

BM5033 Statistical Inference Methods in Bioengineering

Assignment 1

Instructions

1. **Deadline:** Before/at the start of the class on 21st August
2. You have to write down the answers to the question clearly.
3. Assignment has to be submitted on the deadline by submitting the PDF of the answers to [this link](#). **Note that you are allowed to submit the PDF only once. So be careful.**
4. You are expected to work on these problems on your own. **Any reasonable signs of ‘copying/plagiarism’ will attract penalties.**
5. There are additional 5% bonus marks for neatly written answers.
6. The dataset for this assignment can be downloaded from [here](#).

Questions

1. Consider the following data obtained from 4 repetitions of an experiment measuring the rate of cell death due to an anti-cancer drug

Starting cell count	Time taken	Rate of cell death
10^6	2.0hrs	5.0×10^5 cells/hr
5×10^6	8.0hrs	6.3×10^5 cells/hr
2×10^6	4.0hrs	5.0×10^5 cells/hr
2×10^6	5.0hrs	4.0×10^5 cells/hr

What is the average rate of cell death in this experiment? (10)

2. Consider the following three data sets
 - (a) Volume (in mm^3) of AV malformation in the brain (A1Q1a.dat)
 - (b) Elastic modulus (in GPa) of femoral cortical bone (A1Q1b.dat)
 - (c) Hematocrit levels (in %) in dengue patients (A1Q1c.dat)

and answer the following

- (i) Make box plots for these data using R. (3 × 3)
- (ii) Identify and calculate the appropriate measure of central tendency and dispersion you will use for each. Justify your answer in 30 words for each case. (3 × 5)

3. Consider the data

- (a) Degree of anisotropy in compressive strength and age for male and female donors (A1Q2a.dat)
- (b) Hemoglobin (in g/dl) and hematocrit levels (in %) of dengue patients (A1Q2b.dat)

and answer the following

- (i) Make scatter plots of these datasets using R. **(3 × 3)**
 - (ii) Calculate the Pearson's, Spearmann's, and Kendall's correlations for all datasets. **(3 × 4)**
 - (iii) Identify the appropriate correlation measure for each. Justify your answers in 30 words for each case. **(3 × 3)**
4. As we saw in our discussion on the central limit theorem, the distribution of the sample means follows a normal distribution (under certain conditions on sample size and number of repetitions etc.). The mean of this distribution is the same as that of the population and the standard deviation of this distribution is equal to σ/\sqrt{n} (where σ is the population standard deviation, and n is the sample size) and is also called *standard deviation of the mean* or *standard error of the mean* or, sometimes, only *standard error*. Since in most situations, we do not know the population variance beforehand and have to estimate it from the sample itself. In that condition, the best estimate of the standard error is s/\sqrt{n} where s is the standard deviation of the sample.

Consider the dataset given below as a sample of measurements for systolic blood pressure (in mmHg)

121, 125, 128, 134, 136, 138, 139, 141, 144, 145, 149, 151

- (a) Calculate the *standard error of the mean* from this sample. **(6)**
- (b) What is the probability of obtaining a sample with a mean value > 146 for such a population? *Enumerate* all the assumptions you make for this calculation. **(10)**



భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
 भारतीय प्रौद्योगिकी संस्थान हैदराबाद
 Indian Institute of Technology Hyderabad