

Introduction to Data Analysis and Machine Learning with Python

Homework 4

Mon Apr 1 2024

Due: April 8 2024

Please submit only the answers to the questions, not your code.

1. Read the data from the "TopRight_20230803.txt" file into a Pandas DataFrame. This can be done with the `pandas.read_csv` function. You can use `"skiprows = 7"` to skip the first 7 lines in the data file. To use whitespace as the delimiter between data values, use `"delim_whitespace=True"`. Finally set the names of the columns using `"names = ['Event', 'Time', 'Date', 'TimeStamp', 'ADC1', 'ADC2', 'SiPM', 'Temp', 'Pressure', 'DeadTime', 'Coincident', 'ID']"`
2. Approximately, how many rows of data does the DataFrame have?
 - a. 12
 - b. 75000
 - c. 750000
 - d. 9000000
3. Approximately, what is the average value $\langle p \rangle$ for the "Pressure" data series?
 - a. 13
 - b. 500
 - c. 7200
 - d. 99000
 - e. 720000
4. Approximately, what is the uncertainty on the $\langle p \rangle$ value?
 - a. 0.3
 - b. 30
 - c. 300
 - d. 3000
 - e. 60000
5. Use the `dataframe.plot.scatter` command to plot Pressure vs TimeStamp. Is there a correlation between the two variables?
 - a. No, they are independent
 - b. Yes, pressure has a downward trend with time
 - c. Yes, pressure has an upward with time
6. Use the `dataframe.plot.scatter` command to plot SiPM vs TimeStamp. Is there a correlation between the two variables?

- a. No, they are independent
 - b. Yes, SiPM has a downward trend with time
 - c. Yes, SiPM has an upward with time
7. Using the “Coincident” column determine the fraction of so-called coincidence events in the file (i.e., rows with Coincident = 1). The fraction is about:
- a. 0%
 - b. 0.15%
 - c. 1.5%
 - d. 15%
8. Plot histograms of the “SiPM” values for rows with “Coincident” == 0 and rows with “Coincident” ==1 on top of each other. Subsets of the dataframe for certain column values can be selected with the “.loc” command (e.g., “`cw.loc[cw['Coincident'] == 1]`”, where “cw” is the name of the dataframe. For comparing the two histograms, it is useful to set “density = True” in the hist() command. Based on the comparison of the two histograms, would you conclude that:
- a. The SiPM distribution is different for coincident and non-coincident events
 - b. The SiPM distribution is the same for coincident and non-coincident events
 - c. Not enough information to tell