

SISVAR: um programa para análises e ensino de estatística

SISVAR: a program for statistical analysis and teaching

Daniel Furtado FERREIRA

Prof. Associado, Dex/UFLA, Bolsista, CNPq

Contato: danielff@dex.u.br

RESUMO

Este trabalho teve por objetivo apresentar o programa Sisvar, que é um sistema para análises estatísticas, podendo também ser utilizado no ensino. O Sisvar é um dos programas brasileiros de análise estatística mais utilizados no país, seja diretamente nas análises estatísticas de trabalhos científicos das mais diferentes áreas do conhecimento científico ou seja no ensino de estatística básica e experimental. Nesse contexto, as principais rotinas de análises estatísticas existentes no programa são apresentadas nesse trabalho. Além do mais, as principais vantagens do Sisvar em relação a outros programas de mesma natureza são enfatizadas. Também são destacadas as limitações do programa e o que vem sendo feito para sanar tais problemas, para ampliar sua capacidade de análises e para otimizar suas rotinas, tanto no aspecto da velocidade de processamento como na acurácia de suas respostas e com habilidade de seus resultados. Aspectos relativos aos cuidados que o usuário do Sisvar deve tomar para garantir a confiabilidade de seus resultados são apresentados e discutidos. Informações técnicas a respeito do programa são apresentadas e discutidas. Como conclusão do presente trabalho destaca-se a grande importância registrada pelo número de citações do Sisvar de suas versões mais antigas nos âmbitos técnico e científico.

Palavras-chave: Acurácia. Robustez. Análises estatísticas.

ABSTRACT

This study aimed to present the Sisvar program, which is a system for statistical analysis and can also be used for teaching statistics. The Sisvar is one of the Brazilian programs of statistical analysis commonly used in the country, for statistical data analysis of scientific papers from all branches of scientific knowledge or for teaching basic and experimental statistics. In this context, the main statistical analysis routines in the program are presented in this work. Moreover, the main advantages of Sisvar in comparison with other similar programs are emphasized. It is also highlighted the program limitations and what is being done to fix such problems, to expand its analysis capacity and to optimize their routines, both in terms of processing speed and of the accuracy of their responses and reliability of their results. Aspects of care that the Sisvar's user should take to ensure the reliability of their results are presented and discussed. Technical information about the program are presented and discussed. In conclusion of this study it was emphasizes the great importance observed by the number of citations found for Sisvar older versions in the technical and scientific knowledge areas.

Keywords: Accuracy. Robustness. Statistical analysis.

INTRODUÇÃO

O Sisvar é um dos programas brasileiros de análise

estatística mais utilizados. Ele foi desenvolvido em linguagem de programação Pascal e implementado em Delphi. Sua primeira versão foi implementada em Turbo

Pascal 3.0 em 1996, ainda no sistema operacional DOS. Possuía um único módulo de análise, para realizar análises estatísticas de modelos lineares provenientes de experimentos com fatores simples ou combinados em diferentes delineamentos experimentais. Nas versões posteriores, o Sisvar passou a usar linguagem orientada por objetos e teve sua primeira implementação em Delphi versão 1.0 em 2000. A versão atual do Sisvar é a 5.3 construção 74, sendo compilada no Delphi 7.0. Portanto, o programa Sisvar está completando 14 anos de existência em 2010. O Sisvar é um programa gratuito e pode ser baixado em <http://www.dex.ufla.br/danielff/>. Sisvar é uma marca registrada em nome da Universidade Federal de Lavras (Ufla) de 28/04/2006 sob o número 828459851. O principal objetivo da criação do Sisvar foi o aspecto didático, ou seja, sua criação foi motivada pela construção de uma ferramenta que possibilitasse uma melhoria no ensino de estatística no Departamento de Ciências Exatas da Ufla. Posteriormente, esse objetivo foi ampliado, no sentido de possibilitar seu uso nas análises de dados oriundos das pesquisas científicas na instituição e fora dela. Atualmente, o Sisvar é usado em todo o território nacional com essas duas finalidades básicas: análise de dados de pesquisa e ensino de estatística.

A construção de um programa de análise estatística é uma tarefa árdua, que exige muita dedicação e demanda muito tempo. Além dos conhecimentos de informática, que avançam a cada dia em um ritmo avassalador, o desenvolvedor precisa acompanhar os avanços na área de estatística, que embora tenham um ritmo menos acelerado, exige grande esforço intelectual para que seja acompanhado. Dessa forma, a criação e a manutenção do Sisvar em todos esses anos de existência consumiu muito esforço, dedicação, estudo e principalmente tempo, o mais escasso dos recursos. Tudo isso, por outro lado, foi extremamente benéfico, pois possibilitou a criação de inúmeras linhas de pesquisa científica, a resolução de inúmeros problemas científicos, a criação de muitos métodos de análise e o desenvolvimento de inúmeras rotinas, sejam elas novas ou versões mais eficientes de outras já existentes na literatura.

Mais recentemente, com o advento do Windows Vista, o Sisvar apresentou problemas de compatibilidade. Esses problemas podem ser contornados, com o uso do módulo de compatibilidade desse sistema operacional. Assim, basta escolher o modo de compatibilidade para executar o programa como sendo do sistema operacional Windows XP, service pack 2, para se obter plena funcionalidade. A razão da incompatibilidade é que o Sisvar usa banco de dados relacionados ao Dbase e Paradox, que foram considerados obsoletos pela Microsoft na edição 2007. Assim, tanto o Windows Vista quanto o Office 2007 deixaram de dar suporte aos arquivos e tabelas relacionados a esse tipo de banco de dados. O Linux é outra plataforma

que tem tido grande aceitação nas universidades públicas no Brasil. Assim, programas gratuitos para esse sistema operacional são cada vez mais exigidos. O Sisvar pode ser instalado no Linux por meio do Wine, que emula um ambiente de compatibilização de programas do Windows. Assim, instalações em Ubuntu e Mint já foram testadas com sucesso. Embora isso seja verdade, o Sisvar não é um programa genuinamente criado para ser executado no Linux ou em outro sistema operacional que não seja o Windows.

Muito embora esses problemas sejam uma realidade, o grande uso do Sisvar no meio científico é sempre uma motivação muito grande para a busca de soluções e para a ampliação de sua capacidade de análise e do aumento de sua eficiência. Uma outra razão motivadora da manutenção desse contínuo esforço de desenvolvimento do Sisvar é a necessidade de se ter programas gratuitos de análise estatística em língua portuguesa e diminuição da dependência de programas estrangeiros de elevado custo. Sabe-se que hoje se pode contar com alguns programas gratuitos de grande desempenho e qualidade como o R (R Development Core Team, 2010), mas não se tem certeza de quanto tempo a característica de ser gratuito irá perdurar. Além do mais, o programa R exige uma maior especialização para o seu uso por pesquisadores de outras áreas, além da estatística.

Assim, o objetivo desse trabalho é apresentar o Sisvar a toda comunidade científica, divulgar os seus principais recursos, limitações e vantagens competitivas e principalmente mostrar o que tem sido realizado para contornar os problemas de compatibilidade e para ampliar as capacidades de análise do programa e para aumentar a eficiência de suas rotinas.

2 PRINCIPAIS ROTINAS DO SISVAR

O Sisvar é um programa simples de ser utilizado. Ele foi dividido em quatro partes para facilitar suas análises, quais sejam, arquivo de dados, análise, relatório e auxílio. O sistema de arquivo do Sisvar utiliza bancos de dados Dbase e Paradox. Ao acessar o menu de manipulação de arquivos da interface principal do Sisvar (Figura 1). Ao ser escolhido o menu de arquivos, opção manipular, o programa conecta-se com um banco de dados denominado sisvar, permitindo ao usuário criar suas tabelas, que são arquivos para serem analisados. Construído o arquivo, que inclusive pode ser feito em uma planilha eletrônica e exportado no formato dbf (dbase) quando disponível, o usuário passa para o módulo de análise. Realizada a análise o programa abre automaticamente o módulo de relatório, com as saídas apropriadas de cada tipo de análise solicitada. O usuário pode manipular o arquivo, eliminando partes ou acrescentando informações, ou seja, pode editar os resultados conforme sua preferência. Os resultados podem

finalmente ser impressos ou gravados como RTF e utilizados em editores de texto tradicionais.

O módulo de auxílio do Sisvar foi construído usando html e possui conteúdos de hipertexto para auxiliar ao usuário como realizar as análises desde a confecção do arquivo até a obtenção dos relatórios. Possui também informações a respeito dos principais métodos estatísticos utilizados em cada opção de análise disponível no programa. Detalhes específicos de cada tópico desses serão fornecidos a seguir.

O módulo de arquivo permite ao usuário construir seus arquivos, que na verdade são tabelas do banco de dados sisvar, criado na instalação do programa. O usuário não precisa saber nada de banco de dados, pois o programa gerencia tudo nos bastidores. Ao se criar um arquivo, o Sisvar irá solicitar ao usuário as informações do diretório e do nome do arquivo, do número de colunas do arquivo, que corresponde ao número de variáveis. Em seguida ele abre o arquivo para construção, ou seja, para preenchimento dos dados. A gravação é feita automaticamente ao ser digitado. Uma nova linha é inserida após cada digitação seguida do uso da seta do teclado \downarrow . Após a entrada do último dado, usar a seta \leftarrow seguida de \downarrow , para que o programa grave esse último valor, mas não grave uma célula vazia. Se houver um erro e células vazias forem criadas, elas podem ser apagadas usando a opção do menu

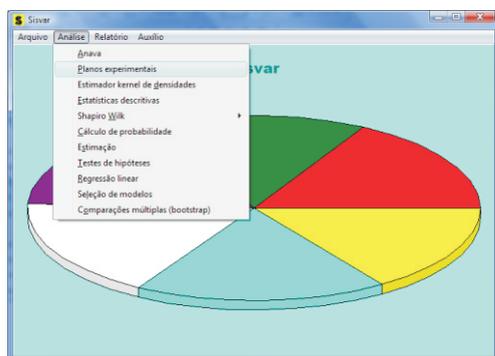


Figura 1: Interface principal do programa Sisvar, destacando seus quatro módulos: arquivo, análise, relatório e auxílio. A opção análise está em destaque, mostrando as principais opções de análises estatísticas disponíveis no Sisvar.

campos, opção deletar. A combinação de teclas Ctrl e Del são usadas para apagar a linha do arquivo desejada, que estará iluminada em azul a partir da escolha usando o mouse para marcar essa linha. O programa não aceita células vazias no arquivo durante o processo de análise.

Na Figura 2 destaca-se um arquivo de um delineamento inteiramente casualizado com um fator com 4 níveis e 5 repetições. O fator em questão são doses de rações

e é um efeito com níveis qualitativos. Ao ser construído, cada variável recebe a indicação de seu tipo, se qualitativa ou quantitativa. A coluna das repetições por sua vez é quantitativa, bem como a coluna da variável resposta, que é o ganho de peso dos animais em kg. Nesse exemplo, o usuário criou um arquivo com 20 observações resultante da combinação dos 4 níveis do fator rações com as 5 repetições de cada um deles. O arquivo em questão e alguns outros são copiados para uma pasta de exemplos, que fica no diretório de instalação do programa, durante o processo de instalação do Sisvar. Arquivos feitos em planilhas podem ser exportados para dbase e utilizados no Sisvar. Arquivos textos (ASCII e Unicode) podem ser importados para serem utilizados no Sisvar.

Após a criação do arquivo o usuário deve passar para a fase de análise dos dados. Os principais módulos de análise do Sisvar são destacados na Figura 1. Ao se escolher uma opção de análise a sequência a ser seguida até a obtenção do resultado é extremamente simples e guiada pelo próprio programa. Assim, a etapa mais difícil da utilização do Sisvar é a manipulação de arquivos. Serão destacadas as rotinas de análise disponíveis no Sisvar na sequência, mas não serão dados detalhes de quais são os passos necessários de cada uma até a obtenção dos resultados. Esses detalhes podem ser obtidos consultando o sistema de auxílio do programa ou o manual do Sisvar disponível na homepage indicada anteriormente.

Rações	Repetições	Ganho de pesos (kg/animal)
B	1	27
B	1	40
C	1	39
A	1	35
B	2	35
D	2	12
C	2	27
A	2	19
C	3	20
D	3	13
A	3	31
B	3	46
D	4	28
A	4	15
C	4	29
B	4	41
D	5	30
A	5	30
B	5	33
C	5	45

Figura 2: Interface de manipulação de arquivos do Sisvar, destacando o arquivo pimen43.db de um delineamento inteiramente casualizado com os fatores rações e repetições e a variável resposta ganho de pesos em kg.

A principal análise do Sisvar é acionada com a opção anava do menu Análise. Essa opção permite ao usuário realizar análises de variâncias do seu arquivo de dados de acordo com o modelo linear apropriado. Assim, os modelos lineares para delineamentos inteiramente casualizados (DIC), blocos casualizados (DBC) e quadrados latinos (DQL) são

facilmente analisados pelo Sisvar, tendo um fator ou mais de um fator com interações entre eles ou efeitos aninhados ou hierarquizados. Na verdade o Sisvar é capaz de lidar com modelos com até 150 efeitos que não sejam fontes de erros e 50 efeitos considerados como fontes de erros, totalizando 200 efeitos no esquema de análise de variância. Assim, pode-se considerar para a maioria das situações encontradas que o programa é praticamente ilimitado. Parcelas subdivididas no tempo e no espaço e combinações desses tipos são facilmente analisados. Ao se analisar um modelo, o Sisvar permite ao usuário que realize testes de comparações múltiplas, contrastes e regressão conforme o caso para os níveis dos fatores que estão sob estudo. Várias opções estão disponíveis, incluindo o teste de Scoot-Knott que não apresenta resultados ambíguos como o de seus tradicionais concorrentes (Tukey, SNK, etc.). Além disso, essas opções podem ser escolhidas quando se está desdobrando os efeitos de um fator dentro dos níveis dos demais, que é o desdobramento de interações e de efeitos aninhados ou hierarquizados. Além disso, algumas opções de transformações de variáveis são disponíveis no programa. Escolhido uma ou mais transformações, o programa realiza a análise da variável original e das transformações escolhidas. Ainda na parte de experimentação, o Sisvar permite por intermédio da opção planos experimentais criar delineamentos experimentais tais como DIC e DBC, com diversas estruturas de tratamentos, como os fatoriais e parcelas subdivididas. Num contexto mais geral são permitidos o arranjo de tratamentos com até 7 fatores distribuídos entre dois níveis de restrição de casualização, as parcelas e as subparcelas. Permite ainda, a criação de delineamentos parcialmente balanceados (PBIB) e os balanceados (BIB). São gerados arquivos com extensão db para serem editorados com o acréscimo dos dados experimentais de acordo com o sorteio realizado e as planilhas de anotação, denominadas books, onde os pesquisadores irão anotar seus dados experimentais e, antes disso, irão utilizar para construir o experimento no campo, uma vez que elas contêm o sorteio realizado.

Vários módulos de análises relacionadas à estatística básica (Ferreira, 2009), tais como estimador kernel de densidades, testes de normalidade, estatísticas descritivas, testes de hipóteses sobre médias de uma população normal e sobre diferenças de duas médias normais independentes e emparelhadas, testes sobre proporções, variâncias e desvios padrões, bem como todos os processos de estimação por intervalo desses mesmos parâmetros mencionados podem ser facilmente aplicados ou obtidos.

Além disso, o Sisvar apresenta módulos de análise envolvendo modelos de regressão linear simples e múltipla. Nesse módulo, o usuário pode estimar parâmetros do modelo

adotado, realizar análises de resíduos e obter os valores preditos com seus intervalos de confiança. Além disso, os métodos de seleção de modelos de regressão stepwise, backward e forward podem ser aplicados. Esses métodos permitem que seja determinado o subconjunto de variáveis regressoras que melhor explicam, de forma significativa, a variação da variável resposta a partir de um grande conjunto de variáveis candidatas a participarem do modelo. Assim, variáveis com informações redundantes podem ser identificadas e eliminadas do estudo e, principalmente, não mensuradas em futuros estudos dessa natureza.

Finalmente, o módulo de análise de variância e comparações múltiplas via testes bootstrap e permutação permite ao usuário realizar tais análises quando os dados experimentais violarem as suposições de resíduos normais e/ou homocedásticos. Dados de contagem, com ou sem excesso de zeros, dados qualitativos ordinais, entre outras possibilidades podem ser submetidos a essa análise. Testes comparando um tratamento com as testemunhas, testes análogos aos testes Tukey, SNK e Scott-Knott estão disponíveis nessa opção. O usuário pode escolher entre milhares e milhões de reamostragens ou permutações na aplicação dos testes. Todas essas opções de análises são executadas com um simples clicar de mouse e as sequências de passos a seguir são altamente intuitivas.

3 PRINCIPAIS CARACTERÍSTICAS

O Sisvar possui uma série de opções de análises, que o tornam bem atrativo. Entretanto, algumas características devem ser mencionadas a respeito dessas opções de análise, para que o usuário saiba exatamente quais as vantagens e desvantagens desse programa. Várias opções de análises ainda não estão disponíveis no Sisvar e, algumas delas, merecem ser destacadas. Nessa seção as vantagens e desvantagens do Sisvar são apontadas. Suas principais características e exigências são também apresentadas.

Como foi dito anteriormente, o principal módulo de análise do Sisvar é a opção anava, para análise de modelos lineares experimentais. Embora seja muito poderosa, essa opção só pode ser utilizada para dados balanceados. A única exceção a essa regra é o caso do delineamento inteiramente casualizado com um fator apenas. O Sisvar não pode realizar análises se houver células vazias, em branco. Quando ocorre perdas de parcelas, o usuário simplesmente deve omitir a linha correspondente à parcela perdida na construção do arquivo. Em variáveis alfanuméricas o usuário deve se atentar para o fato de que o Sisvar diferencia maiúscula de minúscula, e até mesmo a presença de um espaço a mais ou a menos pode determinar um nível da variável diferente. Esse é um erro

comum, pois o que deveria ser o mesmo nível de um fator qualitativo com uma pequena mudança na digitação, representará um novo nível e modificará profundamente os resultados. Variáveis alfanuméricas também tem limitação de tamanho de 30 caracteres.

O Sisvar aceita um máximo de 200 variáveis, colunas, no arquivo de dados e, portanto, é limitado a essa quantidade de variáveis. Um máximo de 100 variáveis regressoras é o limite do Sisvar no módulo de regressão linear e seleção de modelos de regressão. Finalmente, pode-se dizer que o Sisvar até o momento analisa modelos fixos Gauss-Markov apenas. Não há limites de números de observações, a não ser por limitações de espaço no disco rígido do usuário.

3.1 Vantagens

Entre as vantagens que o Sisvar possui em relação aos seus concorrentes está a capacidade de desdobrar os efeitos de interação ou de efeitos aninhados entre fatores de um modelo linear fixo. Além da análise de variância do desdobramento, existe a possibilidade do usuário aplicar testes de médias, incluindo o teste de Scott-Knott ausente na maioria dos concorrentes do Sisvar, além de aplicar contrastes. Para efeitos quantitativos, os desdobramentos desses efeitos devem ser seguidos de análise de regressão. Desconhece-se a existência de algum programa de análise estatística que possa aplicar os testes de regressão, quando ela é realizada dentro do contexto de desdobramento de efeitos de interação, além do Sisvar.

O teste Scott-Knott é uma das grandes vantagens das análises realizadas pelo Sisvar. Esse teste, além de não apresentar resultados ambíguos, possui grande poder e bom controle do erro tipo I por experimento, sob H_0 parcial. O R e o SAS não possuem tais testes implementados entre suas rotinas básicas de análises. Uma outra grande vantagem do Sisvar é sua grande interatividade com o usuário, propiciando um dos ambientes de análises estatísticas mais fáceis de ser usado. Além disso, a grande robustez do programa, fornecendo resultados de altíssima acurácia e de fácil interpretação tem feito do Sisvar um dos mais atrativos programas de análise estatística.

3.2 Desvantagens

A grande desvantagem do Sisvar diz respeito a suas limitações, tanto por não ser capaz de realizar análises de modelos lineares não-balanceados quanto da ausência de algumas importantes técnicas de análises, tais como métodos multivariados, modelos lineares generalizados, modelos não-lineares, tabelas de contingência, entre outras possibilidades. Essas limitações, serão minimizadas em futuras versões do programa. Além do mais, quando o

Sisvar tiver plenamente sendo utilizado na sua nova versão em Java (veja seção posterior), a intenção é abrir todas as suas fontes para receber contribuições de toda a comunidade científica.

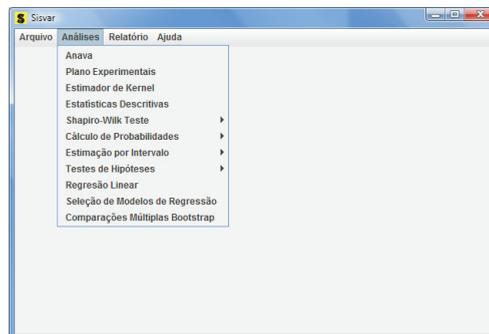


Figura 3: Interface principal do programa Sisvar em Java, destacando seus quatro módulos: arquivo, análise, relatório e auxílio. A opção análise está em destaque, mostrando as principais opções de análises estatísticas disponíveis no novo Sisvar.

Durante a fase de implementação da versão em Java, versão multiplataforma do Sisvar, estão sendo abordados diversos problemas específicos que precisam soluções ou generalizações relacionadas ao aumento da eficiência e da melhoria de técnicas existentes. Dentre elas destacam métodos univariados e multivariados. Serão enfocados nesse sistema métodos multivariados (Johnson & Wichern, 1998; Ferreira, 2008) envolvendo o problema de Behrens-Fisher, testes multivariados de normalidade, além de métodos computacionais para cálculos nas distribuições não-centrais. Além do mais, técnicas tradicionais de análises estatísticas multivariadas e univariadas estão sendo implementadas na nova versão do Sisvar. Muitos métodos estudados durante esse processo foram e estão sendo parcialmente ou integralmente assuntos de dissertações e teses no curso de pós-graduação em estatística e experimentação agropecuária do DEX/UFLA.

A grande barreira existente para o uso de métodos computacionais intensivos na agricultura e no melhoramento de plantas é a inexistência de softwares amigáveis, além da pequena divulgação desses métodos entre os pesquisadores das ciências agrárias e de outras áreas do conhecimento. Devido a essas dificuldades aventadas e a necessidade de que os principais métodos da estatística computacional intensiva aplicados nos mais diferenciados problemas sejam disponibilizados para os cientistas das mais diversas áreas do conhecimento surgiu a ideia de tal inovação no Sisvar. Espera-se que no máximo em 1 ano e meio, a versão Java do Sisvar já esteja disponível.

Atualizações constantes ainda são realizadas na versão em Pascal/Delphi do Sisvar, podendo ser baixado no

endereço: www.dex.ufla.br/ danielff.

5 CONCLUSÕES

O Sisvar possui uma série de vantagens competitivas, sendo a principal delas sua grande interatividade com o usuário propiciando um ambiente simples, eficiente, poderoso, muito robusto e acurado. Essas são as principais razões de seu enorme uso por grande parte da comunidade científica; espera-se que esse artigo seja capaz de ampliar ainda mais sua utilização, apesar das limitações apontadas, que espera-se que estejam minimizadas em futuras edições do programa. As principais ações nesse sentido são a resolução dos problemas de compatibilidade, a ampliação das capacidades de análise do programa, o aumento a eficiência de suas rotinas, bem como a portabilidade para outros sistemas operacionais além do Windows.

Destaca-se a grande importância registrada pelo número de citações do Sisvar de suas versões mais antigas nos âmbitos técnico e científico.

Acknowledgements

O autor gostaria de agradecer ao CNPq, CAPES e à FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais - pelo apoio financeiro concedido, ao longo dos 14 anos de existência do Sisvar.

REFERÊNCIAS

DEITEL, H. M.; DEITEL, P. J. Java: como programar. Pearson Prentice Hall, São Paulo, 6th edition, 2007. 1110 p.

FERREIRA, D. F. Estatística multivariada. Lavras: Ed. Ufla, 2008. 662 p.

FERREIRA, D. F. Estatística básica. Lavras: Ed. Ufla, 2ª ed. ampliada e revisada. 2009. 664 p. JOHNSON, R. A.; WICHERN, D. W. Applied multivariate statistical analysis. Prentice Hall, New Jersey, 4th edition, 1998.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2010. Disponível em: <<http://www.r-project.org>>. Acesso em: 31 Jan. 2010.