**Excercise-6**

**Title: Noise smoothing and sampling**

**Part-A**

1. Create at dataset (excel/vector)/download of your choice having at least 200 observations (numerical).

2. Smooth it by binning as:

a. equal width

Calculate width bin size via Sturge’s Rule, Scott’s Rule, Rice’s Rule, Freedman-Diaconis’s Rule

b. equal depth

3. Observe histogram before and after binning for both methods (2.a and 2.b)

4. observe mean, SD, range before and after binning both methods (2.a and 2.b)

**Part-B**

1. Create at dataset (excel/vector)/download of your choice having at least 200 observations (numerical) and it must have wide range ( Ex. Min value-Max value must be greater than 100). Its hereby called population.
2. Perform random labeling (binary class A and B) with class percentage as A: 60% and B 40%.
3. Perform sampling (sample size 50%) without replacement as:
4. Simple random sampling
5. Stratified sampling
6. Measure Standard Deviation (SD) as:
7. SD (Population all together regardless of class label)
8. SD (for class A and B) – Population
9. SD (sample from 3.a, all together regardless of class label)
10. SD (for class A and B) – sample from 3.a
11. SD (sample from 3.b, all together regardless of class label)
12. SD (for class A and B) – sample from 3.b
13. Repeated steps 3-4 as sampling with replacement.
14. Fill the result summary in below table and observe the working of underlying techniques:

|  |  |
| --- | --- |
| **Criteria** | **SD** |
| Population (4.a) |  |
| Population (class A), 4.b |  |
| Population (class B), 4.b |  |
| Sample all together (4.c) |  |
| Sample class A (4.d) |  |
| Sample class B (4.d) |  |
| Sample all together (4.e) |  |
| Sample class A (4.f) |  |
| Sample class B (4.f) |  |
| Sample all together (4.c) |  |
| Sample class A (4.d) |  |
| Sample class B (4.d) |  |
| Sample all together (4.e) |  |
| Sample class A (4.f) |  |
| Sample class B (4.f) |  |