

Grundlagen und Anwendung der Wahrscheinlichkeitstheorie

Wintersemester 2022/2023

L02 Projekt Datenanalyse

Gruppe 06

Björn Kämpfert

Felix Luis Fritsche

Benedict Casper Satter

Inhalt

Datensatz 4

4.1 Informationen zu den Daten

4.2 Bereinigte Form

4.3 Maßnahmen zur Bereinigung

4.4 Verwendete Software und genutzte Funktionen

4.5 Modus, arithmetischer Mittelwert, Median

4.6 Stichprobenvarianz

4.7 Box-Whisker-Plot

4.8 Zusammenfassung

4.1 Informationen zu den Daten

Unseren Datensatz haben wir über die `Rand()` Funktion in C generieren lassen. Sie liegen zwischen 0 und 100 und sind zufällig generiert.

4.2 / 4.3 Bereinigte Form / Maßnahmen zur Bereinigung

Da unser Datensatz zufällig von einem Programm generiert wurde, gab es keine Notwendigkeit den Datensatz zu bereinigen. Alle Werte befinden sich im Intervall von 0 bis 100.

4.4 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`.

Zum Generieren der Daten wurde ebenfalls in VS Code die `rand()` Funktion in C verwendet.

4.5 Modus, arithmetischer Mittelwert, Median

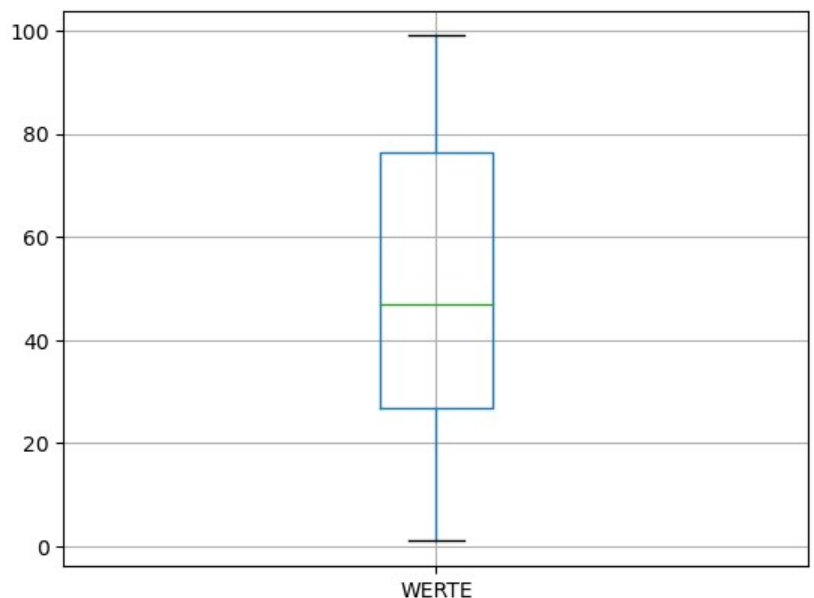
Modus: 45, 86, 99

Arithmetischer Mittelwert: 51.17

Median: 47.0

4.6 Stichprobenvarianz 810.36

4.7 Box-Whisker-Plot



4.8 Zusammenfassung

Unser 100-Werte großer 1D-Datensatz, der Werte zwischen 0 und 100 generiert, hat einen arithmetischen Mittelwert von 51.17, was so ziemlich in der Mitte des Intervalls liegt. Drei Werte kommen am häufigsten vor, also gibt es keinen bestimmten Modus, die Werte 45, 86 und 99 kommen gleich oft vor. Die Stichprobenvarianz beläuft sich auf 810.36, diese gibt uns den quadrierten Abstand zum Mittelwert. Im Boxplot sieht man, dass sich die Werte eher noch oben hin häufen.