First Assignment

Global Warming

Research question

Compare the yearly and monthly temperatures in the reference period 1951-1980 and the more recent period 1981-2010. Is the temperature change in *your country* significant?

Learning goals

- Use Python for function definitions, loops, conditions (→ Lecture 3.2), plotting, and reading and writing to files (→ Lecture 4.1).
- Use descriptive statistics to explore the data (→ Lectures 3.3 and 4.2).
- Use inferential statistics for hypothesis testing (→ Lecture 4.3).

Data

The country assigned to you can be found in the file **student_country_assignment1.csv**.

Temperature anomaly data from Berkeley Earth can be downloaded <u>here</u>. Click on the link to your country, and then on the link to the data table below the top figure. It opens a text file named ***country*-TAVG-Trend.txt**. Right-click on it, and then click on 'save page as'.

Tasks

- Task 1: Read the data into Python. Calculate (a) the number of data records, (b) the share of missing values, and (c) the share of ties (equal ranks, i.e. duplicate values) in the two periods combined. As a sanity check, test if the average of the anomaly in the reference period is close to zero, considering the accuracy of the data, and draw a conclusion. Calculate the intra- and interannual variation by means of the standard deviation. [Intraannual variation is the variation within a year, e.g., the monthly variation where a month is the average of this month across all years. Interannual variation is the variation between years, i.e. of 1951-2010, where each year is the average of the months within this year.]
- Task 2: Plot two histograms and the curves of the theoretical normal distribution, and perform two Kolmogorov-Smirnov or Lilliefors tests including its interpretation (accept / reject) to verify if the two samples are normally distributed. Use the information to choose either a parametric or non-parametric hypothesis test for means, and make the reason for the choice obvious in the code.
- Task 3: Calculate the difference in averages, the standard deviation for both periods, and the p-value of the hypothesis test (e.g., independent or paired T-test, Wilcoxon-Mann-Whitney test, or Wilcoxon signed-rank test) for annual averages and for each month. Write these data to a file together with an additional column that indicates the interpretation of the p-value (i.e. rejection of the null hypothesis or not). Make sure that the data file contains sufficient information (metadata) to understand its content and that it is easily readable.

- Task 4: Plot the time series for both periods as well as horizontal lines at their averages into one figure, including a well-chosen symbology and conventional annotations such as a legend. Both periods should cover the same x-axis range of 360 months, i.e. the x-axis represents relative and not absolute time. Save the figure in a file.
- Task 5: Write code that is nice (→ Lecture 3.2, slide 27) and free of errors and warnings (→ Lecture 5.2).

Deliverables

- A Python script should be named *country*_assignment1.py
- The input data file of your country

When running the script in the same folder as the data input file, no user interaction should be requested, and all requested results should be shown in the console except for those exported. Two additional files should be automatically created and also submitted:

- A comma-separated output data file named *country*_statistics.csv
- An image named *country*_time_series.png

Deadline: Friday, **2 November, 16:00** via Brightspace

Assessment

The assignment will be reviewed by 2 of your fellow students. Until **Friday, 2 November, 23:59**, you will **provide feedback to 2 of your peers** and 2 peers will provide feedback to you. Please be fair and honest in your evaluation and provide constructive feedback where you can. The final grading will be done by the lecturer.

The assignment is worth 1/5 of the final grade, and each of the 5 tasks weighs 20%. The code (syntax and logic), its meaning (semantics), and its output (clear, nice, and relevant) are evaluated.

There will be **no retake** within the same year. So, please take this assignment seriously. There is no reason to fail. If you have difficulties with the assignment, seek help on **Google / Stack Overflow** and attend the **Friday workshops**.