<PKU 통계스터디 5번째 >

Statistical Project training

2018/04/01

이준호

Using pre-defined fuction

import sys

sys.path.append(“PATH OF THE FUNCTION”)

from FILENAME_NAME import FUNCTION_NAME

- For my case…

import sys

sys.path.append("/Users/leejunho/Desktop/git/python3Env/group_study/fruit_team/ROOT/Project/functions/rawTxt_Tree_root")

from Raw_text_to_Tree_root import Raw_text_to_Tree_root

Please check …. (20M)

- 1. “TESTs/functions/c1_basic_statistic.py”

  (1) What kind of function do you see?

  (2) Mind returning value of each function!

  *algorithm ??

- 2. “TESTs/functions/c2_basic_histo_plotting.py”

  (1) What is this for?

Project 1 :: (20~30 Min) “TESTs/project_180401/project1_10K.txt ”

- 1. Using the script “python_script.py” on same directory, see what happens.
- 2. Uncommenting last two line of the script, see what happens.
- 3. Why is this happen? (Think)

- 4. Try to make your own script for inducing more functions. (print out all of the value defined inside of the function, save the plot)

- 5. What is “MEAN/Standard_deviation”?
- 6. What is “Total Bin number”?
- 7. What is “Variance/(Standard_deviation*Standard_deviation)” ?

Project 2 : Real Data from Stock. (25~30M) "TESTs/project_180401/project2_realDATA.txt"

- 1. Plot and take a look on the distribution. Also print to see mean, variance, …
- 2. Take a look on "raw_data" :: " TESTs/project_180401/**sh600000.txt**"

(Think)

- 3. The data pre-processing procedure performance from fruit-team.

(how to apply "cut")

- 4. Future plan..

Project 3. (1 hour) "TESTs/project_180401/project2_1M.txt

- 1. Draw & save the plot. See what does the distribution looks like
- 2. Difference between "Project1's" Plot? (Think)
- 3. Standard deviation value?
- 4. "X-range" of histogram, which height of two bins reach half of the heighest bin's. Compare this "X-range" with "[(mean-Standard_deviation),(mean+Standard_deviation)]"
- 5. Exclude outside of the "X-range", regenerate histogram. (suggest you to regenerate corresponding text file first. (To take advantage of function reusing))
- 6. On the base of 5, generate 25% Entry on both side excluded txt(so only remains 50% of '5's' Entry), and draw the plot.
- 6. Calculate "Total-Entry, MEAN, Standard_deviation" and "Normalize the histogram".