



भारतीय सूचना प्रौद्योगिकी संस्थान गुवाहाटी
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY GUWAHATI

Data Analytics Lab, M.Tech 3rd Semester

Instructions

1. Upload all your codes to Github.
2. You will be called randomly to explain the code based on which marks/grade will be assigned.

Assignment -4

1. Perform the following tasks:
 - (a) In R a simple bar graph can be used to model the probability distribution function. Take a random vector and probability associated with it and plot it with a bar graph.
 - (b) For the above problem create a cumulative frequency table and plot the cumulative frequency against each sample point mentioned in the vector.
 - (c) Using some simple commands in R generate the probability values for the binomial distribution for the number of children in 10 with blue eyes using $p=0.16$. Plot the obtained result.
 - (d) Run the above example with probabilities that a child has blue eyes is 0.05, 0.2, 0.5, and 0.8 and see how this changes the distribution.
 - (e) Consider you have a vector 0:10. Compute poisson distribution with $p=0.2$. Plot it to visualise the distribution.
2. Data whose distribution is close to lognormal are common. Size measurements of biological organisms often have this character. As an example, consider the measurements of body weight (body), in the data frame Animals (MASS). Begin by drawing a histogram of the untransformed values, and overlaying a density curve. Then
 - (a) Draw an estimated density curve for the logarithms of the values.
 - (b) Determine the mean and standard deviation of $\log(\text{Animals\$body})$. Overlay the estimated density with the theoretical density for a normal distribution with the mean and standard deviation just obtained.
3. Take a random sample from the normal distribution, and plot the estimated density function. Now take repeated samples of size 4, calculate the mean for each such sample, and plot the density. Repeat the above: taking samples of size 9, and of size 25.