor.

Nach der Datenbereinigung lässt sich nun nähere Angaben zu dem Datensatz machen. In dem uns gegebenen Datensatz sind die Beschäftigten prozentual zu dem Jahr 2015 angegeben. Dieser Wert ist jeweils für den Monat Januar und Dezember eines Jahres angegeben. Die aufgenommenen Werte in unserem Datensatz erstrecken sich vom Jahr 1994 bis 2024. Da die Quelle besagt, dass die Daten auf dem Stand vom 09.04.2024 beruhen, ist keine Angabe zu dem Wert aus Dezember 2024 vorhanden.

Datenursprung

Die Daten stammen von dem statistischen Bundesamt (Destatis, Genisis-Tabelle: 45212-0002 (Einzelhandel Fleisch und Fleischwaren)) und sind vom Stand 09.04.2024.

Format

Die Daten die uns zur Verfügung gestellt wurden, sind auf 3 unterschiedliche Dateien aufgeteilt (Datensatz, Rohdaten, Quelle).

Datensatz:

- Name: data-2.csv
- Dateiformat: csv-Datei
- Kodierung: UTF-8
- Trennzeichen: Semikolon
- Anzahl Zeilen: 61
- Anzahl Spalten: 3

Rohdaten:

- Name: 45212-0002_\$F_Metzger.csv
- Dateiformat: csv-Datei
- Kodierung: UTF-8
- Trennzeichen: Semikolon
- Anzahl Zeilen: 372
- Anzahl Spalten: 7

Quelle:

- Name: Source.txt
- Dateiformat: txt-Datei

Maßnahmen zur Datenbereinigung

Um die Lücken in den Werten der Übernachtungen zu bereinigen wurde der Wert des naheliegendsten Monat verwendet.

Die Lücken bei den Jahren wurden ergänzt durch Auswertung des Jahres davor und dem danach.

Da wir zu jedem Jahr jeweils zwei Monate gegeben haben, wurde geschaut, welcher Monat zu diesem Jahr fehlt.

Verwendete Python Anweisungen

Für die Auswertung der Datensätze wurde Python 3.13.0 verwendet.

Anweisungen:

aus der Bibliothek statistics:

- median() - Bestimmen des Medians
- mean() - Mittelwert bestimmen
- mode() - Modus bestimmen
- multimode() - Modus bestimmen

aus der Bibliothek matplotlib:

- scatter() - zum Erstellen eines Scatterplots
- plot() - zum Erstellen eines Line Plots
- diverse Funktionen zum individualisieren der Plots wie z.B. anpassen des Titels

aus der Bibliothek numpy:

- `std()` - Standardabweichung
- `cov()` - Kovarianz
- `var()` - Stichprobenvarianz

aus der Bibliothek csv:

- `open()` - Öffnen von Dateien
- `write()` - Schreiben in Dateien
- `close()` - schließen von Dateien

Python built in Funktionen:

- `abs()` - Absolutwert
- `len()` - zum Bestimmen der Länge der Daten
- `sort()` - zum Sortieren der Daten

Grafische Darstellungen

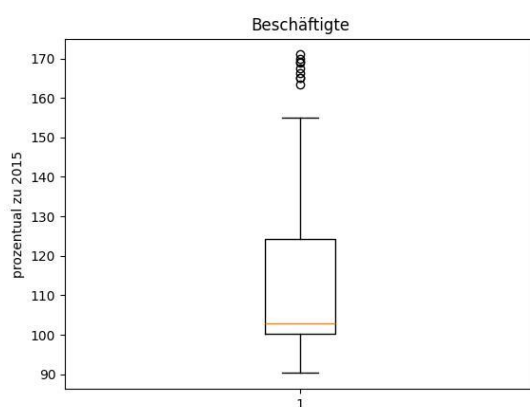
Skalenvarianten

Beschäftigte: prozentual zu 2015, Verhältnisskaliert

Jahr: Originalaufgelöst, Intervallskaliert

Monat: Originalaufgelöst (nach Umwandlung in numerischen Wert), Intervallskaliert

Abbildung 6: Box-Plot-Beschäftigte



26.01.2025

Abbildung 7: Box-Plot-Jahr

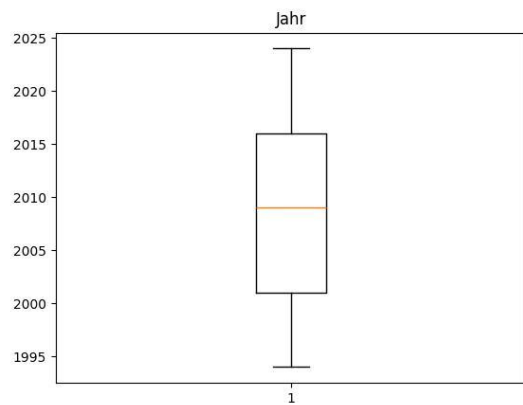


Abbildung 8: Box-Plot-Monat

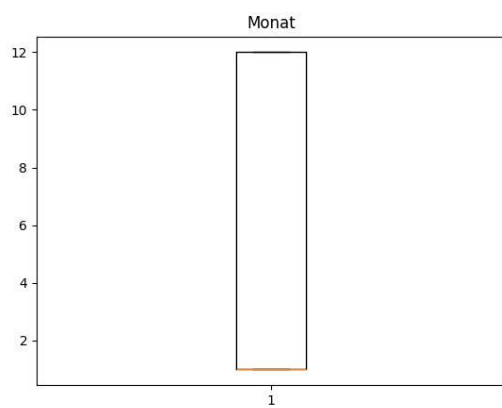


Abbildung 9: Scatterplot

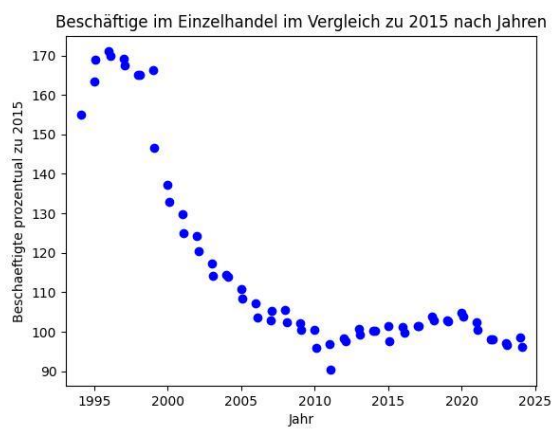


Abbildung 10: Ander grafische Darstellung Line-Plot



Der erstellte Line Plot gibt den Verlauf der Anzahl der Beschäftigten im Vergleich zu 2015 wieder.

Auswertung der Daten

Lage und Streumaße

Alle Werte sind auf drei Nachkommastellen gerundet.

Variableneigenschaften

Beschäftigte (prozentual zu 2015):

- Modus: 165
- arithmetischer Mittelwert: 116.020
- Median: 102.900
- Spannweite: 80
- Mittlere Abweichung vom Median: 16.500
- Stichprobenvarianz: 604.595
- Variationskoeffizient: 0.212
- Kovarianz: 614.671
- Quartile:
 - q1: 100
 - q2: 102
 - q3: 124
 - q4: 171
- Dezile:
 - d1: 97
 - d2: 99
 - d3: 100

26.01.2025

- d4: 102
- d5: 102
- d6: 105
- d7: 114
- d8: 132
- d9: 165
- d10: 171

Jahr:

- Modus: 1994
- arithmetischer Mittelwert: 2008.754
- Median: 2009.000
- Spannweite: 30
- Mittlere Abweichung vom Median: 7.623
- Stichprobenvarianz: 77.562
- Variationskoeffizient: 0.004
- Kovarianz: 78.855
- Quartile:
 - q1: 2001
 - q2: 2009
 - q3: 2016
 - q4: 2024
- Dezile:
 - d1: 1997
 - d2: 2000
 - d3: 2003
 - d4: 2006
 - d5: 2009
 - d6: 2012
 - d7: 2015
 - d8: 2018
 - d9: 2021
 - d10: 2024

Monat:

- Modus: 1
- arithmetischer Mittelwert: 6.410
- Median: 1.000
- Spannweite: 11
- Mittlere Abweichung vom Median: 5.410
- Stichprobenvarianz: 30.242

- Variationskoeffizient: 0.858
- Kovarianz: 30.746
- Quartile:
 - q1: 1
 - q2: 1
 - q3: 12
 - q4: 12
- Dezile:
 - d1: 1
 - d2: 1
 - d3: 1
 - d4: 1
 - d5: 1
 - d6: 12
 - d7: 12
 - d8: 12
 - d9: 12
 - d10: 12

Zusammenfassung

Insgesamt lässt sich aus dem Datensatz erkennen, dass die Zahl der Beschäftigten im Vergleich zu Beginn der Statistik 1994 stark gesunken ist.

Aus dem Line Plot (siehe Abbildung 9) der Daten ist eine exponentielle Abnahme sichtbar. Ab 2010 ist das Niveau gleichbleibend.

Im Januar 1994 begann die Statistik mit einem Wert von 155 Beschäftigten (im Folgenden alle Werte prozentual zu den Werten aus dem Jahr 2015). Dann gab es einen großen Anstieg an Beschäftigten im Dezember 1995.

Im Anschluss sank die Beschäftigten Zahl über die Jahre bis auf ein Minimum von 90,3 im Januar 2011. Danach stieg die Beschäftigtenzahl wieder auf ein lokales Maximum von 104,9 im Dezember 2019. Dann fängt die Beschäftigtenzahl wieder langsam an zu fallen bis sie im Januar 2024 bei einem Wert von 96.2 endet.

Abschließen lässt sich sagen, dass sich seit dem Minimum 2011 die Beschäftigtenzahl nicht stark verändert hat.

Die meisten beschäftigten Personen gab es nach dem Anstieg von 1994 im Dezember 1995 mit einem Wert von 171 Beschäftigten. Am wenigsten gab es in dem historischen Tief vom Januar 2011.

Datensatz 3

Beschreibung

Struktur und Inhalt

Der Datensatz 3 gibt die Anzahl der Ankünfte und Übernachtungen in Beherbergungsbetrieben in Deutschland wieder.

Der Datensatz wurde hier in zwei unterschiedliche Dateien aufgeteilt und muss zunächst richtig zusammengefügt werden.

Aus der bereinigten Form der Daten lässt sich nun genaueres herauslesen. Uns liegt wiederum nur ein Ausschnitt der Rohdaten vor. Die Aufgenommenen Werte für Ankünfte (Spalte C) und Übernachtungen (Spalte D) sind hier jeweils für jeden Monat (Spalte B) eines Jahres (Spalte A) vorhanden. Diese Aufnahme erstreckt sich vom Jahr 1992 bis 1998. Im Jahr 1999 liegen uns lediglich die Werte von den Monaten Januar bis Juni vor.

Datenursprung

Die Daten stammen von dem statistischen Bundesamt (Destatis, Genisis-Tabelle 45412-0014 (Tourismus)) und sind vom Stand 09.04.2024.

Format

Die Daten liegen aufgeteilt in 4 verschiedene Dateien (zwei Datensätze, Rohdaten, Quelle).

Datensatz:

- Name: data-3-a.csv und data-3-b.csv
- Dateiformat: csv-Datei
- Kodierung: UTF-8
- Trennzeichen: Semikolon
- Anzahl Zeilen (3a): 90
- Anzahl Spalten (3a): 2
- Anzahl Zeilen (3b): 90
- Anzahl Spalten (3a und 3b): 3

Rohdaten:

- Name: 454212-0014_\$F_Tourismus.csv
- Dateiformat: csv-Datei
- Kodierung: UTF-8
- Trennzeichen: Semikolon
- Anzahl Zeilen: 396
- Anzahl Spalten: 4

Quelle:

- Name: Source.txt
- Dateiformat: txt-Datei

Maßnahmen zur Datenbereinigung

Um die Daten zu bereinigen wurden lediglich die beiden gesplitteten Dateien wieder richtig zusammengefügt.

Verwendete Python Anweisungen

Für die Auswertung der Datensätze wurde Python 3.13.0 verwendet.

Anweisungen:

aus der Bibliothek statistics:

- median() - Bestimmen des Medians
- mean() - Mittelwert bestimmen
- mode() - Modus bestimmen
- multimode() - Modus bestimmen

aus der Bibliothek matplotlib:

- scatter() - zum Erstellen eines Scatterplots
- plot() - zum Erstellen eines Line Plots
- diverse Funktionen zum individualisieren der Plots wie z.B. anpassen des Titels

aus der Bibliothek numpy:

- std() - Standardabweichung
- cov() - Kovarianz
- var() - Stichprobenvarianz

aus der Bibliothek csv:

- open() - Öffnen von Dateien
- write() - Schreiben in Dateien
- close() - schließen von Dateien

Python built in Funktionen:

- abs() - Absolutwert
- len() - zum Bestimmen der Länge der Daten
- sort() - zum Sortieren der Daten

Grafische Darstellungen

Skalenvarianten

Ankünfte: Originalaufgelöst, Verhältnisskaliert

Übernachtungen: Originalaufgelöst, Verhältnisskaliert

Jahr: Originalaufgelöst, Intervallskaliert

Monat: Originalaufgelöst (nach Umwandlung in numerischen Wert), Intervallskaliert

Abbildung 11: Box-Plot-Ankünfte

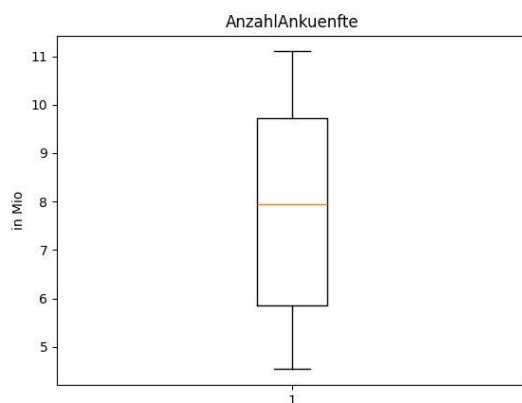


Abbildung 12: Box-Plot-Übernachtungen

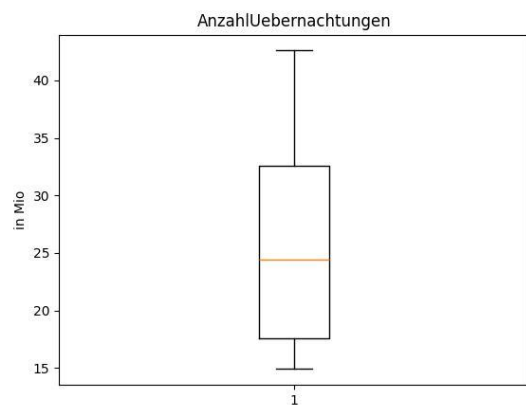


Abbildung 13: Box-Plot-Jahr

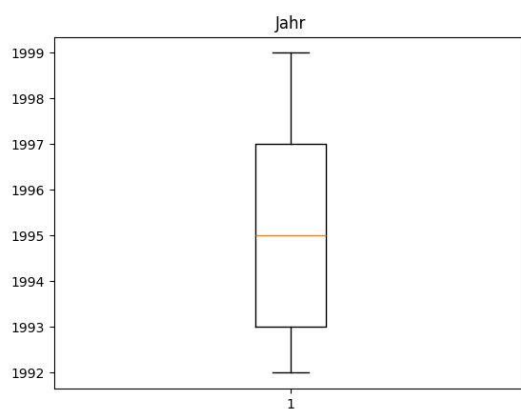


Abbildung 14: Box-Plot-Monat

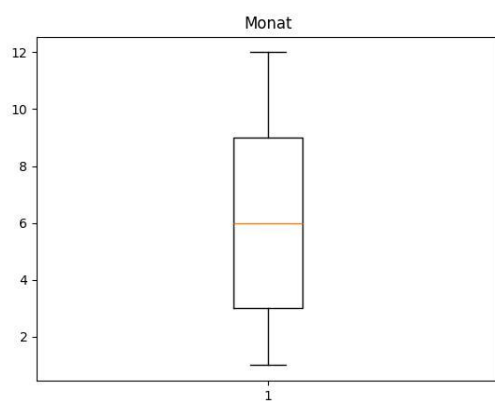
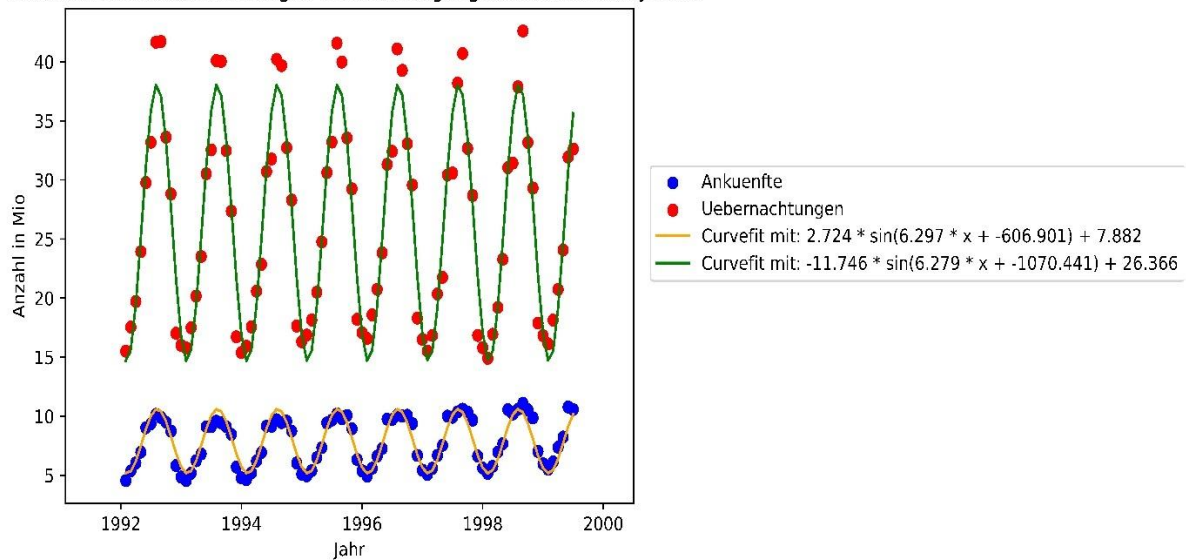


Abbildung 15: Scatterplot

Ankuenfte und Ueberrnachtungen in Beherbergungsbetrieben nach Jahren



Auswertung der Datensätze

Lage und Streumaße

Alle Werte sind auf drei Nachkommastellen gerundet.

Variableneigenschaften

Anzahl der Unterkünfte:

- Modus: 4545298
- arithmetischer Mittelwert: 7840121.300
- Median: 7952325.000
- Spannweite: 6560767
- Mittlere Abweichung vom Median: 1884820.367
- Stichprobenvarianz: 4131273292412.788
- Variationskoeffizient: 0.259
- Kovarianz: 4177692093451.133
- Quartile:
 - q1: 5824884
 - q2: 8230904
 - q3: 9730772
 - q4: 11100085

- Dezile:
 - d1: 5155353
 - d2: 5590958
 - d3: 6208138
 - d4: 6938233
 - d5: 8230904
 - d6: 9139351
 - d7: 9571054
 - d8: 9905191
 - d9: 10187113
 - d10: 11100085

Anzahl der Übernachtungen:

- Modus: 15509574
- arithmetischer Mittelwert: 26177643.478
- Median: 24406795.000
- Spannweite: 27705308
- Mittlere Abweichung vom Median: 7731544.411
- Stichprobenvarianz: 74160126394474.562
- Variationskoeffizient: 0.329
- Kovarianz: 74993386241603.453
- Quartile:
 - q1: 17561842
 - q2: 24751506
 - q3: 32615343
 - q4: 42603257
- Dezile:
 - d1: 16304627
 - d2: 17021150
 - d3: 18197969
 - d4: 20732955
 - d5: 24751506
 - d6: 30421291
 - d7: 31764731
 - d8: 33183942
 - d9: 39970478
 - d10: 42603257

Jahr:

- Modus: 1992
- arithmetischer Mittelwert: 1995.267
- Median: 1995.000
- Spannweite: 7
- Mittlere Abweichung vom Median: 1.867
- Stichprobenvarianz: 4.729
- Variationskoeffizient: 0.001
- Kovarianz: 4.782
- Quartile:
 - q1: 1993
 - q2: 1995
 - q3: 1997
 - q4: 1999
- Dezile:
 - d1: 1992
 - d2: 1993
 - d3: 1994
 - d4: 1995
 - d5: 1995
 - d6: 1996
 - d7: 1997
 - d8: 1998
 - d9: 1998
 - d10: 1999

Monat:

- Modus: 1
- arithmetischer Mittelwert: 6.300
- Median: 6.000
- Spannweite: 11
- Mittlere Abweichung vom Median: 2.967
- Stichprobenvarianz: 11.877
- Variationskoeffizient: 0.547
- Kovarianz: 12.010
- Quartile:
 - q1: 3
 - q2: 6
 - q3: 9
 - q4: 12

- Dezile:
 - d1: 2
 - d2: 3
 - d3: 4
 - d4: 5
 - d5: 6
 - d6: 7
 - d7: 9
 - d8: 10
 - d9: 11
 - d10: 12

Zusammenfassung

Insgesamt lässt sich sagen, dass sich die Anzahl der Ankünfte und Übernachtungen in den Beherbergungsbetrieben über die Jahre nicht stark verändert haben.

Aus den Daten die uns vorliegen wird deutlich, dass die Ankünfte und Übernachtungen im Schnitt in einem direkten Verhältnis stehen. Jedoch ist die Anzahl der Übernachtungen immer um einiges höher als die der Ankünfte. Daraus lässt sich schlussfolgern, dass die Anreisenden Personen im Schnitt mehrere Tage in den Betrieben übernachten. Es lässt sich also behaupten, dass die zwei Ereignisse Ankünfte und Übernachtungen korrelieren. Dies wird vor allem in Abbildung 15 deutlich.

Außerdem lässt sich auf dieser Grundlage (Daten + Abbildung 15) eine Periodizität von Ankünften und Übernachtungen herleiten. In den Anfangsmonaten eines Jahres fängt die Anzahl an Ankünften und Übernachtungen leicht an zu steigen bis sie dann im Juli oder August ihr Maximum erreicht. Zum Jahresende hin fällt dann wieder die Anzahl bis sie im Dezember oder Januar des neuen Jahres ihr Minimum erreicht.

Das Maximum an Übernachtungen gab es im August 1998 und am wenigsten Übernachteten die Menschen im Januar desselben Jahres. Im August 1998 gab es zudem auch die meisten Ankünfte, jedoch nicht die wenigsten Ankünfte in diesem Jahr, die gab es im Januar 1993.

Datensatz 4

Beschreibung

Struktur und Inhalt

Beim Datensatz 4 handelt sich um gemessene Lux Werte am Schreibtisch. Diese Werte wurden mithilfe der App Phyphox aufgezeigt und durch die eingebauten Sensoren eines Smartphones ermittelt.

Die Daten liegen als CSV Datei vor.

Zusätzlich liegen die Meta Daten vor.

Datenerhebungszeitraum: 18.01.2025 12:39:47-12:40:19

Datenursprung

Device Modell: Samsung SM-G990B

App: Phyphox

Format

Die Daten liegen aufgeteilt in 3 verschiedene Dateien (Datensatz, Rohdaten, Quelle).

Datensatz:

- Name: data-4.csv
- Dateiformat: csv-Datei
- Kodierung: UTF-8
- Trennzeichen: Komma
- Anzahl Zeilen (4): 202
- Anzahl Spalten (4): 2

Rohdaten:

- Name: data-4.csv
- Dateiformat: csv-Datei
- Kodierung: UTF-8
- Trennzeichen: Komma
- Anzahl Zeilen: 202
- Anzahl Spalten: 2

Quelle:

- Phyphox

Maßnahmen zur Datenbereinigung

Es wurden keine Maßnahmen zur Bereinigung der Daten benötigt.

Verwendete Python Anweisungen

Für die Auswertung der Datensätze wurde Python 3.13.0 verwendet.

Anweisungen:

aus der Bibliothek statistics:

- median() - Bestimmen des Medians
- mean() - Mittelwert bestimmen
- mode() - Modus bestimmen
- multimode() - Modus bestimmen

aus der Bibliothek matplotlib:

- scatter() - zum Erstellen eines Scatterplots
- plot() - zum Erstellen eines Line Plots
- diverse Funktionen zum individualisieren der Plots wie z.B. anpassen des Titels

aus der Bibliothek numpy:

- std() - Standardabweichung
- cov() - Kovarianz
- var() - Stichprobenvarianz

aus der Bibliothek csv:

- open() - Öffnen von Dateien
- write() - Schreiben in Dateien
- close() - schließen von Dateien

Python built in Funktionen:

- abs() - Absolutwert
- len() - zum Bestimmen der Länge der Daten
- sort() - zum Sortieren der Daten

Grafische Darstellungen

Skalenvarianten

Beleuchtungsstärke (in Lux): Originalaufgelöst, Verhältnisskaliert

Zeit (in Sekunden): Originalaufgelöst, Intervallskaliert

Abbildung 16: Box-Plot-Beleuchtungsstärke

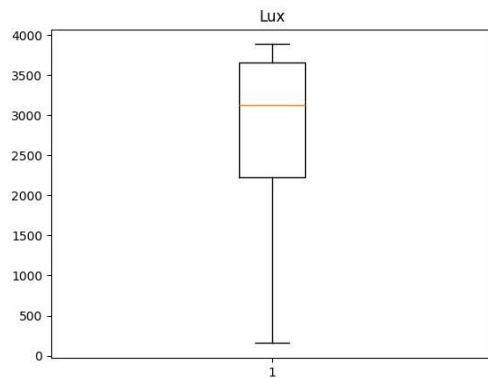
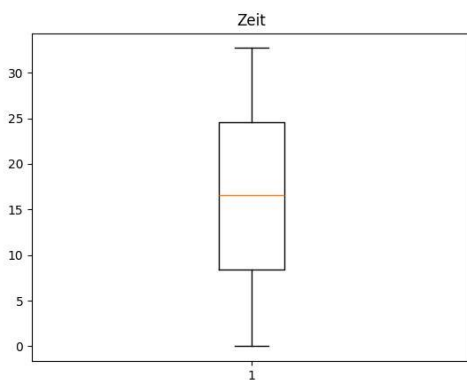


Abbildung 17: Box-Plot-Zeit



Auswertung der Daten

Lage und Streumaße

Alle Werte sind auf drei Nachkommastellen gerundet.

Variableneigenschaften

Beleuchtungsstärke (Lux):

- Modus: 3127
- arithmetischer Mittelwert: 2743.011
- Median: 3127.005
- Spannweite: 3727
- Mittlere Abweichung vom Median: 834.894
- Stichprobenvarianz: 1335300.596
- Variationskoeffizient: 0.421
- Kovarianz: 1341977.099
- Quartile:
 - q1: 2226
 - q2: 3127
 - q3: 3659
 - q4: 3885
- Dezile:
 - d1: 364
 - d2: 2038
 - d3: 2465
 - d4: 3102
 - d5: 3127
 - d6: 3248
 - d7: 3538
 - d8: 3734
 - d9: 3847
 - d10: 3885

Zeit:

- Modus: 0
- arithmetischer Mittelwert: 16.500
- Median: 16.545
- Spannweite: 32
- Mittlere Abweichung vom Median: 8.099
- Stichprobenvarianz: 87.695
- Variationskoeffizient: 0.568

- Kovarianz: 88.133
- Quartile:
 - q1: 8
 - q2: 16
 - q3: 24
 - q4: 32
- Dezile:
 - d1: 3
 - d2: 6
 - d3: 10
 - d4: 13
 - d5: 16
 - d6: 19
 - d7: 22
 - d8: 26
 - d9: 29
 - d10: 32

Zusammenfassung

Zusammenfassend lässt sich aus dem Datensatz die Beleuchtungsstärke am 18.01.2025 von 12:39:47-12:40:19 erkennen. Dabei lässt sich erkennen, dass diese leicht variiert. Dies lässt sich durch die Änderung der Neigung des Smartphones erklären. Die Höheren Lux Werte wurden erzielt, als das Handy mit dem Sensor Richtung Fenster geneigt war. Die niedrigeren als dies nicht der Fall war. Der starke Anstieg am Anfang des Datensatzes lässt sich durch die Ausrichtung des Sensors, bzw. dessen Trägheit erklären. Die Kurze Stagnation lässt sich durch das Anlehnen des Handys an ein Heft erklären und Kleinen Ausreißer als Messungenauigkeiten erklären. Die "großen" Sprünge werden immer schneller, da das Handy immer schneller geneigt wurde.