

ME 794 Statistical Design of Experiments

Assignment 2

Due: February 8, 2023

1. MS Dhoni is a prolific Indian cricketer and former team captain. He is known to be the best finisher of his time based on his average in games chasing a score. Using his performance data provided in the file 'MS Dhoni ODI Batting.csv', complete the following exercise using the Python script. [3+2+3 marks]

- a. Write a function/script that plots the probability distribution for his runs scored in each match, i.e., the y-axis represents the probability, and the x-axis shows the match number (consider data to be discrete). The function should also annotate the mean, median and mode values of the runs he has scored. Based on your plot, comment on the probability distribution of MS Dhoni's performance. Also, based on his performance, comment on the plot's skewness and kurtosis (the plot colour will be blue).
- b. Write another function to plot and compare the mean, median and mode values of his runs scored in each calendar year from 2004-2015 and comment on his performance based on the results. The mean should be represented with error bars considering ± 1 standard deviation. (The plot colour will be blue for mean, red for median and black for mode).
- c. From the data provided, extract his performances in ICC events (sample) over the years and plot and compare it with overall performance based on mean with ± 1 standard deviation. Comment on the results observed.

Comment
←

2. From a population of $N = 10$, shown in the table below, [4+2+1 marks]

65	82	81	67	57
59	66	75	82	70

- a. Write a function/script in python, randomly selecting 3 data points from the population ($N=10$), and evaluate the sample mean and standard deviation. Repeat the code 'a' number of times, where $a \geq 15$.
- b. Plot
- sample mean vs population mean for every repetition
 - sample standard deviation vs population standard deviation for every repetition.
- c. Assign an array $p = [1, 10, 100, 1000, 10000, 100000]$, and plot the cumulative average of the 'p' sample means (of the randomly selected 3 data points).
- d. Comment on the results observed in the three plots.

For both questions,

- *the plots must have x- and y-axis labels and appropriate legends.*
- *Population data should be represented in red, and sample data should be represented in blue.*
- *line thickness should be two wherever applicable.*
- *If **not specified**, you are free to choose any plot style of your choice.*

Documents required to be uploaded while submitting the assignment in teams to receive full credit

1. Include the script file and results in a single pdf file.
2. Separate Jupyter notebook (ipynb format) for each question (two in this case).