Lab 3

Exercise (Parametric and Non-Parametric Statistical Tests)

When evaluating the performance of algorithms, you often need to compare two (or more) data distributions. Statistical tests can help us to make statements about *whether* and *how much* one algorithm differs from another. In this lab, the aim is to implement an Artificial Statistician that automatically answers these questions. Given two or more sequences of data and an α level, the program should first analyze the properties of the data distributions. Then, the program should provide appropriate significance tests and effect measures to compare the data distributions (we restrict the Artificial Statistician to unpaired data and continuous target values).

1. Read the papers of Cohen (1994) and Hentschke & Stüttgen (2011) and summarize the arguments of the authors on the common methodology of Null Hypothesis Significance Testing. Do you agree or disagree with the authors?

2. In `artificial_statistician.py`, we provided two sequences. Extend the function `test_all(sequences,`α`)` so that it computes for the two sequences a and b

   (a) histograms showing the distributions of the data and

   (b) box plots showing the quartiles, outliers, etc.

3. In `test_all(sequences,`α`)`, a lot of statistics are already implemented. Which statistics are needed in which situation? Using pencil and paper sketch a decision flow of an Artificial Statistician. Discuss your design within your group and try to justify your decisions.

4. Adjust `test_all(sequences,`α`)` so that it implements the decision flow from ex. 3 and returns the results needed. Additionally to the generated results, provide appropriate interpretations.

5. Extend your code such that it is able to compare more than two sequences. Compare the best results of the EA from lab 2. Which team wins the competition?