

Name: Rico Metz, Niclas Best, Rooney Hannan

Zeitraum: Wintersemester 2022/23

Modul: Grundlagen und Anwendung der Wahrscheinlichkeitstheorie

Projekt L02

von Gruppe 10

Quellenverzeichnis:

1. "Verwenden Sie Pandas, um Statistiken in Python zu berechnen" von Acervo Lima, abgerufen am 3. Februar 2023 um 14:20 Uhr von <https://de.acervolima.com/verwenden-sie-pandas-um-statistiken-in-python-zu-berechnen/>
2. "Pandas DataFrame.mean() Funktion" von Delft Stack, abgerufen am 10. Januar 2023 um 11:45 Uhr von <https://www.delftstack.com/de/api/python-pandas/pandas-dataframe-dataframe.mean-function/>
3. "Mean Absolute Deviation in Python" von datagy, abgerufen am 28. Januar 2023 um 16:55 Uhr von <https://datagy.io/mean-absolute-deviation-python/>
4. "Variance" von Data Librarian, abgerufen am 5. Februar 2023 um 9:30 Uhr von https://bonartm.github.io/data-librarian/descriptive_statistics/univariate/variance/
5. „Generiere zufällige Datensätze“ von meridianoutpost, abgerufen am 20. Februar 2023 um 17:38 Uhr von <https://www.meridianoutpost.com/resources/etools/calculators/generator-random-integer-real-numbers.php?>

Datensatz 1:

Beschreibungen von Datensatz 1 vor dem bearbeiten:

Struktur und Inhalt des Datensatzes sind bekannt. Der Datensatz enthält Importdaten in Millionen Euro für Deutschland von 1952 bis 2021. Die Daten sind chronologisch sortiert, beginnend mit dem Jahr 1952 und endend mit dem Jahr 2021. Die Quelle der Daten ist das Statistische Bundesamt (Destatis). Der Datensatz liegt in Form einer Tabelle vor, die aus zwei Spalten besteht: "Jahr" und "Importe". Die Jahreszahlen sind als ganze Zahlen angegeben und die Importdaten sind als ganze Zahlen in Millionen Euro angegeben. Die Daten liegen in einer .csv Datei vor und sind nicht codiert.

Skalenvarianten: Die geeignete Skalenvariante für Spalte „Jahr“ ist die Intervallskala und für „Importe“ die Verhältnisskala.

Beschreibung danach:

Die Tabelle enthält Daten über den Import von Waren in Euro über mehrere Jahrzehnte, die vom Statistischen Bundesamt (Destatis), einer deutschen Bundesbehörde für Statistik, bereitgestellt wurden. Sie gibt das Jahr und die entsprechenden Importwerte an. Zusätzlich zu den Daten enthält die Tabelle statistische Informationen wie den Mittelwert, Median, Modus, Bereich und Standardabweichung. Außerdem wird eine Trendlinienformel für die Daten bereitgestellt und ein zugehöriges Scatterplot. Die Daten zeigen, dass die Importwerte im Allgemeinen im Laufe der Zeit gestiegen sind, mit einigen Schwankungen und gelegentlichen Rückgängen. Der Median der Importwerte beträgt 223.400,50 € und der Durchschnitt der Importwerte beträgt 358.286,75 €. Die Trendlinienformel legt nahe, dass die Importwerte auch in Zukunft weiter steigen werden.

Datensatz 2:

Beschreibungen von Datensatz 2:

Der gegebene Datensatz besteht aus einer Tabelle mit 70 Zeilen und 2 Spalten. Die Spalten sind mit "Jahr" und "Importe" überschrieben und enthalten entsprechend das Jahr und den Importwert in Euro. Der Datensatz wurde vom Statistischen Bundesamt (Destatis) bereitgestellt und hat Stand vom 13. Oktober 2022. Es gibt jedoch einige Fehler im Datensatz, wie beispielsweise fehlerhafte oder fehlende Werte, die vor der weiteren Auswertung behoben werden müssen. Die Tabelle liegt als .csv vor und ist nicht codiert.

Beschreibung danach:

Die Tabelle enthält Daten über den Import von Waren in Euro über mehrere Jahrzehnte, die vom Statistischen Bundesamt (Destatis), einer deutschen Bundesbehörde für Statistik, bereitgestellt wurden. Sie gibt das Jahr und die entsprechenden Importwerte an. Zusätzlich zu den Daten enthält die Tabelle statistische Informationen wie den Mittelwert, Median, Modus, Bereich und Standardabweichung. Außerdem wird eine Trendlinienformel für die Daten bereitgestellt und ein zugehöriges Scatterplot. Die Daten zeigen, dass die Importwerte im Allgemeinen im Laufe der Zeit gestiegen sind, mit einigen Schwankungen und gelegentlichen Rückgängen. Der Median der Importwerte beträgt 222032€ und der Durchschnitt der Importwerte beträgt 344595,74 €. Die Trendlinienformel legt nahe, dass die Importwerte auch in Zukunft weiter steigen werden.

Datensatz 3a:

Beschreibungen von Datensatz 3a:

Der vorliegende Datensatz enthält Importdaten für Deutschland für verschiedene Jahre, angefangen vom Jahr 1 bis zum Jahr 70. Die Importzahlen sind in Millionen Tonnen angegeben. Es handelt sich um einen einzigen Datensatz, der aus 70 Zeilen besteht. Der Datensatz wurde vom Statistischen Bundesamt (Destatis) bereitgestellt und die Quellenangabe sowie das Datum der Erhebung sind am Ende des Datensatzes angegeben. Der Datensatz liegt in tabellarischer Form als .csv Datei vor, mit der Jahreszahl als erster Spalte und den Importzahlen als zweiter Spalte.

Beschreibung danach:

Die Tabelle enthält Daten über den Import von Waren in Euro über mehrere Jahrzehnte, die vom Statistischen Bundesamt (Destatis), einer deutschen Bundesbehörde für Statistik, bereitgestellt wurden. Sie gibt das Jahr und die entsprechenden Importwerte an. Zusätzlich zu den Daten enthält die Tabelle statistische Informationen wie den Mittelwert, Median, Modus, Bereich und Standardabweichung. Außerdem wird eine Trendlinienformel für die Daten bereitgestellt und ein zugehöriges Scatterplot. Die Daten zeigen, dass die Importwerte im Allgemeinen im Laufe der Zeit gestiegen sind, mit einigen Schwankungen und gelegentlichen Rückgängen. Der Median der Importwerte beträgt 222032€ und der Durchschnitt der Importwerte beträgt 344595,74 €. Die Trendlinienformel legt nahe, dass die Importwerte auch in Zukunft weiter steigen werden.

Datensatz 3b:

Beschreibungen von Datensatz 3b:

Der gegebene Datensatz enthält eine Liste von Jahren mit dazugehörigen numerischen Schlüsseln. Es ist davon auszugehen, dass diese numerischen Schlüssel für eine weitere Verwendung im Zusammenhang mit den entsprechenden Jahren genutzt werden können. Der Datensatz stammt vom Statistischen Bundesamt (Destatis) und wurde zuletzt am 13. Oktober 2022 aktualisiert. Der Datensatz liegt in tabellarischer Form als .csv Datei vor, mit der Jahreszahl als erster Spalte und den Importzahlen als zweiter Spalte.

Beschreibung danach:

Die Tabelle enthält Daten über den Import von Waren in Euro über mehrere Jahrzehnte, die vom Statistischen Bundesamt (Destatis), einer deutschen Bundesbehörde für Statistik, bereitgestellt wurden. Sie gibt das Jahr und die entsprechenden Importwerte an. Zusätzlich zu den Daten enthält die Tabelle statistische Informationen wie den Mittelwert, Median, Modus, Bereich und Standardabweichung. Außerdem wird eine Trendlinienformel für die Daten bereitgestellt und ein zugehöriges Scatterplot. Die Daten zeigen, dass die Keys im Allgemeinen im Laufe der Zeit streng monoton gefallen sind. Der Median der Keys beträgt 35,5 und der Durchschnitt der Keys beträgt ebenfalls 35,5. Die Trendlinienformel legt nahe, dass die Keys auch in Zukunft weiter fallen werden.

Datensatz 4:

Beschreibungen von Datensatz 4:

Die vorliegenden Daten bestehen aus 100 Werten, die in zwei Spalten aufgeteilt sind. In der ersten Spalte befinden sich die Datenpunkte von 1 bis 100, während in der zweiten Spalte die entsprechenden Werte aufgeführt sind. Es gibt keinen Modus in den Daten, da jeder Wert nur einmal vorkommt. Der Mittelwert der Werte beträgt 23,7325 und der Median ist 33,45. Die Stichprobenvarianz der Werte ist 434,8602.

In Bezug auf die Datenpunkte gibt es ebenfalls keinen Modus, da jeder Punkt nur einmal vorkommt. Der Mittelwert der Datenpunkte ist 50,755102 und der Median beträgt 51,5. Die Stichprobenvarianz der Datenpunkte beträgt 855,403324.

Verwendete Software und Funktionen für alle Datensätze:

Excel: Scatterplot + Trendlinienfunktion + Einfache Rechenoperationen/Anweisungen zur Berechnung von den Koeffizienten, =QUARTILE(X:Y) für quartile.

sciPy,matplotlib,pandas: Alle extensions wurden importiert, `df = pd.read_csv`, `df.mean()`, `df.median()`, `df.mode()`, `df['Importe'].var()`, `df['Importe'].std()`, `df['Importe'].max() - df['Importe'].min()`, `matplotlib.pyplot.boxplot(df['Importe'])`