

# Activities\_Session 5

June 28, 2023

## 1 Activities Session 5

For each of the next activities, create a new **code cell** and write your solution.

For these activities you will need the file `5.1_data_session_5_activities.xlsx`, which contain 2 different sheets: `data1` and `data2`.

1. Open the file `5.1_data_session_5_activities.xlsx` and load the data for sheets `data1` to `data2` in pandas dataframes called `df1` and `df2`.
2. Load the dataset in sheet `data1` and consider it like your population of study. Compute the mean and the standard deviation.
3. Suppose you take samples of size=200 from the previous data set. Using the theory, what would you expect to be the average of all sample means, and what would you expect to be the standard error (that is, the standard deviation of all sample means)?
4. Now, select 5000 of these random samples and for each of them, save the sample mean in an array. Compute empirically the average of the sample means and the standard deviation of the sample means. Is it close to the theoretical results?
5. Do the same as before but now with the data on sheet `data2` and choosing sample size = 15. Do you also obtain similar results between theory and experiment? How would you explain this result?
6. Population is normally distributed with mean = 20 and standard deviation = 3. Compute an 80%, 90% and 95% confidence intervals or the mean of a sample size of size 30.
7. A process in an electronic manufacturing company produces integrated circuits with an error rate of only 0.1% (and the standard deviation is known to be .05%) The company buys a new machine to increase production, but you are afraid that the calibration of the machine might not be right, and that the failure rate in that production line will be higher. You ask the provider for a testing, where 1000 electronic chips are produced and tested, and the failure rate equals 0.101%. If you company typically uses significance levels of 1% to make investment decisions, should you recommend installing the new machine, or not?
8. The production manager of an electronic payment has asked for your help to analyse a risk management process. The process entails sending a large number of controlled hack attempts to their system and measure the slow down on the company servers. So far, the slowdown is only 1% in average when the attacks are sent (and the standard distribution of the slowdowns is about 0.1%). They want to install a new server facility and in their first 5 hack attempts, the average slowdown was already 1.1%. Should they stop the implementation

of the new server facility? Since this decision would have a big economic impact, you want to select a significance level of 0.01.

[ ]: