PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit - 1 - Introduction to Data Science

QUESTION BANK

Sampling

Exercises for Section 1.1

[Text Book Exercise – Section 1.1 – Q. No. [1 – 9] – Pg. No. [12 - 13]]

1. Each of the following processes involves sampling from a population. Define the population, and state whether it is tangible or conceptual.
   a. A chemical process is run 15 times, and the yield is measured each time.
   b. A pollster samples 1000 registered voters in a certain state and asks them which candidate they support for governor.
   c. In a clinical trial to test a new drug that is designed to lower cholesterol, 100 people with high cholesterol levels are recruited to try the new drug.
   d. Eight concrete specimens are constructed from a new formulation, and the compressive strength of each is measured.
   e. A quality engineer needs to estimate the percentage of bolts manufactured on a certain day that meet a strength specification. At 3:00 in the afternoon he samples the last 100 bolts to be manufacture.

2. If you wanted to estimate the mean height of all the students at a university, which one of the following sampling strategies would be best? Why? Note that none of the methods are true simple random samples.
   a. Measure the heights of 50 students found in the gym during basketball intramurals.
   b. Measure the heights of all engineering majors.
   c. Measure the heights of the students selected by choosing the first name on each page of the campus phone book.

3. A sample of 100 college students is selected from all students registered at a certain college, and it turns out that 38 of them participate in intramural sports. True or false:
   a. The proportion of students at this college who participate in intramural sports is 0.38.
   b. The proportion of students at this college who participate in intramural sports is likely to be close to 0.38, but not equal to 0.38.

4. A certain process for manufacturing integrated circuits has been in use for a period of time, and it is known that 12% of the circuits it produces are defective. A new process that is supposed to reduce the proportion of defectives is being tested. In a simple random sample of 100 circuits produced by the new process, 12 were defective.
   a. One of the engineers suggests that the test proves that the new process is no better than the old process, since the proportion of defectives in the sample is the same. Is this conclusion justified? Explain.
   b. Assume that there had been only 11 defective circuits in the sample of 100. Would this have proven that the new process is better? Explain.
   c. Which outcome represents stronger evidence that the new process is better: finding 11 defective circuits in the sample, or finding 2 defective circuits in the sample?

5. Refer to Exercise 5. True or false:
   a. If the proportion of defectives in the sample is less than 12%, it is reasonable to conclude that the new process is better.
   b. If the proportion of defectives in the sample is only slightly less than 12%, the difference could well be due entirely to sampling variation, and it is not reasonable to conclude that the new process is better.
   c. If the proportion of defectives in the sample is a lot less than 12%, it is very unlikely that the difference is due entirely to sampling variation, so it is reasonable to conclude that the new process is better.