

Reconstrução de famílias no microcomputador: o SYGAP

Marco Aurélio Monteiro Pereira*

Introdução

O SYGAP – *Système de Gestion et d'Analyse de Population* – é um *software* construído para o gerenciamento de bases de dados populacionais. É uma poderosa ferramenta auxiliar na reconstrução de famílias e na análise e consolidação de dados e indicadores demográficos, representando um salto qualitativo substancial no que diz respeito à História Demográfica como tradicionalmente concebida, isto é, centrada na reconstrução manual de famílias.

Este salto se configura em dois aspectos específicos e definidos: a economia de tempo e a maior profundidade nas análises demográficas, quer pela possibilidade praticamente ilimitada de recortes, quer pela extensão das análises ao campo da Genética.

Notas preliminares (1)

O SYGAP é fruto da produção conjunta do Programa de Pesquisa em Demografia Histórica do Departamento de Demografia da Universidade de Mon-

treal e do Grupo Rendu-Osler, que aglutina a Maison Rhône-Alpes des Sciences de l'Homme, o Centre Pierre Léon e o Institut Européen des Génomutations.

O *software* é um sistema modular de análise e gestão de registros de população implantado em microcomputadores padrão IBM-PC, dotado de disco rígido. Seu módulo central permite a elaboração e a gestão de um modelo numérico de uma dada população, módulo este referencial aos módulos-satélite de Genética e Demografia.

É um programa de arquitetura aberta, que permite uma junção natural e fácil do módulo central aos módulos-satélite. Sendo compatível com o padrão dBase (Versão III Plus), o SYGAP abre a seus usuários o uso de uma vasta gama de atributos e rotinas de tratamento de dados disponíveis e programáveis neste ambiente ou em Clipper.

Estrutura

O SYGAP pretende a definição, realização e exploração de uma base de dados de população, trabalhando suas duas variáveis fundamentais – os indivíduos e as suas uniões – em dois arquivos básicos (INDIV.DBF para indivíduos e UNION.DBF para uniões).

Este modelo se resume às variáveis fundamentais (datas de nascimento, de óbito e de casamento, dentre outras), mas, como é aberto, deixa liberdade aos usuários de ampliar a composição dos arquivos de indivíduos e casamentos,

* Professor da Universidade Estadual de Ponta Grossa, PR, e doutorando em História Demográfica na Universidade Federal do Paraná (UFPR).

(1) As observações a seguir são extraídas, em sua grande maioria, do *Manual de uso do SYGAP* e da minha experiência pessoal no uso do programa e na consultoria informal prestada na implementação do SYGAP nos trabalhos de outros doutorandos e professores da UFPR.

com a ajuda do dBase, com as variáveis adequadas às suas necessidades (por exemplo, o caso de compadrio).

A gestão das informações básicas (inclusão, exclusão e ordenamento dos dados) é assegurada pelo módulo central (SGFP.EXE), cujas possibilidades de acesso foram intencionalmente limitadas à consulta e à manutenção da integridade dos dados. Os módulos de análise, que atuam de forma integrada e independente do módulo central, são em número de cinco e responsáveis por análises de fecundidade, nupcialidade, mortalidade, genética e, finalmente, um módulo de análises diversas (recenseamentos, movimentos de população e estatísticas gerais).

O módulo central (SGFP.EXE)

O módulo central do SYGAP, contido no subprograma SGFP.EXE, realiza a manutenção, integração e ordenamento da base de dados populacionais através de dois arquivos básicos: o de indivíduos e o de uniões.

A base de dados de população

Uma base de dados de população pode ser definida como um sistema de informações que contém os dados pertinentes à história de um grupo de indivíduos, concebido para estudos demográficos ou genéticos. Corresponde a um conjunto de arquivos, tendo seus registros ligados logicamente por regras de coerência e limitadores de integridade dos dados, contidos no próprio sistema.

A base de dados compõe-se dos seguintes arquivos permanentes: o registro de indivíduos e seus índices; o registro das uniões e seus índices; o dicionário

das variáveis da base de dados; o arquivo de parâmetros de funcionamento do sistema.

A base de dados também pode possuir arquivos temporários auxiliares, que são apagados após o uso: o arquivo das biografias individuais; o arquivo das descendências individuais; o arquivo das ascendências individuais.

O registro de indivíduos

O registro de indivíduos contém dois tipos de informações: os elos genealógicos entre os indivíduos e as variáveis que os caracterizam. Cada indivíduo é identificado por um código numérico único, que permite recuperá-lo sem qualquer ambigüidade. É um sistema de identificação com uma lógica diferente da alfabética, tradicionalmente usada na reconstrução manual de famílias, que, embora de assimilação um pouco mais complicada, possui um grau maior de eficiência. O código numérico é atribuído pelo usuário e sua unicidade é controlada pelo sistema. A quantidade das variáveis e suas características são deixadas a critério do usuário, exceto as variáveis de base impostas por razões de controle de integridade.

O arquivo de registro de indivíduos contém, para cada indivíduo, a informação necessária para elaborar e manter a estrutura genealógica do arquivo de população, para assegurar a integridade dos registros biológicos do mesmo e, enfim, para veicular a informação necessária à pesquisa. É composto em formato dBase (.DBF), podendo conter, além das variáveis básicas fornecidas pelo programa, as variáveis que o usuário tiver necessidade. Não existem restrições quanto ao número de variáveis residentes fora do sistema.

A estrutura do arquivo (variáveis básicas)

Nome da Variável	Tipo da Variável	Descrição da Variável
IDI	N*6	Código de identificação do indivíduo
SEXE	C*1	Sexo (M, F ou X (2))
NOMINATIF	C*30	Nome completo do indivíduo
DATEN	D	Data de nascimento do indivíduo
XDATEN	C*2	<i>Drapeau</i> da data de nascimento
LIEUN	C*3	Código do local de nascimento do indivíduo
DATED	D	Data do óbito do indivíduo
XDATED	C*2	<i>Drapeau</i> da data de óbito
LIEUD	C*3	Código do local do óbito do indivíduo
PERE	N*6	Código de identificação do pai do indivíduo
MERE	N*6	Código de identificação da mãe do indivíduo
CNGCOEF	N*9,7	Coefficiente de consaguinidade

Os limitadores de integridade são verificados com os parâmetros definidos pelo usuário através do sub-item *Parâmetros* do item *Procedures Utilitaires* do menu principal do módulo de controle. São eles:

- Unicidade dos pais: um indivíduo não pode ter mais do que um pai ou mais do que uma mãe, ou aparecer como ascendente de seus pais. Um

dos genitores pode, entretanto, ser desconhecido.

- Definição de sexo: o sexo de um indivíduo pode ser somente masculino (M), feminino (F) ou desconhecido (X).
- Paternidade e maternidade: um indivíduo inscrito como pai deve ser obrigatoriamente do sexo masculino e uma mãe deve ser obrigatoriamente do sexo feminino.
- Elo genealógico: um indivíduo não pode encontrar-se na sua própria ascendência ou descendência.
- Idade no parto: uma mulher não pode dar à luz com mais de 49 anos completos (salvo especificação do usuário).
- Intervalo de nascimento: um indivíduo não pode nascer a menos de 14 anos depois do nascimento de seu pai e a menos de 12 anos do nascimento de sua mãe, salvo se o usuário precisa outros valores. Ele não pode nascer mais de nove meses após o falecimento de seu pai, nem após o falecimento de sua mãe.

O registro de uniões

O registro de uniões completa o arquivo precedente, dando conta dos casamentos entre os indivíduos. Cada união entre os indivíduos da população é identificada por um código numérico único que permite recuperá-la sem ambigüidade. Entretanto, a base de dados de população pode ser elaborada na ausência de um arquivo de uniões. Da mesma forma que o arquivo de indivíduos, o arquivo de uniões contém dois tipos de informações: laços matrimoniais e variáveis descritivas da união. O usuário pode igualmente

(2) Usa-se "X" para designar sexo indeterminado.

anexar novas variáveis ao arquivo, servindo-se do dBase.

A existência deste registro permite o acesso por unidade familiar. O usuário pode, então, tomar conhecimento dos sobrenomes dos cônjuges, suas características individuais, e das diferentes variáveis concernentes à união.

A estrutura do arquivo (variáveis básicas)

Nome da Variável	Tipo da Variável	Descrição da Variável
UNION	N°6	Número da união
HOMME	N°6	Código individual do marido
FEMME	N°6	Código individual da mulher
DATEM	D	Data do casamento
XDATEM	C*2	<i>Drapeau</i> da data do casamento
LIEUM	C*3	Código do local do casamento
DATEFU	D	Data do fim da união
XDATEFU	C*2	<i>Drapeau</i> da data do fim da união
ENFILL	N°2	Identificação dos casais em que pelo menos uma criança nasceu antes do casamento
ENFPRENUP	N°2	Identificação dos casais cujo primeiro filho foi concebido antes do casamento
CODEM	C*1	Código de categoria do casamento
CODER	C*1	Código do fim da união
SEXES	C*1	Sexo do cônjuge sobrevivente

Os limitadores de integridade são verificados com os parâmetros definidos pelo usuário através do sub-item *Parâmetros* do item *Procedures Utilitaires* do

menu principal do módulo de controle. São eles:

- Monogamia: um indivíduo não pode estar casado com várias pessoas ao mesmo tempo, salvo especificação em contrário pelo usuário.
- Existência dos cônjuges: o sistema se assegura que os números de identificação dos cônjuges declarados no registro de união existem no registro de indivíduos. É impossível o registro de um casal em que um ou os dois cônjuges não estejam contidos no registro de indivíduos.
- Possibilidade de divórcio: o sistema considera que o fim de uma união ocorre com o óbito de um dos cônjuges, salvo se o usuário especificar a possibilidade de a desunião ocorrer por divórcio quando da definição dos parâmetros.

O arquivo dicionário

O arquivo dicionário da base de dados contém a descrição das variáveis do sistema, vital para o correto funcionamento do programa. Estas variáveis podem ser provenientes da definição dos arquivos permanentes ou dos temporários. É o arquivo dicionário que permite a abertura do registro de população sem nenhum procedimento de instalação.

A estrutura do arquivo

Nome da Variável	Tipo da Variável	Descrição da Variável
FIELD_NAME	C*10	Nome da variável
FIELD_TYPE	C*1	Tipo da variável
FIELD_LEN	N°3	Comprimento da variável
FIELD_DEC	N°3	Número de casas decimais
FIELD_TAB1	N°3	Nome do arquivo da variável

O arquivo dos parâmetros de funcionamento

O arquivo de parâmetros de funcionamento e integridade guarda, de forma permanente, as variáveis que asseguram a conservação e a coerência do programa. O valor destes parâmetros pode ser definido e ajustado pelo usuário.

Descrição dos parâmetros

BANTEXT	Identificação do sistema (texto afixado ao alto de cada uma das telas)
DELTAP	Idade mínima de paternidade (15 anos)
DELTAM	Idade mínima de maternidade (12 anos)
DELTAMX	Idade máxima de maternidade (49 anos)
PLY	Indicador de poligamia
DIV	Indicador de divórcio

Os módulos de análise

O SYGAP não é somente um instrumento de consulta e gestão de populações. Ele se constitui, igualmente, num instrumento de análise simples e de alto desempenho em dois domínios intimamente ligados aos estudos de população, compreendidos nos seguintes módulos: os módulos de análise demográfica e o módulo de análise genética.

Estes instrumentos de análise são construídos segundo os mesmos princípios do módulo de gerenciamento central do SYGAP: todas as funções e comandos são claros e simples. A escolha dos indivíduos ou das uniões pelo usuário é feita com o auxílio do dBase.

A razão de ser destes módulos reside na preocupação de dar ao usuário instrumentos sofisticados de análise, rápidos e eficazes, que não exigem conhe-

cimentos aprofundados em Demografia e em Genética.

Módulos de análise demográfica

Estes módulos apresentam os três eixos básicos de análise demográfica (fecundidade, nupcialidade e mortalidade) e um módulo com enfoques demográficos globais diversos (recenseamentos, movimentos de população e estatísticas populacionais). Sua abrangência de enfoque de análise é vista a seguir, por módulo:

Fecundidade (FECOND.EXE)

- Cálculo das taxas de fecundidade legítima das mulheres;
- Cálculo das taxas de fecundidade das mulheres anteriormente fecundas;
- Medida do intervalo entre o casamento e o primeiro nascimento e dos intervalos entre os nascimentos;
- Medida de incidência de gêmeos, da ilegitimidade e das concepções pré-nupciais;
- Distribuição absoluta e percentual das uniões segundo as categorias (MF, MO, EF, EO);
- Cálculo da idade média da mãe quando do nascimento do último filho;
- Cálculo das probabilidades de aumento da fecundidade segundo o tamanho das famílias;
- Distribuição das famílias segundo o tamanho;
- Medida da razão de masculinidade no nascimento;
- Distribuição mensal das concepções e dos nascimentos.

Nupcialidade (NUPTIA.EXE)

- Cálculo da tábua de nupcialidade de adultos solteiros;
- Cálculo das idades médias dos cônjuges ao primeiro casamento;
- Estado matrimonial dos cônjuges ao casar;
- Idades conjugadas dos cônjuges ao casar, no primeiro e no último nascimento;
- Medida da frequência do recasamento em função da idade ao enviuvar e medida da duração da viuvez;
- Medida da duração do casamento e data do fim da união, entre o primeiro nascimento e o último nascimento e entre o casamento e o último nascimento;
- Sexo do cônjuge sobrevivente em função da diferença de idade entre os cônjuges;
- Medida do movimento mensal dos casamentos.

Mortalidade (MORTAL.EXE)

- Cálculo das tábuas mínima e máxima de mortalidade de adultos (dez anos e mais);
- Cálculo das tábuas mínima e máxima de mortalidade (0 a 100 anos e mais);
- Estado matrimonial ao falecer;
- Medidas do movimento mensal dos óbitos em função do grupo de idade.

Outras funções demográficas (DEMAUTRE.EXE)

- Recenseamentos: medida dos efetivos mínimo e máximo em função do grupo de idade quinquenal e do sexo, para uma data precisa;
- Movimentos de população: compilação dos nascimentos, óbitos, ca-

samentos, dos emigrantes e imigrantes, por período;

- Estatísticas gerais: número de indivíduos e de famílias e distribuição das variáveis por categoria para o conjunto da população.

Módulo de análise genética (ASCEND.EXE)

Este módulo centra sua análise nas genealogias ascendentes. A análise pode se limitar a um só indivíduo ou ao conjunto de uma população. O módulo se fundamenta em quatro funções de cálculo específicas: compilação das genealogias ascendentes; descrição da genealogia ascendente de um indivíduo; cálculo de coeficientes de consangüinidade; cálculo dos coeficientes de parentesco.

A agregação dos módulos analíticos

O SYGAP, contrariamente a muitos outros programas comerciais, não possui uma estrutura fechada. Seu módulo central de gestão e os módulos de análise permanecem inacessíveis em sua forma compilada, mas nada impede que o usuário enxerte no sistema seus próprios programas em linguagem dBase ou compilados em Clipper para efetuar diversas operações na base de dados. O usuário deve somente respeitar o nome dos diferentes índices para poder acessá-los em seus próprios programas.

Os módulos de gerenciamento e análise são construídos de forma independente e executáveis de forma independente e autônoma, agindo sobre a mesma base de dados.

Modo de funcionamento

O sistema é concebido para funcionar de maneira interativa e amigável com o usuário. Assim, dá resposta imediata a

toda solicitação, o que se traduz, por exemplo, pela retenção instantânea de uma informação, a responsabilidade pelo controle de validade de uma correção, a emissão de mensagem de erro quando for o caso etc.

Um sistema que funciona de modo interativo, por meio de *menus*, poupa o usuário do cansativo aprendizado de inúmeros comandos e o auxilia na própria utilização do sistema, possibilitando que pessoas com pouca experiência, quer em Demografia, quer em Informática, possam utilizar o SYGAP facilmente.

Um *menu* é, de maneira geral, uma tela com várias opções de escolha entre procedimentos diversos, mostradas no monitor. Frequentemente conduz a um *submenu* e assim por diante, até se chegar ao procedimento desejado.

Considerações complementares

O SYGAP foi utilizado experimentalmente no tratamento dos dados demográficos da população compreendida na Paróquia de Sant'Anna de Abranches, em Curitiba, de imigração principalmente polonesa, para a elaboração da tese de doutorado do autor, em fase de conclusão. Seu uso possibilitou a paulatina definição de seus limites de aplicabilidade e a detecção de alguns problemas inerentes à concepção lógica do programa, voltado para populações europeias.

O Manual de Referência é sucinto e claro, porém não define alguns procedimentos básicos para a constituição da população como, por exemplo, a necessidade obrigatória das datas de nascimento e óbito, mesmo que fictícias, para cada indivíduo, e de casamento e fechamento de cada união, indispensáveis para o bom andamento das análises demográficas. Isto se dá devido à arquitetura do programa ser elaborada para o tratamento de bases de dados completas, comuns em levantamentos de populações europeias, mas bastante difíceis de obtenção

em populações brasileiras, mesmo de paróquias imigrantes, principalmente por causa do alto grau de intensidade dos fluxos migratórios característicos de sociedades coloniais de exploração.

O método de codificação dos indivíduos, numérico, é sem dúvida nenhuma a grande vantagem do programa. Mediante uma revolução na lógica da codificação dos indivíduos e uniões, normalmente alfabética na reconstrução de famílias tradicional, o código numérico permite solucionar parcialmente um dos grandes problemas para a reconstrução de famílias em populações brasileiras, marcadamente nos séculos XVII a XIX: a variação onomástica. Com a codificação numérica, as relações familiares se dão através da ligação dos códigos numéricos de seus componentes e não mais de uma relação patronímica fundada no sobrenome. Assim, se os registros contiverem a relação familiar entre pessoas com sobrenomes distintos, sua ligação através dos códigos numéricos não gerará dificuldades para a reconstrução da família do indivíduo. Evidentemente, o SYGAP não resolve totalmente o problema das variações onomásticas de um indivíduo no transcorrer de seu ciclo vital, embora o cruzamento de códigos possa, potencialmente, apresentar soluções parciais para isto.

O fato de o programa-fonte estar em francês ocasiona alguns contratempos para o usuário brasileiro comum, que tem seu equipamento configurado para a execução de rotinas fundadas no idioma inglês ou português. Por exemplo, o caractere "è", muito recorrente no idioma francês, ocasiona, nas impressoras configuradas para inglês ou português, um retorno do cabeçote de impressão, provocando uma quebra de linha, o que prejudica sobremaneira a emissão de tabelas e relatórios pelo programa. Isto, porém, é facilmente contornável pela geração de relatórios em dBase.

A função de impressão do programa demanda aperfeiçoamento. Alguns módulos muito importantes não possuem comandos de impressão, forçando o usuário a imprimir a tela por meio da tecla "print screen", o que gera a decodificação das molduras em caracteres em formato texto e o conseqüente desalinhamento do texto.

Porém, considerando-se no cômputo geral, o SYGAP representa um investi-

mento de tempo e equipamento extremamente rentável, agilizando sobremaneira o trabalho do historiador de populações, marcadamente no que diz respeito à relação entre o tempo gasto e o retorno obtido, substancialmente mais positiva do que na reconstrução manual de famílias, além da agregação de módulos que ampliam e aprofundam numa nova dimensão o trabalho em Demografia Histórica.

(Recebido para publicação em setembro de 1995)