

Praktikum 1: Installation Python und Deskriptive Statistik

Installation Python und Arbeitsumgebung

- Finden Sie eine:n Praktikumpartner:in! Fragen Sie dazu z.B. im Forum und in der Vorlesung.
- Laden Sie von <https://www.anaconda.com/download> Python 3.10 für Ihr Betriebssystem herunter, und installieren Sie es wie dort angegeben.
- Achtung:** Dies ist eine vollständige Python-Distribution mit Python, Bibliotheken, Entwicklungsumgebung, Jupyter, etc. Daher ist die Installationsdatei sehr groß, was Sie beim Download berücksichtigen sollten. Sie sollten noch mindestens **3GB** Platz auf Ihrem System haben.
- Sie können Python auch auf andere Art installieren oder das möglicherweise auf Ihrem Rechner schon installierte Python3 (**nicht Python2!**) benutzen. Wichtig ist aber, dass Sie auch eine Arbeitsumgebung haben. Dazu reicht im einfachsten Fall die Kommandozeile und ein Texteditor. Im Folgenden wird davon ausgegangen, dass Sie *anaconda* mit der Entwicklungsumgebung *spyder* benutzen.
- Falls Sie mit Python nicht vertraut sind oder eine Auffrischung benötigen, schauen Sie sich die Einführung auf der Kursseite an.

Aufgaben

Schreiben Sie ein Programm `main1.py`, in dem Sie folgende Funktionen mit den angegebenen Rückgabewerten (keine Ausgabe nötig) definieren.

- `mittel(x)`: Berechnung und Rückgabe des arithmetischen Mittels der Zahlen in x . Dabei soll x ein *iterable*, also etwa eine Liste oder ein *numpy*-Vector von Zahlen sein.
- `quartil(x, p)`: Berechnung des Quartils gemäß Vorlesung. Dabei ist x gegeben wie oben und p eine Zahl zwischen 0 und 1.
- `median(x)`: Berechnung des Medians.
- `var(x)`: Berechnung der unkorrigierten Stichprobenvarianz.
- `regres(x, y)`: Rückgabewert soll ein Dreitupel sein, bestehend aus Steigung und Achsenabschnitt der Regressionsgeraden sowie dem quadratischen Fehler.
- `pca(X)`: Rückgabewert soll ein Dreitupel sein, bestehend aus der Transformationsmatrix Q , dem Vektor der Eigenwerte (in fallender Reihenfolge), sowie dem transformierten Datensatz in Matrixform, siehe Vorlesung.

Bearbeiten Sie die folgenden Aufgaben.

- Bestimmen Sie den Median der Zahlen $2^{3^{5^7}}$, $2^{3^{7^5}}$ und $2^{3^{6^6}}$, sowie denjenigen der Zahlen in $\{z^{3^{5^7}} \mid z \in \{2, 3, -2, -3\}\}$.
- Finden Sie für die Prozentzahlen aus der zweiten Tabelle von Beispiel 1.3.2 Anzahlen von Bewerberinnen für die beiden Fächergruppen A und B, so dass sich eine Gesamtzulassungsrate von 50% für Bewerberinnen ergibt.
- Diagonalisieren Sie die Matrix $S = \begin{pmatrix} 363 & 60 \\ 60 & 482 \end{pmatrix}$, und geben Sie dabei die Einträge der in der Vorlesung definierten Matrizen Q und D als ganze oder als Bruchzahlen an.
- Führen Sie per Hand eine Hauptkomponentenanalyse für den Datensatz

$$X = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \\ 4 & 5 \\ 5 & 4 \end{pmatrix}$$

durch. Geben Sie Q und D an, und zeichnen Sie die Originaldaten X und die transformierten Daten.

Abgabe

- Laden Sie das Archiv `P1vorgabe.zip` von [moodle](#) herunter, entpacken Sie es, und testen Sie Ihre Programme, indem Sie `test1.py` im gleichen Verzeichnis mit `python` ausführen. Erhalten Sie `ERROR`, so entspricht Ihr Programm nicht der Spezifikation von oben. Erhalten Sie `FAIL`, so ist Ihr Programm zwar lauffähig, aber die berechneten Werte sind fehlerhaft.
- **Abgaben, bei denen der Test gar nicht durchläuft oder mit `ERROR`, werden nicht akzeptiert**, `FAIL` führt nur zu Punktabzug.
- Sie finden in obigem Archiv auch die Datei `info1.md` mit anzugebenden Informationem zu Ihrem Team und Ihrer Abgabe, **bitte füllen Sie diese nach dortiger Anleitung aus, und vergessen Sie nicht die Quellenangabe**. Abgaben mit unvollständiger Datei `info1.md` können nicht gewertet werden. **Belassen Sie diese Datei in der UTF-8-Kodierung**.
- Komprimieren und bündeln Sie alle oben erzeugten oder geänderten Dateien, indem Sie ein ZIP-Archiv erstellen. Sollten Sie nicht wissen, wie das geht, konsultieren Sie dazu die Dokumentation Ihres Betriebssystems.
- Nennen Sie Ihr ZIP-Archiv `P1.zip`.
- Schreiben Sie eine Email an rosehr@hm.edu mit Betreff **Abgabe Datenanalyse**, Dateianhang **P1.zip** und irgendwelchem sonstigen Inhalt. **Der Automat akzeptiert die Abgabe nur, wenn diese Angaben (Betreff, Dateiname), sowie Ihre persönlichen Angaben (Nach-, Vor- und Teamname) korrekt sind**. Sie erhalten innerhalb der nächsten 24 Stunden (meistens aber deutlich schneller) eine Bestätigungsemail, wenn alles korrekt war.

Praktikumstermine P1: 03.04. (Gruppe a) und 17.04. (Gruppe b), **Abgabetermin:** 19.04.