

RPTU Kaiserslautern

DATENPROJEKT

Datensatz 4

verfasst von

Jens Bleymehl

427741 (Matrikelnummer)

Isaac Lee Zen Xue

427008 (Matrikelnummer)

Grundlagen und Anwendungen der Wahrscheinlichkeitstheorie

Wintersemester 2023 / 2024

Dozent: Christian De Schryver

Quellenverzeichnis

Experiment:

- <https://humanbenchmark.com/tests/reactiontime>

Webseiten

- <https://miniwebtool.com/de/variance-calculator/>

Vorwort

In diesem Datensatz wird die Reaktionszeit einer Person anhand mehrerer erhobener Daten dargestellt. Diese beschreiben, mit welcher Zeitverzögerung die Person nach einem Bildschirmfarbenwechsel von rot nach grün ihre Maustaste betätigte. Dieses Experiment wurde insgesamt 120 Mal von ihr durchgeführt (= 120 Datenpunkt der Urliste), wobei nach jeder Stichprobe das entsprechende Ergebnis in eine Tabelle eingetragen wurde.

R4.3 - Maßnahmen zur Datenbereinigung

Es wurden alle Daten entfernt, die mit „Too soon“ markiert waren. Diese sagen aus, dass der Test ungültig war und dementsprechend keine Reaktionszeit gemessen werden konnte (in diesem Fall durch Betätigen der Maustaste bei rotem Bildschirm). Außerdem wurden alle Daten über 1000 ms entfernt, da diese zu weit von dem Erwartungsbereich entfernt lagen (1876 ms, 1116 ms, 1322 ms). Die Werte 131 ms, 424 ms, 510 ms, 379 ms und 348 ms weisen zwar auch eine erhöhte Abweichung bezüglich der anderen Werte auf, jedoch haben wir uns dennoch dazu entschieden, diese zu inkludieren. Wir taten dies, um zu illustrieren, dass mögliche Konzentrationsschwankungen Einfluss auf das Ergebnis des Tests gehabt haben könnten. Schließlich wurde der Wert 308 ms aus der Analyse ausgeschlossen, da die Person den Test nicht auf gleiche Weise durchführte (sie klickte hier durch das Laptop-Touchpad statt durch die vorgesehene Maus).

Für ein besseres Verständnis der Datenbereinigung haben wir folgende Übersicht konstruiert:

- Entfernter Datenpunkt, bei dem die Person statt einer Maus ein Touchpad zum Klicken benutzte:
 - Wert: 308 ms

- Alle über 1000 ms gelegenen entfernten Datenpunkte, welche zu weit außerhalb des Erwartungsbereichs lagen:
 - Wert: 1876 ms
 - Wert: 1116 ms
 - Wert: 1322 ms
- Inkludierte Datenpunkte zur Illustrierung des möglichen Konzentrationsschwankungszusammenhangs:
 - Wert: 131 ms
 - Wert: 424 ms
 - Wert: 510 ms
 - Wert: 348 ms
 - Wert: 379 ms

R4.4 - Benutzte Softwares und Funktionen

Die Analyse dieses Datensatzes wurde unter Verwendung der Software GitHub, zusammen mit der Programmiersprache Python und -umgebung Visual Studio Code, sowie verschiedenen Tools von Google (wie Google Sheets und Google Docs) durchgeführt. Zur Datensatzerhebung wurde die im Quellenverzeichnis hinterlegte Website “humanbenchmark.com” verwendet.

R4.5 - Modus, Mittelwert und Median

Modus: In diesem Datensatz gibt es zwei Werte, die am häufigsten vorkommen:

- 195 ms und 198 ms jeweils 5 mal

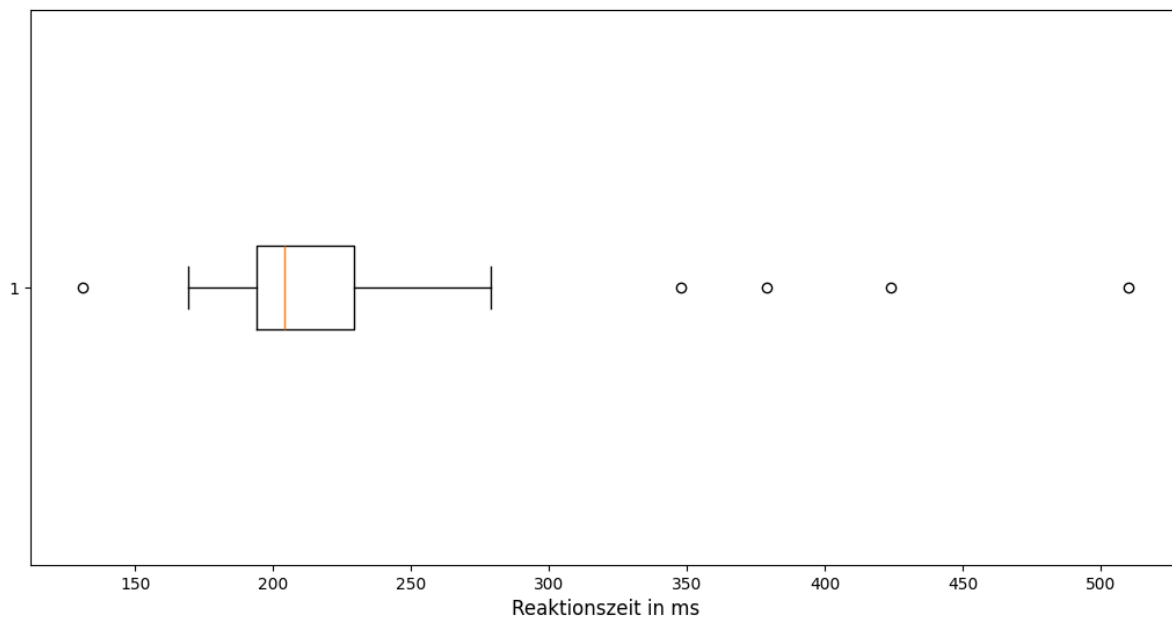
Mittelwert der Reaktionszeit ≈ 217.63 ms

Median: 204 ms

R4.6 - Stichprobenvarianz

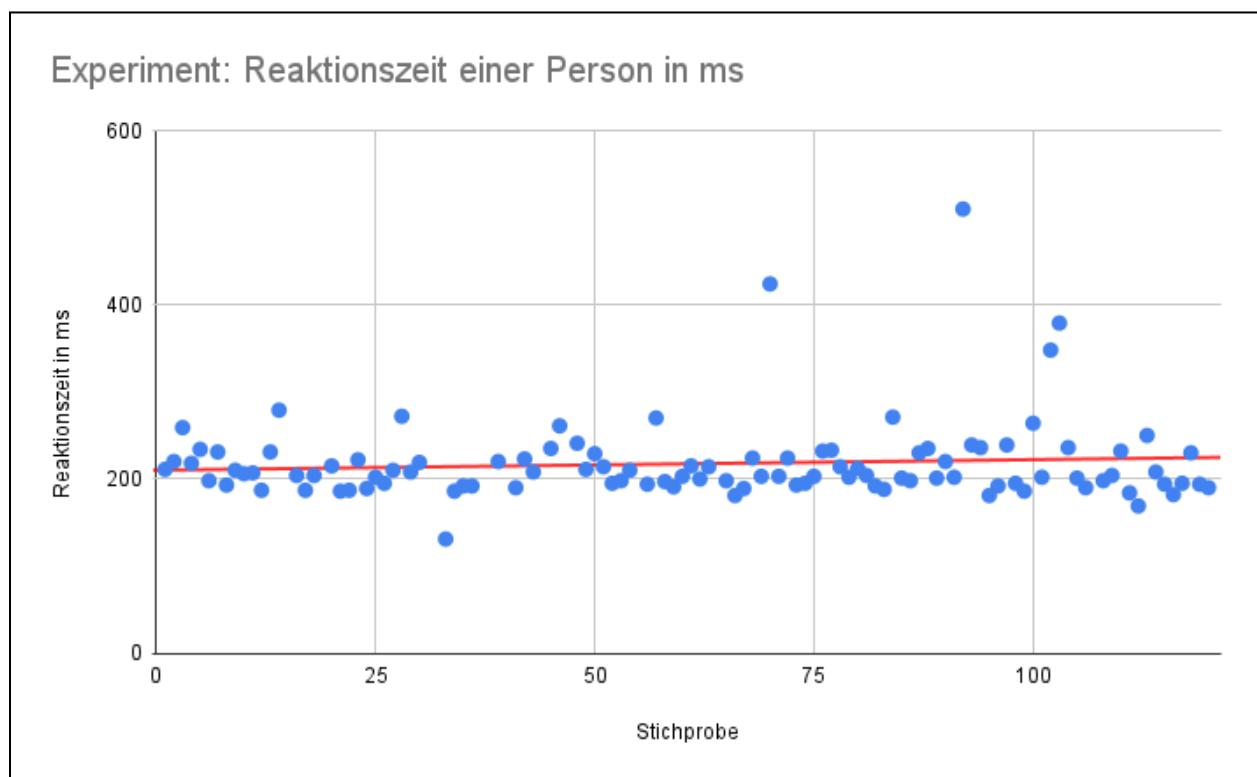
Die Stichprobenvarianz der Reaktionszeit beträgt $s^2 \approx 2206.6466 \Rightarrow$ Standardabweichung $s \approx 46.975$.

R4.7 - Box-Whisker-Plot



Bemerkung: Die durch Kreise markierten Stellen des Plots stellen die zuvor genannten stärker abweichenden Datenpunkte dar.

Scatterplot



R4.8 - Fazit

Da der Mittelwert des Datensatzes doch deutlich höher als die nahe beieinander liegenden Modi ist, wird klar dass hier eine Abweichtendenz nach oben vorliegt. Diese ist in Verbindung mit den inkludierten, stärker abweichenden Datenpunkten zu bringen, die mit einer Konzentrationsschwankungsvermutung versehen wurden.

Durch das Betrachten des beigefügten Scatterplots wird visualisiert, dass die meisten Datenpunkte nahe beieinanderliegen. Dies führt zu der Vermutung, dass diese Person eine klare Reaktionszeit bei diesem Experiment besitzt und die Werte wahrscheinlich nicht zufällig entstehen. Der Mittelwert ist hier also ein gutes Maß für die tatsächliche Reaktionszeit der Person bei diesem Experiment.