

 04

Transformando uma data

Transcrição

[00:00] Vamos continuar trabalhando no nosso projeto para Alura Play. Dessa vez, o que eles querem saber é o perfil de idade dos nossos clientes. Eles querem fazer a estimativa e ter uma informação melhor sobre as características das pessoas que resolveram trabalhar com eles.

[00:20] Antes de começar, vamos voltar para o SAS. Já fiz a declaração da biblioteca. Vou dar uma dica bem interessante. Sempre que abrimos nossa seção de SAS, tínhamos que lembrar de fazer a declaração da biblioteca porque ele perdia a declaração quando encerrávamos a seção. Existe uma forma de deixarmos isso dentro de um arquivo auto executável, em que sempre que iniciarmos a seção, um trecho de código pré-definido é executado, ao mesmo tempo em que nossa seção de SAS é iniciada. Isso facilita um pouco nossa vida, porque não precisamos ficar lembrando sempre de declarar a biblioteca.

[01:23] Vou selecionar o código. Na barra de menu superior, do lado do botão de programador, existe o botão que são mais opções de aplicação. Vamos clicar nele. Depois, na primeira opção, que é editar arquivo auto executável. Nós podemos escrever nosso código. Copio a declaração da biblioteca. Podemos inclusive validar nosso código clicando no botão run. Ele mostra o que acontece quando executamos o código. Tudo funcionando, então podemos salvar.

[02:12] Sempre que iniciarmos o SAS, nossa biblioteca da Alura já é declarada automaticamente também e isso facilita bastante. Mas alguns cuidados precisam ser tomados. Por exemplo, vou manter essa declaração de biblioteca ainda no meu código, porque se eu passar para outra pessoa, talvez ela não tenha essa declaração automática da biblioteca no SAS dela e perde a informação.

[02:48] Como configurar esses arquivos muda de acordo com cada versão do SAS que você usar. Fique atento a isso.

[03:00] Vamos começar a trabalhar com nossa base de cadastros dos clientes. Vamos começar fazendo o PROC CONTENTS para relembrar o conteúdo da nossa base de clientes. Clientes_v2 é a base que criamos na parte um do curso, é a base de cadastro do cliente, mas a variável que contém o Estado em que o cliente mora, baseado na variável que criamos a partir do CEP.

[03:41] Executando o código, temos a variável de CEP, e-mail, estado civil, nascimento, nome, sexo, telefone. O Estado é a variável que nós criamos e temos a variável de nascimento, que é a data de nascimento do cliente. Observe que essa variável é depoimento tipo caractere. Ou seja, a data de nascimento nessa base está como um texto. Calcular a idade nada mais é do que fazer uma conta. Sua idade é a data de hoje menos a data de nascimento. Esse intervalo de tempo é sua idade.

[04:22] Precisamos fazer uma conta com essa idade. Vamos transformar essa variável de texto em numérica. Para isso, vimos na primeira parte que podemos usar um DATA. Vou nomear. Depois vou declarar a base de cadastro de cliente. Para fazer a transformação de texto para caractere, vou criar uma variável chamada data_nascimento. Ela vai receber o comando input, com os parâmetros de nascimento, a variável de texto que quero transformar em caractere, e depois o formato de numérico, por exemplo, best. Fecho o DATA com um comando de run. Seleciono tudo e executo.

[06:05] Não funcionou. Vamos olhar a log. Significa que o SAS está tentando fazer uma conversão, não conseguiu e desistiu. O problema é que estamos usando um formato que não é adequado. Queremos transformar um texto em número, mas não é um texto qualquer, e nós não queremos também um número qualquer. Apesar do SAS ter dois tipos de variáveis, existem particularidades. As variáveis numéricas podem ser configuradas de certa forma que seja um número sim, mas um número que o SAS consegue entender como se fosse uma data.

[07:04] Vamos ver nossa base de cadastro do cliente, usando o PROCPRINT. Essa variável de nascimento tem um ano, com um separador, depois o mês, outro separador, e o dia do nascimento. Ela é um texto que está com ano, mês e dia com separadores. Fica difícil o SAS simplesmente converter para um número se eu não falar para ele como ele tem que converter. Só falando que é um texto não adianta.

[08:14] Precisamos nesse parâmetro passar um formato numérico de data. Assim o SAS vai entender que o texto contém uma data e ele tem que transformar para um número, que tenha uma data.

[08:34] Antes de falar dos formatos numéricos de data, vamos aprender um pouco mais do inglês. Como escrevo ano em inglês? Year. Como escrevo mês? Month. Como escrevo dia? Day. Isso não é tão importante. O importante é o que deixei em maiúsculo, a letra que começa cada termo. Para ano é Y, mês é M, e dia D. Os formatos de data do SAS geralmente usam essa estrutura. Nós usamos o formato de data dia mês e ano. Transformando, sendo que o SAS utiliza duas vezes a inicial, para dia escrevo DD, para mês escrevo MM, o ano fica YY. DDMMMYY já é um formato de data do SAS que ele sabe que é dia, mês e ano.

[10:13] Nossa variável de nascimento era ano, mês e dia. Ficaria YYMMDD. Então, temos um formato, lembrando que ele termina com ponto. Tentando executar novamente, deu um aviso, porque o formato não é suficiente. A variável de data é ano, mês e dia, mas o padrão do SAS é colocar dois dígitos para cada termo. Nosso formato tem quatro dígitos no ano. Como falar para o SAS que nossa data tem quatro dígitos no ano?

[11:30] Vamos fazer uma conta rápida. Temos dois dígitos para o ano, dois para o mês e dois para o dia. Ele costuma colocar separadores entre eles. Quantos dígitos estamos usando ao todo? Dois para cada um, mais dois separadores, oito. Ou seja, o formato padrão do SAS usa oito dígitos ao todo. O nosso, como ano tem quatro dígitos, ficaria YYYY-MM-DD. Dez dígitos ao todo. No formato do SAS, escrevemos sempre só duas vezes cada letra.

[12:26] Nós dizemos que tem dez dígitos nesse formato colocando um dez no final. Então fica YYMMDD10. Esse é um formato de SAS que começa com ano, vem para o mês, dia, o ano com quatro dígitos e mais dois separadores.

[12:52] Agora não dá mais problema. Usando o PROCPRINT para ver a base, temos a variável de data de nascimento, que é numérica. Mas o preenchimento é meio maluco. Ele transformou em números estranhos. O que significa isso?