

1. INTRODUÇÃO

O R é um software famoso entre os pesquisadores para análises estatística, e é uma ferramenta muito utilizada para o ensino das técnicas.

Nesse artigo será apresentado o software estatístico livre R e sua utilidade em diversas áreas principalmente na área de bioestatística.

2. O USO DA ESTATÍSTICA NA BIOLOGIA

Sabe-se que há centenas de anos atrás os profissionais da área da saúde e da área biológicas começaram a utilizar conceitos estatísticos para analisar fenômenos biológicos. Desse modo, as duas áreas evoluíram bastante quase que simultaneamente. Na estatística, novas técnicas para análise de dados foram desenvolvidas de acordo com a necessidade, em muitos casos a demanda de novas técnicas vinha da área biológica. Já na área biológicas os profissionais passaram a utilizar mais análises quantitativas em pesquisas científicas, utilizando mensurações. Com a facilidade suportada pela tecnologia, como computadores, as duas áreas evoluíram de uma forma surpreendente nas últimas décadas.

3. O SOFTWARE R

Em 1984 Richard Stallman iniciou um projeto que tem como objetivo criar um sistema operacional totalmente livre, o GNU. Um desses projetos é o R-cran, criado por Ihaka & Gentleman em 1996, que é um ambiente, uma linguagem utilizados principalmente para análises estatísticas. Nele é possível fazer manipulação de dados, realização de cálculos e construção de gráficos com pleno controle do usuário. O *software* possui uma imensa quantidade de pacotes que permite a realização de uma grande variedade de técnicas estatística, como modelagem, testes estatísticos, séries temporais, agrupamento, análise multivariadas entre várias outras. Além dessas, outras facilidades do *software* são: facilidade para manipular e armazenar dados; operadores sobre matrizes e outros; linguagem simples, eficiente e bem desenvolvida e facilidade para entrada e saídas.

É possível utilizar o R em muitas plataformas como Linux, Unix, Windows e Macintosh, sua instalação pode ser feita através de arquivos binários ou executáveis.

4. VANTAGENS EM UTILIZAR O R-CRAN

Como o R é um software livre é liberada sua distribuição, instalação, utilização como a alteração e criação de novos comandos. Por isso o R é um *software* flexível, pois permite que o usuário acrescente novas funções, o que não acontece com outros *softwares*. Dessa

forma, existe uma variedade de pacotes disponíveis que permite o usuário aplicar técnicas modernas, muitas das vezes bem específicas em sua área de estudo.

Além disso, o *software* possui código aberto que são gerenciados pela ‘*Core Development Team*’. Isso permite a livre visualização dos comandos implementados pelos usuários, possibilitando a detecção de falhas mais rapidamente, além de uma resolução do problema mais eficiente e rápida, o que na maioria das vezes não acontecem com os *softwares* comerciais. Para esses *softwares* as falhas demoram muito, as vezes anos, para serem percebidas e mais um tempo para lançar uma nova versão com as devidas alterações e além do mais o usuário deve pagar um preço bem alto pela nova versão atualizada.

Alguns *softwares* comerciais e bancos de dados oferecem a sua integração com o R durante a utilização, dentre eles está o Excel da Microsoft, que disponibiliza um add-in *free* do R, que integra análises e gráficos do R no Excel.

Todos os pacotes possuem uma documentação com explicação e exemplos sobre sua utilização, que são criados em LATEX. Além dessa documentação, existem vários fóruns e comunidades do Brasil todo o mundo em que são postadas dúvidas para serem esclarecidas por usuários mais experientes e até mesmo professores.

O *software* R é o *software* mais utilizado por acadêmicos da estatística, uma vez que as linhas de comandos e o código aberto tornaram o software muito mais didático. Por isso muitas pesquisas são feitas e implementadas nele. Com isso, o conhecimento do *software* contribui para um melhor conhecimento e entendimento dessas pesquisas.

Por último, usuários mais avançados tem a opção de escrever manipular objetos R diretamente através de códigos em C, C++ e Fortran.

5. CONCLUSÃO

Como foi citado no texto acima a utilização da estatística está se tornando cada vez mais freqüente na área biológica, com isso a utilização de um bom *software* está se tornando uma necessidade dos profissionais da área.

São vários os *softwares* comerciais que são utilizados para tratar dados, porém o custo deles é bem grande e durante a maioria do tempo o *software* não será utilizado. Assim um *software* livre, como o R, seria a escolha mais viável. Além do mais, como foi mostrado nesse artigo, o R é um *software* eficiente que possui muitas opções para análises. Além disso, o R é uma plataforma independente, assim, se você utilizar o *windows*, por exemplo, não terá grandes problemas em trabalhar com alguém que utiliza o Linux.

Queixas de alunos do curso de estatística e de usuários de outras áreas sobre a dificuldade do uso do software, devido a utilização de linhas de comandos são freqüentes. Entretanto

já existem pacotes que criam uma interface de “*point and click*” que torna as análises de dados no *software* muito mais simples e visualmente mais sofisticadas.

Enfim, o R é um software completo e muito viável ao se tratar de análise de dados, por isso sua utilização na bioestatística é uma boa alternativa.

6. REFERÊNCIA BIBLIOGRÁFICA

Paes A - Itens Essenciais em Bioestatística - São Paulo, SP

Beasley C. – Apostila de exemplos para o biólogo –UFP- Bragança PA

Souza E., Peternelli L., Mello M. - Software Livre R: aplicação estatística

labstats.net/articles/why_r.htmlEm cache [acessado em 31/10/2011]

<http://monkeysuncle.stanford.edu/?p=367> [acessado em 31/10/2011]