

Introdução ao SQL

Transcrição

[00:00] Só por curiosidade, conseguimos gerar nossa base, mas voltando para o código temos também um procedimento muito parecido com o MEANS, chamado procedimento de sumarização de informações, que é basicamente o que fizemos. Esse é o PROCSUMMARY. É muito parecido com o MEANS, inclusive só vou copiar e colar. É só trocar o nome na chamada do procedimento.

[00:46] Uma característica dele é que ao contrário do MEANS ele não tem uma saída padrão nos resultados. Ele não precisa do no print que usávamos antes. Se eu executo, temos a mesma base, sem fazer grandes alterações. Se algum dia vocês se depararem com o PROCSUMMARY, é assim que ele funciona e é isso que ele faz.

[01:15] Também possuímos outro procedimento que pode ser usado para fazer isso de agrupar uma base dentro de determinada variável, fazer essa sumarização. Esse é um dos procedimentos mais importantes do SAS. Vocês vão perceber mais para a frente o motivo. Esse procedimento nos permite usar trechos de SQL, a linguagem SQL de manipulação de base de dados. Usamos PROCSQL.

[02:08] Fechamos a chamada dele com ponto e vírgula, porque a partir daqui o que vamos fazer é usar a linguagem SQL para fazer as manipulações na base de dados. Ele segue um padrão, uma forma de uso diferente dos procedimentos e dos data steps que vimos anteriormente. Vamos usar aos poucos. Tanto que começamos falando quais são as variáveis que queremos que ele selecione. Por exemplo, select faixa_idade. Ao invés de começar passando o nome da base que queremos usar como entrada. Depois, separamos as variáveis com vírgulas. Posso colocar a contagem da idade com count(idade).

[04:11] Agora vou falar de onde essas informações virão. Começo colocando o FROM. O nome da base é base_ranks. Vamos começar com esse procedimento simples para ver o que acontece.

[04:35] Esse PROCSQL é tão único que não encerramos com run, mas sim com encerrar, quit. Executando, ele gera algo difícil de entender. Tenho a variável de faixa de idade e outra que não tem nome, porque do mesmo modo que colocávamos antes como queríamos que essa coluna chamasse, aqui também temos que fazer isso.

[05:48] No final, o que queríamos fazer não era uma sumarização também? Não queremos pegar nossa base e agrupar de acordo com as classificações que temos dentro da variável de faixa de idade? Além de falar que base queremos passar de entrada, vamos falar para ele agrupar a base a partir da variável faixa de idade. Para agrupar, escrevemos group by e a variável. Se executarmos, temos uma saída bem melhor, com exatamente o que vimos antes. As mesmas informações geradas com PROCSQL.