

Grundlagen und Anwendung der Wahrscheinlichkeitstheorie

Wintersemester 2022/2023

L02 Projekt Datenanalyse

Gruppe 06

Björn Kämpfert

Felix Luis Fritsche

Benedict Casper Satter

Inhalt

Datensatz 1 (raw)

1.1 Struktur und Inhalt

1.2 Skalenvariante

1.3 Verwendete Software und genutzte Funktionen

1.4 – 1.6 Urliste, Rangliste und Excel-Datei (Anhang)

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

1.15 / 1.16 Quartile und Dezile, Quartilabstand

1.17 Kovarianz

1.18 Korrelationskoeffizient

1.21 Rangkorrelationskoeffizient nach Spearman

1.1 Struktur und Inhalt

Der gegebene Datensatz befasst sich mit Druckwerten während einer Aufzugsfahrt. Es liegen somit die Druck- sowie die entsprechenden Zeitwerte vor.

1.2 Skalenvariante

Die verwendete Einheit für den Zeitwert ist Sekunde, für den Druckwert ist es Hektopascal, die Skalenvariante beider ist eine Verhältnisskala.

1.3 Verwendete Software und genutzte Funktionen

Die Werte wurden von Christian De Schryver selbst erstellt beziehungsweise mit Phyphox getrackt.

Es liegt eine Datei mit Datensatz in UTF 8-Kodierung vor. Der Datensatz ist in einer CSV-Datei mit einer Größe von 5.118 Bytes vorhanden. Die Trennung in den Urdateien verwendet als Dezimalzeichen ein Komma und als Trennzeichen ein Semikolon.

Zum Bearbeiten des Datensatzes wird Visual Studio Code benutzt, mit der Programmiersprache Python, dem Jupyter Notebook und den extra Bibliotheken pandas, numpy, matplotlib.

1.7 Modus, arithmetischer Mittelwert, Median

Modus: n.A., da es jeden Wert nur einmal gibt

Arithmetischer Mittelwert

Time (s)	13.553961
Pressure (hPa)	981.369634

Median

Time (s)	13.554693
Pressure (hPa)	981.321289

1.8 Spannweite

Time (s) 26.698753
Pressure (hPa) 1.553223

1.9 mittlere Abweichung vom Median

Time (s) 6.699672
Pressure (hPa) 0.429906

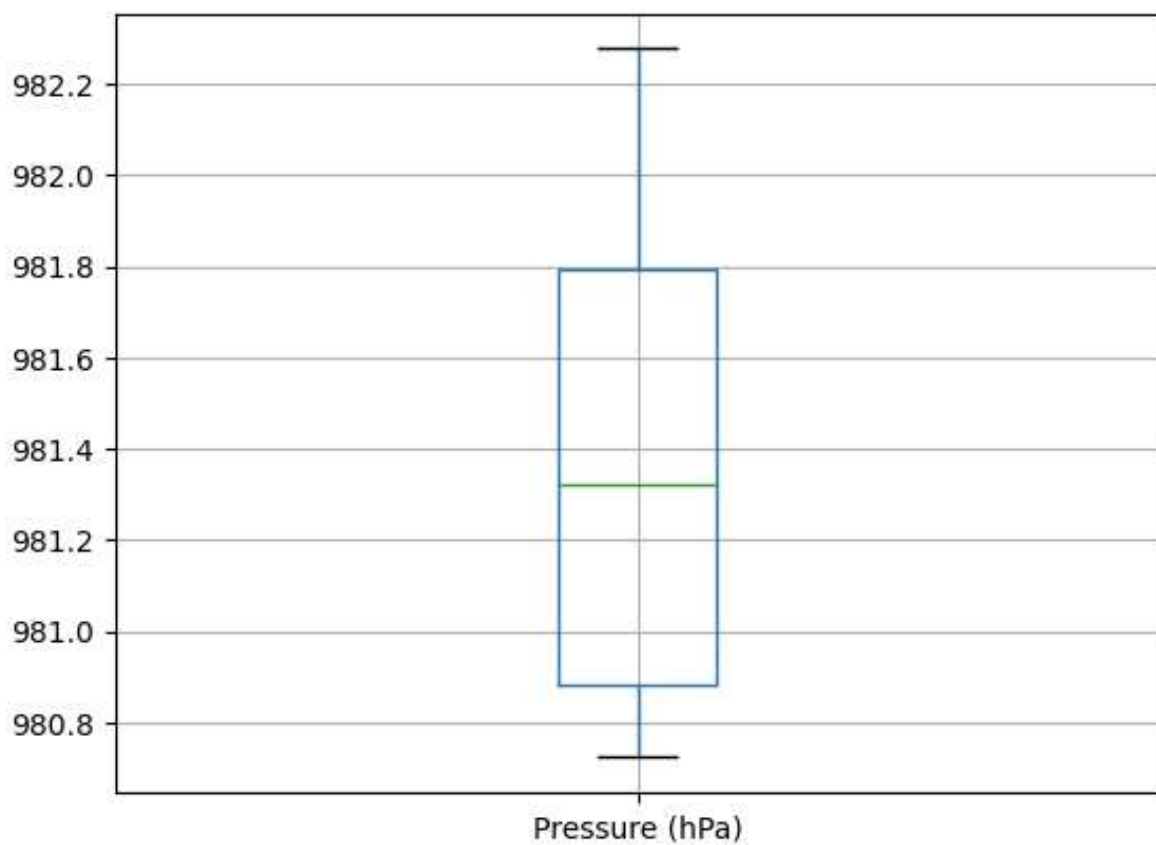
1.10 Stichprobenvarianz

Time (s) 60.073123
Pressure (hPa) 0.234525

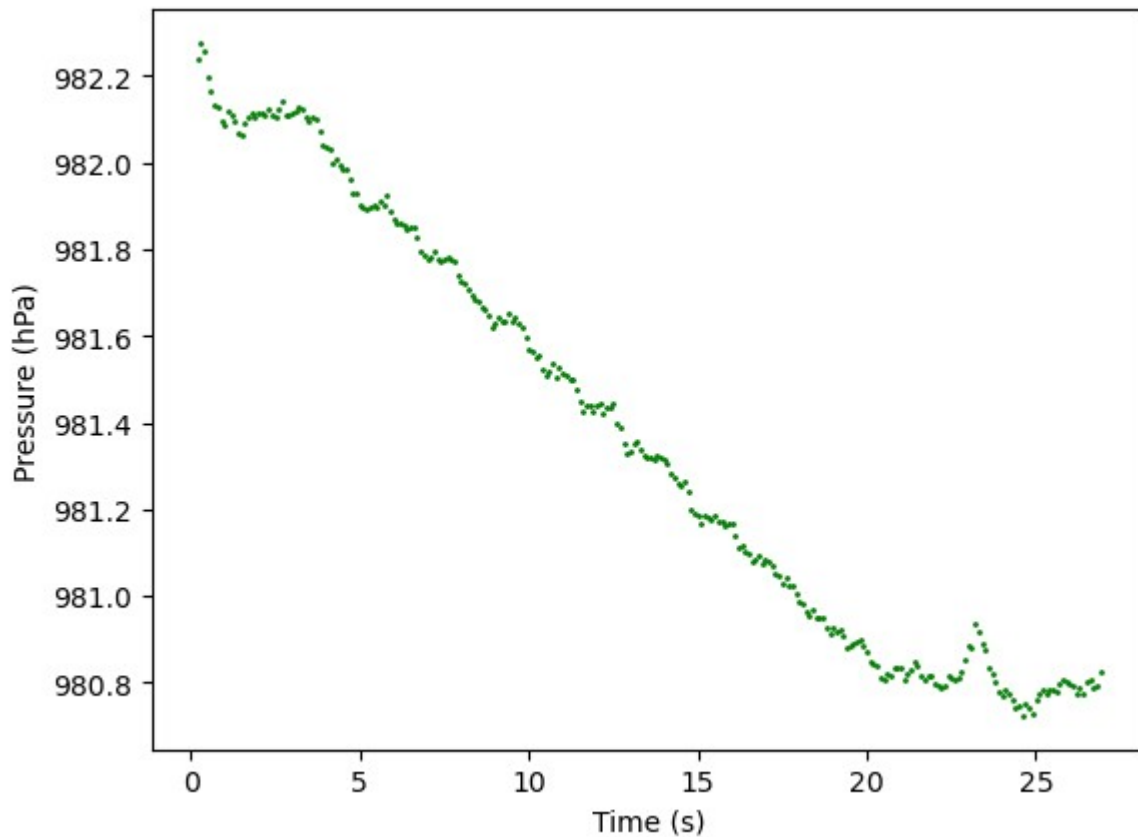
1.11 Variationskoeffizient

Time (s) 57.183913
Pressure (hPa) 0.049347

1.12 Box-Whisker-Plot



1.13 Scatterplot



1.14 Zusammenfassung

Es lässt sich erkennen, dass der arithmetische Mittelwert nicht weit vom Median entfernt liegt, woran man ausmachen kann, dass die Verteilung der Druckwerte ziemlich symmetrisch um den Median liegen, was auch zusätzlich im Scatterplot zu beobachten ist. Die Spannweite, also die Dauer des Experiments beträgt ungefähr 26.699 Sekunden, der Druck ändert sich im ganzen Experiment um maximal 1.55 hPa. Die Stichprobenvarianz von 0.235 gibt uns den quadrierten Abstand

Der Variationskoeffizient von 4.9% sagt aus, dass die Druckwerte sich nicht allzu stark vom Mittelwert streuen, ebenso erkennen kann man das an der Stichprobenvarianz, also dem quadrierten Abstand vom Mittelwert mit, 0.235.

1.15 / 1.16 Quartile und Dezile, Quartilabstand

Quartilsabstand: $Q3 - Q1 = 13.35$

	Time (s)	Pressure (hPa)
<i>min</i>	0.203140	980.723144
<i>25%</i>	6.879581	980.883545
<i>50%</i>	13.554693	981.321289
<i>75%</i>	20.229424	981.793762
<i>max</i>	26.901893	982.276367

	Time (s)	Pressure (hPa)
<i>0.1</i>	2.873140	980.796997
<i>0.2</i>	5.544581	980.832471
<i>0.3</i>	8.214590	980.927686
<i>0.4</i>	10.884693	981.113184
<i>0.5</i>	13.554693	981.321289
<i>0.6</i>	16.224692	981.510010
<i>0.7</i>	18.892181	981.705835
<i>0.8</i>	21.563979	981.898193
<i>0.9</i>	24.233499	982.104614
<i>1.0</i>	26.901893	982.276367

1.17 Kovarianz

-3.697132

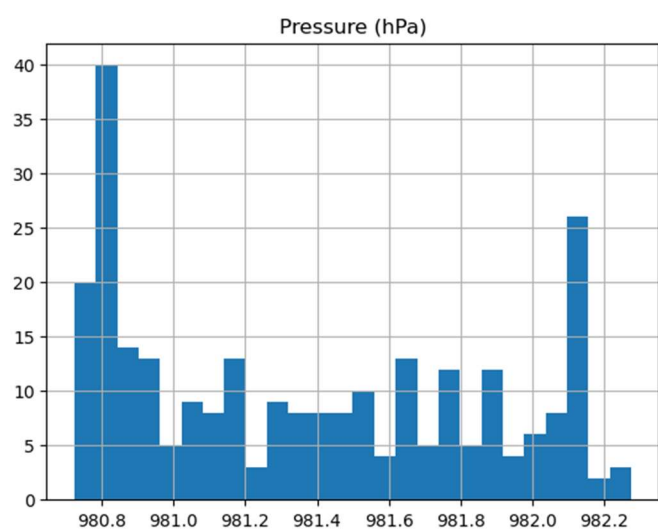
1.18 Korrelationskoeffizient

-0.984987

1.19 Klasseneinteilung

Wir haben den Druck in 26 Klassen eingeteilt, da 268 Werte vorliegen.

Histogramm:



1.21 Rangkorrelationskoeffizient nach Spearman

-0.989849

Quellenverzeichnis

Folien aus Vorlesung

<https://pandas.pydata.org/>