

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**Disciplina: **Agrupamento de Dados**Curso: **Ciência da Computação (graduação e pós-graduação)**Professora: **Elaine Ribeiro Faria****Aula Prática no 3****Alunos:**

Danilo Augusto Nunes - 11611BCC021

João Gomes - 11611BCC043

Perguntas:

1- Faça a binarização, usando codificação 1-de-n, dos seguintes valores: amarelo, vermelho, verde, azul, laranja, branco

Valor categórico	Valor inteiro	x1	x2	x3	x4	x5	x6
Amarelo	0	1	0	0	0	0	0
Vermelho	1	0	1	0	0	0	0
Verde	2	0	0	1	0	0	0
Azul	3	0	0	0	1	0	0
Laranja	4	0	0	0	0	1	0
Branco	5	0	0	0	0	0	1

2- Discretizar em 3 intervalos, o atributo que possui os valores a seguir: 0, 1, 3, 6, 6, 9, 10, 10, 10, 13, 18, 20, 21,21, 25

Usar:

a) Larguras iguais

Largura do intervalo = (valor máximo - valor mínimo)/ n° de intervalos

Largura do intervalo =  $(25-0)/3 = 8,3$ 

Intervalo	[0, 8.3]	[8.4, 16.6]	[16.7, 25]
Valores	0,1,3,6,6	9,10,10,10,13	18,20,21,21,25

b) Frequências iguais

Largura do intervalo = # de elementos/ n° de intervalos

Largura do intervalo =  $15/3 = 5$ 

Intervalo	1	2	3
Valores	0,1,3,6,6	9,10,10,10,13	18,20,21,21,25

3- Escreva um pseudo-código para representar a técnica de amostragem aleatória sem reposição

DataFrame é uma estrutura bidimensional de dados, como uma planilha.  
Caracteres entre chaves({}) representam comentários.

**Variáveis:**

população: **String** { arquivo de entrada dos dados que constituem a população }

amostra: **String** { arquivo de saída com os dados da amostra }

iterador, índice, tamanho\_amostra, tamanho\_populacao: **Integer**

p\_df, a\_df, registro: **DataFrame**

**leia** (população, amostra) { lê os nomes dos arquivos }

**leia** (tamanho\_amostra) { lê o tamanho que terá a amostra }

{ gera um data frame a partir da leitura do arquivo .csv }

p\_df <- le\_csv(população)

{ DataFrame tem uma função que retorna a quantidade de registros: tamanho() }

tamanho\_população <- p\_df.tamanho()

iterador <- 0

**enquanto** (iterador < tamanho\_amostra) **faça**

    { gera um número aleatório em um intervalo de 0 até o tamanho da população }

    índice <- número\_aleatório(tamanho\_população)

    { seleciona um registro da população pelo índice }

    registro <- p\_df.localiza(índice)

    { adiciona o registro na tabela de amostra }

    a\_df.adiciona(registro)

    { remove da população o registro selecionado anteriormente }

    p\_df.remove(índice)

    tamanho\_população <- tamanho\_população - 1

    iterador <- iterador + 1

**fim enquanto**

{ escreve os dados que constituem a amostra para um arquivo .csv }

escreve\_csv(a\_df, amostra)