

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

EP219 Data analysis and interpretation

Assignment 2

Dated : 8 - 23 - 2017

Instructions

1. Take a look at the csv file facebookdata.csv that shows a sample of (anonymous) facebook user data. Each row is the record for an individual facebook user. The first column shows the number of facebook friends of the user. The second column shows the number of posts by the user over a 10 year period, and the third column shows the total number of likes for all posts.

Your Assignment

1. Make 1D histograms to show the distribution of number of friends, number of posts and number of likes. Calculate the mean and standard deviation of each of these quantities.
2. Make **all** possible 2D histograms of pairs of these variables. For example show the distribution of number of friends vs number of posts. Comment on each of the 2D histogram as to what intuitive inference you can draw about user behavior. Also try to give some kind of educated guess (with reasoning from your histograms) about what kind of sample of users might have been selected for this data set.

Notes: (as before try to follow these guidelines)

- Make sure python 2.7/2.8 is installed. We will prefer this to python 3 for this course.
- Make sure to label all your plots, axes etc. Install latex so that you can use latex symbols in the plot legends.
- Try to experiment with histogram bins, axes range, colors, linestyle, plot markers, displaying multiple plots on the same image, saving plots to pdfs etc.
- Comment your code with detailed comments! Uncommented code will receive no credit.
- Try to follow best programming practices in python. <https://gist.github.com/slوريا/7001839>