

COMO USAR A BASE DE DADOS

Descontada uma única dificuldade, o uso da base de dados é trivial. Mas essa dificuldade é tão importante, que justifica a produção de um texto para seu enfrentamento. Trata-se do problema de ordenamento e comparabilidade das tabelas de informações.

Ao longo do tempo, não foram apenas os números de municípios que variaram, mas sua grafia e o padrão de indexação dos mesmos. Herval já foi grafado sem **H**, e Santana do Livramento, com apóstrofe separando o Sant do Ana. Quando importamos tabelas com municípios grafados à moda antiga, a ordem dos mesmos altera-se completamente, e a comparação de tabelas torna-se inviável sem o reordenamento dos municípios.

Ainda mais importante do que isso é o fato de que existem distintos padrões de ordenamento alfabético. O padrão oficial, gramaticalmente correto, ignora o espaço em branco entre as palavras (de forma que Arroio do Tigre vem depois de Arroio dos Ratos, pois o **s** vem antes do **t**). Mas o Excel (e todo o sistema computadorizado, a princípio) toma o espaço em branco como um caractere que antecede qualquer letra e inverte a ordem alfabética oficial e consagrada.

Por outro lado, o IBGE define códigos identificadores para os municípios. E, atualmente, a ordem numérica desses códigos segue a ordem alfabética gramaticalmente correta dos municípios. **O resultado é que, se definimos a ordem dos municípios de uma Tabela qualquer pelo Código do IBGE ou pelo nome dos mesmos, impomos ordenamentos distintos.** E este é um problema grave, pois se o pesquisador faz testes de correlação e regressão com ordens inconsistentes — mesmo que essa inconsistência seja imposta por um único município desalinhado —, todas as conclusões serão inconsistentes.

E, como se não bastassem os problemas já apontados, há ainda outros três: (a) os microdados do **Censo Demográfico de 1970** geram resultados, a partir dos códigos do IBGE vigentes no período do referido recenseamento, que não correspondem aos códigos atuais e seguem um ordenamento muito

particular¹; (b) os microdados do **Censo Demográfico de 2000** geram resultados ordenados pelo código do IBGE de 2000, que corresponde à ordem alfabética gramaticalmente correta²; (c) o *software* do Sistema de Conversão produzido por nós gera planilhas em que os municípios estão ordenados alfabeticamente pelo Excel.³

Para evitar problemas de comparação, uniformizamos o padrão de indexação alfabética dos municípios: ele é feito de acordo com a ordem alfabética dos dois principais *softwares* de processamento de dados do sistema: o Excel e o Sistema de Conversão Municipal⁴.

Não obstante o nosso esforço de uniformização interna, a expectativa é de que o usuário deste sistema venha a enriquecê-lo com outras e diversificadas fontes de dados. Ao incorporá-los é preciso ter cuidado para garantir a preservação da homogeneidade do indexador alfabético. O que pode ser conseguido através de um procedimento muito simples: ao recortar (partes de) tabelas e colar em outras, carregue sempre os nomes (que explicitam a ordem) dos municípios da tabela recortada, e faça um Teste de Correspondência com a ordem alfabética da tabela que recebe a informação⁵.

¹ Na época, o código do município levava em conta a região em que o mesmo se encontrava (de forma que a proximidade geográfica era levada em consideração). Atualmente, apenas a ordem alfabética da denominação do município é considerada na definição do referido código.

² Arroio do Tigre vem depois de Arroio dos Ratos.

³ Arroio dos Ratos vem depois de Arroio do Tigre.

⁴ Como diversas tabulações foram resgatadas de outras fontes (IBGE, IPEA, Tribunal de Contas, etc.) em sua indexação original e só depois reindexadas, não é impossível (ainda que seja muito pouco provável) que alguma tabela tenha escapado ao nosso controle interno, e continue indexada pela ordem gramatical rigorosa ou por outro critério qualquer. Em caso de resultados teoricamente surpreendentes e/ou destoantes de uma série, não deve ser descartada a hipótese das variáveis estarem indexadas por critérios distintos.

⁵ O Teste de Correspondência no Excel: escolha uma célula vazia e digite =, clique no nome do primeiro município da tabulação original, digite = novamente e clique na linha correspondente com o nome do município da tabela recém incorporada. Se as células contiverem sinais com os mesmos dígitos (nomes ou números), aparecerá a expressão verdadeiro (ou *true*). Reproduzindo o exercício para a coluna toda, fica fácil perceber qualquer discrepância. Depois, basta alterar a ordem e adequar as discrepâncias, o que, via de regra, se obtém reindexando a tabela importada no padrão alfabético do Excel.