

Testat Aufgabe 2

Untersuchen Sie die Bevölkerungsentwicklung der USA. In der Datei sr_aufg_2_xy.txt im Unterverzeichnis Endziffer_xy („xy“ steht für die beiden Endziffern Ihrer Matrikelnummer) finden Sie die dafür erforderlichen Daten. Diese Datei enthält zwei Spalten fester Breite: Jahreszahl und Anzahl der Einwohner der USA in Millionen (zu einem festen Stichtag in dem angegebenen Jahr).

1. Lesen Sie Daten in **Python** bzw. ein **Jupyter Notebook** ein. Speichern Sie sie in einem **Pandas DataFrame** ab.
2. Stellen Sie die Bevölkerungsentwicklung in einem Streudiagramm dar.
3. Untersuchen Sie die folgenden acht Funktionen um eine Regressionskurve durch die Punkte zu legen.

- Lineare Funktion: $y = ax + b$
- Polynom 2. Grades: $y = ax^2 + bx + c$
- Polynom 3. Grades: $y = ax^3 + bx^2 + cx + d$
- Polynom 4. Grades: $y = ax^4 + bx^3 + cx^2 + dx + e$
- Polynom 5. Grades: $y = ax^5 + bx^4 + cx^3 + dx^2 + ex + u$
- Polynom 6. Grades: $y = ax^6 + bx^5 + cx^4 + dx^3 + ex^2 + ux + v$
- Exponentialfunktion: $y = ae^{bx}$
- Logarithmusfunktion: $y = a \ln(x) - b$

Stellen Sie in einer Tabelle den Regressionstyp (z.B. lineare Funktion, Polynom 2. Grades, ...), die geschätzten Koeffizienten und das Bestimmtheitsmaß dar.

Gibt es eine Funktion für die eine Regressionskurve nicht geschätzt werden kann? Wenn ja, so geben Sie diese Funktion an.

Verwenden Sie für die Schätzung der Regressionskurve und des Bestimmtheitsmaßes normalisierte Eingangswerte (Jahreszahlen)

4. Wählen Sie aus den von Ihnen untersuchten Möglichkeiten die am besten geeignete Kurve. Erläutern Sie Ihre Wahl.

5. Zeichnen Sie in das Diagramm die optimale Regressionskurve aus 4) ein. Dabei sollen auf der x-Achse die Jahreszahlen und auf der y-Achse die Bevölkerungszahl angegeben sein.
6. Berechnen Sie aus den Ihnen vorgelegten Daten eine Prognose über die Bevölkerungszahl der USA im Jahr 2020. Erläutern Sie Ihr Vorgehen.

Hinweise:

- ⚠ Beim Wert für die Prognose sollten Sie überlegen, ob Ihr Wert stimmen kann oder ob er nicht vielleicht völlig unrealistisch ist. Der Grund für unrealistische Werte ist dabei evtl. gar nicht in der Wahl der Regressionsvariante zu suchen, sondern evtl. darin, dass numerische Probleme vorliegen und die geschätzten Koeffizienten mit zu wenigen Nachkommastellen verwendet werden!
- ⚠ Bitte überprüfen Sie Ihre Prognose vor der Abgabe mit dem Teilergebnis aus der Datei Teilergebnisse_xy.txt.

Die Unterlagen sind in als pdf-Format in Moodle hochgeladen abzugeben. Richtige Lösungen werden nicht mehr an Sie zurückgegeben. Eine Abgabe per Email ist nicht möglich.

Bitte drucken Sie das folgende Deckblatt aus und geben es mit Ihrer Lösung zusammen ab:

Statistik Rechnerübungen SWB 2 Sommersemester 2022, Aufgabe 2

Nachname:

Vorname:

Matrikelnummer:

Semester:

Email-Adresse:

Abgabe-Schlusstermin: Donnerstag, 12.05.2022