

Prüfung 3 Mathematik

Prüfungsinhalt: Statistik und Wahrscheinlichkeit mit Python: Lage- und Streumasse, Visualisierungen (Histogramm, Boxplot, Scatter-Plot) und Zusammenhänge zwischen Merkmalen (lineare Regression, Korrelationskoeffizient)

Datum: Mi, 18.09.2024

Dauer: 90 Minuten; anschliessend Hochladen der Lösungen als Jupyter-Notebook-Dateien in «Teams-Aufgaben»

Regeln:

- **Für jede Aufgabe ist eine neue Jupyter-Notebookdatei zu erstellen.**
- Falls eine **KI-Software** benutzt worden ist, ist dies beim entsprechenden Abschnitt oder Aufgabe zu **vermerken**.
- Für die **Bestnote** müssen **nicht alle Punkte erreicht werden**.
- **Nicht erlaubt sind:** Interne und/oder externe Unterhaltungen jeglicher Art

Aufgabe 1 – Daten generieren, speichern, auswerten, vergleichen (Münze)

Eine Münze soll n -mal geworfen werden. Dabei soll jedes Ereignis «Kopf» oder «Zahl» gleichwahrscheinlich auftreten.

- a) Erstelle ein Pythonprogramm, welches dir diese Daten generiert und in die Datei «**daten_p3_1_out.csv**» speichert.
- b) Analysiere diese Daten, zähle, wie oft die beiden Ereignisse «Kopf» und «Zahl» aufgetreten sind.
- c) Erstelle ein Histogramm, welches diese Häufigkeiten visualisiert.
- d) Vergleiche diese Häufigkeiten mit der erwarteten (theoretischen Wahrscheinlichkeit) absolut und relativ. Gibt es Unterschiede – warum?

Aufgabe 2 – Daten generieren, speichern, einlesen, auswerten und visualisieren

- a) Es sollen 1'000 zufällige, normalverteilte Zahlen generiert werden. Damit wir alle die gleichen Zufallszahlen erhalten, setzen wir ein sogenanntes «Korn» oder «seed», welches mit einem Initialwert diesen Sachverhalt sicherstellt: «`np.random.seed(99)`». Du kannst beispielsweise die in «numpy» eingebaute Methode «`np.random.normal(loc, scale, size)`» für das Generieren der Zufallszahlen benutzen. Dabei bedeuten: *loc*: Mittelwert (Sollwert), *scale*: Standardabweichung und *size*: Anzahl Zufallszahlen (n). Als Beispiel kannst du folgende Vorgabewerte verwenden: *loc* = 17.5 und *scale* = 1.8. Die Anzahl ist 1'000.
- b) Speichere diese Zufallszahlen in der Textdatei: «**daten_p3_2.txt**» ab. Verwende dazu die Methode «`np.savetxt`» aus «numpy». Vergewissere dich, mit einem Editor, dass diese 1'000 Zufallszahlen in die Datei geschrieben worden sind.
- c) Berechne nun das Minimum, das Maximum, die Spannweite, den Mittelwert und die Standardabweichung dieser Daten.
- d) Stelle die Daten als Histogramm dar. Nutze dabei die Methode «`plt.hist()`» aus «matplotlib.pyplot»

Aufgabe 3 – Daten einlesen, auswerten und visualisieren

- a) Die in Aufgabe 2 generierten Daten sollen mit Hilfe der Methode «`pd.read_csv()`» aus dem Modul/Bibliothek «pandas», eingelesen werden.
- b) Füge nun den Daten eine Spaltenüberschrift «ZufallsZahl» hinzu und gib die ersten sechs Zeilen der Daten aus.

- c) Da die Daten mit Hilfe der «pandas»-Methode «read_csv()» eingelesen worden sind, gibt es nun eine Möglichkeit, sich schnell einen Überblick über die Daten zu beschaffen. Der Befehl heisst «describe». Lass dir dadurch die wichtigsten Parameter anzeigen. Welche Bedeutung haben die einzelnen Parameter?
- d) Erstelle mit Hilfe der Pandas-Methode ein Histogramm. Die Anzahl der Klassen k wird in der Regel durch die Formel $k = \lfloor \sqrt{n} \rfloor$ festgelegt.

Aufgabe 4 – Zusammenhang berechnen

Gegeben sind die untenstehenden Daten der Firma T-Maschinen AG. Es ist jeweils die maximale Temperatur innerhalb eines Zeitintervalls $[0, x_i]$ an einer bestimmten Maschine abgelesen worden:

Betriebsstunden in [h]	x_i	1.5	3.6	4.1	4.8	5.6	6.3	9.2	10.1
Temperatur in [°C]	y_i	32.1	35.4	36.2	36.5	37.2	39.7	41.3	43.5

- a) Berechne die beiden Koeffizienten der Regressions- respektive Ausgleichgeraden.
- b) Wie gross ist der Zusammenhang zwischen den Betriebsstunden und den Temperaturen? Berechne dazu den Korrelationskoeffizienten und interpretiere das Resultat.

Aufgabe 5 – Daten visualisieren und interpretieren

Gegeben sind die gleichen Daten, wie in Aufgabe 4.

- a) Stelle diese Daten in einem geeigneten Diagramm dar. Begründe deine Wahl.
- b) Ergänze deine Grafik mit der Regressionsgeraden.
- c) Erkennst du aus dieser Grafik einen Zusammenhang zwischen den beiden Merkmalen? Wenn ja, wie ist das zu interpretieren?
- d) Lies aus der Grafik (Interpolation) den ungefähren Wert für die Temperatur nach sieben Betriebsstunden ab.

Aufgabe 6 – Daten einlesen, Zusammenhang berechnen

Gegeben sind die Daten, welche in der Datei «daten_p1_6.csv» vorliegen. Die beiden Merkmale sind durch das Trennzeichen «;» voneinander getrennt.

- a) Lies diese Daten ein und lass dir die ersten fünf Datensätze anzeigen, damit du dir ein Bild über die Struktur der Daten und die Merkmale machen kannst
- b) Berechne die beiden Koeffizienten der Regressions- respektive Ausgleichsgeraden.
- c) Wie gross ist der Zusammenhang (Korrelation) zwischen den beiden Merkmalen? Berechne dazu den Korrelationskoeffizienten und interpretiere das Resultat.

Aufgabe 7 – Daten visualisieren

Gegeben sind die gleichen Daten, wie in Aufgabe 6.

- a) Stelle diese Daten in einem geeigneten Diagramm dar. Begründe deine Wahl.
- b) Ergänze deine Grafik mit der Regressionsgeraden.
- c) Erkennst du einen Zusammenhang zwischen den beiden Merkmalen? Wenn ja, wie ist das zu interpretieren?