

Informações das classes

Transcrição

[00:00] Geramos os resultados do UNIVARIATE, mas tem um problema. Geramos reformações demais com ele. Só queríamos a quantidade, máximo e mínimo dessas nossas faixas, seria suficiente. Mas temos muito mais. Fica até difícil encontrar as informações.

[00:33] Voltando para o código, vamos ver um procedimento bem parecido com o UNIVARIATE, mas por padrão gera uma saída mais simples. Conseguimos executar ele só trocando o nome para MEANS. Executando, a saída dele é muito mais simples. Ele tem como saída só a quantidade de observações na faixa, a média, o desvio padrão, mínimo e máximo.

[01:42] Na verdade, queremos que apareça só a quantidade, o mínimo e o máximo. Podemos voltar para o código e personalizar esse resultado falando que queremos só o N o MIN e o MAX, antes do ponto e vírgula.

[02:20] Voltando para o código, podemos inclusive colocar outras informações estatísticas. Se quisermos, podemos colocar várias, na ordem em que pedirmos.

[02:57] Outro detalhe é que queremos salvar essa base com as informações da faixa de idade, para que se alguém tiver interesse, podemos checar as informações nessa base e já vai ter tudo isso salvo. Não precisamos executar um PROCMEANS quando quisermos essas informações.

[03:27] Como fazer uma base de saída com as informações desse PROCMEAN? Tínhamos visto anteriormente que se queremos colocar uma base de saída escrevemos OUT igual o nome da base que queremos. Uma particularidade do PROCMEANS é que a saída dele possui diversos parâmetros que podemos personalizar a saída. Inclusive, fazer uma base de saída é uma das personalizações. Precisamos primeiro falar para o PROCMEANS que queremos que ele gere uma saída diferente. Escrevo output. E aí a base que vai ser gerada com uma saída é um desses parâmetros que passamos dentro de todos os outputs que queremos gerar. Se executarmos, vamos ver que continua imprimindo as nossas tabelas nas saídas, na aba de resultados, mas o que queremos olhar mesmo é a base que ele gerou.

[04:52] O que está acontecendo aqui é que temos a nossa faixa de idade com diversas linhas e temos uma coluna chamada STAT, que tem N, MIN, MAX, MEAN e STD. E depois temos a coluna com a idade e quais os valores que essa variável tem em cada uma dessas informações.

[05:20] Essa não é uma forma tão intuitiva de ver a base. O que queríamos é ter uma base com a faixa de idade, e cada uma das colunas que vêm a seguir seriam as estatísticas que queremos daquela faixa de idade. Não queremos que ela fique empilhada dessa forma, mas sim que ela fique mais horizontal. Por exemplo, uma variável com a faixa de idade, outra com o mínimo, outra com o máximo.

[05:57] Temos que voltar para o código para personalizar um pouco mais. Inclusive, não quero que fique imprimindo essas tabelas no resultado, porque já estou gerando uma base. Quero cortar essa parte. Vimos que o PROCMEANS por padrão sempre gera um resultado com as informações que tínhamos visto. Precisamos suprimir a opção de impressão automática do PROCMEANS. A gente faz isso falando que não queremos imprimir, ou seja, no print. Isso já faz com que não apareça nenhuma tabela automática nos nossos resultados.

[06:35] Também precisamos otimizar a mais o nosso output. Queremos que essas informações apareçam por colunas, precisamos dizer quais as colunas que queremos colocar e qual informação vai estar em cada uma delas. Por exemplo, queremos uma informação com o N, que é a quantidade. Colocamos N igual, ele vai criar uma coluna com essa

informação do N e nós colocamos depois o nome da coluna que queremos colocar. No caso, quantidade. Depois, podemos colocar as outras informações que queremos que apareça. Além da quantidade, queremos o mínimo, o máximo. Lembrando que no nome das variáveis não pode ter caracteres especiais.

[08:13] Tem duas variáveis que ele gera automaticamente que não quero aparecendo na minha base. Para tirar elas, podemos usar o comando de descartar as variáveis, o drop. Tínhamos visto nas opções que fazemos nas chamadas de base, tanto na entrada quanto na saída, abrimos o parênteses, escrevemos drop, colocamos o igual, quais as variáveis que queremos descartar. A primeira, é a *TYPE*, que só é importante se fôssemos fazer estatísticas com variáveis cruzadas, como por exemplo faixa de idade por Estado. Como estamos fazendo só por faixa de idade, não precisamos. A variável *FREQ* fica obsoleta porque já escolhemos que queremos que o N apareça numa coluna que chamamos de quantidade.

[09:20] Se executarmos, temos uma base com faixa de idade, a quantidade dessa faixa de idade, o mínimo e o máximo. Vemos nessa base já pronta tem faixa de idade zero, com 21 pessoas, de 18 a 21 anos.