



Estatística Descritiva com o uso do software R: Pacote Rcmdr

Robson dos Santos Ferreira
Universidade Bandeirante de São Paulo
Brasil

robson1santos@yahoo.com.br

Verônica Yumi Kataoka
Universidade Bandeirante de São Paulo
Brasil

veronicayumi@terra.com.br

Vanderlei Toledo Severino
Universidade Bandeirante de São Paulo
Brasil

vandertoledo@ig.com.br

Jefferson Adevaldo Santos
Universidade Bandeirante de São Paulo
Brasil

profjeffmat@hotmail.com

Resumo

Esta oficina tem como objetivo apresentar o pacote Rcmdr do software R (R, 2010) para o ensino de conceitos estatísticos básicos, utilizando como exemplo um banco de dados de uma turma do 3º ano do ensino médio de uma escola pública do estado de São Paulo. A oficina será desenvolvida em quatro etapas: instalar o pacote Rcmdr, importar o banco de dados, determinar medidas estatísticas e tabelas; gerar gráficos. Nas etapas três e quatro serão exploradas quais as medidas estatísticas e gráficos estão disponíveis para as variáveis qualitativas, bem como para as quantitativas.

Palavras chave: Ensino, Estatística, software R, pacote Rcmdr.

Introdução

Vivemos em uma sociedade em que prevalece a rapidez na atualização das informações, e, por conseguinte, se exige uma maior habilidade para se trabalhar de forma mais rápida e eficiente na interpretação e assimilação das mesmas em prol de um melhor entendimento da realidade que nos cerca. Geralmente essas informações são transmitidas após algum tipo de tratamento estatístico, sendo representadas em forma de tabelas, gráficos, medidas, intervalos de confiança.

Assim, nesse processo acelerado de geração e repasse de informações, nota-se uma necessidade crescente de que o cidadão se aproprie de alguns conceitos básicos de Estatística, para auxiliá-lo na interpretação e compreensão de grande parte dessas informações, além de ser imprescindível no contexto inferencial para a tomada de decisões.

Os Parâmetros Curriculares Nacionais (PCN) apontam que a demanda social é o que leva a destacar este tema como um bloco de conteúdos específico, evidenciando a sua importância em função do seu uso na sociedade atual. Mais especificamente, os PCN recomendam que ao final do ensino fundamental os alunos tenham a compreensão da importância da Estatística na atividade humana, observando que ela pode induzir a erros de julgamento, tanto pela manipulação de dados, como pela apresentação incorreta das informações, como por exemplo, na ausência da frequência relativa ou em gráficos com escalas inadequadas (BRASIL, 1998). Nesta etapa de ensino recomenda-se que o trabalho com conceitos estatísticos possibilite o desenvolvimento de formas particulares de pensamento e raciocínio de forma que os alunos, além de ler e interpretar gráficos, realizar cálculos de algumas medidas estatísticas, estejam habituados a fazer observações que permitam confiar ou não nos resultados encontrados. Já no ensino médio o objetivo é “capacitar o aluno para compreender e interpretar situações, para se apropriar de linguagens específicas, argumentar, analisar e avaliar, tirar conclusões próprias, tomar decisões, generalizar” (BRASIL, 2002, p.111).

Atualmente contamos com uma importante ferramenta para atingir os objetivos supra citados, o uso da tecnologia, que de acordo com Penteadó (2004) vem a cada dia ocupando um maior espaço em nossa sociedade, sobretudo no cotidiano das pessoas. Nesse sentido, é importante integrar ao processo de ensino e aprendizagem de Estatística, além do ambiente papel e lápis, o uso do recurso computacional.

De acordo com os Parâmetros Curriculares Nacionais “as tecnologias, em suas diferentes formas e usos, constituem um dos principais agentes de transformação da sociedade, pelas modificações que exercem nos meios de produção e por suas conseqüências no cotidiano das pessoas” (BRASIL, 1998, p.43). Destaca-se alguns pontos importantes como: relativizar a importância de cálculos mecânicos, evidenciar a importância do papel da linguagem gráfica e de novas formas de representação, entre outros.

Vale ressaltar que, quando apontamos para as vantagens da utilização do recurso computacional para o ensino de Estatística não estamos descartando a importância do ambiente papel e lápis no desenvolvimento dos conceitos. O trabalho realizado apenas no ambiente computacional pode não assegurar uma aprendizagem ativa, uma vez que se corre o risco dos alunos assumirem uma posição de observadores passivos, tendo desta forma uma baixa assimilação dos conceitos. Assim, é recomendável que o professor desenvolva paralelamente ao ambiente computacional, atividades no ambiente papel e lápis explorando desta forma os conceitos de maneira mais ampla e significativa.

Notamos ainda algumas barreiras no trabalho de Estatística no ambiente computacional, quer seja pela dificuldade tanto por parte dos alunos como dos professores no uso do recurso computacional ou pelo fato de muitos softwares estatísticos não serem gratuitos ou por exigirem conhecimento de linguagem de programação.

Nesse contexto, esta oficina tem como objetivo apresentar o pacote Rcmdr do software R (R, 2010) para o ensino de conceitos estatísticos básicos, utilizando como exemplo um banco de dados de uma turma do 3º ano do ensino médio de uma escola pública do estado de São Paulo.

Software R: Pacote Rcmdr

O software R é um ambiente computacional integrado para manipulação, análise, e representação gráfica de dados baseado em linguagem de programação orientada por objetos e que disponibiliza uma grande variedade de métodos estatísticos.

Um dos atrativos desse software é estar disponível sob os termos da GNU *General Public License* da Free Software Foundation, na forma de código aberto (open source), podendo ser compilado e rodado em um grande número de plataformas UNIX e similar (incluindo FreeBSD e Linux), além do Windows 9x/NT/2000 e MacOS. Outra característica importante do R a destacar, é a possibilidade de utilização em computadores com configurações simples, uma vez que escolas públicas que não tenham sala de informática possam conseguir, por meio de doações, computadores com capacidade similar.

De fato, uma possível limitação inicial no uso do software R é que necessário conhecer e digitar comandos para realizar as operações. Mas, essa primeira dificuldade pode ser contornada com o uso do pacote Rcmdr (chamado também de R Commander), que foi desenvolvido em 2000, em vários idiomas (tais como, inglês, português, francês, italiano, alemão, espanhol) e disponibiliza diversos métodos estatísticos usuais, tanto para auxiliar no ensino da Estatística Descritiva como da Inferência Estatística.

Uma das vantagens desse pacote é que as operações podem ser feitas por meio de *menus* e não diretamente por linhas de comando. Mas, por outro lado, a cada operação realizada é possível que o usuário visualize na janela do script qual seria a linha de comando necessário, ou seja, este pacote permite que o usuário possa aprender passo a passo quais são as linhas de comando que estão sendo executadas no R.

Existe também uma janela para saída dos resultados, sendo que os gráficos são gerados numa janela separada e no rodapé são exibidas algumas mensagens que o sistema julga relevante, como por exemplo, mensagens de erro na execução, número de variáveis e linha de um banco de dados (figura 1).



Figura 1. Tela inicial do pacote Rcmdr.

Para começar a trabalhar com esse pacote o usuário pode encontrar diversos manuais disponíveis gratuitamente na web, tais como, Fox (2005, 2006, 2011), e Mignozzetti (2009).

Desenvolvimento da Oficina

Por conta da limitação do tempo, para o desenvolvimento dessa oficina, será utilizado como exemplo um banco de dados pronto de uma turma de 3º ano do ensino médio, em que foram investigadas as seguintes variáveis: gênero, idade, altura, peso, nota em Matemática, nota em Física, nota em Português, quantidade de irmãos, preferência musical, time favorito, religião, lazer de preferência e gosto pela Matemática (Apêndice A).

Vale salientar que numa situação de aula para trabalhar os conceitos básicos de Estatística é interessante que o professor possa seguir as etapas de uma pesquisa científica, a saber: elaboração de uma questão de pesquisa (escolha de um tema); formulação de hipóteses, construção do instrumento, coleta, tratamento e interpretação dos dados e comunicação dos resultados. Outro aspecto a ser observado, é que na escolha das variáveis, preferencialmente, elas sejam de natureza diferentes (qualitativa nominal, qualitativa ordinal, quantitativa discreta, quantitativa contínua), para potencializar a exploração dos diversos conceitos estatísticos e tipos de gráficos, uma vez que dependendo da natureza o tratamento é diferenciado.

Essa oficina será desenvolvida em quatro etapas: instalar o pacote Rcmdr, importar o banco de dados, determinar medidas estatísticas e tabelas; gerar gráficos. Nas etapas três e quatro serão exploradas quais as medidas estatísticas e gráficos estão disponíveis para as variáveis qualitativas, bem como para as quantitativas.

1-Instalação do pacote Rcmdr

Para usar o pacote Rcmdr é necessário inicialmente instalá-lo. Ao abrir o R vá ao *menu pacotes* e depois em **instalar pacotes**, aparecerá um *menu* com várias nomes de países, que representam onde os pacotes do R e o próprio software estão disponíveis. Selecione qualquer CRAN e esta tela se fechará abrindo outra em que estará disponível todos os pacotes, nesta tela escolha o pacote **Rcmdr** e clique em OK para instalá-lo (figura 2).

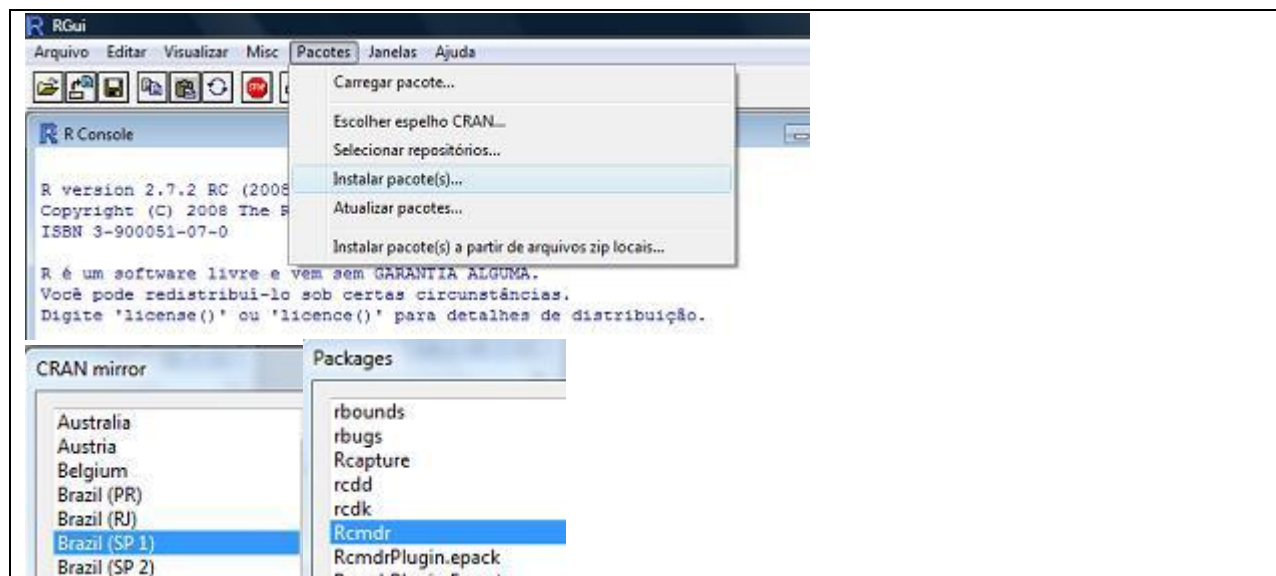


Figura 2. Etapas para a instalação do pacote Rcmdr.

Depois de instalado basta digitar no console do R: **require(Rcmdr)** e teclar enter, que uma janela se abrirá disponibilizando o programa. Ressalta-se que, na primeira vez que o pacote for usado, o R abrirá uma janela, perguntado se você deseja instalar alguns pacotes adicionais, a resposta deve ser SIM.

2-Importando o banco de dados

Para importar um banco de dados devemos utilizar o **menu Dados** e escolher a opção **importar arquivos de dados**, e posteriormente escolher o programa de onde os dados serão importados. Para nossa oficina importaremos os dados do Excel (figura 3).

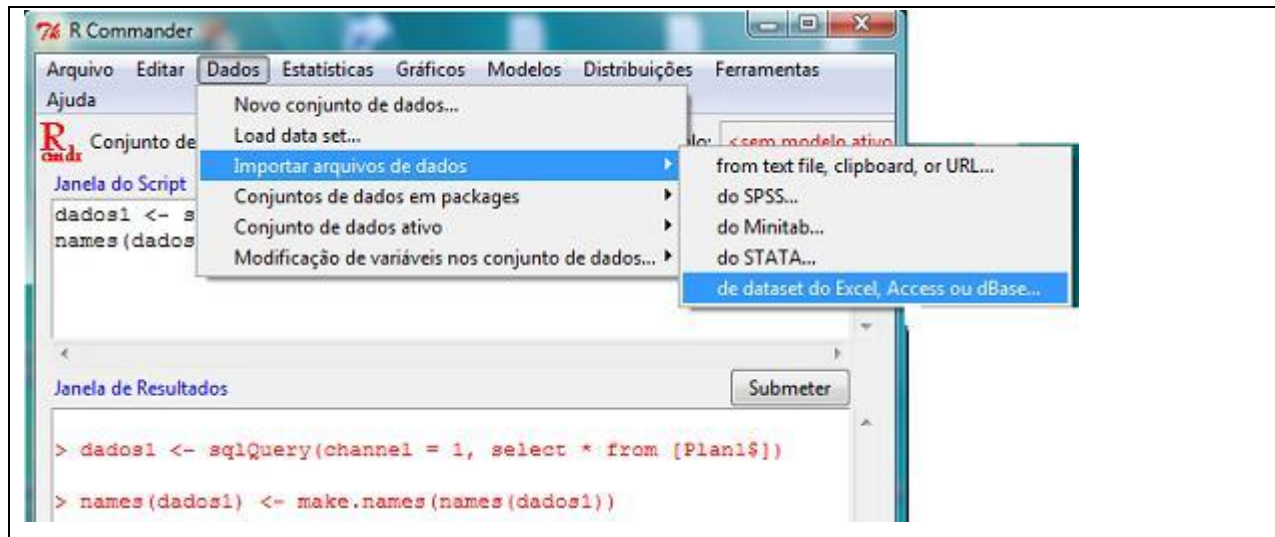


Figura 3. Importando dados do Excel.

No **menu ver dados** podemos visualizar os dados importados (figura 4).

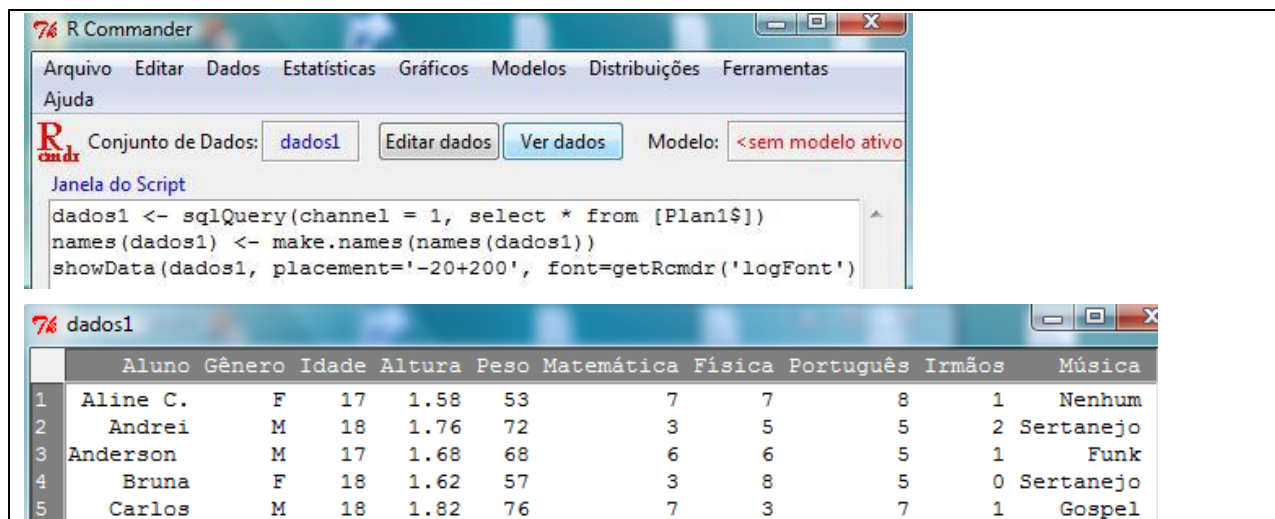


Figura 4. Visualização dos dados importados.

3-Obtenção das medidas estatísticas.

No **menu Estatísticas** é possível obter várias medidas estatísticas e clicando na opção **Resumos- Conjunto de dados ativos**, se obtém o resumo de medidas, que no caso das variáveis

qualitativas seria o número de observações de cada categoria e para as variáveis quantitativas o valor mínimo, 1º quartil, 2º quartil – mediana, média, 3º quartil e valor máximo (figura 5).

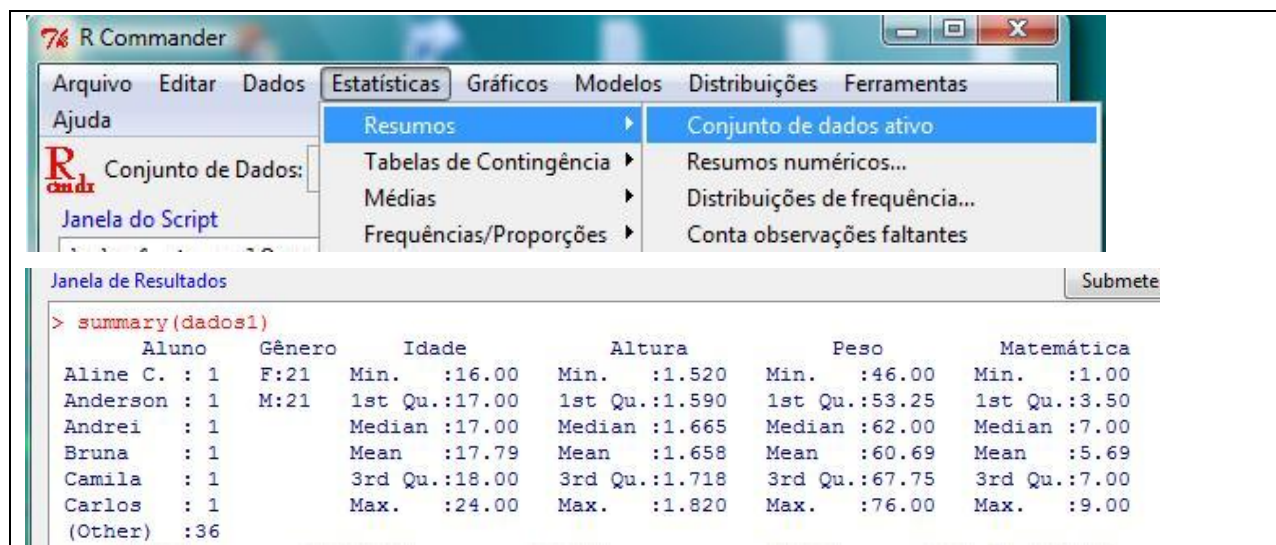


Figura 5. Resumo das medidas para todas as variáveis.

Para obter o desvio padrão ou um quantil de qualquer variável quantitativa, ao invés de optar por **Conjunto de dados ativos**, basta clicar em **Resumos numéricos** (figura 6). É possível calcular também cada medida estatística de uma variável quantitativa de acordo com as categorias de uma variável qualitativa (figura 7).

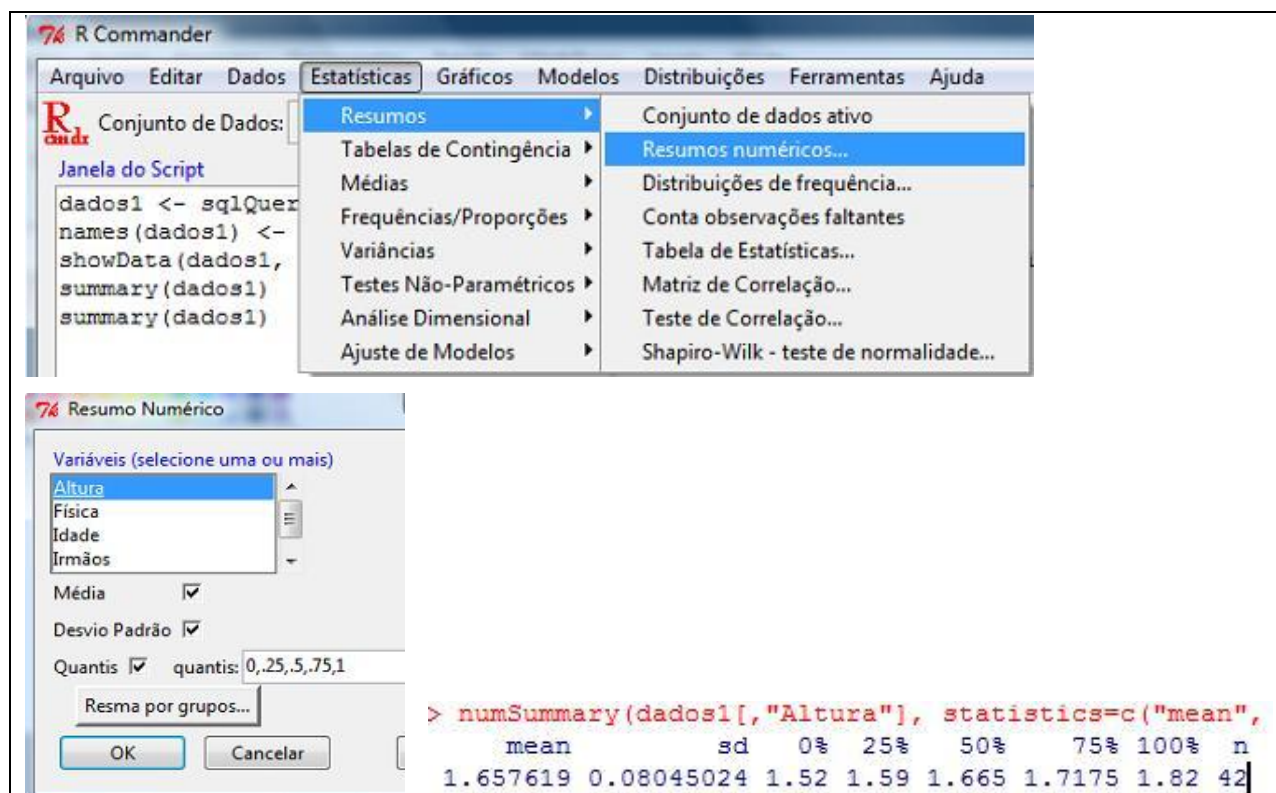


Figura 6. Resumo numérico para a variável altura.



Figura 7. Resumo numérico para a variável altura de acordo com o gênero.

Para formar uma tabela de dupla entrada é só clicar no *menu Tabelas estatísticas* e escolher as duas variáveis qualitativas que serão relacionadas (figura 8).

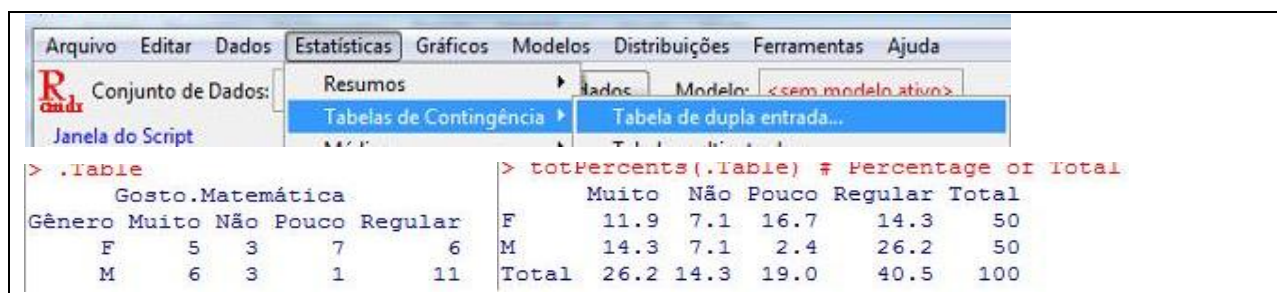


Figura 8. Tabela de dupla entrada para as variáveis gênero e gosto pela Matemática.

4-Gráficos.

No *menu Gráficos* é apresentado todas as opções de gráficos disponíveis no pacote e quando escolhemos, por exemplo, a opção **Gráfico de Barras** é apresentado quais são as variáveis disponíveis para este tipo de gráfico (apenas as variáveis qualitativas), escolher a variável desejada e clicar em **OK** que o gráfico será gerado (figura 9).

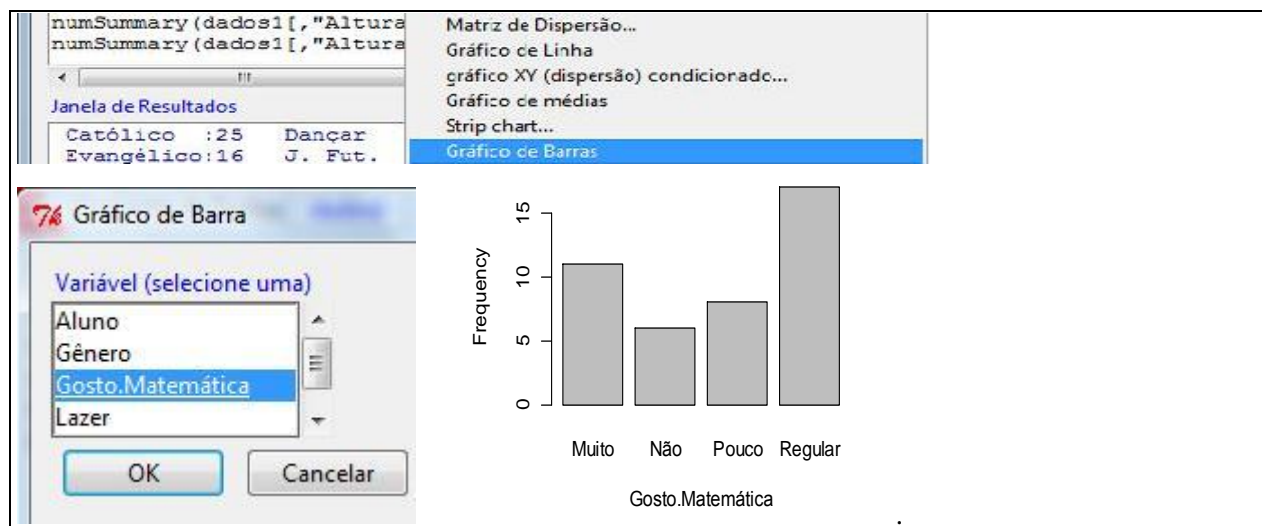


Figura 9. Gráfico de barras da variável gosto pela Matemática.

Vale destacar que, o professor pode aproveitar o momento da geração dos gráficos para estimular aos alunos a modificar as linhas de comando do *default*, por exemplo, no caso do gráfico de barras da figura 9, o comando é

```
barplot(table(dados1$Gosto.Matemática), xlab="Gosto.Matemática",ylab="Frequency")
```

O aluno pode mudar o nome dos eixos trocando as expressões em xlab e ylab, pode modificar também a cor das barras, adicionando o comando col após o ylab, escrevendo as cores em inglês separadas por aspas, submeter e gerar assim um novo gráfico (figura 10):

```
barplot(table(dados1$Gosto.Matemática), xlab="Gosto pela Matemática",ylab="Frequência", col = c("red","green","blue","black"))
```

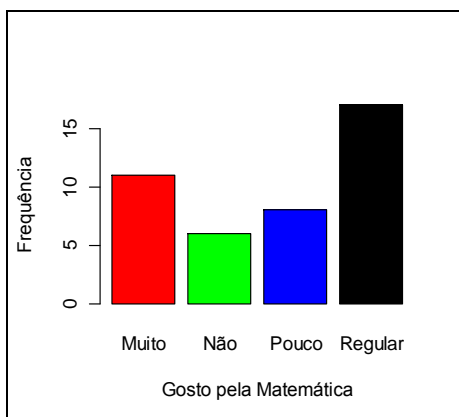


Figura 10. Gráfico de barras da variável gosto pela Matemática.

Quando optamos, por exemplo, pela opção Histograma são apresentadas as opções de variáveis disponíveis, neste caso as variáveis quantitativas (figura 11).

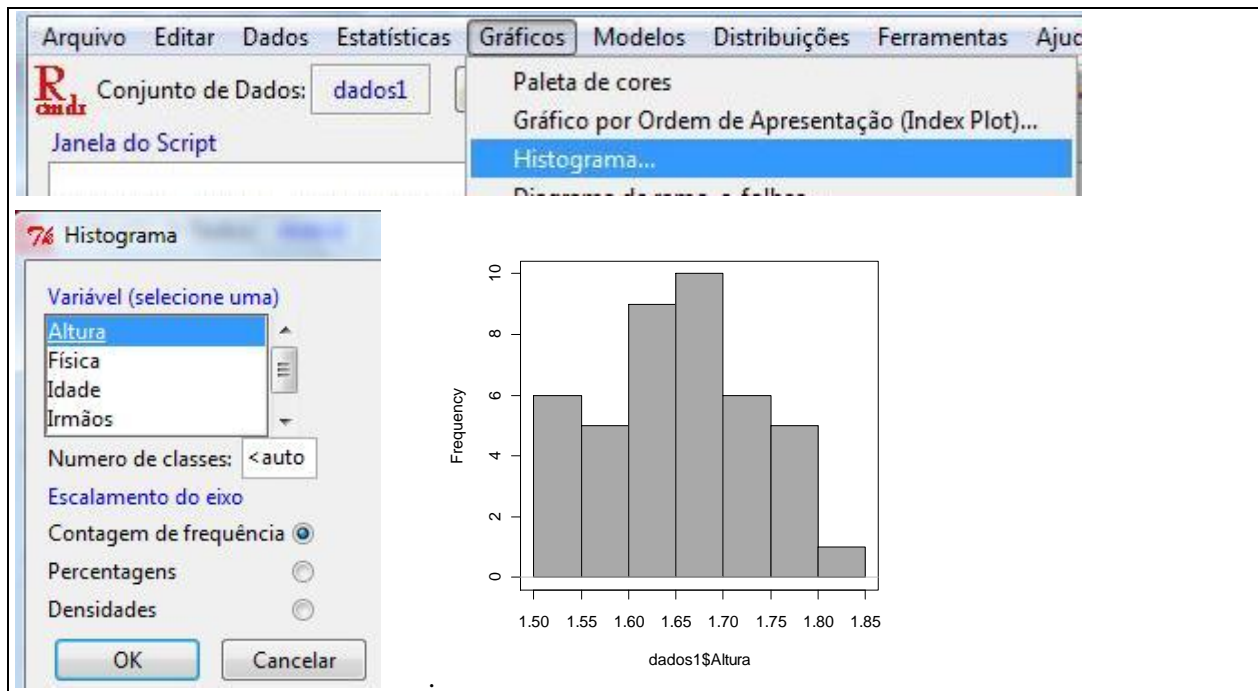


Figura 11. Histograma da variável altura.

É possível gerar um gráfico para uma variável quantitativa de acordo com as categorias de uma variável qualitativa, como por exemplo, um boxplot da variável altura de acordo com o gênero (figura 12). Pode ser gerado também o diagrama de dispersão que relaciona duas variáveis quantitativas (figura 13).

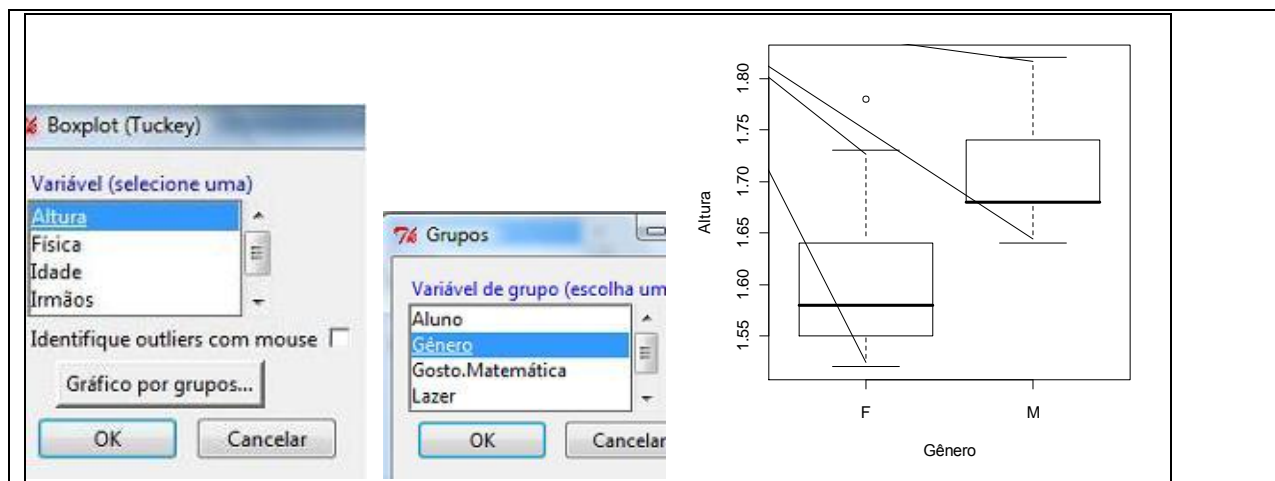


Figura 12. Boxplot da variável altura de acordo com o gênero.

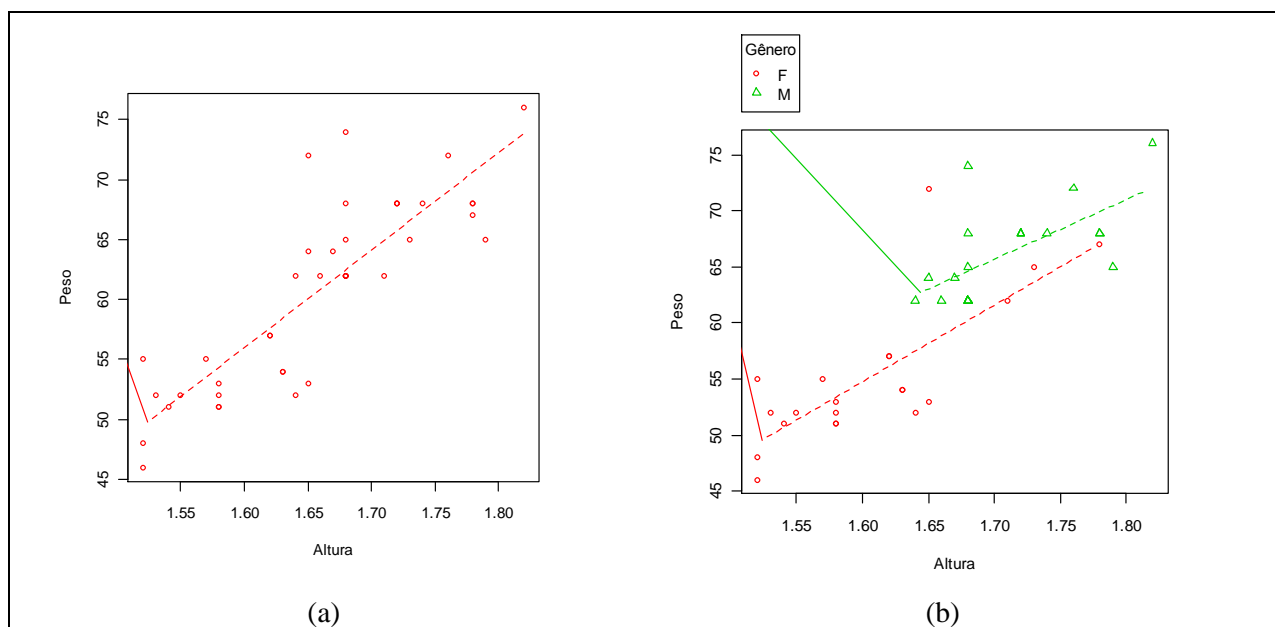


Figura 13. Diagrama de dispersão entre as variáveis altura e peso (a) e de acordo com o gênero (b).

Considerações Finais

Tem-se por hipótese inicial que o trabalho com o software R por meio do pacote Rcmdr proporcionará novas formas de compreensão de conceitos estatísticos, uma vez que o mesmo permite explorações diferenciadas daquelas usualmente tratadas no ambiente papel e lápis, tais como o acesso rápido e nítido, tanto nas medidas estatísticas como dos gráficos.

Por outro lado, em uma primeira análise, poderíamos pensar que esta oficina seria inviável para a aplicação em sala de aula devido à precariedade dos laboratórios de informática de uma grande parte das escolas de Educação Básica e também da falta de familiaridade de um

número considerável de professores de matemática com o uso do computador. Contudo, devemos considerar que segundo Penteado (2004), o movimento, a velocidade, o ritmo acelerado com que a informática imprime novos arranjos na vida fora da escola caminham rumo à escola, causando transformações no ambiente escolar e exigindo transformações nos sistemas de hierarquias e prioridades tradicionalmente estabelecidas na profissão docente.

Desta maneira, é importante novas propostas de trabalho que contemplem essas necessidades, e foi pensando nestas dificuldades que para esta oficina optamos pela utilização de um software free de baixa configuração e a estruturamos de forma a orientar o seu desenvolvimento passo a passo, desde a instalação do pacote Rcmdr até a construção dos gráficos, para que mesmo pessoas com pouca familiaridade com o uso do computador, possam utilizá-lo em suas práticas docentes para o ensino de conceitos básicos de Estatística.

Bibliografia e referências

- Brasil, Ministério da Educação (1998). Secretaria de Educação Fundamental. *Parâmetros Curriculares Nacionais: Matemática*. Brasília: Ministério da Educação/Secretaria de Educação Fundamental.
- Brasil, Ministério da Educação (2002). Secretaria de Educação Média e Tecnológica. *PCN Ensino Médio: Orientações Educacionais complementares aos Parâmetros Curriculares Nacionais – Ciências da Natureza, Matemática e suas Tecnologias*. Brasília: Ministério da Educação/ Secretaria de Educação Média e Tecnológica.
- Fox, J (2004). The R Commander: A Basic-Statistics Graphical User Interface to R. *Journal of Statistical Software*, 14 (9), 1-42.
- Fox, J. (2006) *Getting started with R Commander*. Disponível em: <http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/Getting-Started-with-the-Rcmdr.pdf>
- Fox, J. (2011). *Package Rcmdr*. Disponível em: <http://cran.r-project.org/web/packages/Rcmdr/Rcmdr.pdf>
- Mignozzetti, U.G. (2009). *Introdução ao R Commander*. Disponível em: <http://www.nadd.prp.usp.br/cis/arqs/aprcmdr.pdf>
- Penteado, M. G. (2004). *Educação matemática: pesquisa em movimento*. São Paulo: Cortez, p. 77-91.
- R.Development Core Team (2010). *R: A language and environment for statistical computing*. R. Foundation for Statistical Computing, Vienna, Austria, v. 0.02.

Apêndice A

Banco de dados de uma turma de 3º ano do ensino médio

Aluno	Gênero	Idade	Altura	Peso	Matemática	Física	Português	Irmãos	Música	Time de Futebol	Religião	Lazer	Gosto Matemática
Aline C.	F	17	1,58	53	7	7	8	1	Nenhum	Palmeiras	Evangélico	Tocar	Muito
Andrei	M	18	1,76	72	3	5	5	2	Sertanejo	Corinthians	Católico	J. Fut.	Não
Anderson	M	17	1,68	68	6	6	5	1	Funk	S.Paulo	Católico	J. Fut.	Regular
Bruna	F	18	1,62	57	3	8	5	0	Sertanejo	S. Paulo	Católico	Passear	Pouco
Carlos	M	18	1,82	76	7	3	7	1	Gospel	Nenhum	Evangélico	Cantar	Regular
Carolina	F	19	1,53	52	2	5	5	3	Funk	S. Paulo	Católico	Dançar	Pouco
Camila	F	16	1,65	53	6	9	5	2	Sertanejo	Nenhum	Evangélico	Cantar	Regular
Cristiano	M	17	1,68	62	8	4	5	1	Rock	Santos	Nenhum	Tocar	Regular
Celso	M	18	1,72	68	3	5	3	2	Pagode	Corinthians	Católico	J. Fut.	Pouco
Daniel	M	17	1,77	64	6	7	5	1	Todos	Corinthians	Católico	Natação	Regular
Daniela	F	19	1,54	51	7	8	7	2	Gospel	Palmeiras	Evangélico	Cantar	Não
Elion	M	18	1,74	68	2	6	4	1	Funk	S.Paulo	Católico	Dançar	Não
Fabiana	F	17	1,52	55	7	7	7	2	Sertanejo	Corinthians	Evangélico	Dançar	Regular
Henrique	M	17	1,79	65	5	1	6	1	Todos	S.Paulo	Católico	J. Fut.	Regular
Juliana	F	18	1,62	57	1	8	8	0	Funk	S.Paulo	Católico	Passear	Não
Julio	M	21	1,68	74	6	9	8	2	Todos	Corinthians	Evangélico	J. Fut.	Regular
Katia	F	17	1,57	55	2	6	2	4	Sertanejo	Corinthians	Católico	J. Fut.	Pouco
Kelly	F	17	1,52	48	8	3	5	1	Classica	Palmeiras	Evangélico	Leitura	Muito
Laysla	F	18	1,78	67	5	5	6	2	Funk	Corinthians	Católico	Dançar	Pouco
Leandro	M	17	1,64	62	8	5	7	3	Todos	S.Paulo	Católico	Dançar	Muito
Lincon	M	19	1,65	64	7	6	7	2	Rap	Palmeiras	Católico	Dançar	Regular
Luis	M	17	1,68	62	8	8	7	1	Forró	Corinthians	Católico	Correr	Muito
Lilian	F	24	1,71	62	8	7	6	0	Todos	S.Paulo	Evangélico	Ouve Mus.	Regular
Maria Ap.	F	17	1,52	46	3	6	8	2	Nenhum	Nenhum	Evangélico	Cantar	Regular
Mariana	F	17	1,63	54	1	7	6	3	Funk	Corinthians	Católico	Passear	Pouco
Meire	F	16	1,55	52	7	7	9	2	Sertanejo	S.Paulo	Católico	Dançar	Pouco
Marcos	M	17	1,68	65	7	5	6	1	Gospel	Corinthians	Evangélico	Tocar	Regular
Mirian	F	18	1,58	51	2	6	2	3	Funk	S.Paulo	Católico	Ouve Mus.	Pouco
Murilo	M	17	1,78	68	8	2	9	2	Pagode	S.Paulo	Católico	Ouve Mus.	Não
Nilson	M	17	1,68	62	6	9	6	1	Pagode	Palmeiras	Católico	J. Fut.	Regular
Odair	M	18	1,72	68	7	7	5	2	Nenhum	Nenhum	Evangélico	J. Fut.	Muito
Paula	F	19	1,65	72	1	5	1	3	Funk	S.Paulo	Católico	Dançar	Não
Priscila	F	17	1,58	52	7	2	8	2	Sertanejo	S.Paulo	Católico	J. Fut.	Regular
Plinio	M	17	1,68	62	6	6	8	1	Todos	Palmeiras	Evangélico	Passear	Regular
Kledson	M	17	1,78	68	5	8	6	2	Nenhum	Corinthians	Evangélico	J. Fut.	Regular
Rayssa	F	18	1,63	54	8	9	8	1	Todos	S.Paulo	Católico	Dançar	Muito
Rayanna	F	18	1,58	51	7	6	6	2	Sertanejo	Corinthians	Católico	Ouve Mus.	Regular
Rita	F	17	1,73	65	8	8	6	4	Todos	S.Paulo	Evangélico	Ouve Mus.	Muito
Robson	M	17	1,68	62	7	5	7	2	Sertanejo	Palmeiras	Católico	Passear	Muito
Ruth	F	17	1,64	52	9	5	3	1	Todos	Corinthians	Evangélico	Passear	Muito
Rhuan	M	18	1,72	68	8	6	8	3	Funk	S.Paulo	Católico	Dançar	Muito
Sergio	M	21	1,66	62	7	8	8	2	Todos	S.Paulo	Evangélico	Ouve Mus.	Muito

Apêndice B

• Informações gerais sobre a oficina

Título da oficina	Estatística Descritiva com o uso do software R: pacote Rcmdr
Nome dos autores	Robson dos Santos Ferreira , robson1santos@yahoo.com.br Verônica Yumi Kataoka , veronicayumi@terra.com.br Vanderlei Toledo Severino, vandertoledo@ig.com.br Jefferson Adevaldo Santos, profjeffmat@hotmail.com
Instituições dos autores	UNIBAN - Universidade Bandeirantes de São Paulo
País	Brasil
Número de horas mais convenientes	2 horas
Nível de escolarização	Ensino Fundamental, Ensino Médio, Ensino Superior
Número máximo de pessoas	20
Equipamentos audiovisuais	Datashow, laboratório de informática com o software R e o pacote Rcmdr instalados.