

---

## Análise da Atenção Básica em João Pessoa Através da Integração de Bases de Dados de Saúde

**Alfredo A. A. Malheiros Filho , Lucas Vieira de Souza, Patrícia S. Nascimento, Renata G. S. Gomes, Liliane S. Machado e Ronei M. Moraes**

Laboratório de Estatística Aplicada ao Processamento de Imagens e Geoprocessamento  
Depto de Estatística, Universidade Federal da Paraíba,  
58051-900, João Pessoa, PB

malheirosf@yahoo.com.br, lucasvsdsi@yahoo.com.br, patriciapb1408@yahoo.com.br, renatagrigoio@yahoo.com.br,  
liliane@di.ufpb.br, ronei@de.ufpb.br

**Resumo:** *A saúde é uma das áreas prioritárias do desenvolvimento humano. Atenção Básica é um conjunto de medidas no âmbito da saúde pública que visam melhorias na saúde. Hoje no Brasil não existe um padrão único para armazenamento dos dados da saúde pública fornecidos pelas secretarias de saúde. Este fato dificulta a análise integrada de dados e os dados acabam sendo usados apenas para simples consultas. Neste caso, existe uma preocupação em integrar as diversas bases de dados visando utilizar os dados em análises que incluam todos eles ao mesmo tempo. Este artigo tem o intuito de mostrar metodologias de avaliação e análise da atenção básica em João Pessoa, a partir da integração das bases de dados disponíveis na Secretaria de Saúde do Município. Neste trabalho foi modelado um sistema para a integração de diversos sistemas de informações para permitir a análise dos dados mutuamente com a capacidade de fornecer análises estatísticas dos mesmos. Definiram-se, também, análises estatísticas para a previsão, controle e mapeamento de eventos específicos de saúde para tomadas de decisão (geoprocessamento). O sistema desenvolvido apresenta uma estrutura dividida em três camadas: acesso aos dados, negócio, e apresentação.*

**Palavras-chave:** saúde pública, banco de dados, integração, estatística.

### Introdução

A saúde é uma das áreas prioritárias do desenvolvimento humano. Junto com a educação básica, saneamento básico e planejamento familiar, os cuidados primários de saúde podem ser considerados indicadores de desenvolvimento para um país ou região [12].

Atenção Básica é um conjunto de medidas no âmbito da saúde pública que visam melhorias na saúde [13]. Ela é caracterizada por ações democráticas e participativas, considerando a heterogeneidade populacional do local. Através dela é possível analisar um indivíduo em toda sua singularidade e inserção social e buscar a promoção da saúde através da

prevenção, tratamento, redução de danos e sofrimentos que prejudiquem uma vida saudável.

A partir de 1997 com a vigência da NOB/96 (Norma Operacional Básica do Sistema Único de Saúde-SUS), o plano estratégico de saúde pública no Brasil teve mudanças consideráveis. O Ministério da Saúde mudou o seu foco de investimentos onde ao invés de priorizar o pagamento direto ao prestador do serviço de saúde passou a enfatizar a transferência de recursos aos municípios. Concomitantemente, escolheu a Atenção Básica como prioridade de ação [8].

Os resultados obtidos pelas ações da gestão são refletidos na saúde da população, fato que torna as decisões adotadas de extrema responsabilidade e cuidado. Há também a necessidade de uma maior velocidade na análise dos resultados para reparos na estratégia adotada. Com a necessidade de um retorno mensurável das ações adotadas pela atenção básica, mostra-se convincente que todos os dados relevantes sejam armazenados em locais de fácil e rápido acesso e manipulação. Uma análise sobre o estado de saúde de uma população requer rever índices e dados anteriores e adicionar dados atuais para prover meios com os quais os gestores possam alterar ou dar continuidade a um trabalho em andamento. Informações sobre saúde pública sempre são objetos de estudo em pesquisas para melhorias nas decisões tomadas e devem ser armazenadas de forma responsável e ininterrupta, com uma preocupação constante em sua integridade e veracidade. Tais fatos estão de acordo com uma das diretrizes do governo federal que é executar a gestão pública por resultados mensuráveis [13].

Existe hoje no Brasil um grande volume de dados sobre a saúde pública que são coletados pelas secretarias de saúde municipais e estaduais. O armazenamento destas informações não segue um padrão único e faz com que estes sejam utilizados apenas para simples consultas [15]. Entretanto, observa-se a necessidade dos gestores de saúde pública em identificar pontos falhos em suas estratégias, o que encontra barreiras nas diferentes formas de implementação e uso das informações. As

incompatibilidades e falta de padronização prejudicam a análise e reduz a eficiência dos resultados, muitas vezes inconsistentes ou irrealistas. Tal fato restringe o uso dos dados para análises mais robustas. Apesar da falta de um padrão de armazenamento, existe hoje uma preocupação em integrar as diversas bases de dados de vários sistemas [6].

Este artigo tem o objetivo de mostrar metodologias de avaliação e análise da Atenção Básica à saúde na cidade de João Pessoa a partir da integração das bases de dados disponíveis na Secretaria de Saúde do Município. A integração de bases de dados pode ser definida como a atividade de conciliar esquemas de bases de dados existentes ou propostos em um esquema global e unificado [2].

## Concepção

O sistema em desenvolvimento visa integrar todos os outros sistemas de informações, ser capaz de analisar os dados mutuamente e ter a capacidade de fornecer análises estatísticas. Neste sistema deve existir a possibilidade de atualização dos dados com um fluxo constante, proveniente dos atuais sistemas de informações já implantados nas secretarias de saúde. Neste caso, utiliza dados dos sistemas SIM, SINASC, SINAN, SIAB existentes nas secretarias municipais. A Figura 1 mostra um diagrama da concepção deste sistema, onde podem ser vistas as partes representativas dos bancos de dados originais e o sistema de forma geral. Pode-se observar que os diversos bancos de dados são integrados e seus dados utilizados para a realização de análises estatísticas e geração de relatórios através de uma única interface. A interface também é responsável pela comunicação com o *software R*.

O uso de *software* livre vem abrindo portas para um novo futuro em tecnologia da informação no nosso país. A cada dia temos mais programas e aplicações sendo desenvolvidos com base na filosofia de distribuição de conhecimento, utilização gratuita e livre de pagamentos de qualquer natureza. Isso mostra que a compra de programas que de um modo geral são bastante custosos pode ser evitado através do uso de similares livres. Pensando nisso o sistema em questão está sendo desenvolvido em *software* livre e com código aberto, ou seja, o código fonte pode ser posteriormente alterado por qualquer pessoa, facilitando sua manutenção e livre de custos com direitos de propriedade. O sistema está sendo desenvolvido em um ambiente também livre, o que facilita sua adoção pelas secretarias de saúde de todo Brasil.

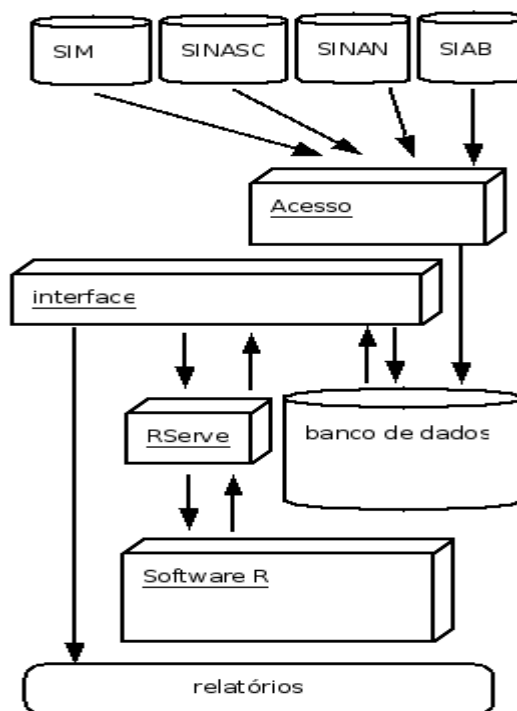


Figura 1: diagrama de concepção do sistema

O sistema computacional desenvolvido para atender os requisitos do problema em questão tem a necessidade de conversão e integração das bases de dados de saúde, como as do SIAB, SIM, SINASC e SINAN. O Sistema de Informação de Agravos de Notificação (SINAN) foi implantado no Brasil no início da década de 90. Utiliza-se para suscitar a notificação e investigação de casos de doenças e agravos que constam na lista nacional de doenças de notificação compulsória. Sua utilização efetiva permite a realização do diagnóstico dinâmico da ocorrência de um evento na população, podendo fornecer subsídios para explicações causais dos agravos de notificação compulsória. O SINAN também indica riscos aos quais as pessoas estão sujeitas, contribuindo assim, para a identificação da realidade epidemiológica de determinada área geográfica [16]. O Sistema de Informações sobre Nascidos Vivos (SINASC) foi implantado em 1990 com o objetivo de reunir informações epidemiológicas referentes aos nascimentos informados em todo território nacional. Por intermédio desses registros é possível subsidiar as intervenções relacionadas à saúde da mulher e da criança para todos os níveis do Sistema Único de Saúde (SUS). O acompanhamento da evolução das séries históricas do SINASC permite a identificação de prioridades de intervenção, o que contribui para efetiva melhoria do sistema [17]. O Sistema de Informações sobre Mortalidade (SIM) foi criado pelo Ministério da Saúde em 1975 para a obtenção regular de dados sobre mortalidade no país. A partir da criação do SIM foi

possível a captação de dados sobre mortalidade, de forma abrangente e confiável, para subsidiar as diversas esferas de gestão na saúde pública. O SIM proporciona a produção de estatísticas de mortalidade e a construção dos principais indicadores de saúde. A análise dessas informações permite estudos não apenas do ponto de vista estatístico e epidemiológico, mas também sócio-demográfico [18]. O Sistema de Informação da Atenção Básica (SIAB) foi implantado em 1998. Por meio do SIAB obtêm-se informações sobre cadastros de famílias, condições de moradia e saneamento, situação de saúde, produção e composição das equipes de saúde [19]. As informações são coletadas em âmbito domiciliar e em unidades básicas nas áreas cobertas pelo Programa Saúde da Família e pelo Programa Agentes Comunitários de Saúde [20].

A partir das informações contidas nestes bancos de dados, o sistema será capaz de criar relatórios utilizando os dados integrados destes bancos ou os resultados de análises estatísticas aplicadas a estes dados.

## Desenvolvimento

Ao analisar o sistema proposto, foi verificada a necessidade de previsão a futuras expansões do sistema. Assim, ele deve ser robusto o suficiente para permitir que novos sistemas sejam adicionados a integração atual. Também deve ser expansível o suficiente para permitir que novos tipos de análises sejam adicionados quando necessário. Quando se pensa em um sistema expansível, vê-se a necessidade de modificar o seu código-fonte. Isso só é possível se este estiver disponível para modificação.

Uma vantagem em se trabalhar com programas livres é a potencialidade de modificação do seu código-fonte, permitindo adequá-lo e até expandi-lo quando for necessário. Outra vantagem dos programas livres é o custo zero de aquisição. Com este requisito estabelecido, desenvolveu-se uma tecnologia baseada em *software* livre, com bibliotecas capazes de criar mecanismos de leitura dos diversos bancos provenientes dos sistemas nacionais de monitoramento à saúde. Para a construção do sistema, a partir das bases de dados da Secretaria Municipal de Saúde de João Pessoa, escolheu-se a tecnologia JAVA por ser uma linguagem que possui um ótimo suporte a bancos de dados e ser bem portátil a vários sistemas operacionais.

A *interface* de comunicação do sistema comunica-se com ferramentas estatísticas (*software R*) para a análise e geoprocessamento dos dados. O *software R* [24] é um conjunto integrado de sistemas computacionais para manipulação de dados, cálculos, visualização gráfica, programação, entre outras funcionalidades, oferecendo o desenvolvimento de diversas metodologias em muitas áreas do conhecimento humano, uma delas a estatística. Um dos pontos fortes do *R* é a de ser um sistema computacional livre distribuído sob a Licença Pública Geral (GPL)

podendo ser livremente copiado e distribuído entre os usuários e instalado em qualquer computador livremente, desta forma, ele é gratuito. O *software R* nas versões *Linux*, *Windows*, *MacOS X* pode ser obtido no endereço eletrônico: <http://www.r-project.org>.

A interação entre a aplicação Java e o *software R* pode ser realizada através do sistema *Rserve* [11][23] que funciona como um tradutor de solicitações de operações estatísticas entre os sistemas, permitindo a execução de várias requisições simultaneamente e em rede. O *Rserve* pode ser obtido no endereço eletrônico: <http://www.rforge.net/Rserve>. Por sua vez, essas solicitações oriundas da aplicação, necessitam da presença de um cliente capaz de efetuar essa comunicação. Por isso será utilizado o *JRclient*, um *client framework* específico para a tecnologia JAVA que provê a tradução automática de tipos como *int*, *double*, *arrays*, *String* e *Vector* para objetos especiais do *R*, como *Rbool* e *RList*. A Figura 2 ilustra a arquitetura de comunicação do sistema de consulta ao banco de dados e análise dos dados.

Este modelo de arquitetura [14] pode ser utilizado em qualquer outro sistema que queira fazer a comunicação de uma aplicação com o *software R* e com bases de dados. A aplicação JAVA é responsável pelo acesso aos dados, pela interação com os usuários, pela troca de informações com o *software R* e o *RServer* através do framework *JRclient* e por todas as funcionalidades do sistema. Do outro lado observa-se também a necessidade de um sistema de gerenciamento de banco de dados (SGBD) para o armazenamento de todos os dados, além do próprio *R* e *RServer*.

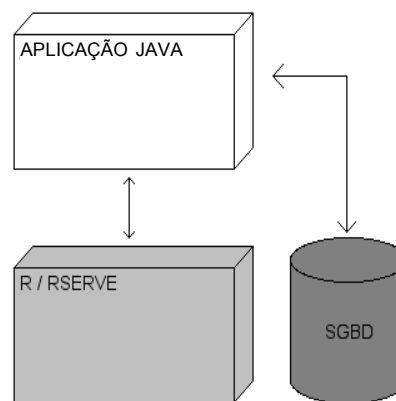


Figura 2: Arquitetura de comunicação do sistema de consulta e análise.

O SGBD escolhido deve ser capaz de suportar grandes volumes de dados sem perda de integridade das informações. Também deve ser robusto o suficiente para aceitar múltiplas conexões e possuir abordagem relacional de dados. Por apresentar tais características, escolheu-se o PostgreSQL[21], para ser a ferramenta responsável pelo armazenamento e manipulação de todos os dados. Outro ponto decisivo

dessa escolha foi o suporte a bancos de dados geográficos, pois como há a possibilidade de crescimento do sistema, o PostgreSQL estará pronto para receber mais uma tarefa.

Definiram-se análises estatísticas para a previsão, controle e mapeamento de eventos específicos de saúde para tomadas de decisão. A metodologia adotada consta de estudos de adequação de modelos nas áreas de geoprocessamento, indicadores de saúde, testes não-paramétricos, coeficientes de correlação, séries temporais, análise de regressão linear e de *Poisson*, análise de sobrevivência e de classificação e controle estatístico de qualidade. As metodologias estatísticas utilizadas para desenvolver as funções do *R* foram as seguintes: testes não paramétricos, coeficientes de correlação, análise de regressão, modelos lineares generalizados, análise de séries temporais, análise de sobrevivência e controle estatístico de qualidade.

Geoprocessamento é uma técnica que analisa os dados espacialmente através de mapas cadastrais. A análise espacial serve para examinar dados para conhecer suas funções e relações [3]. O mapa cadastral é um mapa utilizado na análise geo-estatística, onde cada um de seus elementos são registrados numa tabela em forma de cadastro. Nestas tabelas ocorre a associação de atributos não gráficos, os dados, a atributos gráficos, o mapa político. Na análise geoestatística é possível mensurar as propriedades e as relações do fenômeno em estudo, levando em consideração sua localização espacial na interpretação dos resultados. Ou seja, a idéia central é incorporar o espaço à análise que se deseja saber, sendo esta análise feita sobre geo-campos [3]. Um dos principais conceitos de geoprocessamento é o de espaço geográfico, definido como uma coleção de localizações na superfície da terra onde ocorrem os fenômenos geográficos. O espaço geográfico pode ser modelado segundo duas visões: geo-campo representando a distribuição espacial de uma variável, que possui valores em todos os pontos pertencentes a uma região do espaço geográfico, num dado tempo  $t$ ; e geo-objeto ou objeto geográfico que são entidades distintas e localizáveis que representam uma determinada região do espaço geográfico [3].

Indicadores de saúde têm o objetivo de avaliar, o estado de saúde dos indivíduos, bem como fornecer subsídios aos planejamentos de saúde [5]. Existem vários indicadores de saúde utilizados para medir a mortalidade, natalidade, morbidade e avaliação hospitalar. A OMS (Organização Mundial de Saúde), em seu Informe Técnico 137 (21), sugere a divisão dos indicadores de saúde em três grupos [5]:

- Os que traduzem diretamente a saúde (ou a sua falta) em um grupo populacional, estão separados em 2 tipos: Globais: Razão de mortalidade proporcional; coeficiente geral de mortalidade; esperança de vida ao nascer; Específicos: Coeficiente de mortalidade infantil; coeficiente de mortalidade por doenças transmissíveis.
- Os que se referem às condições do meio e que

têm influência sobre a saúde: abastecimento de água, rede de esgotos.

- Os que medem os recursos materiais e humanos relacionados às atividades de saúde, tais como rede de postos de saúde, número de profissionais de saúde, em relação à população.

Os testes não-paramétricos são utilizados para comparar duas ou mais populações independentes ou relacionadas. Em todos os testes: *Friedman* (duas ou mais populações relacionadas), *Wilcoxon* (duas populações relacionadas), *Kruskal-Wallis* (duas ou mais populações independentes) e *Mann-Whitney* (duas populações independentes), o valor  $p$  é a probabilidade de que os grupos sejam iguais. Esta probabilidade determina a aceitação ou rejeição da hipótese de similaridade entre os grupos em comparação [22].

Os Coeficientes de Correlação são úteis para verificar a existência de correlação entre as variáveis. Para isso, um gráfico de dispersão deve ser feito antes da análise numérica dos dados, construído com conjuntos de pontos formados por pares de valores  $(x,y)$ . O gráfico de dispersão pode indicar correlação linear positiva, negativa ou inexistência de correlação entre duas variáveis. Quando as variáveis são quantitativas, pode-se fazer um gráfico de dispersão, e medir a associação calculando um coeficiente de correlação. São denotados por  $r$ , e têm as seguintes propriedades:  $r$  varia entre -1 e 1;  $r = 0$  corresponde a não associação; quanto maior o valor de  $|r|$ , mais forte a associação;  $r > 0$  ambas variáveis crescem juntas;  $r < 0$  uma variável diminui à medida que a outra aumenta [22].

A Análise de Regressão é uma metodologia estatística que utiliza a relação entre duas ou mais variáveis quantitativas (ou qualitativas) de tal forma que uma variável pode ser predita a partir de outras. O caso mais simples é regressão linear simples que tem duas variáveis e a relação entre elas é representada por uma linha reta. A idéia básica é escolher a reta  $y = a + bx$  que minimiza a soma de quadrados de desvios verticais dos pontos até a reta. Na análise de regressão assume-se que as variáveis têm distribuição de probabilidade Normal e uma variância constante [7].

Modelos lineares generalizados (*MLG*) são usados quando se quer estimar o valor da variável dependente  $Y$  como uma combinação linear de variáveis explicativas  $X_1, X_2, \dots$ . A distribuição de *Poisson* descreve a probabilidade de que um evento ocorra  $k$  vezes num período fixo dado que cada ocorrência é independente e tem uma probabilidade constante. Em regressão de *Poisson*, a distribuição do erro é *Poisson*, o preditor linear é o logaritmo natural do valor predito, assim a função de ligação é o logaritmo [7]. A regressão de *Poisson* é apropriada quando a variável dependente é contagem. No modelo de regressão de *Poisson* o valor predito da variável dependente para  $i$  é a probabilidade máxima estimada  $X_i$  da média de uma variável dependente  $Y$  com distribuição de *Poisson*. O logaritmo natural deste estimador é igual a uma combinação linear dos valores das variáveis

independentes [1].

A análise de séries temporais constitui em um método estatístico que pode ser empregado para o planejamento de ações e políticas públicas, pois permite realizar previsões de eventos futuros, a partir de dados passados. Precisa-se que os dados possuam distribuição Normal, o que pode ser verificado através do teste de *Lilliefors* [22]. A análise de séries temporais, segundo o enfoque de *Box-Jenkins*, tem como principal objetivo fazer previsões. Essa metodologia permite prever valores futuros tomando como base apenas seus valores presentes e passados [10]. Uma das suposições mais freqüentes que se faz a respeito de uma série temporal é que ela seja estacionária, isto é, a média e a variância não apresentam variação ao longo do tempo. As etapas da Metodologia de *Box-Jenkins* [10] são: a Identificação, que consiste em determinar quais dos modelos AR e MA compõem o processo gerador da série, bem como, quais são suas respectivas ordens; a Estimção, que diz que escolhidos os valores adequados para  $p$  e  $q$ , o próximo passo será a estimção dos parâmetros do modelo, em geral, por mínimos quadrados; a Verificação, que consiste em verificar se o modelo identificado é adequado - em caso positivo, pode-se adotá-lo para fazer previsões e em caso negativo, outra especificação deve ser escolhida para modelar a série, o que implica em refazer as etapas de identificação e estimção; a Previsão, considerada a última e mais importante etapa da metodologia de *Box-Jenkins*, que consiste na realização de previsões para a série em instantes de tempo posteriores ao amostrado.

A Análise de Sobrevivência consiste em um conjunto de técnicas onde a variável resposta é o tempo até a ocorrência de um evento de interesse, denominado tempo de falha. A principal característica de dados neste tipo de análise é a presença de censura, que consiste na observação parcial da resposta interrompida por alguma razão e que não permite a observação completa do tempo de falha. Análise de dados de sobrevivência está geralmente relacionada à identificação de fatores ou comparação de diversos métodos controlados por diversos fatores [4]. O principal componente da análise descritiva é a função de sobrevivência. Existem diversos estimadores não-paramétricos para a função de sobrevivência. Dentre eles os mais conhecidos são o estimador da tabela de vida, *Kaplan-Meier* e *Nelson-Aalen*. Em análises onde existem variáveis que podem estar relacionadas com o tempo de sobrevivência existem duas classes de modelos, chamados modelos de regressão para dados censurados: os modelos paramétricos e os modelos semi-paramétricos [4].

O Controle Estatístico de Qualidade (CEQ) tem como objetivo estabelecer, melhorar e assegurar a qualidade de uma produção, atuando em todas as fases do processo produtivo. Um processo está sob controle quando a variação da qualidade for atribuída ao acaso e não a causas identificáveis, ou seja, quando as causas da modificação podem ser detectadas. A principal ferramenta utilizada na identificação destas variações é

o gráfico de controle [9]. Introduzido por W. A. Shewhart em 1924, os gráficos de controle tratam de comparações gráficas do desempenho do processo com determinados limites de controle, verificando se os pontos do gráfico distribuem-se segundo padrões aleatórios (naturais) [9]. Existem duas classes de gráficos de controle de qualidade: de controle por variáveis e de controle de atributos. Os gráficos de controle por variáveis baseiam-se em mensurações da característica de qualidade, sendo de três tipos: da média, do desvio padrão e da amplitude. Os gráficos de controle de atributos baseiam-se na verificação da presença ou ausência de um atributo, em especial com relação a uma lista de defeitos. São de quatro tipos: de fração defeituosa, do número  $np$  de itens defeituosos na amostra de tamanho  $n$ , do número  $u$  de defeitos por unidade e do número  $c$  de defeitos em uma amostra [9]. Além dos Gráficos de Shewhart existem outros tipos de gráficos que são úteis quando existe o interesse em avaliar pequenos desvios no processo. São exemplos destes o gráfico de Soma Cumulativa (CUSUM) e o EWMA [9]. Para a utilização dos gráficos de controle, é necessária a verificação de algumas restrições. Dentre elas, está a verificação da normalidade dos dados. A verificação da normalidade pode ser feita por meios gráficos (Q-Qplots) e de testes estatísticos para inferir sobre a normalidade. Dentre os testes mais utilizados, estão o teste Qui-Quadrado; Kolmogorov-Smirnov; Lilliefors, Anderson-Darling e Shapiro-Wilk. Em casos onde a normalidade não é satisfeita, existem famílias de transformações utilizadas para tornar os dados normais. As transformações mais utilizadas são as da família logarítmica, raiz enésima e BoxCox [22].

## Resultados

O sistema desenvolvido apresenta uma estrutura dividida em três camadas. A primeira camada é a de acesso aos dados, responsável por toda troca de informação com as bases de dados vindas da secretaria de saúde. A segunda é a de negócio, a qual responde por todas as funcionalidades do sistema, desde a autenticação até toda a comunicação com o RServer, através do *JRclient*. A terceira camada, é a de apresentação, trata de toda interação com o usuário.

O fato de lidar com informações sigilosas como de mortalidade e histórico de famílias, obriga o sistema a não fornecer ao seu operador permissões ou formas de alterar esses dados. Assim, uma cópia de todos os dados é feita para a base de dados do sistema. A base de dados principal adotada foi o PostgreSQL, um sistema gerenciador de bancos de dados objeto-relacional de código aberto e livre. O PostgreSQL possui características que o torna interessante para o seu uso nesta aplicação tais como a capacidade de armazenar imagens e dados geográficos além de permitir o uso de tabelas relacionais.

A base de dados do SIAB, que está em sua totalidade armazenada no formato DBASEIII (.dbf), foi integrada a base do sistema. Essa base possui

características peculiares como tipo de caracteres para vários dados numéricos e tipo data somente para alguns valores de data. Fez-se necessário adotar um padrão para toda a base e o escolhido foi que todos os campos fossem importados como caracteres, minimizando o problema de conversão de tipos. Para acessar os arquivos dbf em formato DBASEIII, foi utilizado um pacote sob licença LGPL denominado “com.linuxense.JavaDBF”, que acessa os arquivos e retorna os seus valores em forma de registros. Durante a conversão dos arquivos .dbf em tabelas na base principal foram encontrados diversos problemas, dentre eles: dados incompletos e caracteres inválidos. Estes erros foram contornados em algoritmos desenvolvidos em JAVA antes de inserir os registros na base de dados principal. Observou-se certa demora para incluir a base de dados do SIAB ao sistema devido à necessidade de excluir todos os dados antigos para a atualização da base principal. Isto se dá porque as tabelas da base SIAB não possuem chave primária, o que poderia acarretar na duplicidade dos dados. O diagrama de classes do acesso aos arquivos dbf pode ser visualizado na Figura 3.

A base de dados do SIM e SINASC são bem parecidas com a do SIAB. Parte de seus dados está na forma de arquivos dbf em formato DBASEIII, é a parte mais antiga, e a parte mais nova está em formato INTERBASE. Para tratar isso, foi utilizado o mesmo pacote para leitura dos arquivos dbf, o “com.linuxense.JavaDBF”. Nestes bancos o tratamento de tipos é utilizado para tornar os arquivos dbf iguais a parte que está em INTERBASE, unificando os bancos internamente. Após isso os bancos são convertidos em PostgreSQL.

A base de dados do SINAN é formada por vários arquivos cada um contendo um agravo específico. Estes arquivos são dbf em formato DBASEIII e é utilizado o mesmo pacote “com.linuxense.JavaDBF” para acessá-los. Nestes bancos os tipos são mantidos como estão nos arquivos dbf, por não ser encontrado ainda nenhuma inconsistência nos tipos utilizados. Para cada arquivo é criado uma tabela em PostgreSQL que contém as mesmas informações.

Espera-se que cada um dos sistemas citados passe a utilizar um sistema gerenciador de banco de dados mais eficiente, implantados pelo próprio Ministério da Saúde. Com isso o sistema desenvolvido será expandido e passará a utilizar tanto os dados antigos quanto os novos. O acesso à base de dados mais novas é mais fácil de ser feito, podendo contar também com várias soluções já desenvolvidas em *software* livre para este uso.

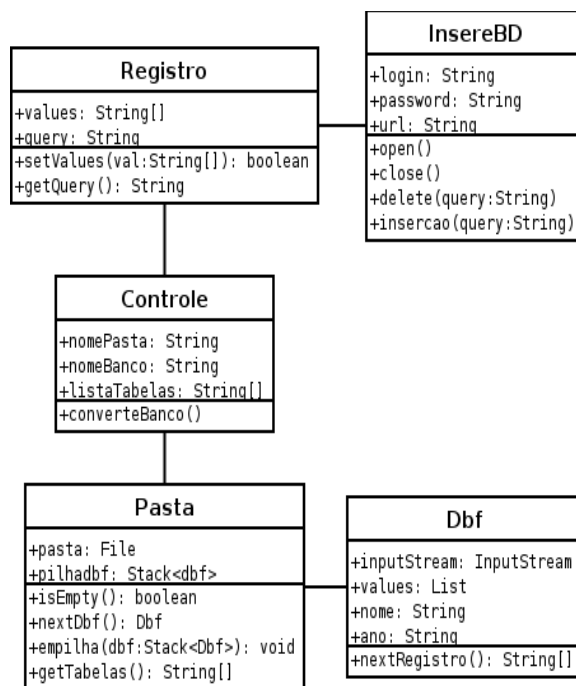


Figura 3: Diagrama de classes do acesso aos arquivos dbf.

Para a aplicação das metodologias estatísticas, foram desenvolvidas funções com o R para serem inseridas no sistema. Com a aplicação destas funções estatísticas obtém-se os relatórios. Um exemplo da criação das funções do R pode ser vista a seguir.

```

RetaReg = function(A, B)
{
  lm.x = lm(B ~ A, data=x)
  abline(lm.x, lwd = 2)
}
  
```

Posteriormente, fez-se a documentação das funções do R com as devidas referências. A documentação está no formato:

- Nome da Função
- Descrição - Para que serve a função e devida referência.
- Parâmetros - parâmetros utilizados na função.
- Retorno - o que a função fornece como resposta.

## Conclusão

A atenção básica à saúde é uma área de extrema importância social. Medidas tomadas para o avanço dos cuidados com a atenção básica, podem significar evolução dos indicadores de desenvolvimento de uma região. O governo brasileiro já mostrou isso claramente ao mudar o foco dos seus investimentos em saúde pública.

A falta de padronização dos sistemas de informação de saúde resultou no problema da não

integração das bases de dados utilizadas. O grande volume de dados não produziu a quantidade de informações esperadas devido a essa falta de padronização e da não integração de seus sistemas e respectivas bases de dados.

Com base nessas dificuldades, este artigo apresentou metodologias de avaliação e análise da atenção básica em João Pessoa a partir da integração das bases de dados disponíveis na Secretaria de Saúde do Município. Para isso, está sendo desenvolvido um sistema baseado em bibliotecas, capaz de criar mecanismos de leitura e comunicação com os diversos bancos provenientes dos sistemas nacionais de monitoramento à saúde e com softwares estatísticos. Para a construção do sistema é utilizada a tecnologia JAVA. A integração dos módulos de leitura e análise com o *software* estatístico R está sendo realizada através do sistema Rserve, que faz a tradução de solicitações de operações estatísticas, permitindo a execução de várias requisições simultâneas em rede. Além disso, foram implementadas rotinas para permitir o cálculo de índices e análises estatísticas para a previsão, controle e mapeamento de eventos específicos de saúde. A metodologia adotada constou de estudos de adequação de modelos nas áreas de geoprocessamento, indicadores de saúde, testes não-paramétricos, coeficientes de correlação, séries temporais, análise de regressão linear e de *Poisson*, análise de sobrevivência e de classificação e controle estatístico de qualidade.

Atualmente, as rotinas de cálculo de índices e análises estatísticas estão sendo integradas ao sistema. Assim, as informações contidas nos bancos de dados farão com que o sistema seja capaz de criar relatórios dos resultados das análises estatísticas utilizando os dados integrados dos bancos.

## Agradecimentos

Este trabalho é financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq pelo Projeto CNPq-MCT/MS-DAB 401978/2005-2.

## Referências

- [1] A. Agreste, An introduction to categorical data analysis. Wiley-Interscience Publication, 1996.
- [2] C. Batini, M. Lenzerini, S. B. Navathe, A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys*, Vol. 18, No 4, 1986, pp. 323-364.
- [3] G. Câmara, A. M. Monteiro, S. Fuks, E Camargo, C. Felgueiras, Análise Espacial de Dados Geográficos, 2ª edição, São José dos Campos: INPE, 2001, [online], disponível em <http://www.dpi.inpe.br/gilberto/livro/analise/>, Setembro, 2002.
- [4] E. A. Colosimo, S. R. Giolo, *Análise de Sobrevivência Aplicada*, 1ª Edição, Edgard Blücher, São Paulo, 2006.
- [5] R. Laurenti, M. H. P. Mello Jorge, M. L. Lebrão, S. L. D. Gotlieb, Estatísticas de Saúde. Editora EPU, 2ª edição, São Paulo, 1987.
- [6] S. Loureiro, Sistema Único de Informação em Saúde?: Integração dos dados da Assistência Suplementar à Saúde ao Sistema SUS, *ANS*, 2003.
- [7] A. Lovett, R. Flowerdew, Analysis of count data using Poisson Regression. *University of Lancaster. Professional Geographer*, Vol. 41, No 2, 1989, pp. 190-198.
- [8] R. M. Marques, A. Mendes, Atenção Básica e Programa de Saúde da Família (PSF): novos rumos para a política de saúde e seu financiamento?, *Ciencia & Saude Coletiva*, Vol. 8, No 2, 2003, pp. 403-415.
- [9] D. C. Montgomery, "Introdução ao Controle Estatístico de Qualidade", 4ª Edição, LTC, 2004.
- [10] P. A. Morettin, C. M. C. Toloi, Análise de Séries Temporais, 2ª edição, revista e ampliada, ABE – Projeto Fisher, Editora Edgard Blücher, 2006.
- [11] F. Murtagh, Correspondence Analysis and Data Coding with Java and R. Chap. Hall, 2005.
- [12] M. A. Pieruccini, O.C. P. Tschuá, S. Iwake, Evolução e caracterização dos indicadores de saúde, Vol. 3, 2001, pp. 359-417. [online] disponível em <http://www.unioeste.br/projetos/oraculus/PMOP/su-mario.htm>, 2007
- [13] Política nacional de atenção básica, Departamento de Atenção à Saúde, Ministério da Saúde, 2006.
- [14] R. Pressman, Engenharia de Software, 5ª Ed. McGraw-Hill, 2005.
- [15] R. S. Santos, M. A. Gutierrez, S. Furuie, U. Tachinardi, Projeto de um Data Warehouse para a Saúde Pública, *Proceedings of the X Congresso Brasileiro de Informática em Saúde*, 2006.
- [16] Sistemas de Informação - SINAN, Ministério da Saúde, [online] disponível em [http://portal.saude.gov.br/portal/saude/visualizar\\_texto.cfm?idtxt=21383](http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=21383), 2007.
- [17] Sistemas de Informação - SINASC, Ministério da Saúde, [online] disponível em [http://portal.saude.gov.br/portal/saude/visualizar\\_texto.cfm?idtxt=21379](http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=21379), 2007.
- [18] Sistemas de Informação - SIM, Ministério da Saúde, [online] disponível em [http://portal.saude.gov.br/portal/saude/visualizar\\_texto.cfm?idtxt=21377](http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=21377), 2007.

- [19] Sistema de Informação da Atenção Básica – SIAB, Ministério da Saúde, [online] disponível em <http://www.datasus.gov.br/siab/siab.htm>, 2007.
- [20] Sistema de Informação da Atenção Básica – SIAB, Ministério da Saúde, Secretaria de Atenção a Saúde, Departamento de Atenção a Saúde, 2006. Série G. Estatística e Informação em Saúde. Brasília, [online] disponível em <http://dtr2004.saude.gov.br/dab/caadab/publicacoes.php>, 2007
- [21] The PostgreSQL Global Development Group 2007, The PostgreSQL Reference Manual, Network Theory LTD, First Edition., Vol. 1. University of California, 2007.
- [22] S. Siegel, NonParametric Statistics for the behaviorel sciences. McGram-Hill Book Company, Inc, 1956.
- [23] S. Urbanek, Rserve – a Fast Way to Provide R functionality to Applications, *Proceedings of the 3<sup>rd</sup> International Workshop on Distributed Statistical Computing*, 2003.
- [24] B. Venables, D. Smith, R. Gentleman, R. Ihaka, Notes on R: A Programming Environment for Data Analysis and Graphics, [online] disponível em <http://bayes.math.montana.edu/Rweb/Rnotes/R.html>, 1998.