



# Bioestatística quantitativa aplicada

Edison Capp  
Otto Henrique Nienov  
Organizadores

Caroline Darski  
Charles Francisco Ferreira  
Cristiana Palma Kuhl  
Fernanda Dapper Machado  
Fernanda Vargas Ferreira  
Hellen Meiry Grosskopf Werka  
Johanna Ovalle Diaz  
Marina Petter Rodrigues  
Michele Strelow Moreira  
Nadine de Souza Ziegler  
Paula Barros Terraciano  
Pedro Henrique Comerlato  
Sinara Santos

Universidade Federal do Rio Grande do Sul  
Faculdade de Medicina  
Programa de Pós-Graduação em Ciências da Saúde:  
Ginecologia e Obstetrícia

# Bioestatística Quantitativa Aplicada

Porto Alegre 2020  
UFRGS

U58b Universidade Federal do Rio Grande do Sul. Faculdade de Medicina. Programa de Pós-Graduação em Ciências da Saúde: Ginecologia e Obstetrícia Bioestatística quantitativa aplicada/ Universidade Federal do Rio Grande do Sul; organizadores: Edison Capp e Otto Henrique Nienov – Porto Alegre: UFRGS, 2020.

260p.

ISBN: 978-65-86232-43-1

E-Book: 978-65-86232-44-8

1. Epidemiologia e Bioestatística 2. Estatística 3. SPSS I. Capp, Edison, org. II. Nienov, Otto Henrique, org. III Título.

NLM: WA950

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
(Bibliotecária Shirlei Galarça Salort – CRB10/1929)

Endereço:

PPG em Ciências da Saúde: Ginecologia e Obstetrícia

FAMED – UFRGS

Rua Ramiro Barcellos, 2400/2º andar

CEP 900035-003 – Porto Alegre – RS

Telefone: +55 51 3308 5607

E-mail: ppggo@ufrgs.br

Editoração e diagramação: Edison Capp

Capa: Edison Capp, imagens: [www.freepik.com/starline](http://www.freepik.com/starline)

Edison Capp  
Otto Henrique Nienov  
Organizadores

Caroline Darski  
Charles Francisco Ferreira  
Cristiana Palma Kuhl  
Fernanda Dapper Machado  
Fernanda Vargas Ferreira  
Hellen Meiry Grosskopf Werka  
Johanna Ovalle Diaz  
Marina Petter Rodrigues  
Michele Strelow Moreira  
Nadine de Souza Ziegler  
Paula Barros Terraciano  
Pedro Henrique Comerlato  
Sinara Santos

## 3 Primeiros passos no SPSS

Hellen Meiry Grosskopf Werka  
Johanna Ovalle Diaz  
Edison Capp  
Otto Henrique Nienov

### Conhecendo o SPSS

O *software* IBM® SPSS (*Statistical Package for Social Sciences*) é um pacote estatístico amplamente utilizado no ambiente acadêmico. É uma ferramenta para análise de dados, que utiliza técnicas estatísticas básicas e avançadas. Além disso, permite executar estatísticas descritivas, correlação e regressão, categorizar variáveis, manipular dados e criar gráficos e tabelas.

É uma ferramenta paga, adquirida no site da IBM (<https://www.ibm.com>), mas também está disponível para a comunidade acadêmica da UFRGS (<https://www.ufrgs.br/documentacaoti/spss-tutorial-de-instalacao/>), sendo necessária a instalação e ativação da VPN (<https://www.ufrgs.br/documentacaoti/vpn/>). Através da VPN (*Virtual Private Network*) é criado um canal que permite o tráfego de dados de forma segura e o acesso a rede interna da UFRGS, possibilitando o acesso em casa, por exemplo.

Apesar de ser um *software* estatístico de fácil manuseio, usar o SPSS sem conhecimento algum de estatística pode ser arriscado. O SPSS apresenta um complexo conjunto de opções para cada teste, muitos dos quais necessitam de atenção na hora de serem utilizados e interpretados. Pensando nisso, este capítulo busca descrever aspectos básicos para aqueles que não têm experiência com o SPSS.

### Construindo um banco de dados no SPSS

O primeiro passo é conhecer o SPSS. Para isso, vamos construir um banco de dados. Abra o SPSS, clicando no atalho da versão instalada na área de trabalho do seu computador. Na caixa

de diálogo é possível abrir arquivos existentes que, por padrão possuem a extensão “.sav”, ou podemos inserir novos dados criando um banco de dados.

Selecione a opção “Inserir Dados” e, em seguida, clique em “Ok”. Quando você executa o SPSS desta forma, ele irá apresentar uma planilha em branco. Nessa janela de edição de dados, onde inserimos os dados e executamos as funções estatísticas, há duas planilhas: uma chamada “Visualização da variável” e outra “Visualização dos dados”. A primeira corresponde a um dicionário das variáveis, onde iremos criar as variáveis do banco de dados e inserir as informações que a definem, como sua descrição, identificação dos rótulos e sua natureza (quantitativa ou qualitativa); enquanto a segunda corresponde ao banco de dados. Lembre-se que o banco de dados é uma estrutura organizada que permite a extração de informações. As linhas correspondem aos dados de uma unidade experimental (que pode ser uma pessoa, animal ou objeto) e as colunas correspondem as variáveis criadas na “Visualização da variável”. Caso você esteja utilizando o SPSS, é possível abrir um novo banco de dados clicando em “Arquivo”, “Novo” e, em seguida, em “Dados”.

Desta forma, o primeiro passo para inserir os dados no SPSS é criar a variável utilizando a planilha “Visualização da variável” do editor de dados e então entrar com os dados na planilha “Visualização dos dados”. Na planilha de “Visualização dos dados”, clicando duas vezes sobre a coluna “var”, você será direcionado para a primeira linha da planilha “Visualização da variável”. Isso significa que, a primeira coluna da “Visualização dos dados” corresponde à primeira linha na “Visualização da variável”. Portanto, está será nossa primeira variável. É recomendado que a primeira coluna da matriz corresponda a uma variável de identificação. Pode ser o número do questionário, número do caso, ou ainda código do registro. Isso facilita a localização de informações no caso de serem identificados equívocos de digitação.

Nesta planilha de “Visualização da variável”, vamos detalhar como a variável será criada. Na primeira coluna (“Nome”), digita-se o nome da variável desejada (por exemplo, “ID”). Esse nome será apresentado no início da coluna na planilha “Visualização dos

dados” quando você for inserir os dados e serve para identificar as variáveis. Existem algumas regras gerais para o nome das variáveis, como utilizar oito dígitos no máximo e não utilizar espaço em branco, nem os símbolos “-”, “. ” e “/”. Também, o SPSS não permite que o nome de uma variável comece com um número. Por isso, caso você tenha variáveis repetidas, insira o número ao final do nome da variável, por exemplo. Se for violada alguma dessas regras, o SPSS informará que o nome da variável é inválido.

Seguindo na descrição da variável, na coluna “Tipo”, é definido o tipo de dado. Você pode ter diferentes tipos de dados. Clique na célula “Tipo” e, na janela “Tipo de variável”, você pode selecionar o tipo de dados que constará naquela variável. O tipo mais comum (padrão) é o “Numérico”. Caso a variável seja numérica, como por exemplo o número de prontuário, selecione a opção “Numérico”. Importante! As variáveis qualitativas, em que as categorias foram descritas em códigos, também são do tipo “Numérico”. Caso a variável seja alfa-numérica (texto) selecione a opção “Sequência”. Uma variável de texto é simplesmente uma linha de caracteres que representa uma informação que não se deseja analisar, como o nome do participante. Há outras opções específicas, como “Data”. Esta deve ser classificada, na coluna “Medir”, como quantitativa (“Escala”) quando se deseja calcular diferença entre datas para obter-se outras variáveis, como idade, tempo de acompanhamento, etc. Preferencialmente use sempre a modalidade “Numérico” para variáveis categóricas (qualitativas), como sexo, estado civil, etc., atribuindo um código para as categorias. O SPSS não reconhece caracteres (palavras), sendo necessário codificar as categorias da variável, que serão identificadas na coluna “Valores”. Você deve atribuir a cada categoria da variável um número e, então, informar ao SPSS que número representa cada categoria. Selecione o tipo de variável desejado e clique em “Ok”.

Na coluna “Largura”, é possível definir o tamanho do valor da variável. Por exemplo, se o número do caso tiver quatro dígitos, definimos como a largura sendo igual a quatro. Na coluna “Decimais”, é possível definir o número de casas decimais após a vírgula que iremos atribuir à variável. A largura não pode ser menor do que o decimal. Na coluna “Colunas”, definimos o tamanho do nome da variável.

Como o nome de uma variável é restrito a oito caracteres, o SPSS fornece uma maneira para que possamos identificar as variáveis de uma forma mais detalhada. Na coluna "Rótulo", é possível digitar um nome mais completo (por extenso) para a variável, identificando a variável e especificando a unidade de medida. Por exemplo, para a variável "ID" é possível digitar a expressão "Identificação do caso". Em "Valores", atribuímos os valores da variável de interesse, ou seja, o valor das categorias (códigos). Clique na célula "Valor" e, em seguida, na janela "Rótulos de valor", insira o valor da categoria ("Valor") e o nome da categoria ("Rótulo"). Clique em "Adicionar" para inserir o valor do rótulo e, em seguida, em "Ok". É conveniente registrar no banco de dados os nomes das categorias de variáveis categóricas. Por exemplo, para a variável "SEXO", poderíamos atribuir o valor "1" para o rótulo "masculino" e "2" para "feminino". Esses códigos informam ao SPSS que todos os casos da variável atribuídos ao código "1" pertencem ao sexo masculino, da mesma forma que os casos atribuídos ao código "2" pertencem ao sexo feminino. Importante! A utilização de códigos numéricos para as categorias de uma variável qualitativa não a tornam quantitativa. Também é possível alterar as informações de valor e rótulo em "Alterar" ou removê-los em "Remover".

Existem situações em que não temos todos os dados coletados (*missing value*) para os casos ou estes se apresentam como não-resposta ou, ainda, respostas que não desejamos considerar para o tratamento estatístico, como por exemplo, respostas não corretas, que não se aplicam, etc. A omissão de dados pode ocorrer por vários motivos: em questionários longos onde os participantes podem acidentalmente esquecer de responder alguma questão; em fichas de coleta onde o investigador pode esquecer de registrar algum dado; em estudos que utilizam registros de prontuários onde os dados podem não estar apresentados ou; em um experimento que falha para determinado caso e não se obtém o valor do dado. Para isso, utilizamos a coluna "Ausente". Na janela "Valores ausentes", pode-se registrar o código de não-resposta na opção "Valores ausentes discretos", utilizando preferencialmente 9, 99, 999, etc. Clique em "Ok" para adicionar o valor. Entretanto, a melhor opção para não-resposta é deixar o espaço em branco no banco de dados,

selecionando a opção “Sem valores ausentes”, pois variáveis podem assumir o código de não-resposta como um valor válido.

Em “Alinhar”, pode-se selecionar o alinhamento dos dados na planilha à direita, à esquerda ou no centro. Em “Medir”, deve-se selecionar como serão analisadas estatisticamente as variáveis: a) “Escala” refere-se a variáveis quantitativas, sejam discretas ou contínuas; b) “Ordinal” refere-se a variáveis qualitativas ordinais e; c) “Nominal” a variáveis qualitativas nominais. Por fim, na coluna “Função”, é definida a funcionalidade da variável. Como padrão, as variáveis são classificadas como “Entrada”.

Além da janela de edição de dados, o SPSS fornece uma janela de visualização de resultados, onde irão ser apresentados os resultados das análises, e a janela de edição de sintaxe, que trará os comandos das funções estatísticas.

Quando você for construir um novo conjunto de dados, você deve entrar com os dados de uma forma lógica. Para exemplificar, vamos construir um banco de dados fictício, com características demográficas de uma amostra aleatória. Crie as variáveis na planilha de “Visualização da variável” e, uma vez criada a variável, insira os dados na planilha de “Visualização dos dados” (Quadro 1). Neste banco de dados, a variável “ID” corresponde à identificação dos casos, “NOME” ao nome dos sujeitos, “SEXO” ao sexo (onde, 1 = sexo masculino e 2 = sexo feminino), “DN” à data de nascimento, “IDADE” à idade (em anos) e “ESTATURA” à estatura (em metros). Essas informações devem ser inseridas na planilha de “Visualização da variável”. Lembre-se que cada linha na planilha de “Visualização da variável” representa uma variável e, para particularizar uma variável, basta configurar as informações nas colunas.

À medida que as informações são registradas no banco de dados, é importante salvá-las. Para isso, clique em “Arquivo” e, em seguida, em “Salvar como...”. Na janela “Salvar dados como”, se escolhe o caminho desejado para salvar o arquivo e cria-se um nome para o banco de dados, que terá automaticamente a terminação “.sav”. Entretanto, você poderá salvar dados em outros formatos, como o do Microsoft Office Excel. Para isso, clique em “Salvar como tipo” e localize o formato “.xls”.

Quadro 1. Exemplo de banco de dados.

ID	NOME	SEXO*	DN	IDADE	ESTATURA
1	A	1	12/11/1980	39	1,65
2	B	2	06/01/1975	45	1,92
3	C	1	13/07/1988	31	1,54
4	D	2	01/04/2001	19	1,74
5	E	2	13/10/1976	43	1,68
6	F	1	12/09/1978	41	1,72

\*1 = sexo masculino; 2 = sexo feminino.

## Acessando um banco de dados existente

Se você irá trabalhar com um banco de dados existente, é importante que você saiba como abrir esse arquivo de dados no SPSS. O procedimento é bastante simples. Para isso, faça o download do "Banco de dados 1.sav" em seu computador (disponível em <https://bit.ly/bancosdedados>). Este banco de dados nos fornece dados de alunos de uma disciplina. Ao abrir o SPSS, na caixa de diálogo, é possível abrir arquivos existentes que, por padrão, possuem a extensão ".sav", na opção "Abrir uma origem de dados existente". Localize o arquivo "Banco de dados 1.sav" e clique em "Abrir" para abrir o banco de dados. Também pode-se "Colar" o banco de dados, gerando um arquivo de sintaxe, que corresponde ao comando da ação. Para executar o comando, selecione-o no arquivo de sintaxe e clique em "Executar seleção". Caso o SPSS esteja aberto, clique em "Arquivo", "Abrir" e, em seguida, clique em "Dados" para abrir o arquivo "Banco de dados 1.sav". No mesmo caminho é possível abrir arquivos de sintaxe e de saída.

Importante! Ao acessar um banco de dados existente, verifique se as variáveis estão corretamente classificadas antes de iniciar qualquer análise estatística. Observe se as variáveis estão classificadas corretamente e se as variáveis qualitativas possuem os valores dos rótulos atribuídos. Variáveis classificadas incorretamente ou sem seus rótulos identificados, podem impossibilitar as análises e/ou dificultar a sua interpretação.

## O que é um arquivo de sintaxe?

Um arquivo de sintaxe é simplesmente um arquivo de texto que contém uma linguagem de comandos para executar análises estatísticas e manipulação de dados no SPSS, ou seja, é uma criptografia do comando. Você pode abrir uma janela de sintaxe e digitar comandos diretamente, mas é mais fácil deixar que as caixas de diálogo façam uma parte ou todo o trabalho para você. A sintaxe é muito útil quando você ainda está coletando o seu banco de dados ou realizando análises semelhantes muitas vezes. Nessas situações, ao criar e salvar o arquivo de sintaxe, você poderá executá-la futuramente com o banco de dados atualizado, não necessitando executar comando por comando. É interessante, por exemplo, informar a função de cada comando, caso você venha executar diferentes análises. Desta forma, você também pode editar o arquivo de sintaxe para descrever os comandos. Basta clicar em uma das linhas da sintaxe, inserir um asterisco (\*), escrever a sentença e inserir um asterisco novamente para fechar a sentença. Uma vez que os comandos tenham sido gerados ou digitados, é necessário executá-los. Para isso, selecione o comando e clique em "Executar seleção".

Assim como o banco de dados, é possível abrir e salvar um arquivo de sintaxe. Para abrir um arquivo de sintaxe salvo, clique em "Arquivo", "Abrir" e, em seguida, "Sintaxe...". Uma caixa de diálogo padrão para abrir arquivos é exibida. Selecione um arquivo de sintaxe. Se nenhum arquivo de sintaxe for exibido, certifique-se de que a opção "Sintaxe (\*.sps)" esteja selecionada como o "Tipo de arquivo" que você deseja visualizar. Selecione o arquivo e clique em "Abrir". Para salvar o arquivo de sintaxe, clique em "Arquivo", "Salvar como" e escolha o destino para salvá-lo, dando origem a um arquivo do tipo ".sps".

Para exemplificar, no "Banco de dados 1.sav", vamos analisar a frequência da variável "SEXO". No menu "Analisar", em "Estatísticas descritivas", clique em "Frequências...". Selecione a variável "SEXO" e mova-a para a lista "Variável(is)". Clique em "Gráficos". Na caixa de diálogo "Gráficos", selecione "Gráficos de pizza". No grupo "Valores do Gráfico", selecione "Frequências". Clique em "Continuar" e, em seguida, em "Colar" para copiar a

sintaxe criada como resultado das seleções da caixa de diálogo para o arquivo de Sintaxe. Na janela de edição de sintaxe estará o comando para executar a análise estatística solicitada. Para executar a sintaxe, selecione o comando e clique em "Executar seleção". Abrirá uma nova janela contendo o arquivo de saída, ou seja, o resultado da distribuição da frequência de sexo entre os alunos da disciplina.

## Os resultados no visualizador de saídas

O objetivo ao executar o arquivo de sintaxe é gerar um arquivo de saída, que trará os gráficos, tabelas e resultados da análise estatística. Portanto, os resultados da análise estatística aparecem em uma nova janela de resultados, que poderá ser salva, dando origem a um arquivo do tipo ".spv". Também podemos editar o arquivo de saída. Para isso, clique em "Inserir" e, em seguida, em "Novo texto".

É possível também editar os gráficos no arquivo de saída. Para isso, basta dar um duplo clique naquele que se deseja editar. No lado esquerdo da janela de saída existe um diagrama de árvore ilustrando a estrutura da saída. Esse diagrama é útil quando são realizadas inúmeras análises, pois nos fornece uma forma fácil e rápida de acessar partes específicas da saída. A estrutura de árvore é praticamente autoexplicativa, listando cada procedimento executado como um cabeçalho principal.

## Executando uma análise

Da mesma forma que precisamos conhecer o banco de dados no Microsoft Office Excel, precisamos também conhecer no SPSS. Tome algum tempo para examinar o banco de dados e explorar os dados. Uma boa forma de checar o banco de dados é realizar algumas estatísticas descritivas e construir alguns gráficos.

## Estatísticas descritivas no SPSS

Uma boa forma de explorar os dados é realizar uma análise descritiva. Existem duas formas de realizá-la: 1) utilizando o "Banco de dados 1.sav", clique em "Analisar", "Estatísticas descritivas" e, em seguida, em "Frequências...". Localize no painel da esquerda a variável de interesse (por exemplo, "IDADE") e clique na seta para mover para a lista de "Variável(is)". Qualquer análise escolhida será realizada em todas as variáveis listadas na lista de "Variável(is)". Clique em "Estatísticas...". Como a idade é uma variável quantitativa, iremos assinalar as medidas de tendência central ("Média", "Mediana" e "Moda"), de dispersão ("Desvio padrão", "Variância", valores "Mínimo" e "Máximo") de variabilidade ("Quartis") e de forma ("Assimetria" e "Curtose"). Clique em "Continuar" e, após, em "OK" para gerar o arquivo de saída, ou em "Colar" para gerar o arquivo de sintaxe. Os resultados da análise estatística serão apresentados na janela de resultados (arquivo de saída); 2) clique em "Analisar", "Estatísticas descritivas" e, em seguida, em "Descritivos...". Localize no painel da esquerda a variável de interesse (por exemplo, "IDADE") e clique na seta para mover para a lista de "Variável(is)". Clique em "Opções...". Assinale as mesmas medidas de análise ("Média", "Desvio padrão", "Variância", "Mínimo", "Máximo", "Curtose" e "Assimetria"). Clique em "Continuar" e, após, em "OK" para gerar o arquivo de saída, ou em "Colar" para gerar o arquivo de sintaxe. Os resultados da análise estatística serão apresentados na janela de resultados (arquivo de saída).

Agora, com os resultados obtidos, é preciso observá-los e interpretá-los. Primeiro, observe o número de casos válidos e omissos (*missing value*). Em seguida, observe os valores máximo e mínimo. Verifique se há alguma inconsistência nos valores encontrados. Essa inspeção visual é importante para localizarmos possíveis erros de digitação no banco de dados. Compare também os valores de média, mediana e moda obtidos. Note que o SPSS nos indica quando há mais de uma moda na amostra, enquanto que o Microsoft Office Excel nos fornece apenas um valor de moda. Para verificar se uma distribuição é normal, podemos olhar para os valores da assimetria e da curtose (os valores da assimetria e da curtose deverão ser zero em uma distribuição normal), comparar os

valores de média e mediana ou visualizar graficamente a distribuição dos dados através de um histograma. Mas, como mencionado no Capítulo 1 e, abordado no Capítulo 7, existem testes que avaliam de forma mais confiável a normalidade de distribuição dos dados.

Abaixo da tabela de análise estatística, também é apresentada uma tabela com as frequências das idades. Para isso, na opção “Analisar”, “Estatísticas descritivas”, “Frequências...”, deve-se selecionar “Exibir tabelas de frequências”. Por padrão, o SPSS produz uma distribuição de frequências de todos os escores no formato de tabela.

Após observar minuciosamente os dados, é hora de interpretá-los. Se considerarmos uma distribuição normal dos dados, utilizaremos os valores de média e desvio padrão; assim, a interpretação dos dados seria que os alunos da disciplina apresentam idade média de 30 anos, ou então,  $30 \pm 6$  anos de idade. Se assumirmos uma distribuição não normal dos dados, utilizaremos os valores de mediana e quartis ou de mediana e valores mínimo e máximo; assim, a interpretação dos dados seria que os alunos da disciplina apresentam mediana de idade de 29 (25-33) anos ou de 29 (21-55) anos.

## Representação gráfica das análises descritivas

Podemos também apresentar os resultados das análises descritivas através de gráficos. O SPSS tem recursos para construir diferentes tipos de gráficos. Um destes gráficos, muito utilizado é o histograma. Um histograma apresenta barras ao longo de uma escala de mesmo intervalo e a altura de cada barra é a contagem de valores de uma variável quantitativa que corresponde ao intervalo.

Existem duas formas de obtermos o histograma: 1) em “Analisar”, “Estatísticas descritivas”, “Frequências”, ao solicitarmos a estatística descritiva (por exemplo, para a variável “IDADE”), clique na opção “Gráficos...”. Neste caso, como temos uma variável quantitativa, podemos construir um “Histograma”. Lembre-se que o histograma é um gráfico de frequência que tem como objetivo ilustrar como uma determinada amostra ou população de dados está distribuída. Na mesma janela, selecione

a opção "Mostrar curva de normalidade no histograma". Uma curva normal sobreposta em um histograma ajuda a avaliar se os dados são normalmente distribuídos. Clique em "Continuar" e, em seguida, em "OK" para gerar o arquivo de saída, ou em "Colar" para gerar o arquivo de sintaxe; 2) na janela de edição de dados, clique em "Gráficos" e, em seguida, em "Criador de gráfico...". Clique em "Ok". Na janela "Criador de gráfico", na "Galeria", existem diferentes tipos de gráficos. Selecione a opção do "Histograma" e, escolha o primeiro modelo gráfico, arrastando-o para o local indicado (área grande acima da "Galeria"). Use o modelo como seu ponto inicial. A visualização prévia do gráfico usa dados de exemplo. Para definir as variáveis dos eixos "X" e "Y", selecione a variável (por exemplo, "IDADE") em "Variáveis" e arraste-a para o eixo "X". Adicionalmente, você pode adicionar um título ao gráfico, clicando em "Títulos/Notas de rodapé". Basta inseri-lo em "Conteúdo" e clicar em "Aplicar". Por fim, clique em "Ok" para gerar o arquivo de saída ou em "Colar" para gerar o arquivo de sintaxe.

Em seguida, verifique no arquivo de saída o gráfico gerado. É possível editar o gráfico, clicando duas vezes sobre o gráfico no arquivo de saída. Na janela "Editor de gráfico", você poderá fazer edições no gráfico. Clicando com o botão direito do mouse, você pode acrescentar a curva de distribuição dos dados, clicando em "Mostrar curva de distribuição". Além disso, é possível alterar o nome dos rótulos ("X" e "Y") e formatar a área do gráfico, inclusive as cores e o plano de fundo.

Outra forma de representar os resultados é através de um gráfico do tipo *box plot*. Através deste recurso gráfico é possível observar como as variáveis estão distribuídas em relação à homogeneidade dos dados, valores de tendência central (Q1: primeiro quartil, Q2: mediana e Q3: terceiro quartil), valores máximo e mínimo e valores atípicos (*outlier*), se existirem. Lembre-se que os quartis dividem as observações em quatro grupos de tamanhos iguais. Quando a caixinha (*box*) é muito pequena, significa que os dados estão muito concentrados em torno da mediana, e se a caixinha for grande, significa que os dados são mais heterogêneos.

Para isso, clique em “Gráficos” e, em seguida, em “Criador de gráfico...”. É possível redefinir o tipo de gráfico clicando em “Redefinir”. Na galeria, selecione a opção “Diagrama em caixa”. Em seguida, selecione o primeiro modelo de gráfico e arraste para o local indicado. Selecione a variável (por exemplo, “IDADE”) e arraste para o eixo “Y”.

Clique em “Ok” para gerar o arquivo de saída ou em “Colar” para gerar o arquivo de sintaxe. Verifique o arquivo de saída. Há um valor atípico representado pelo número “100”. Isso indica que há um caso, que corresponde ao caso 100, que se diferencia de todos os outros em relação à variável “IDADE”.

### **Outliers: o que fazer?**

Valores atípicos distorcem a média e inflacionam o desvio padrão, sendo necessária uma análise criteriosa dos dados para detectá-los. Entender os *outliers* é fundamental em uma análise de dados por pelo menos dois aspectos: 1) os *outliers* podem viesar negativamente todo o resultado de uma análise e 2) o comportamento dos *outliers* pode representar justamente o que está sendo procurado. Mas, o que fazer com os *outliers*? Pode-se excluir as observações discrepantes da amostra de dados, quando o dado discrepante é fruto de um erro de imputação de dados (por isso é importante verificar os valores mínimo e máximo), então ele precisa ser removido ou corrigido da amostra; recomenda-se realizar a análise com e sem o *outlier* para ver seu impacto sobre os resultados e; realize, se possível, uma análise separada apenas com os *outliers*: esta abordagem é útil quando se quer investigar casos extremos.

### **Distribuição de frequências no SPSS**

Como já mencionado, por padrão, o SPSS produz uma distribuição de frequências de todos os escores no formato de tabela. Essa distribuição de frequências fornece exibições estatísticas e gráficas que são úteis para descrever as variáveis, incluindo variáveis quantitativas (escala) e qualitativas (nominal ou ordinal). As tabulações e porcentagens fornecem uma descrição útil dos dados, especialmente para as variáveis com categorias. Lembre-se que é necessário utilizar códigos numéricos para codificar variáveis categóricas.

Para isso, na janela de edição de dados do "Banco de dados 1.sav", clique em "Analisar", "Estatísticas Descritivas" e, em seguida, em "Frequências...". É possível redefinir as opções de teste clicando em "Redefinir". Localize no painel da esquerda a variável de interesse (por exemplo, "SEXO") e clique na seta para mover para a lista "Variável(is)". Mantenha somente a opção "Exibir tabelas de frequência" selecionada. Clique em "Ok" para gerar o arquivo de saída ou em "Colar" para gerar o arquivo de sintaxe. Verifique no arquivo de saída a análise de frequência gerada.

Na primeira tabela, temos o número de casos válidos e omissos (*missing value*). Na segunda tabela, constam as frequências absoluta e relativa (percentual), além das frequências percentuais válida e cumulativa para as categorias da variável "SEXO" e o total da variável. Observados os resultados, podemos interpretá-los: dos 115 alunos da disciplina, 82,6% ( $n = 95$ ) são do sexo feminino

## Representação gráfica das frequências

Gráficos são uma boa forma de apresentar, ilustrar e resumir resultados. No entanto, lembre-se que a forma de apresentar as variáveis qualitativas é através de gráficos em barra ou de pizza. Um gráfico de pizza exhibe a contribuição de partes para um todo e cada fatia corresponde a um grupo (categoria) que é definido por uma variável de agrupamento. Um gráfico de barras exhibe a contagem para cada valor ou categoria distinta como uma barra separada, permitindo comparar categorias visualmente.

Da mesma forma que obtivemos os gráficos para a variável quantitativa ("IDADE"), existem duas formas de obtermos os gráficos para as variáveis qualitativas, correspondendo aos mesmos caminhos. Na primeira forma de obtermos o gráfico, seguindo com a variável "SEXO", é através da opção "Analisar", "Estatísticas Descritivas", "Frequências...". Na opção "Gráficos", é possível selecionar "Gráfico de barras" ou "Gráfico de pizza", expressando os valores em "Frequências" ou "Porcentagens". Clique em "Continuar", e em seguida em "Ok" para gerar o arquivo de saída ou em "Colar" para gerar o arquivo de sintaxe.

Na segunda opção para obtermos o gráfico, na janela de edição de dados, clique em "Gráficos" e, em seguida, em

“Criador de gráfico...”. Clique em “Ok”. Clique em “Redefinir”, se necessário. Na galeria, selecione a opção “Barra” e selecione o primeiro modelo de gráfico, arrastando-o para o local indicado. Selecione a variável (por exemplo, “SEXO”) e arraste-a para o eixo “X”. Na opção “Propriedades do elemento...”, em “Estatístico”, selecione a opção “Contagem” para expressar o valor absoluto ou “Porcentagem” para expressar o valor percentual. Clique em “Aplicar”. Ao final, clique em “Ok” para gerar o arquivo de saída ou em “Colar” para gerar o arquivo de sintaxe. Para alterar o tipo de gráfico, como por exemplo, um gráfico de “Pizza/Polár”, basta “Redefinir” as opções no “Criador de gráfico” e selecionar a nova opção gráfica. Da mesma forma, no arquivo de saída, é possível editar o gráfico clicando duas vezes sobre o gráfico.

## Exportação do arquivo de saída para Word

É possível exportar as tabelas e gráficos gerados no arquivo de saída para o Word, além da opção de copiar e colar. Para isso, no arquivo de saída, clique em “Arquivo” e, em seguida, em “Exportar”. Na janela “Exportar saída”, em “Objetos a exportar”, é possível escolher exportar tudo, apenas uma seleção ou tudo que está visível. Em “Documento”, “Tipo”, escolha o tipo de arquivo para o qual se deseja exportar os resultados (neste caso, no formato Word). Em “Nome do arquivo”, digite o local e o nome do arquivo que será salvo. Clique em “Ok” para gerar o arquivo de saída ou em “Colar” para gerar o arquivo de sintaxe. Pronto! O arquivo com as análises estatísticas será gerado no Word.

## Exportando um banco de dados do SPSS para o Excel

Para exportar um banco de dados do SPSS para o Excel, clique em “Arquivo” e, em seguida, em “Salvar como...”. Na janela “Salvar dados como”, crie um nome para o banco de dados, em “Nome do arquivo” e, em “Salvar como tipo”, escolha “Excel (.xls)”. Clique em “Salvar”. Um novo arquivo com a terminação “.xls” será salvo.

## Importando um banco de dados do Excel para o SPSS

Antes de importar um banco de dados do Microsoft Office Excel para o SPSS, é importante revisarmos alguns aspectos:

lembre-se que cada linha corresponde a um caso e que cada coluna corresponde a uma variável; o nome da variável pode ter acento, mas não pode conter espaços (pode-se substituir por “\_”) e deve conter 8 dígitos no máximo; o SPSS não entende as variáveis enquanto nomes, devendo ser substituídas por códigos numéricos. Como visto no Capítulo 2, é possível substituir as categorias da variável por códigos utilizando a ferramenta “Localizar e selecionar” do Microsoft Office Excel.

Para exemplificar a importação do banco de dados, vamos utilizar o “Banco de dados 1.xls” (disponível em <https://bit.ly/bancosdedados>). Ao abrir o SPSS, na caixa de diálogo, é possível abrir arquivos existentes na opção “Abrir uma origem de dados existente”. Na janela seguinte (“Abrir dados”), é possível selecionar em “Arquivos do tipo” a terminação “Excel (\*.xls, \*.xlsx, \*.xlsm)” e importar bancos de dados nestes formatos. Selecione o arquivo “Banco de dados 1.xls” e clique em “Abrir”. Na janela “Abrindo origem de dados do Excel”, caso a primeira linha da planilha do Excel corresponda aos nomes das variáveis (como é o caso), é importante você indicar a opção “Ler nomes de variável a partir da primeira linha de dados”. Após, clique em “Ok”. Pronto! O banco de dados será importado do Microsoft Office Excel para o SPSS. Na mesma janela, é possível selecionar ainda a aba da planilha do Excel que se deseja importar (caso haja mais de uma) e qual a faixa de valores dessa planilha (intervalo de casos), indicando em “Planilha” e “Faixa”, respectivamente. Caso o SPSS esteja aberto, esse mesmo processo de importação pode ser feito clicando em “Arquivo”, “Abrir” e, em seguida, em “Dados”, seguindo os mesmos passos descritos.

Após a importação dos dados, verifique a organização e a forma de apresentação dos dados. Note que na planilha “Visualização da variável” é necessário preencher as informações das colunas referentes às variáveis, como tipo de variável, número de casas decimais após a vírgula, descrição do rótulo e valores dos rótulos. Lembre-se também de verificar se a variável está classificada corretamente (ordinal, nominal ou escala).

Deixamos aqui uma sugestão. Como mencionado no Capítulo 2, o Excel é um software de planilha, enquanto o SPSS

é um software de análise estatística. Desta forma, é indicado que você construa o banco de dados diretamente no SPSS, pois 1) evita erros de importação, 2) considere seu tempo precioso (com a importação será necessário revisar e especificar todas as informações das variáveis) e, além disso, 3) você irá utilizar o SPSS de qualquer forma para análise dos dados.

## Referências

IBM®. IBM SPSS software. Disponível em <https://www.ibm.com/analytics/spss-statistics-software>.

Field A. Descobrimos a estatística usando o SPSS. Tradução: Lorí Viali. 2. ed. Porto Alegre: Artmed, 2009. 684 p.

## Exercícios sugeridos

1. Este primeiro exercício propõe, com base nas variáveis do seu projeto de pesquisa descritas no exercício 9 do Capítulo 1, a elaboração do banco de dados no SPSS. Este é um bom exercício prático para você se familiarizar e conhecer o software estatístico.

2. Com relação ao “Banco de dados 1.sav”, que fornece informações de alunos de uma disciplina:

- a) Qual a frequência do nível mestrado entre os alunos da disciplina?
- b) Dentre os programas de pós-graduação, qual(is) o(s) mais frequente(s) entre os alunos da disciplina?
- c) Como é a distribuição do conhecimento dos alunos em estatística? E em epidemiologia?
- d) A maioria dos alunos tem experiência no uso do SPSS?

3. No “Banco de dados 2.sav”, que traz dados antropométricos e de pressão arterial de sujeitos com obesidade, realize a análise descritiva para as variáveis idade, Índice de Massa Corporal (IMC), circunferência da cintura e medidas de pressão arterial. Construa os histogramas e interprete os resultados obtidos supondo distribuição não normal dos dados.

4. No “Banco de dados 2.sav”, realize a análise de frequências para as variáveis grau de obesidade e hipertensão. Construa os gráficos e interprete os resultados.