PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit-1 - Introduction to Data Science

QUESTION BANK – SOLVED

Sampling

Exercises for Section 1.1 **[Text Book Exercise – Pg. No. [12]]**

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 manufactured.
[Text Book Exercise – Section 1.1 – Q. No.1 – Pg. No. 12]

Solution:
Step 1:

- Here for the given processes that involves sampling from a population, we have to define the population and then we have to state whether it is tangible or conceptual.
- Population can be defined as the collection of all units or the pool of units from which samples are drawn.
- Tangible means something that can be felt by touch.
- Conceptual means something that is based on some idea or some concept.

Step 2:

(a) Here it is given that a chemical process is run 15 times and the yield is measured each time. The population consists of all the times the process could be run. It is conceptual.

Thus, this is conceptual.

Step 3:

(b) Here the population is group of all registered voters in a certain state.
This process is tangible since the pollster get the response in numbers.

Step 4:

(c) Here the population consists of all people with high cholesterol levels.
This process is tangible.

Step 5:

(d) Here the population consists of all concrete specimens that could be made from the new formulation.
This process is conceptual.

Step 6:

(e) Here the population consists of all bolts manufactured that day. It is tangible.

2. Each of the following processes involves sampling from a population. Define the population, and state whether it is tangible or conceptual.
(a) A shipment of bolts is received from a vendor. To check whether the shipment is acceptable with regard to shear strength, an engineer reaches into the container and selects 10 bolts, one by one to test.
(b) The resistance of a certain resistor is measured 5 times with the same ohmmeter.
[Other Sources]

Solution:

Step 1:

(a) The population in the process consists of all the bolts in the shipment received from the vendor because it is a entire collection of objects about which information is sought.
Thus, all the bolts in the shipment is tangible.

Step 2:

(b) Here the population consists of all the values in the resistance of a certain resistor that might possibly have been observed. This process is conceptual because all measurements that could be made on that resistor with that ohmmeter.

3. 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.

[Text Book Exercise – Section 1.1 – Q. No.2 – Pg. No. 12]

Solution:

To take a sample to estimate the mean height of all students at a university and that the value you reach is statistically valid you need the sampling method to be random and representative of the whole population, in this example, all university students.

a. Measure the heights of 50 students found in the gym during basketball intramurals.

This method is not the best because you would be sampling only basketball players leaving all other students of the university outside, i.e. your sample will not be representative of all the students, just the ones that play basketball.

b. Measure the heights of all engineering majors.

This method is not good, the sample only represents engineering mayors meaning that it does not include the students of any other subjects.

c. Measure the heights of the students selected by choosing the first name on each page of the campus phone book.

With this method you choose students regardless of the sport or major they're are taking, it is more representative of the population of university students, of the three options, this is the best one.