

Fundamentos de mineração de dados e ciência de dados

Coleta de dados - 08/10

Exercícios:

Considere o conjunto de dados sobre o molusco Abalone, que pode ser encontrado na página: <https://archive.ics.uci.edu/dataset/1/abalone>.

1 - A partir do conjunto de dados, amostrar 20% dos dados via:

- amostragem aleatória simples, sem reposição;
- amostragem aleatória simples, com reposição;
- amostragem sistemática.

2 - Escolher uma variável quantitativa e:

- ordenar;
- coletar um exemplo a cada $N/5$ (N é o tamanho do conjunto);
- fazer amostragem estratificada para 20% do conjunto.

3 - Escolher uma variável qualitativa e, para cada categoria, fazer amostragem aleatória simples (com e sem reposição) dentro dessa categoria, na proporção dessa categoria

Exercício da aula:

Escolher uma da lista e implementar