

 04

Ordenando uma base

Transcrição

[00:00] Classificamos nossa variável de idade em cinco faixas, mas o PROCFREQ não está me dando muitas informações. Por exemplo, eu sei que tenho 21 pessoas nessa faixa de idade, mas quais idades estão dentro dessa primeira faixa? Isso eu não estou sabendo. Quais estão na segunda faixa? E assim por diante.

[00:26] Preciso de um outro procedimento que me mostre essas informações. O que precisamos ver agora são as informações da variável idade em si, mas não simplesmente da variável idade. Precisamos dela dentro de cada uma das classes que definimos na outra variável de faixas de idade.

[00:46] Voltando para o código, tínhamos olhado um procedimento na aula anterior em que podíamos obter diversas informações estatísticas sobre uma variável, que era o PROCUNIVARIATE, ele faz uma análise uni variada. Vamos tentar usar da mesma forma.

[01:22] Começamos passando nossa base de entrada, que no caso é o ranks, depois passamos um próximo parâmetro da variável que queremos analisar. Nós queremos analisar a idade. Vamos usar as faixas de idade só para separar a variável.

[01:45] Como falamos para o PROCUNIVARIATE que queremos fazer uma análise da variável separada por outras classes? Uma análise da variável de idade como se fosse um cruzamento dela. Escrevemos usando o by. Podemos fechar e executar.

[02:22] Ele deu um erro na log, porque ele diz que a base rank não está ordenada pela nossa variável que estamos querendo fazer pelo by. Precisamos voltar para o código e ordenar a base de ranks. Ordenar uma base é algo muito importante que é feito constantemente. Existe um proc já preparado para isso, que é o SORT. Não é muito diferente de outros procedimentos que vimos, em que começamos passando a base em que queremos trabalhar em cima e o que queremos fazer nele.

[03:18] Da mesma forma que usamos by faixas de idade, vai ser a mesma coisa. Queremos ordenar by nossa variável faixa idade. Fechamos com o run. Quando executarmos esse código, o que teremos como saída iremos mudar essa base rank para essa mesma base, só que agora ela vai estar ordenada não pela ordenação inicial que ela tinha, mas pela faixa de idade.

[03:52] Temos uma base, que é a mesma de ranks, mas se olharmos a variável faixa idade, vemos que ela começa com todos os valores em zero e vai crescendo até quatro. Se voltarmos para o código e executarmos novamente o UNIVARIATE, não precisamos alterar nada nele.

[04:28] O que tínhamos antes na nossa saída do UNIVARIATE é grandes informações estatísticas sobre a variável. Só que se formos analisar bem, o que temos agora não é uma só saída, um só conjunto de informações. Temos esse mesmo conjunto de informação repetido para cada uma das classes da nossa variável da faixa de idades.

[05:05] Se analisarmos a variável de faixa de idade, rank for variable idade é o label dessa variável. Lá, ele simplesmente está colocando o label da nossa variável. Voltando aos resultados, quando nossa variável de faixa de idade for igual a zero, ele solta toda a tabela de informações. Quando for igual a um, ele solta o mesmo tipo de informação para a nova faixa. E assim por diante. Ele mostra todos os resultados, mas cinco vezes, porque temos cinco faixas.

