jamovi Stats. Open Now.

Apoio ao uso do software Jamovi

(versão 2.3.28)

Matheus Souza Ferreira
Breno Augusto Gonçalves
Gizelton Pereira Alencar
Denise Pimentel Bergamaschi

USP

jamovi Stats. Open Now.

Apoio ao uso do software Jamovi

(versão 2.3.28)

DOI 10.11606/9786588304235

Matheus Souza Ferreira
Breno Augusto Gonçalves
Gizelton Pereira Alencar
Denise Pimentel Bergamaschi

Universidade de São Paulo
Faculdade de Saúde Pública
São Paulo

2025

Autores
Matheus Souza Ferreira
Breno Augusto Gonçalves
Gizelton Pereira Alencar
Denise Pimentel Bergamaschi

Produção e Realização
Programa de Pós-Graduação em Saúde Pública

Apoio
O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Código de Financiamento 001.

Produção Editorial
Edu Ambiental Serviços
Soraia Fernandes

Apoio técnico:
Equipe da Biblioteca da Faculdade de Saúde Pública da USP
Av. Dr. Arnaldo, 715
01246-904 – Cerqueira César – São Paulo – SP
http://www.biblioteca.fsp.usp.br
markt@fsp.usp.br

Catalogação na Publicação
Universidade de São Paulo. Faculdade de Saúde Pública

Apoio ao uso do software Jamovi (versão 2.3.28) [recurso eletrônico] / Matheus Souza Ferreira ... [et al.]. -- São Paulo : Faculdade de Saúde Pública da USP, 2025.
83 p. : il. color. PDF

ISBN 978-65-88304-23-5 (eletrônico)
DOI 10.11606/9786588304235

1. Análise Estatística de Dados. 2. Bioestatística. 3. Softwares. I. Ferreira, Matheus Souza. II. Gonçalves, Breno Augusto. III. Alencar, Gizelton Pereira. IV. Bergamaschi, Denise Pimentel.

CDD 570.15195

Elaborada por Alice Mari Miyazaki de Souza - CRB 8/6238

SUMÁRIO

APRESENTAÇÃO

Este documento constitui um trabalho coletivo da equipe de organização da disciplina HEP0184 Bioestatística II - 2023, obrigatória, do curso de graduação em Saúde Pública da Faculdade de Saúde Pública da Universidade de São Paulo e visa apresentar um material didático orientador para apoiar usuários do pacote de análise estatística de dados **Jamovi**. Constituem potenciais usuários, estudantes de cursos de graduação, pós-graduação e pesquisadores que necessitem realizar análise de dados estatísticos.

É esperado que os usuários contribuam com a melhoria do material, e, para tanto, comentários e sugestões serão muito bem-vindos. Nesses casos, solicita-se o uso dos endereços eletrônicos dos docentes envolvidos no projeto: Profa Denise Pimentel Bergamaschi (denisepb@usp.br) e Prof Gizelton Pereira Alencar (gizelton@usp.br) ou do aluno de doutorado Matheus Souza Ferreira (matheus.souza.ferreira.sp@gmail.com), que disponibilizou códigos e bancos de dados usados neste material.

O software **Jamovi** (https://www.jamovi.org) é uma plataforma de análise estatística de dados, de código aberto e gratuita, que oferece uma interface acessível e amigável para a análise de dados e é disponível para diversos sistemas operacionais. É um pacote que pode atender tanto usuários experientes quanto iniciantes na área de análise estatística. Ressalta-se, entretanto, a necessidade de conhecimentos sobre estatística aplicada para que a utilização do pacote seja adequada.

Sua interface gráfica intuitiva contém várias técnicas e recursos de análise o que tem impulsionado sua aplicação, principalmente nos últimos anos, no ensino nas disciplinas na área de saúde, especialmente por não serem necessários conhecimentos prévios em linguagens e lógica de programação para seu uso.

Este material contempla os principais tópicos relacionados ao uso da estatística para análises em saúde. Entretanto, o software **Jamovi** apresenta uma série de recursos que podem ser empregados para outras aplicações e sugerimos ao usuário consultar também os diferentes materiais e mídias disponíveis na página do projeto (https://www.jamovi.org).

1. INSTALAÇÃO DO SOFTWARE

O instalador do **Jamovi** pode ser obtido diretamente em sua página oficial https://www.jamovi.org/download.html, por meio do recurso "*Download*". Recomenda-se a escolha do pacote "*solid*", sendo que, nesta primeira versão deste material, será utilizada a versão 2.3.28. Para instalação é necessário escolher a versão apropriada para o sistema operacional do hardware (Figura 1).

Figura 1. Página eletrônica para fazer o download e utilizar o instalador do Jamovi

Em seguida, deve-se executar o instalador que foi carregado no computador (Figura 2).

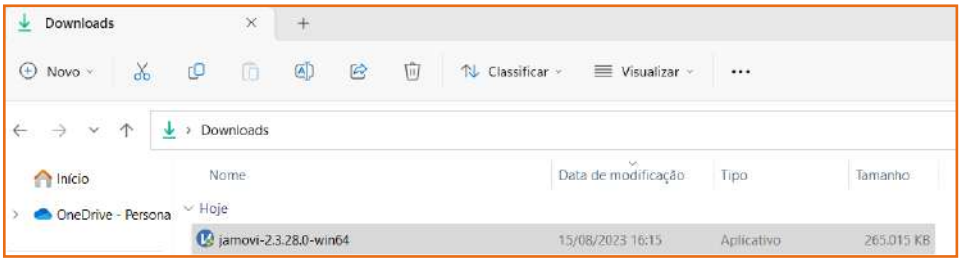


Figura 2. Instalador do Jamovi carregado em computador com sistema operacional Windows 11

Após a instalação, ao abrir o **Jamovi** no computador, observa-se a tela apresentada na Figura 3.

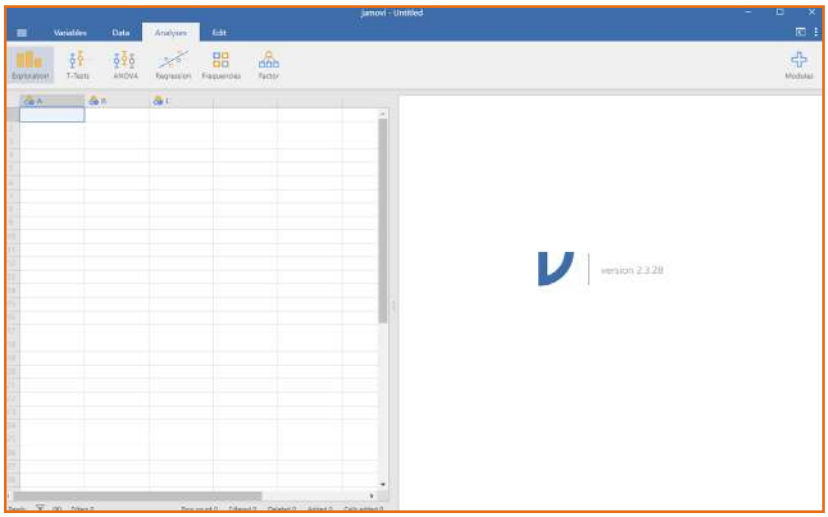


Figura 3. Tela inicial do Jamovi

Note que o **Jamovi** apresenta um conjunto de funcionalidades padrão e que módulos/pacotes específicos podem ser carregados na sessão "*Modules*" cujo ícone é , de acordo com a necessidade do usuário (Figura 4).

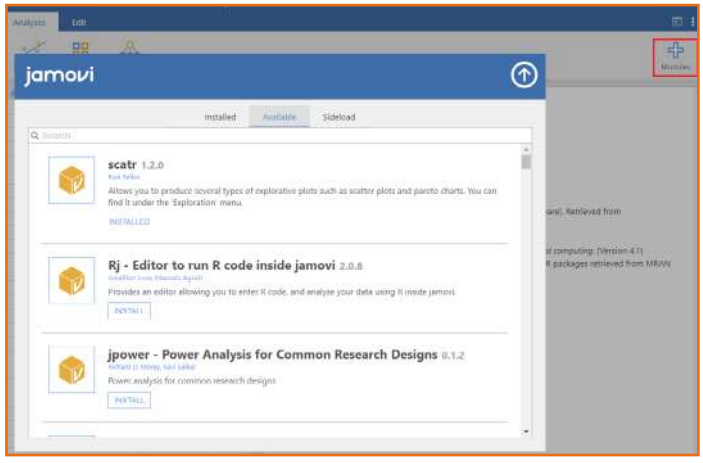


Figura 4. Módulos adicionais que podem ser carregados à sessão de trabalho no Jamovi

Na Figura 5, notar que estão disponíveis os Módulos Exploration, T-tests, ANOVA, Regression, Frequencies.

Figura 5 - Alguns Módulos instalados

2. ENTRADA DE DADOS

2.1. Entrada de dados pelo Jamovi

No **Jamovi** é possível carregar bases de dados em formato relacional – linhas contendo os registros, sendo que cada linha corresponde a uma observação da unidade de análise (e.g. indivíduo, unidade da federação, etc.) e cada coluna, uma variável.

Ao escolher o ícone  , a tela (Figura 6) apresentará as opções **New**, **Open**, **Import**, **Save**, **Save as** e **Export**.

Figura 6. Opções para manuseio de arquivos

Na opção **New** será possível digitar os dados diretamente na planilha do **Jamovi** e, ao salvar, a extensão do arquivo será .omv; na opção **Open**, é possível abrir um arquivo .omv ou um arquivo salvo em .csv (valores separados por vírgulas) que esteja disponível em alguma pasta de uma unidade do computador. Na opção **Import** é possível trazer os arquivos com extensão .xlsx (salvo em csv) ou .omv que também estejam armazenados.

Na opção **Export** é possível exportar os dados (banco de dados) em arquivos PDF Document (.pdf); Web Page (.html; .htm); **Jamovi** template (.omv); CSV (Comma delimeted) (.csv); LaTeX bundle (.zip); R object (.rds); R object (.RData); SPSS sav (.sav); SAS 7bdat (.sas7bdat); SAS xpt (.xpt) e Stata (.dta) (Figura 7).

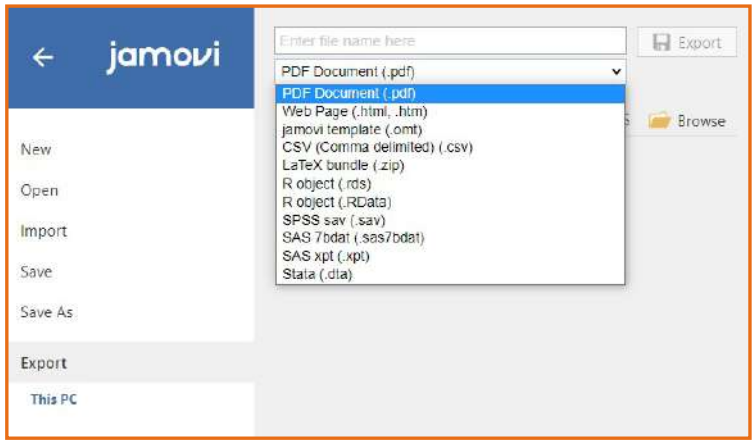


Figura 7. Opções para exportar banco de dados

2.2. Entrada de dados com extração de outros repositórios

No item 9. Saiba Mais é possível encontrar as informações de como obter uma base de dados de interesse. Será utilizado como exemplo a extração dados do DATASUS referente ao Sistema de Informação de Nascidos Vivos (SINASC) de uma localidade em um período, utilizando o script em linguagem R para retirada dos dados brutos por meio do pacote "microdatasus".

3. MANIPULAÇÃO DE DADOS

3.1. Entrada e salvamento de dados

Na tela '*Data*', verifica-se, por padrão, três colunas (variáveis vazias) "A", "B" e "C", indicando onde será armazenado o nome da variável (Figura 8).

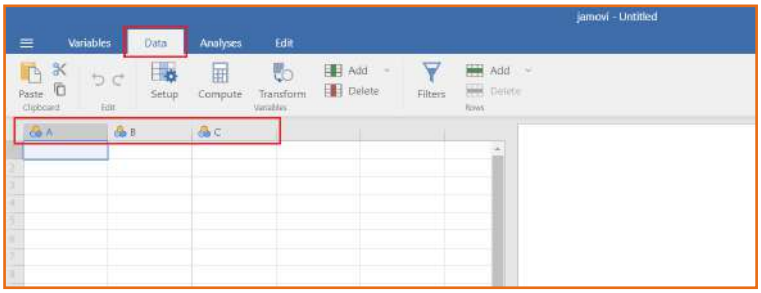


Figura 8. Tela 'Data' com formato de base de dados relacional

Para iniciar o exemplo, iremos digitar os dados apresentados no Quadro 1 ilustrando uma situação fictícia em que se obteve a idade, o peso e a altura de um grupo de usuários de um determinado serviço de saúde.

Quadro 1. Idade, peso e altura de 10 usuários de um serviço de saúde

Número do usuário	Idade (em anos)	Peso (kg)	Altura (cm)
1	24	60	147
2	31	50	159
3	35	54	154
4	18	71	155
5	32	89	163
6	33	68	171
7	27	60	181
8	27	52	157
9	27	53	181
10	26	81	153

Digitar os dados na planilha do **Jamovi** (Figura 9), salvar o banco de dados (Figura 10) e alterar o nome e tipo de cada variável.

Salvar o banco de dados como banco1.omv utilizando-se o ícone  e, em seguida, escolher “guardar como” e guardar o arquivo na pasta de sua preferência (Figura 9).

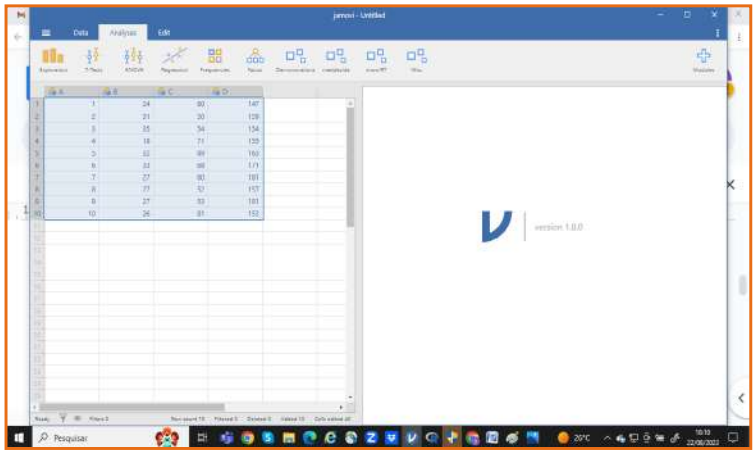


Figura 9. Dados inseridos na planilha (digitados ou copiados)

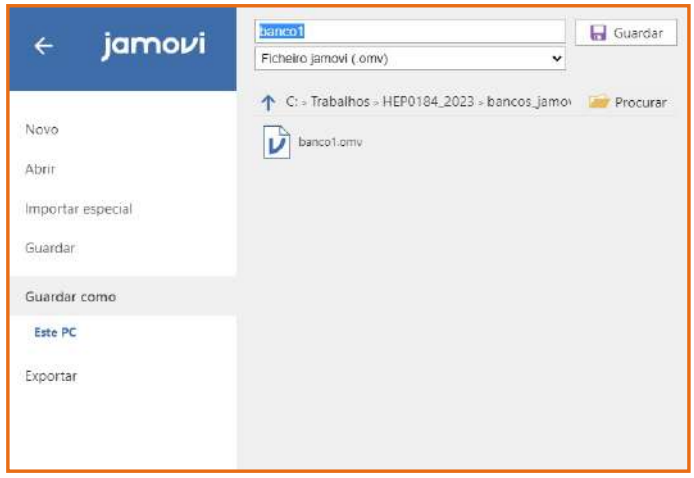


Figura 10. Salvando um arquivo com extensão .omv (digitados ou copiados)

Para editar (inserir o nome ou definir o tipo da variável), posicionar o cursor sobre a primeira célula da coluna e clicar duas vezes (Figuras 11 e 12).

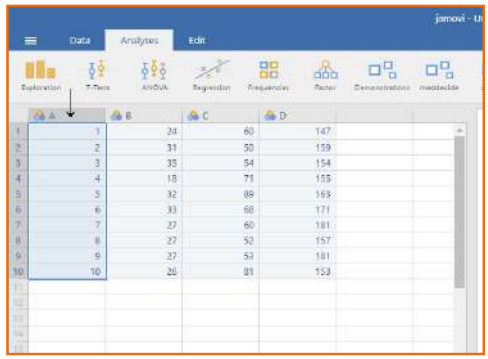


Figura 11. Selecionado uma coluna

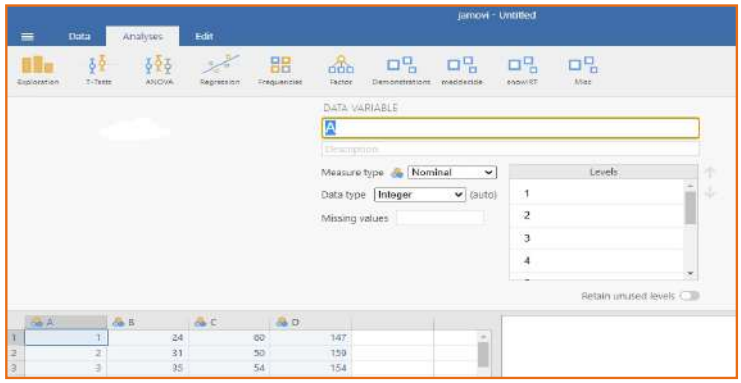


Figura 12. Definindo nome e tipo de variável

Na visualização que se abre é possível definir nome e tipo da variável:

- No primeiro campo é possível alterar o nome da variável de "A" para "id_usuario"; de "B" para "idade_anos" e de "C" para "peso_kg".
- No campo "*Description*" é possível fazer uma descrição da variável como documentação. Manter um dicionário de variáveis é uma boa prática de manejo de bases de dados.
- No campo "*Measure type*" é necessário especificar o tipo da variável se é nominal (categórica), ordinal (categórica) ou contínua (numérica); o padrão (*default*) deste campo é *"Nominal"*, mas, a não ser que a variável seja categórica ordinal, não é preciso preencher o campo porque será definido automaticamente ao se preencher o campo "*Data type*".
- O campo "*Data type*" define o formato do dado (valor inteiro "*Integer*", "*Decimal*", ou de texto "*Text*"). Caso definido como "*Integer*" ou "*De-*

cimal", o campo "*Measure type*" automaticamente será definido como "*Continuous*".

- Em "*missing values*", é possível definir os códigos para valores faltantes ou ignorados, por exemplo, em branco, texto como "NA", codificado como "999", etc.

Na Figura 13 é possível visualizar as primeiras três variáveis do Quadro 1 inseridas no **Jamovi** e as informações da variável "peso_kg" correspondente à antiga variável "C".

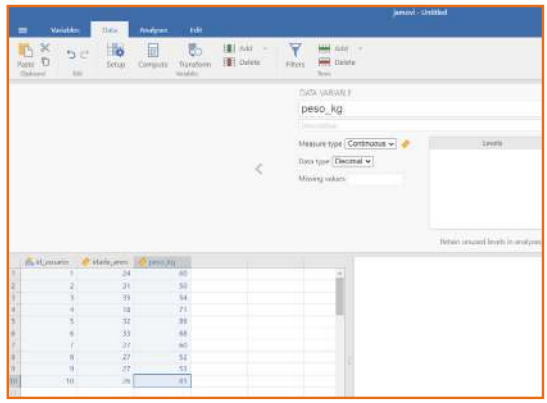


Figura 13. Inserindo dados diretamente no Jamovi e modificando as informações das variáveis padrão "A", "B" e "C"

Para inserir a variável "altura cm" (cm) clicar na próxima coluna que está vazia ao lado de "peso kg" e a seguinte tela se abrirá (Figura 14).

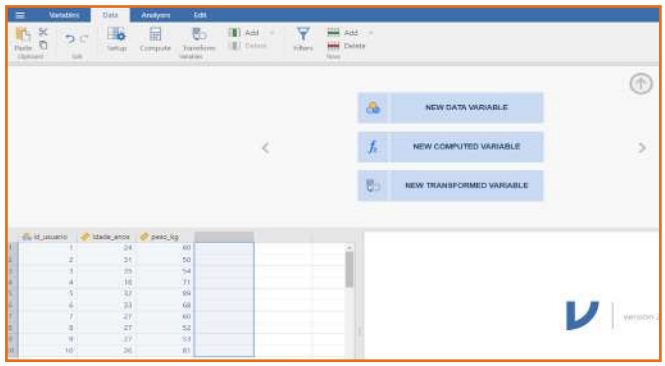


Figura 14. Opções ao se criar uma variável diretamente no Jamovi

Os dados da variável altura serão inseridos diretamente, assim, é necessário selecionar a opção "*new data variable*". A Figura 15 apresenta a base de dados completa conforme o Quadro 1 e as informações da variável "altura_cm".

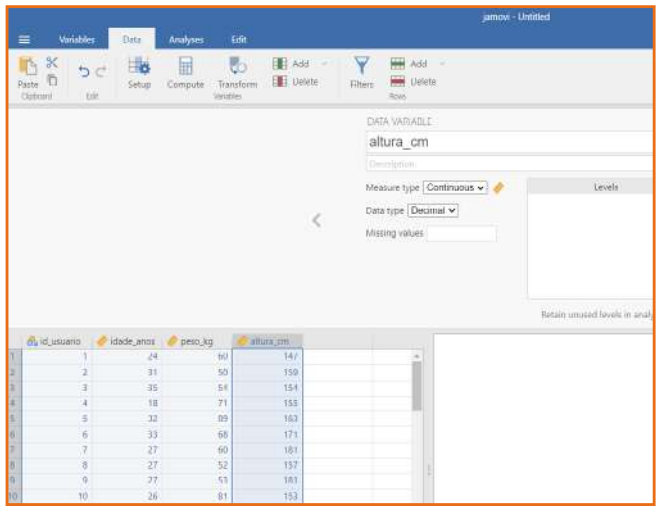


Figura 15. Base de dados inseridos no Jamovi e informações da variável "altura_cm"

3.2. Transformação de variáveis

É possível criar variáveis a partir de expressões e operações envolvendo variáveis existentes na base de dados, como, por exemplo, o cálculo do índice de massa corpórea (IMC), obtido por meio da divisão do peso (individual) em quilos pela altura (individual) em metros, elevada ao quadrado, a partir das variáveis peso e altura existentes na base de dados.

Na tela de criação de nova variável (Figura 14) deve-se optar por "*new computed variable*", que permite a criação de variáveis a partir de operações matemáticas realizadas com variáveis existentes no banco de dados.

Na Figura 16 observamos no painel "*Functions*" que várias operações matemáticas podem ser realizadas, como a obtenção da média ("`mean`"), extrair a raiz quadrada ("`sqrt`"), obter o log na base 10 ("`log10`"), entre outras, e, ao lado pode-se visualizar o painel "*Variables*", trazendo as variáveis existentes na base de dados.

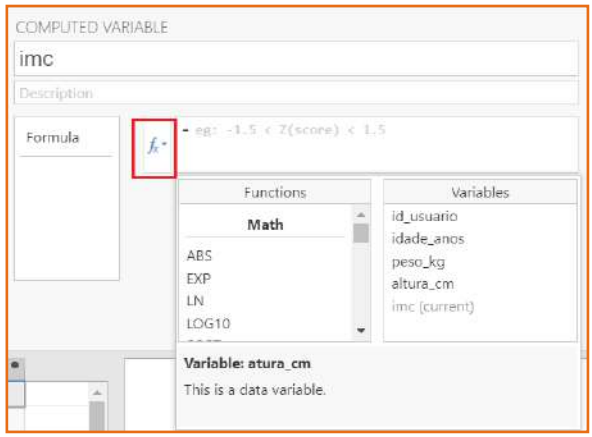


Figura 16. Transformação de variáveis no Jamovi por meio de operações e funções

Para o cálculo do IMC, é necessário inserir a fórmula "`= peso_kg / (altura_cm/100)^2`", conforme a Figura 17. Note que para incluir o nome de uma variável da base no campo de fórmula, basta clicar duas vezes na mesma no painel "*Variables*". A "altura_cm" é dividida por 100 para convertê-la de centímetros, em metros, segundo definição do IMC.

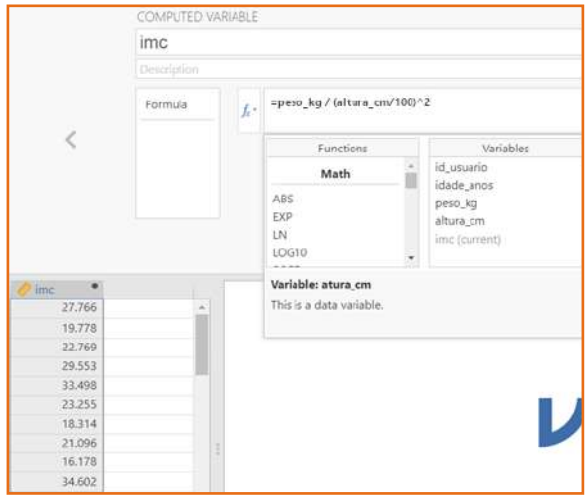


Figura 17. Expressão para criação da variável IMC

É possível criar uma variável categórica nominal ou ordinal a partir da criação de categorias obtidas pela aplicação de condições ou critérios. Deseja-se categorizar a variável IMC em categorias "baixo peso", "eutrófico",

“sobrepeso” e “obesidade”, a partir da classificação utilizada para adultos (SISVAN, 2024), criando-se uma variável categórica ordinal.

Segundo a referência, os critérios para categorizar a variável IMC são:

(categoria 1) baixo peso: adultos com valores de IMC abaixo de 18,5 kg/m^2

(categoria 2) eutrófico: adultos com valores de IMC maiores ou iguais a 18,5 kg/m^2 e menores que 25,0 kg/m^2

(categoria 3) sobrepeso: adultos com valores de IMC maiores ou iguais a 25,0 kg/m^2 e menores que 30,0 kg/m^2

(categoria 4) obesidade: adultos com valores de IMC maiores ou iguais a 30,0 kg/m^2

A criação da nova variável pode ser realizada ao se acessar o painel apresentado na Figura 14 e selecionar a opção “*new transformed variable*”.

Em seguida irá aparecer a tela apresentada na Figura 18 referente à criação de variáveis categóricas a partir de variáveis existentes na base de dados.

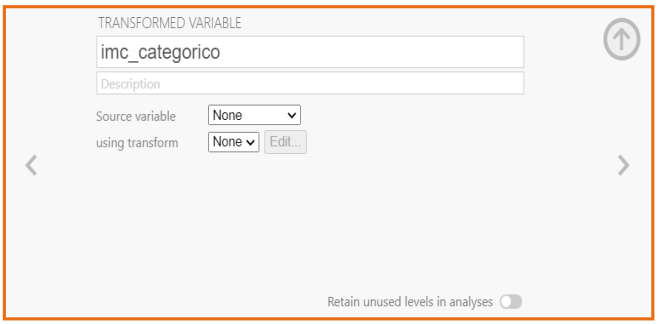


Figura 18. Tela do Jamovi referente à transformação de variáveis

Para definir as condições envolvendo a classificação de um indivíduo em uma categoria a partir do IMC, primeiramente seleciona-se a variável “imc” no campo “*Source variable*” (Figura 19).

Figura 19. Variável fonte que permitirá a criação de outra variável com categorias do IMC

Em seguida, em "*using transform*", selecionar "*Create New Transform...*" para criação de condições que estabelecerão como serão categorizados os indivíduos da base de dados a partir da "*source variable*" (IMC) (Figura 20).

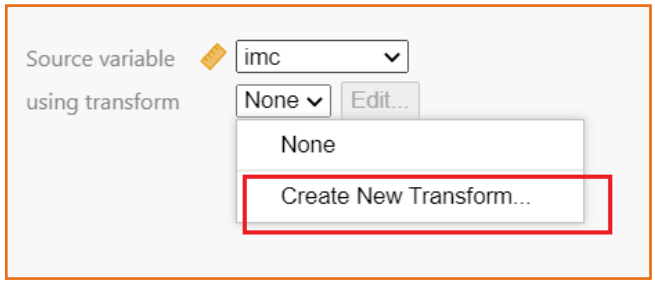


Figura 20. Definição de condições para categorização da variável fonte (IMC)

Na nova tela que se abre observamos um campo vazio onde será digitada a condição que se deseja (Figura 21).

Figura 21. Campo de nova condição durante criação de variáveis categóricas

Aqui, deve-se incluir as condições para criação das categorias. No exemplo, pretende-se criar a variável categórica ordinal com quatro categorias "baixo_peso" (IMC<18,5 kg/m^2), "eutrófico" (18,5kg/m^2 IMC<25,0 kg/m^2), "sobrepeso" (25,0 kg/m^{2}IMC<30 kg/m^2) e "obeso" (IMC 30 kg/m^2).

Para definir essas regras deve-se utilizar os operadores lógicos e as fórmulas.

condição 1 (fórmula 1): `if $source < 18.5 use 'baixo_peso'`

A fórmula significa: se a variável fonte (imc) for < (menor) que 18,5 criar categoria 'baixo_peso'.

Note que no **Jamovi** o separador de decimais é o ponto (.), devido a convenção do formato estadunidense (inglês), diferente do formato brasileiro (francês). Repare também que o nome da categoria deve ser colocado entre aspas simples ' e não ‘.

Deve-se clicar no símbolo de “+” para adicionar mais condições

condição 2 (fórmula 2): `if $source < 25 use 'eutrofico'`

condição 3 (fórmula 3): `if $source < 30 use 'sobrepeso'`

condição 4 (fórmula 4): `else use 'obeso'` (caso o indivíduo não tenha sido classificado ainda, considere-o na categoria ‘obeso’, isto é, se o indivíduo não tiver sido classificado em nenhuma categoria anterior, então deve ser classificado na última categoria).

Na Figura 22 é possível visualizar a definição das regras acima e o resultado da variável nova.

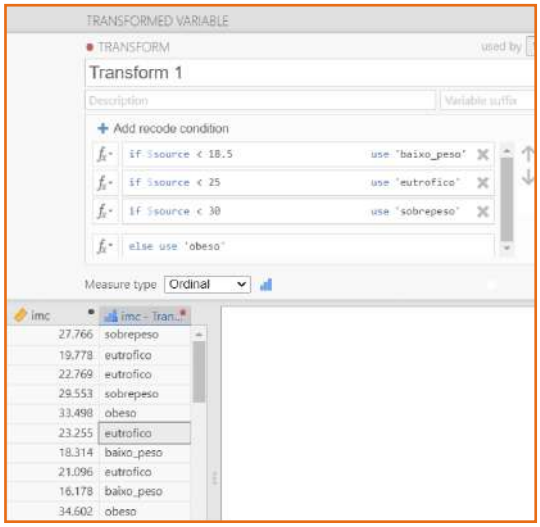


Figura 22. Definição das categorias da variável IMC

Se for de interesse criar uma variável binária a partir de outra variável existente.

Deseja-se criar uma variável binária com categorias: eutrófico; outro equivalente a (eutrófico - sim; eutrófico - não) a partir da variável IMC categorizada “imc_categorico”.

No painel de criação de variáveis (Figura 9) selecionar a opção “*new transformed variable*”:

Em “*Source variable*” selecionar a variável “imc_categorico”;
Em “*using transform*” selecionar “*Create New Transform*”;

A condição será: “se 'imc_categorico' for igual a 'eutrofico' a categoria será 'eutrofico', caso contrário, 'outro'.

No **Jamovi**:

condição 1 (fórmula 1): `if $source == 'eutrofico' use 'eutrofico'` (note que o operador lógico de igualdade no **Jamovi** é `'=='` porque já existia esse conteúdo antes)

condição 2 (fórmula 2): `else use 'outro'`

Essa operação pode ser visualizada na Figura 23.

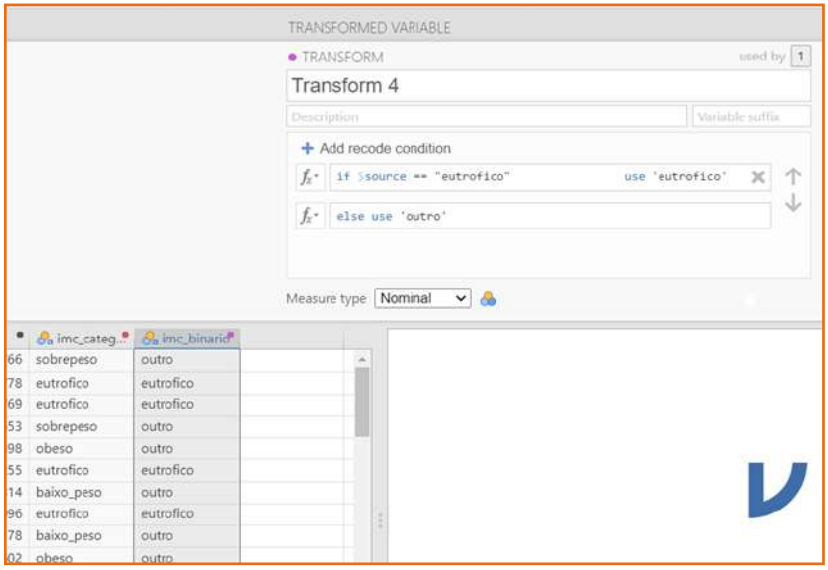


Figura 23. Definição variável binária

3.3. Exercícios

Neste tópico será utilizado o banco de dados “dados_exercicio_sinasc_jamovi.csv” (Bergamaschi, 2024). Também é possível reproduzir o exercício de geração desta base de dados seguindo o tópico 9 “Saiba Mais” da apostila. Banco proveniente de SINASC/DATASUS (https://datasus.saude.gov.br) com 3.616 observações e 21 variáveis, referente ao Município de Osvaldo

Cruz, São Paulo, no período de 2012-2022. Extraído em 12 de dezembro de 2024. O Quadro E3.3 apresenta o dicionário das variáveis.

Quadro E3.3. Dicionário de variáveis contendo o nome original, o nome da variável, a definição e os códigos das categorias

Nome original	Definição	Códigos
DTNASC	Data de nascimento	dd/mm/aa
LOCNASC	Local de nascimento	Hospital, Outro estabelecimento de saúde, Domicílio, Outros, Ignorado (NA)
SEXO	Sexo do recém-nascido	Masculino, Feminino, Ignorado (NA)
PESO	Peso ao nascer (gramas)	numérico
APGAR1	APGAR no primeiro minuto	numérico, de 0 a 10
APGAR5	APGAR no quinto minuto	numérico, de 0 a 10
IDANOMAL	Anormalidade congênita	Sim, Não, Ignorado (NA)
SEMAGESTAC	Semanas de gestação	numérico
GESTACAO	Faixas de idade gestacional (semanas)	22-27, 28-31, 32-36, 37-41, 42+
IDADEMAE	Idade da mãe (anos)	numérico
RACACORMAE	Raça/cor da mãe	Branca, Preta, Amarela, Parda, Indígena, Ignorado (NA)
ESCMAE	Faixas de anos de escolaridade da mãe	Nenhum, 1-3, 4-7, 8-11, 12+, Ignorado (NA)
ESTCIVMAE	Estado civil da mãe	Casada, Solteira, Separada judicialmente, Viúva, União consensual, União estável, Ignorado (NA)
QTDFILMORT	Número de perdas fetais/ abortos anteriores	numérico
QTDFILVIVO	Número de nascidos vivos anteriores	numérico
QTDGESTANT	Número de gestações anteriores	numérico
PARIDADE	Número de partos anteriores	numérico
PARTO	Tipo de parto	Cesáreo, Vaginal, Ignorado (NA)
GRAVIDEZ	Tipo de gravidez	única, dupla, tripla ou +, Ignorado (NA)
CONSULTAS	Faixas de número de consultas de pré-natal	Nenhuma, 1-3, 4-6, 7 ou +, Ignorado (NA)
CONSPRENAT	Número de consultas de pré-natal	numérico

3.3.1. Exercício

Criando a variável ano de nascimento a partir da variável data de nascimento (DTNASC). Ver Figura E3.3.1.

Variables>Transform>ano_nasc>Source Variable: DTNASC>Create New Transform> “SPLIT(DTNASC, '-',1)”).

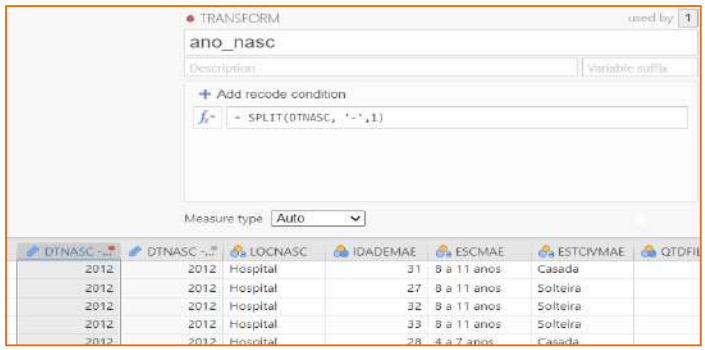


Figura E3.3.1. Criação do ano nascimento

3.3.2. Exercício

Definindo intervalos para a variável de idade gestacional (“SEMAGESTAC”). Ver Figura E3.3.2.

Variables>Transform>ig_faixas>Source Variable: SEMAGESTAC>Create New Transform>

condição 1 (fórmula 1): `if $source < 37 use 'pre-termo'`

condição 2 (fórmula 2): `if $source < 42  use 'termo'`

condição 3 (fórmula 3): `else use 'pos_termo'`

Figura E3.3.2. Seleção da função “create new transform”

3.3.3. Exercício

Definindo intervalos para a variável de peso ao nascer (“PESO”). As faixas escolhidas foram definidas arbitrariamente. Ver Figura E3.3.3.

Variables>Transform>peso_faixas>Source Variable: PESO>Create New Transform>

condição 1 (fórmula 1): `if $source < 2500 use 'baixo_peso'`

condição 2 (fórmula 2): `if $source < 4000  use 'eutrofic'`

condição 3 (fórmula 3): `else use 'excesso_peso'`

Figura E3.3.3. Seleção da função “create new transform”

3.3.4. Exercício

Declarando valores faltantes (NA). Ver Figura E3.3.4.

Data > Clicar duas vezes no nome da variável (‘RACACORMAE’)>Selecionar missing values > Add Missing Value > when $source == 'NA'. Notar que os campos com NA passam a ser considerados como caselas em branco.

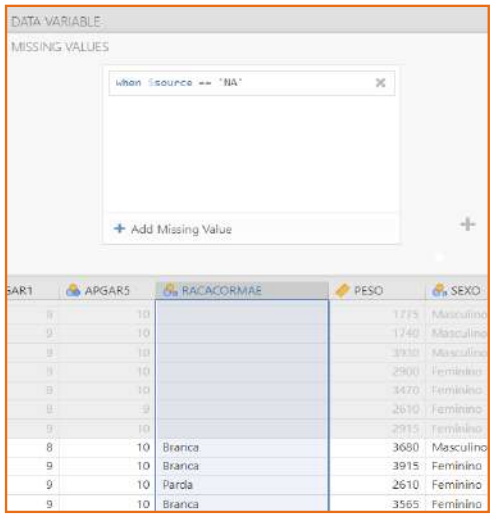


Figura E3.3.4. Declarando valores faltantes (missing values)

3.3.5. Exercício

Considere os dados apresentados no Quadro E3.3.5.2. (Rosner, 2016, p. 34). Digite os dados no **Jamovi** e salve o banco de dados nomeando-o <hospital>. Utilize os nomes das variáveis como sugerido. Defina o tipo das variáveis como apresentado no Quadro E3.3.5.1.

Quadro E3.3.5.1. Nome das variáveis, tipo e rótulo dos níveis (level)

Nome da variável	Tipo	Rótulos	Observações
dias	contínua, inteiro		tempo de internação
idade	contínua, inteiro		
sexo	nominal	1-masculino 2-feminino	
temperatura	contínua, decimal		
hem	contínua, inteiro		hemograma
antib	nominal	1-sim, 2-não	antibiótico
cult	nominal	1-sim, 2-não	cultura bacteriana
atend	nominal	1- clínico 2 - cirúrgico	atendimento

Quadro E3.3.5.2. Dias de hospitalização (dias), idade, sexo (1 - masculino; 2-feminino), primeira temperatura após a internação (temperatura), primeiro hemograma após a internação (hem), recebeu antibiótico (1-sim, 2-não) (antib), recebeu cultura bacteriana (1- sim, 2- não) (cult), atendimento (1- clínico, 2 - cirúrgico) (atend)

id	dias	idade	sexo	temp.	hem	antib	cult	atend
1	5	30	2	37.2	8	2	2	1
2	10	73	2	36.6	5	2	1	1
3	6	40	2	37.2	12	2	2	2
4	11	47	2	36.7	4	2	2	2
5	5	25	2	36.9	11	2	2	2
6	14	82	1	36.0	6	1	2	2
7	30	60	1	37.5	8	1	1	1
8	11	56	2	37.0	7	2	2	1
9	17	43	2	36.6	7	2	2	1
10	3	50	1	36.6	12	2	1	2
11	9	59	2	36.4	7	2	1	1
12	3	4	1	36.5	3	2	2	2
13	8	22	2	37.5	11	1	2	2
14	8	33	2	36.8	14	1	1	2
15	5	20	2	36.8	11	2	1	2
16	5	32	1	37.2	9	2	2	2
17	7	36	1	37.3	6	1	2	2
18	4	69	1	36.6	6	2	2	2
19	3	47	1	36.1	5	1	2	1
20	7	22	1	36.7	6	2	2	2
21	9	11	1	36.7	10	2	2	2
22	11	17	1	37.0	14	1	2	2
23	11	67	2	36.4	4	2	2	1
24	9	43	2	37.0	5	2	2	2
25	4	41	2	36.6	5	2	2	1

3.3.6. Exercício

Considere os dados apresentados no Quadro E3.3.6, retirados de Rosner (2016, p. 35).

Digite os dados no **Jamovi** e salve o banco de dados com o nome colesantesdepois. Utilize os nomes das variáveis como sugerido. Defina o tipo das variáveis: antes - contínua, inteiro; depois - contínua, inteiro.

Quadro E3.3.6. Níveis de colesterol (mg/dL) antes e depois de dieta vegetariana

id	antes	depois
1	195	146
2	145	155
3	205	178
4	159	146
5	244	208
6	166	147
7	250	202
8	236	215
9	192	184
10	224	208
11	238	206
12	197	169
13	169	182
14	158	127
15	151	149
16	197	178
17	180	161
18	222	187
19	168	176
20	168	145
21	167	154
22	161	153
23	178	137
24	137	125

3.3.7. Exercício

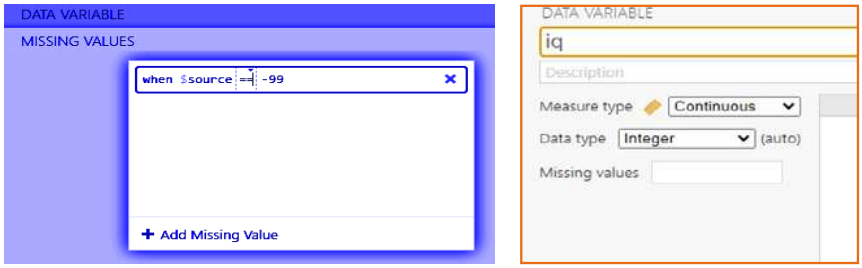
Considerar o banco os dados a seguir:

Os dados são de 118 pacientes psiquiátricas do sexo feminino disponíveis em Hand et al. (1994). As variáveis são:

idade: idade (anos)
iq: escore de quociente de inteligência
ans: ansiedade (1-nenhuma, 2- leve, 3-moderada, 4-severa)
depre: depressão (1-nenhuma, 2- leve, 3-moderada, 4-severa)
sono: você pode dormir normalmente? (1- sim, 2- não)
sexo: você perdeu o interesse em sexo? (1- não, 2- sim)
vida: você teve pensamento suicida recentemente? (1- não, 2- sim)
ganho: ganho de peso nos últimos seis meses (pounds)

Os valores -99 são valores faltantes ou ignorados (missing)

Digitar (copiar) os dados do Quadro E3.3.7, no **Jamovi**, especificando os nomes e rótulos das categorias das variáveis segundo indicado na descrição dos dados; codificar os valores missing (-99) (Figura E3.3.7). Salvar o banco de dados como fem.omv.



1- clicar em missing value 2- incluir operador (==) e o valor do missing (-99)

Figura E3.3.7. Definição de codificação de valores faltantes (missing)

Quadro E3.3.7. Dados de 118 pacientes psiquiátricas

id	idade	iq	ans	depre	sono	sexo	vida	ganho
1	39	94	2	2	2	2	2	4.9
2	41	89	2	2	2	2	2	2.2

continua...

id	idade	iq	ans	depre	sono	sexo	vida	ganho
3	42	83	3	-99	3	2	2	4.0
4	30	99	2	2	2	2	2	-2.6
5	35	94	2	1	1	2	1	-0.3
6	44	90	-99	1	2	1	1	0.9
7	31	94	2	2	-99	2	2	-1.5
8	39	87	3	2	2	2	1	3.5
9	35	-99	3	2	2	2	2	-1.2
10	33	92	2	2	2	2	2	0.8
11	38	92	2	1	1	1	1	-1.9
12	31	94	2	2	2	-99	1	5.5
13	40	91	3	2	2	2	1	2.7
14	44	86	2	2	2	2	2	4.4
15	43	90	3	2	2	2	2	3.2
16	32	-99	1	1	1	2	1	-1.5
17	32	91	1	2	2	-99	1	-1.9
18	43	82	4	3	2	2	2	8.3
19	46	86	3	2	2	2	2	3.6
20	30	88	2	2	2	2	1	1.4
21	34	97	3	3	-99	2	2	-99.0
22	37	96	3	2	2	2	1	-99.0
23	35	95	2	1	2	2	1	-1.0
24	45	87	2	2	2	2	2	6.5
25	35	103	2	2	2	2	1	-2.1
26	31	-99	2	2	2	2	1	-0.4
27	32	91	2	2	2	2	1	-1.9
28	44	87	2	2	2	2	2	3.7
29	40	91	3	3	2	2	2	4.5
30	42	89	3	3	2	2	2	4.2
31	36	92	3	-99	2	2	2	-99.0
32	42	84	3	3	2	2	2	1.7
33	46	94	2	-99	2	2	2	4.8
34	41	92	2	1	2	2	1	1.7

continua...

id	idade	iq	ans	depre	sono	sexo	vida	ganho
35	30	96	-99	2	2	2	2	-3.0
36	39	96	2	2	2	1	1	0.8
37	40	86	2	3	2	2	2	1.5
38	42	92	3	2	2	2	1	1.3
39	35	102	2	2	2	2	2	3.0
40	31	82	2	2	2	2	1	1.0
41	33	92	3	3	2	2	2	1.5
42	43	90	-99	-99	2	2	2	3.4
43	37	92	2	1	1	1	1	-99.0
44	32	88	4	2	2	2	1	-99.0
45	34	98	2	2	2	2	-99	0.6
46	34	93	3	2	2	2	2	0.6
47	42	90	2	1	1	2	1	3.3
48	41	91	2	1	1	1	1	4.8
49	31	-99	3	1	2	2	1	-2.2
50	32	92	3	2	2	2	2	1.0
51	29	92	2	2	2	1	2	-1.2
52	41	91	2	2	2	2	2	4.0
53	39	91	2	2	2	2	2	5.9
54	41	86	2	1	1	2	1	0.2
55	34	95	2	1	1	2	1	3.5
56	39	91	1	1	2	1	1	2.9
57	35	96	3	2	2	1	1	-0.6
58	31	100	2	2	2	2	2	-0.6
59	32	99	4	3	2	2	2	-2.5
60	41	89	2	1	2	1	1	3.2
61	41	89	3	2	2	2	2	2.1
62	44	98	3	2	2	2	2	3.8
63	35	98	2	2	2	2	1	-2.4
64	41	103	2	2	2	2	2	-0.8
65	41	91	3	1	2	2	1	5.8
66	42	91	4	3	-99	-99	2	2.5

continua...

id	idade	iq	ans	depre	sono	sexo	vida	ganho
67	33	94	2	2	2	2	1	-1.8
68	41	91	2	1	2	2	1	4.3
69	43	85	2	2	2	1	1	-99.0
70	37	92	1	1	2	2	1	1.0
71	36	96	3	3	2	2	2	3.5
72	44	90	2	-99	2	2	2	3.3
73	42	87	2	2	2	1	2	-0.7
74	31	95	2	3	2	2	2	-1.6
75	29	95	3	3	2	2	2	-0.2
76	32	87	1	1	2	2	1	-3.7
77	35	95	2	2	2	2	2	3.8
78	42	88	1	1	1	2	1	-1.0
79	32	94	2	2	2	2	1	4.7
80	39	-99	3	2	2	2	2	-4.9
81	34	-99	3	-99	2	2	1	-99.0
82	34	87	3	3	2	2	1	2.2
83	42	92	1	1	2	1	1	5.0
84	43	86	2	3	2	2	2	0.4
85	31	93	-99	2	2	2	2	-4.2
86	31	92	2	2	2	2	1	-1.1
87	36	106	2	2	2	1	2	-1.0
88	37	93	2	2	2	2	2	4.2
89	43	95	2	2	2	2	1	2.4
90	32	95	3	2	2	2	2	4.9
91	32	92	-99	-99	-99	2	2	3.0
92	32	98	2	2	2	2	2	-0.3
93	43	92	2	2	2	2	2	1.2
94	41	88	2	2	2	2	1	2.6
95	43	85	1	1	2	2	1	1.9
96	39	92	2	2	2	2	1	3.5
97	41	84	2	2	2	2	2	-0.6
98	41	92	2	1	2	2	1	1.4

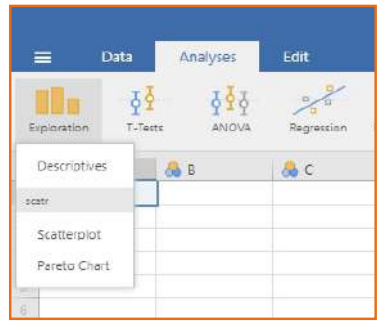
continua...

id	idade	iq	ans	depre	sono	sexo	vida	ganho
99	32	91	2	2	2	2	2	5.7
100	44	86	3	2	2	2	2	4.6
101	42	92	3	2	2	2	1	-99.0
102	39	89	2	2	2	2	1	2.0
103	45	-99	2	2	2	2	2	0.6
104	39	96	3	-99	2	2	2	-99.0
105	31	97	2	-99	-99	-99	2	2.8
106	34	92	3	2	2	2	2	-2.1
107	41	92	2	2	2	2	2	-2.5
108	33	98	3	2	2	2	2	2.5
109	34	91	2	1	1	2	1	5.7
110	42	91	3	3	2	2	2	2.4
111	40	89	3	1	1	1	1	1.5
112	35	94	3	3	2	2	2	1.7
113	41	90	3	2	2	2	2	2.5
114	32	96	2	1	1	2	1	-99.0
115	39	87	2	2	2	1	2	-99.0
116	41	86	3	2	1	1	2	-1.0
117	33	89	1	1	1	1	1	6.5
118	42	-99	3	2	2	2	2	4.9

4. ANÁLISE EXPLORATÓRIA DE DADOS

Toda análise de dados deve ser iniciada pela exploração dos dados. Para isso, utilizar o módulo “Exploration” que é composto por estatísticas descritivas e gráficos incluindo o de dispersão (scatterplot) e de Pareto, após instalação do módulo scatr (Figuras 24A, 24B).

(A)



(B)

Figura 24. Instalação do módulo scatr

Se for de interesse descrever as variáveis quantitativas segundo medidas de resumo (média, mediana, variância, desvio padrão, valores mínimo e máximo) selecionar as variáveis de interesse e direcioná-las (Figura 25), por meio da seta, para a caixa de seleção "variables". Também é possível apresentar a descrição mencionada acima estratificada nas categorias de uma variável qualitativa. Isso pode ser feito utilizando-se a opção "split by" (separado por).

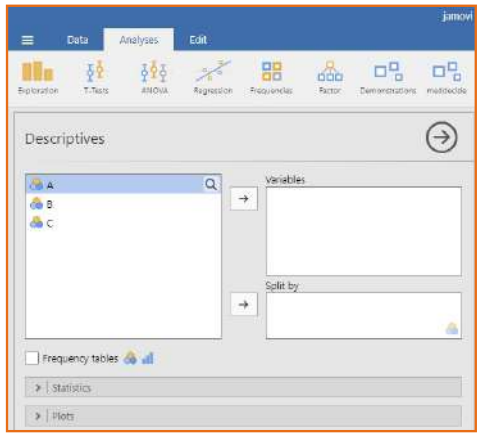


Figura 25. Direcionando as variáveis a serem descritas

4.1. Tabelas de frequência simples

Para variáveis qualitativas é possível fazer as tabelas de frequência ao selecionar a opção "frequency table". Se a variável estiver definida como

contínua, o **Jamovi** não irá fornecer a tabela de frequência pois variáveis em escala de razões contínuas devem ser representadas em intervalos de classe e não em valores pontuais. Uma alternativa é construir os intervalos de classe como indicado anteriormente (item 3.1). O Quadro 2 apresenta as funcionalidades disponíveis no módulo "Exploration" "Statistics".

4.2. Gráficos

Quadro 2. Estatísticas disponíveis no módulo "Exploration" "Statistics"

<table>
<tr><th colspan="2">Estatísticas disponíveis no módulo "Exploration" "Statistics"</th></tr>
<tr><td>Tamanho da amostra (sample size)
Número de observações (N)
Número de valores faltantes (Missing)

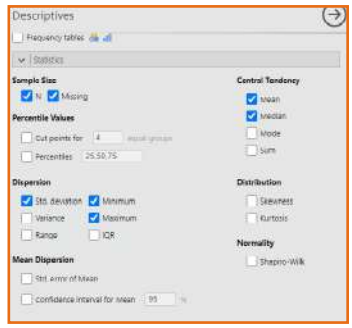
Tendência Central
Média (Mean)
Mediana (Median)
Moda (Mode)
Soma de valores (Sum)</td><td> 

Figura 26. Estatísticas descritivas no Módulo "Exploration", tópico "Statistics"</td></tr>
<tr><td>Valores de posição centis (Percentile values)
Ponto de corte para ____ grupos (Cut point for ____ groups)
utilizar 2 (metade), ... , x (centis)</td><td>Distribuição (Distribution)
Simetria (Skewness)
Curtose (Kurtosis)</td></tr>
<tr><td rowspan="3">Dispersão (Dispersion)
Variância (Variance)
Desvio padrão (Standard Deviation)
Amplitude de variação (Range)
Valores mínimo e máximo
Intervalo Inter-Quartil (IIQ)
(Interquartile range - IQR)</td><td>Dispersão da média (Mean Dispersion)
Desvio padrão da média (Standard error of Mean)</td></tr>
<tr><td>Intervalo de confiança para a Média (Confidence Interval for Mean)
especificar o grau de confiança.</td></tr>
<tr><td>Teste para normalidade (Normality)
Teste de normalidade Shapiro-Wilk</td></tr>
</table>

É possível fazer o gráfico da distribuição de frequências de variáveis contínuas (Histograma e Densidade); a investigação de existência de outlier e investigar dispersão e simetria (box plot); o gráfico de barras para representar variáveis qualitativas (Bar plot); gráfico QQ plot para investigar se a variável segue uma distribuição normal (Figuras 26 e 27).

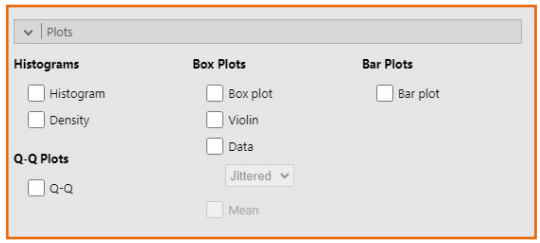


Figura 27. Gráficos a serem utilizados na exploração de dados

4.3. Exercícios

Com base nos dados expostos no Quadro 1, apresentar um resumo, utilizando gráficos e medidas estatísticas de resumo (Figuras E4.3.1, E4.3.2, E4.3.3) .

Resumo dos dados

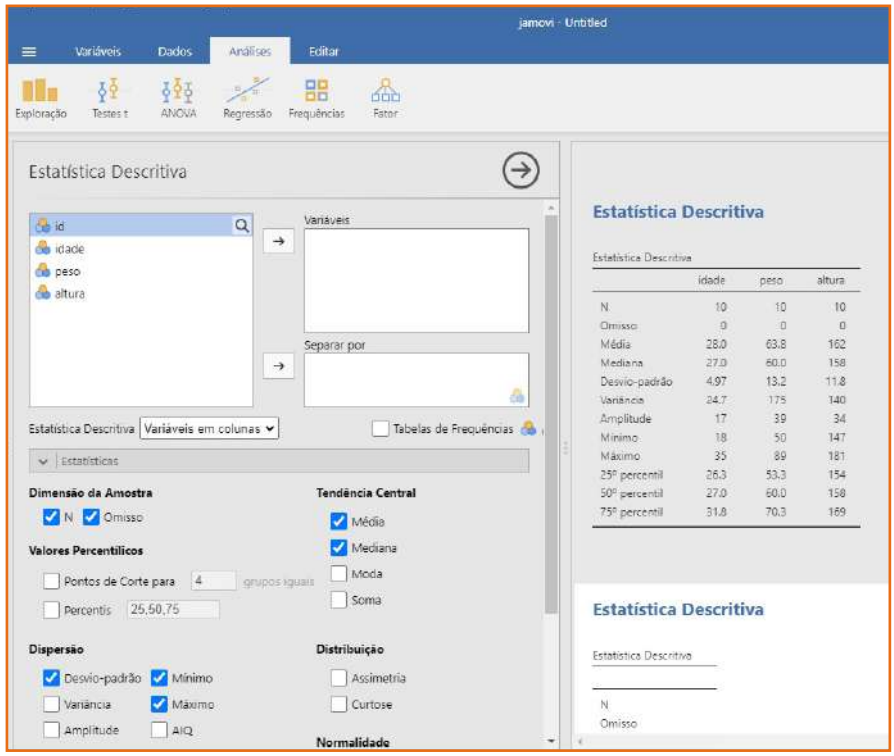


Figura E4.3.1. Resumo estatístico de dados de idade (anos), peso (kg) e altura (cm)

Gráficos:

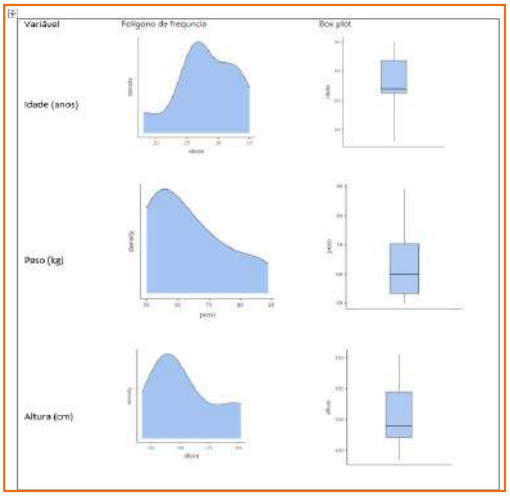


Figura E4.3.2. Distribuição de pessoas e box plot segundo idade (anos), peso (kg) e altura (cm)

Gráfico de dispersão

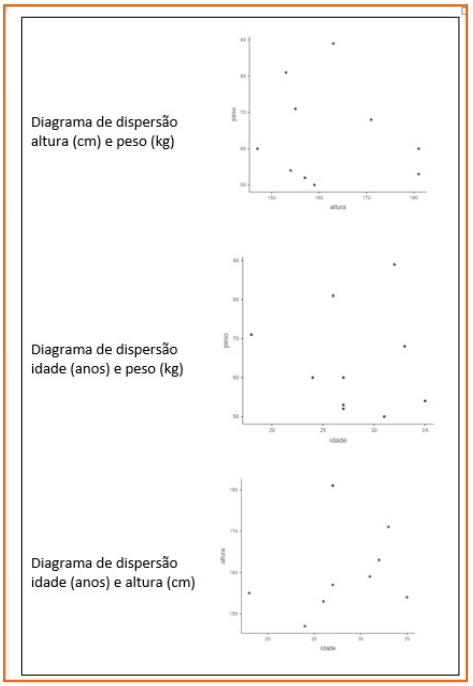


Figura E4.3.3. Gráficos de dispersão entre idade (anos), peso (kg) e altura (cm)

5. TESTE T-STUDENT

Para comparar duas médias populacionais pode-se utilizar o teste t de Student.

5.1. Teste para duas médias com amostras independentes

Deve-se testar as pressuposições do teste t de Sudeste para variância iguais: a homogeneidade das variâncias e a normalidade da variável quantitativa (Figura 28).

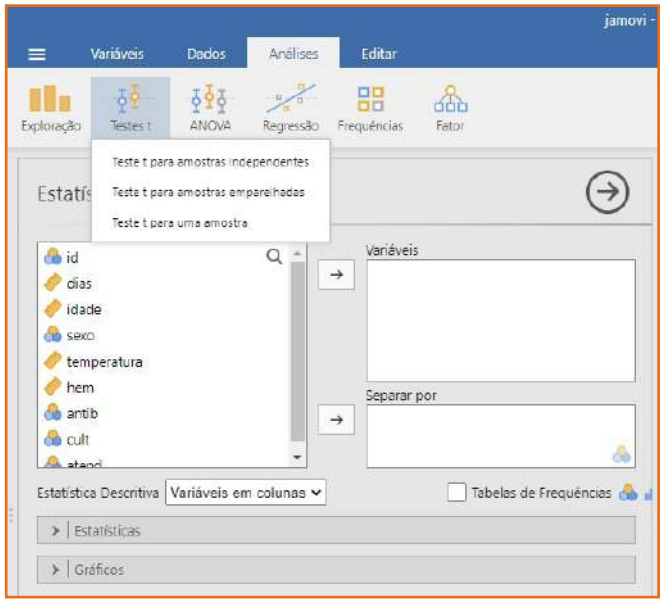


Figura 28. Módulo "Testes t"

Considere o banco de dados hospital.omv, com dados apresentados no Quadro E3.3.5.2.

A apresentação do teste t será feita utilizando-se a variável temperatura. O objetivo é investigar se a temperatura média de cada população de onde os pacientes são provenientes possui igualdade de médias.

Inicialmente realiza-se uma análise exploratória da variável temperatura segundo as categorias da variável sexo (Figura 29).

Note que pelo Teste Shapiro-Wilk, pode-se dizer que a variável temperatura segue distribuição normal tanto para o sexo masculino (p=0,740) como para o sexo feminino (p=0,611).

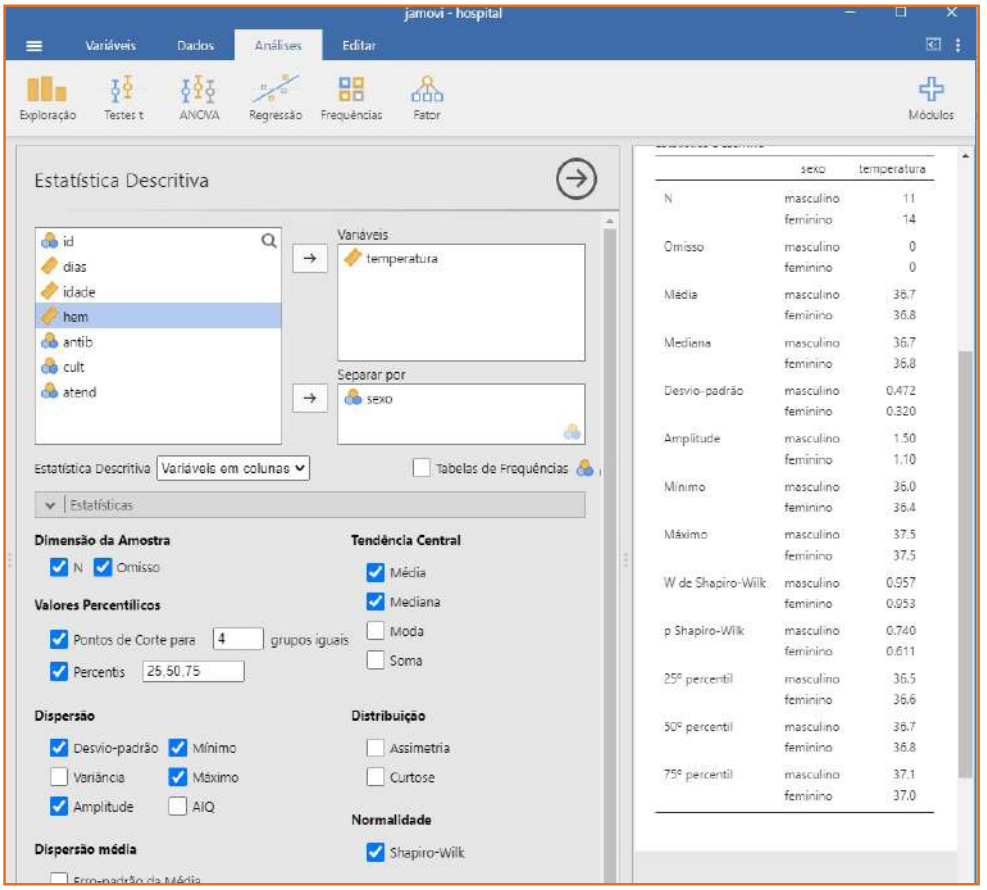


Figura 29. Análise exploratória da variável temperatura segundo a variável sexo

Para testar os pressupostos do teste t, ative as alternativas (Figura 30).

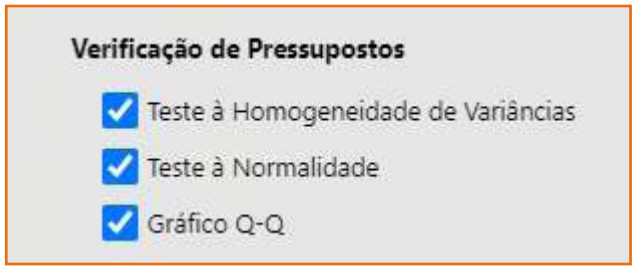


Figura 30. Opções para testar as pressuposições do teste t para amostras independentes e variâncias iguais

Por meio dos testes para normalidade Shapiro-Wilk (p=0,816); de Levene para homogeneidade de variâncias (p=0,230) e QQ plot, pode-se afirmar que não há violação de pressupostos. Também apresenta-se um resumo dos dados de temperatura em cada sexo (Figuras 31 e 32).

Pressupostos

Teste à Normalidade (Shapiro-Wilk)

	W	p
temperatura	0.977	0.816

Nota. Um p-value pequeno sugere a violação do pressuposto da normalidade

Teste à Homogeneidade de Variâncias (Levene)

	F	gl	gl2	p
temperatura	1.52	1	23	0.230

Nota. Um p-value pequeno sugere a violação do pressuposto da homogeneidade de variâncias

Descritivas de Grupo

	Grupo	N	Média	Mediana	Desvio-padrão	Erro-padrão
temperatura	masculino	11	36.7	36.7	0.472	0.142
	feminino	14	36.8	36.8	0.320	0.0856

Figura 31. Resultados da verificação de pressupostos para o teste t de Student para duas médias, com igualdade de variâncias

Note que o QQ plot também indica normalidade para a variável temperatura.

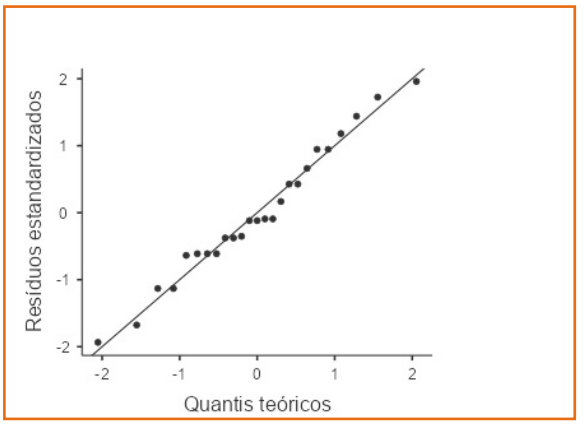


Figura 32. QQ plot para a variável temperatura

Decisão: existe evidência que a temperatura média de homens e mulheres são provenientes de uma mesma população. Portanto, não existe diferença estatística entre as médias populacionais de temperatura, para homens e mulheres (p=0,575) (Figura 33).

Teste t para amostras independentes

		Estatística	gl	p
temperatura	t de Student	-0.569	23.0	0.575

Nota. H_a $\mu_{masculino} \neq \mu_{feminino}$

Figura 33. Resultado do teste t de Student para duas amostras independentes

O **Jamovi** também fornece o Gráfico dos Intervalos de Confiança (IC95%) para as médias populacionais de temperatura, segundo sexo (Figura 34).

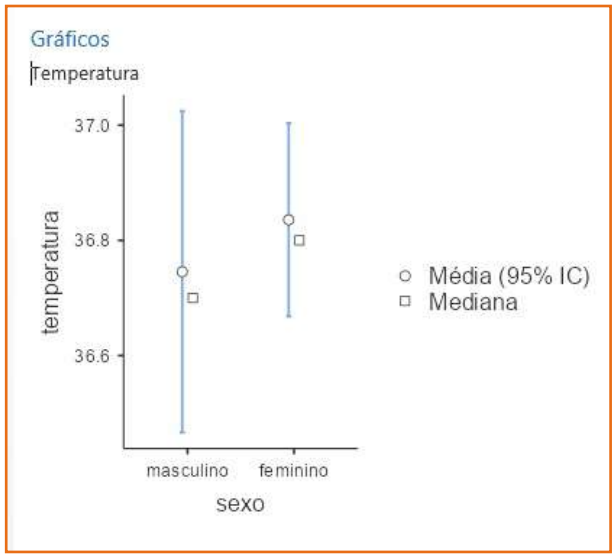


Figura 34. Intervalo de Confiança de 95% para as médias populacionais de temperatura entre pacientes do sexo masculino e feminino

5.2. Teste para duas médias com amostras dependentes

Utilizando-se o banco de dados apresentado no Quadro 3, pode-se observar que o desenho de pesquisa é de duas amostras dependentes. Caso exista interesse em comparar os níveis de colesterol (mg/dL) antes e depois de dieta vegetariana, deve-se utilizar o teste t de Student para amostras dependentes (pareadas ou emparelhadas) (Figura 35).

Selecionar, na opção teste t para amostras emparelhadas, as variáveis que contém os dados a serem comparado (antes e depois) e as opções t de Student, tipo de teste (mono ou bicaudal), verificação de pressuposições (normalidade para a distribuição das diferenças) verificada pelo teste de Shapiro-Wilk e pelo QQ-plot e, caso seja de interesse, o intervalo de confiança para a diferença (Figura 35).

O teste revela que existe diferença nos níveis de colesterol antes e após ($p<0,001$) sendo que após os níveis são menores.

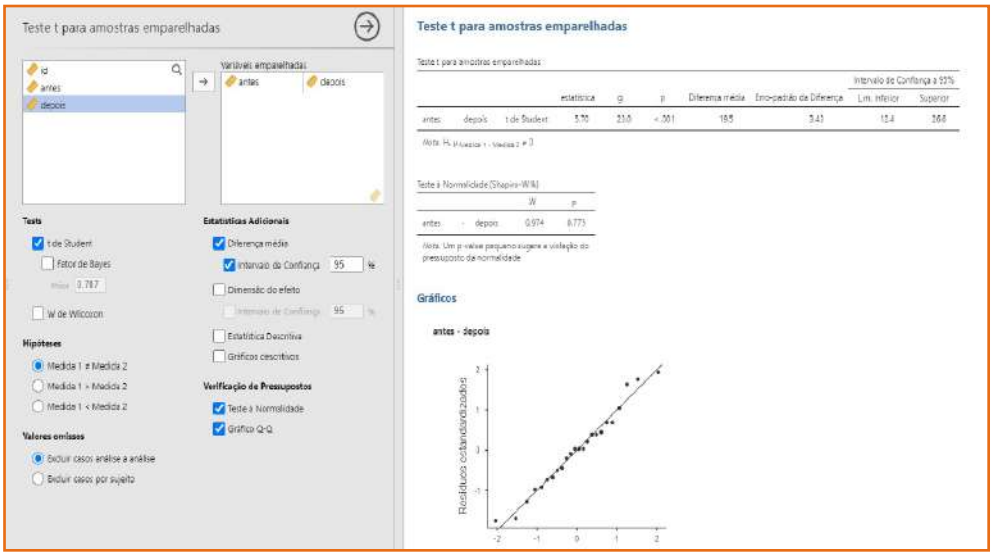


Figura 35. Teste t para amostras emparelhadas

5.3. Teste T de Student: uma média com variância desconhecida

O teste para uma amostra deve ser utilizado nas situações em que se deseja comparar a média populacional desconhecida, com uma média populacional conhecida.

Utilizar a seguinte situação para aplicação do teste:

O conteúdo de iodo em pacotes de sal é recomendado que seja igual, em média a 30 mg (15-45mg) para cada quilograma de sal. As autoridades sanitárias, tendo recebido reclamações de que o teor médio de iodo em sal de consumo humano estava abaixo do recomendado, realizou um estudo com dosagem de iodo em amostras de sal em 15 fábricas. Os resultados das quantidades de iodo são apresentados a seguir. Realize um teste de hipóteses para verificar se a reclamação procede, utilizando o valor de p.

25.5	15.7	23.6	26.4	25.8	15.7	18.6	15.9	27.9	14.3	31.0	6.7	22.5	34.0	35.9

Digite os dados e salve o banco como iodotesteumamedia.omv.

Selecione no Módulo T-tets, a opção One Sample T-Teste (Figura 36).

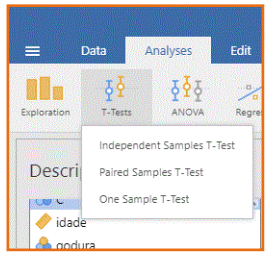


Figura 36. Opções para teste T de duas e uma média

Direcione a variável iodo para a caixa "Dependent variables"; selecione em Tests, Student´s; especificar o valor da média populacional conhecida e que serve de comparação; selecione H_a monocaudal (Teste value<30) uma vez que o enunciado indica que o teste de hipóteses deve ser monocaudal à esquerda; solicitar a verificação da pressuposição do teste (normalidade da variável concentração de iodo no sal de consumo) (Figura 37) ; solicitar o intervalo de confiança da diferença entre a média de iodo no sal de consumo humano de fábricas e o determinado por autoridades sanitárias (Figura 38).

Verificando-se a pressuposição de normalidade do teste observa-se que a variável iodo segue distribuição normal (p=0,872) (Figura 37).

Normality Test (Shapiro-Wilk)

	W	p
iodo	0.971	0.872

Note. A low p-value suggests a violation of the assumption of normality

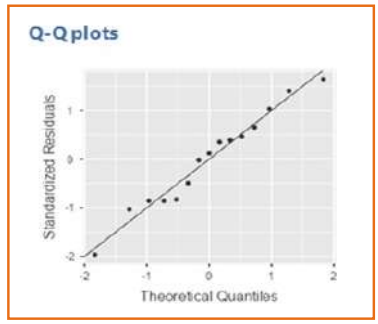


Figura 37. Checagem da pressuposição de normalidade

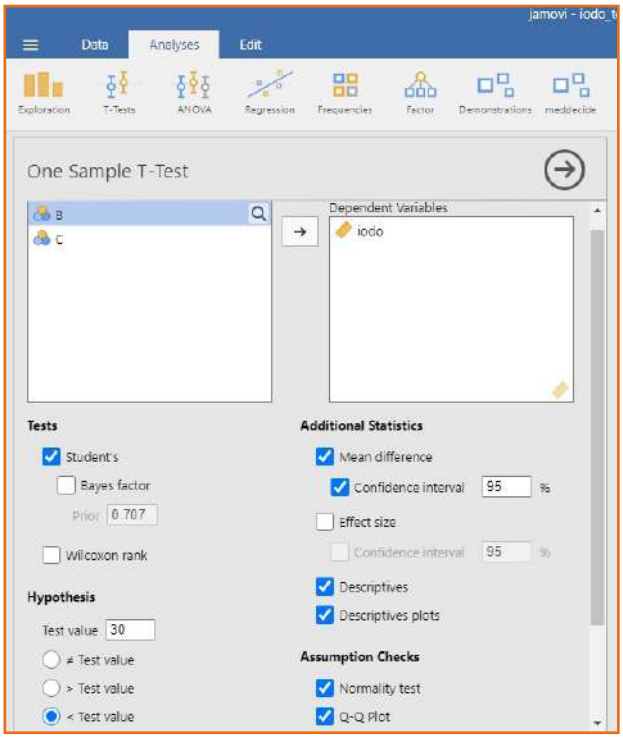


Figura 38. Teste t de Student para uma amostra, com variância desconhecida

O resultado do teste indica que existe diferença entre a quantidade média de iodo diluído no sal de consumo humano, comparado ao especificado por autoridades sanitárias (p=0,002). A mesma conclusão pode ser feita com base no IC95% para a diferença entre as médias populacionais das fábricas e a especificada por autoridades públicas. Como o intervalo de confiança inclui o valor zero, então a diferença média é diferente de zero e negativa indicando que a quantidade média de iodo em sal é menor que 30 mg por quilograma (Figura 39).

One Sample T-Test

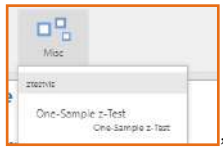
						95% Confidence Interval	
		Statistic	df	p	Mean difference	Lower	Upper
iodo	Student's t	−3.525	14.000	0.002	−7.367	−Inf	−3.685

Note. H_a population mean < 30

Figura 39. Resultado do teste t para uma média populacional com desvio padrão desconhecido

5.4. Teste T de Student: uma média com variância conhecida

Também é possível realizar o teste de uma média para situações em que a variância populacional é conhecida. Nesse caso, o teste a ser utilizado é o teste z (distribuição normal).

Deve-se instalar o módulo Misc 

, segundo apresentado na Figura 40.

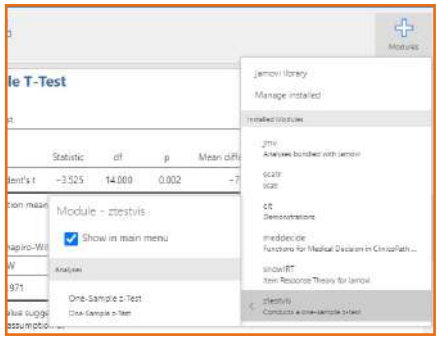


Figura 40. Instalação do pacote Misc que contém o teste z para uma amostra com variância conhecida. Selecione em "Modules" a opção ztestivis

Supondo-se os dados de iodo no sal de consumo humano mas com a variância populacional conhecida X: concentração de iodo ~ N(μ=30; σ=2.5)

As estatísticas descritivas das 15 amostras de quantidade de iodo no sal são apresentadas na Figura 41.

Descriptives

	N	Mean	Median	SD	SE
iodo	15	22.633	23.600	8.095	2.090

Figura 41. Resumo dos dados de quantidade de iodo no sal em 15 amostras

O teste para uma média, com variância conhecida deve ser feito utilizando-se o teste z. O **Jamovi** solicita que sejam informados: a média amostral (observada); o valor especificado na hipótese H_0 (30); o valor da variância populacional conhecida e que por pressuposição também é o valor da variância da população de estudo ($\sigma = 2.5$); o tamanho da amostra (n=15); o valor do nível de significância do teste ($\alpha = 0.05$) (Figura 42).

Notar que o **Jamovi** utiliza a estratégia clássica para realizar o teste de hipóteses para uma média com variância conhecida, especificando o nível de significância do teste; estabelecendo a região de rejeição-aceitação de H0, apresentando a distribuição de $\overline{X}$: Concentração de iodo e a distribuição de que segue uma distribuição normal com média 30 e erro padrão $\frac{\sigma}{\sqrt{n}}$=0.645.

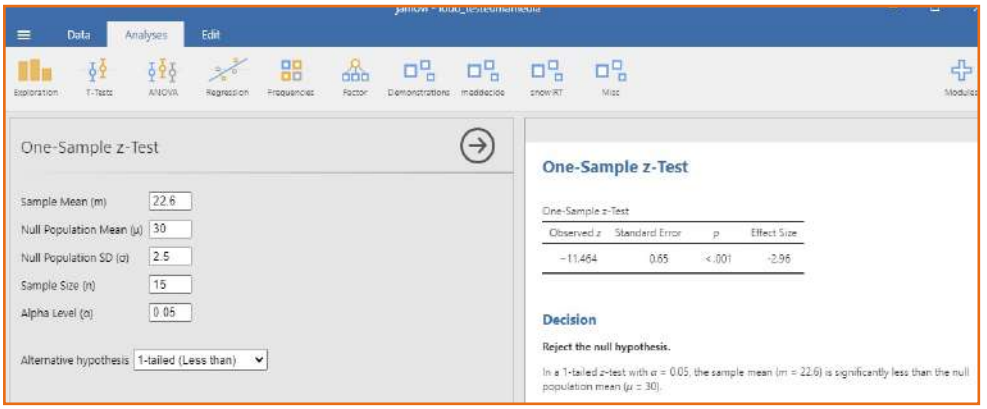


Figura 42. Especificação de valores para a realização do teste z para uma média populacional com variância conhecida

O valor da estatística do teste ($\bar{x}$=22.6) cai na região de rejeição de H_0 indicando que foi encontrada diferença estatisticamente significante entre a média especificada em H_0 e da população onde as amostras são provenientes, para nível de significância de 5% (Figura 43).

Notar que o **Jamovi** realiza o teste com a estatística $\bar{x}_{observado}$ e não com a estatística z.

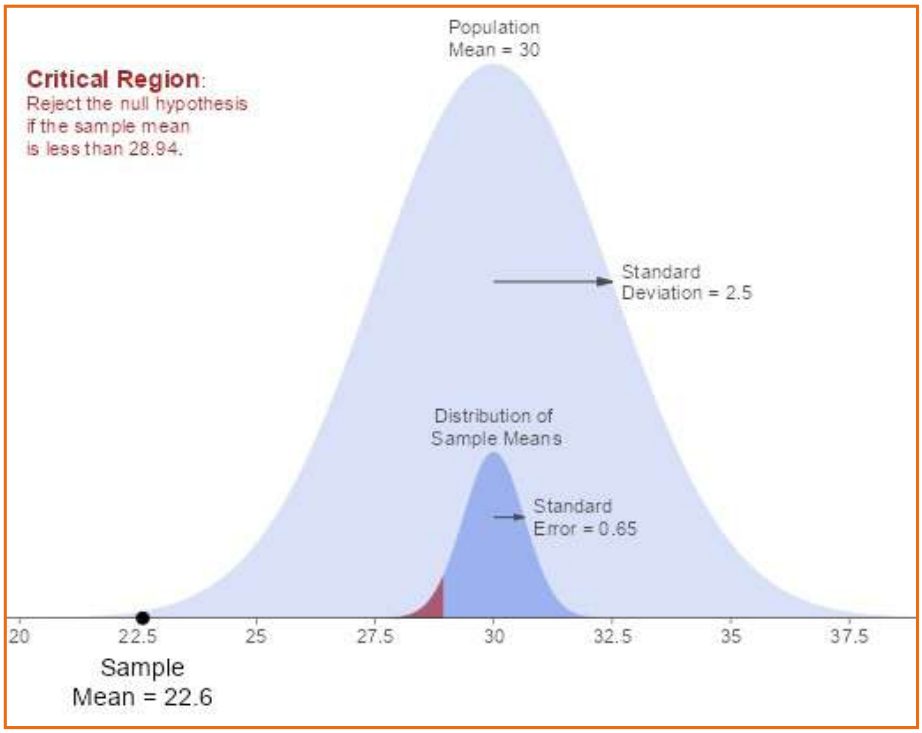


Figura 43. Distribuição da concentração de iodo (gráfico maior) e das médias amostrais especificando a região de rejeição de H_0

5.5. Exercícios

5.5.1. Teste de hipóteses para duas médias com amostras independentes

Considere o banco de dados "dados_exercicio_sinasc_jamovi.csv" (Bergamaschi, 2024). Também é possível reproduzir o exercício de geração desta base de dados seguindo o tópico 9 "Saiba Mais" da apostila. Banco proveniente de SINASC/DATASUS (https://datasus.saude.gov.br) com 3.616 observações e 21 variáveis, referente ao Município de Osvaldo Cruz, São Paulo, no período de 2012-2022. Extraído em 12 de dezembro de 2024. O Quadro E5.5.1. apresenta o dicionário das variáveis.

Quadro E5.5.1. Dicionário de variáveis contendo o nome original, o nome da variável, a definição e os códigos das categorias

Nome original	Definição	Códigos
DTNASC	Data de nascimento	dd/mm/aa
LOCNASC	Local de nascimento	Hospital, Outro estabelecimento de saúde, Domicílio, Outros, Ignorado (NA)
SEXO	Sexo do recém-nascido	Masculino, Feminino, Ignorado (NA)
PESO	Peso ao nascer (gramas)	numérico
APGAR1	APGAR no primeiro minuto	numérico, de 0 a 10
APGAR5	APGAR no quinto minuto	numérico, de 0 a 10
IDANOMAL	Anormalidade congênita	Sim, Não, Ignorado (NA)
SEMAGESTAC	Semanas de gestação	numérico
GESTACAO	Faixas de idade gestacional (semanas)	22-27, 28-31, 32-36, 37-41, 42+
IDADEMAE	Idade da mãe (anos)	numérico
RACACORMAE	Raça/cor da mãe	Branca, Preta, Amarela, Parda, Indígena, Ignorado (NA)
ESCMAE	Faixas de anos de escolaridade da mãe	Nenhum, 1-3, 4-7, 8-11, 12+, Ignorado (NA)
ESTCIVMAE	Estado civil da mãe	Casada, Solteira, Separada judicialmente, Viúva, União consensual, União estável, Ignorado (NA)
QTDFILMORT	Número de perdas fetais/ abortos anteriores	numérico
QTDFILVIVO	Número de nascidos vivos anteriores	numérico
QTDGESTANT	Número de gestações anteriores	numérico
PARIDADE	Número de partos anteriores	numérico
PARTO	Tipo de parto	Cesáreo,Vaginal, Ignorado (NA)
GRAVIDEZ	Tipo de gravidez	única, dupla, tripla ou +, Ignorado (NA)
CONSULTAS	Faixas de número de consultas de pré-natal	Nenhuma, 1-3, 4-6, 7 ou +, Ignorado (NA)
CONSPRENAT	Número de consultas de pré-natal	numérico

5.5.1.1. Investigar existência de diferença em número de consultas pré-natal entre recém-nascidos com baixo peso e sem baixo peso.

Inicialmente, é necessário criar a variável dicotômica para peso ao nascer e adicionar filtro para partos únicos. Para a variável de número de consultas, é necessário remover os valores faltantes.

Criação da variável dicotômica:

Variables > Transform > Variable: baixo_peso > Source variable: PESO > Create new transform > Fórmula: `if $source < 2500 use 'bp' else use 'nao_bp'`

Criação dos filtros:

Data > Filters > Add new filter (+) > Fórmula: `GRAVIDEZ == 'única'`

Data > Filters > Add new filter (+) > Fórmula: `CONSPRENAT!= NA`

Teste-t:

Analyses > T-Tests > Independent Samples T-Test > Dependent Variables: CONSPRENAT > Grouping Variable: baixo_peso > Tests: Student's > Additional Statistcs: selecionar "Descriptives" e selecionar "Descriptives plots" > Hypothesis: selecionar Group 1 Group 2

Os resultados para o exercício podem ser observados na Figura E5.5.1.1.

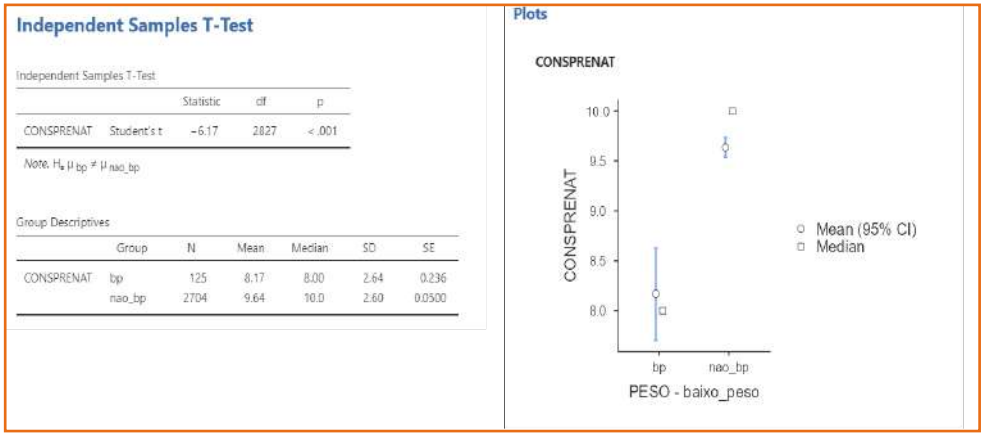


Figura E5.5.1.1. Resultados para o exercício do teste-t

6. ANÁLISE DE VARIÂNCIA – ANOVA

Abrir o módulo da ANOVA selecionando o ícone 

. Se necessário, instalar o módulo clicando em 

.

As opções de modelos de ANOVA podem ser vistos no menu disponível (Figura 44).

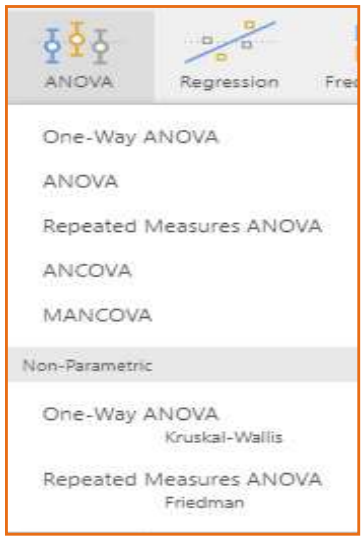


Figura 44. Opções de modelos de ANOVA

Será apresentado, nesta primeira versão do material de apoio ao uso do **Jamovi**, o modelo mais simples de ANOVA, paramétrico, com um fator fixo (One-way ANOVA).

Considerar os dados a seguir utilizados na exemplificação do módulo.

São apresentados dados de hipertensão (mmHg) para três tipos de tratamentos (Quadro 3). Deseja-se saber se existe diferença entre as drogas.

Quadro 3. Pressão arterial (mmHg) em 24 pacientes em três tratamentos

Droga 1	Droga 2	Droga 3
170	186	180
175	194	187
165	201	199
180	215	170
160	219	204
158	209	194
161	164	162
173	166	184
157	159	183
152	182	156
181	187	180
190	174	173
173	189	202
194	194	228
197	217	190
190	206	206
176	199	224
198	195	204
164	171	205
190	173	199
169	196	170
164	199	160
176	180	179
175	203	179

Fonte: adaptado de Rabe-Hesketh e Everitt (2004, p. 78).

a) Construir o banco de dados da seguinte forma: os dados de pressão arterial devem ser digitados (copiados) para a primeira coluna e na segunda coluna, para os primeiros 24 valores, digitar 1, para os 24 valores seguintes digitar 2 e para os últimos 24 valores, digitar 3 de tal forma que a coluna indique a droga que o paciente tomou. Nomear as variáveis pressao e droga. A primeira deve ser definida como contínua, inteiro e a segunda, nominal. Salvar o banco como hipertensao3D.omv (Figura 45).

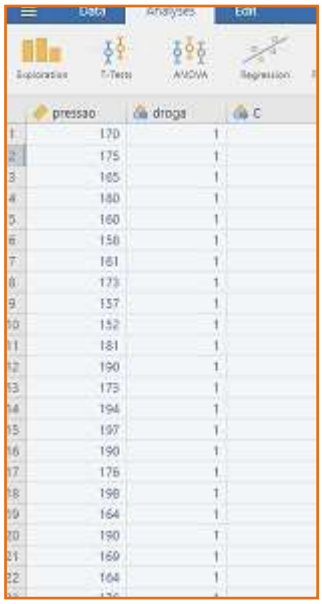


Figura 45. Banco de dados hipertensao3D

b) Abrir o módulo Oneway-ANOVA; indicar a variável dependente (pressao) e a variável de agrupamento (droga); iniciar verificando as pressuposições da ANOVA: selecionar teste de homogeneidade de variâncias (Homogeneity test), teste de normalidade para a variável dependente (Normality test e QQ plot); selecione em Additional Statistics - Descriptives tables e Descriptives plots. Após verificar o resultado do teste de homogeneidade, selecione a opção correta em Variances (Figura 46).

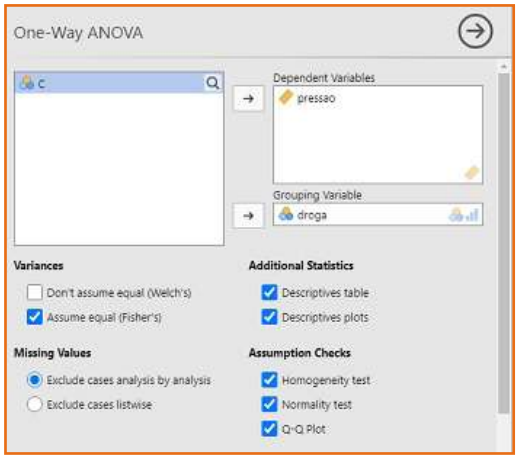


Figura 46. One-way ANOVA com sugestão de itens a serem selecionados

Resultados

Pelo valor de p pode-se concluir que a variável pressão segue distribuição normal (p=0,799) (Figura 47). A normalidade pode ser verificada no QQ--plot (Figura 48).

Normality Test (Shapiro-Wilk)

	W	p
pressao	0.989	0.799

Note. A low p-value suggests a violation of the assumption of normality

Figura 47. Resultados para o teste de normalidade (Shapiro-Wilk)

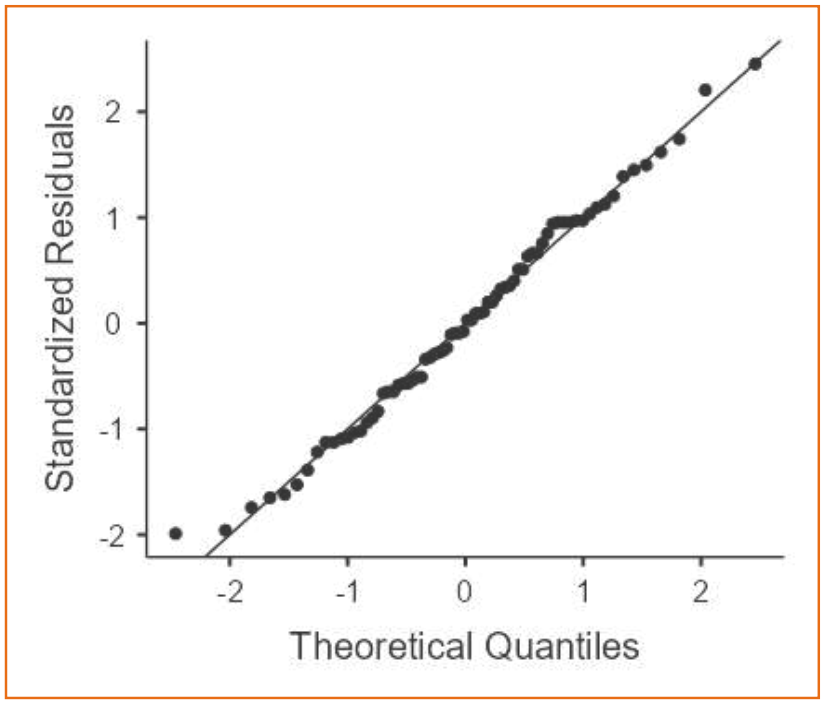


Figura 48. QQ-plot da variável dependente pressão arterial

Pelo valor de p pode-se concluir que a existe homogeneidade de variâncias (p=0,203) (Figura 49).

Homogeneity of Variances Test (Levene's)

	F	df1	df2	p
pressao	1.631	2	69	0.203

Figura 49. Resultados para o teste de homogeneidade de variâncias (Levene's)

Pode-se verificar que pelo menos uma média é diferente (p=0,002) (Figura 50).

One-Way ANOVA (Fisher's)

	F	df1	df2	p
pressao	6.702	2	69	0.002

Figura 50. Resultados para One-way ANOVA (Fisher's)

Apresenta-se o resumo dos dados de pressão, nos três grupos de tratamento (Figura 51).

Group descriptives

	droga	N	Mean	SD	SE
pressao	1	24	174.500	13.319	2.719
	2	24	190.750	17.007	3.472
	3	24	188.250	18.866	3.851

Figura 51. Resultados para estatísticas-resumo dos grupos

Intervalo de confiança de 95% para a pressão média nos três grupos de tratamento (Figura 52).

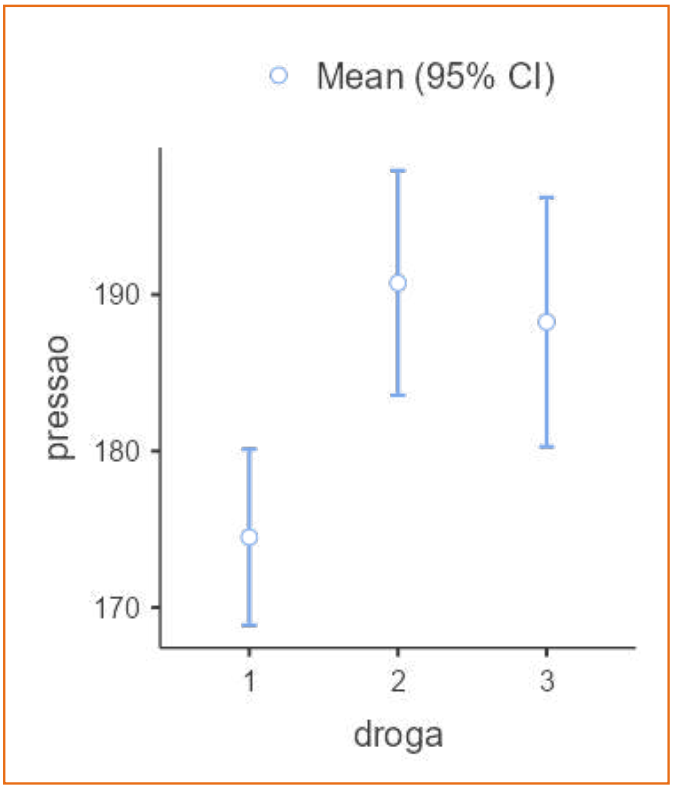


Figura 52. Gráfico contendo a média e intervalo de confiança 95% da pressão arterial segundo grupo de tratamento

A partir dos resultados do teste de Tukey para comparações de médias, pode-se verificar que μ_1 é diferente de μ_2 (p=0,003) e de μ_3 (0,015) (Figura 53).

Tukey Post-Hoc Test — pressao

	1	2	3
1 Mean difference	—	-16.250	-13.750
p-value	—	0.003	0.015
2 Mean difference		—	2.500
p-value		—	0.860
3 Mean difference			—
p-value			—

Figura 53. Resultados para o Tukey Post-Hoc Test da variável pressão segundo grupo de tratamento

6.1. Exercícios

6.1.1

Utilizando os dados apresentados no Quadro E3.3.7, que resultou no banco de dados fem.omv, investigar a existência de diferença entre a média de ganho de peso segundo depressão (depre). Utilize o subtópico ANOVA.

Resultado

As pressuposições da ANOVA estão satisfeitas, a variável ganho segue distribuição normal (p=0,359) e existe homogeneidade de variâncias (p=0,437). A ANOVA indica que não existe evidência de que pelo menos uma média seja diferente (p=0,443) (Figura E6.1.1).

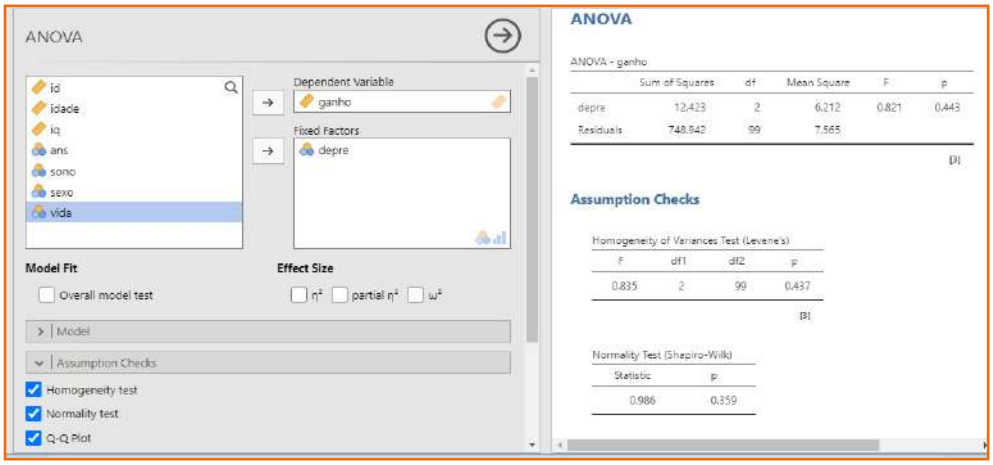


Figura E6.1.1. Resultados da ANOVA

Exercícios 6.1.2 a 6.1.5

Considere o banco de dados "dados_exercicio_sinasc_jamovi.csv" (Bergamaschi, 2024). Também é possível reproduzir o exercício de geração desta base de dados seguindo o tópico 9 "Saiba Mais" da apostila. Banco proveniente de SINASC/DATASUS (https://datasus.saude.gov.br) com 3.616 observações e 21 variáveis, referente ao Município de Osvaldo Cruz, São Paulo, no período de 2012-2022. Extraído em 12 de dezembro de 2024. O Quadro E6.1.2 apresenta o dicionário das variáveis.

Quadro E6.1.2. Dicionário de variáveis contendo o nome original, o nome da variável, a definição e os códigos das categorias

Nome original	Definição	Códigos
DTNASC	Data de nascimento	dd/mm/aa
LOCNASC	Local de nascimento	Hospital, Outro estabelecimento de saúde, Domicílio, Outros, Ignorado (NA)
SEXO	Sexo do recém-nascido	Masculino, Feminino, Ignorado (NA)
PESO	Peso ao nascer (gramas)	numérico
APGAR1	APGAR no primeiro minuto	numérico, de 0 a 10
APGAR5	APGAR no quinto minuto	numérico, de 0 a 10
IDANOMAL	Anormalidade congênita	Sim, Não, Ignorado (NA)
SEMAGESTAC	Semanas de gestação	numérico
GESTACAO	Faixas de idade gestacional (semanas)	22-27, 28-31, 32-36, 37-41, 42+
IDADEMAE	Idade da mãe (anos)	numérico
RACACORMAE	Raça/cor da mãe	Branca, Preta, Amarela, Parda, Indígena, Ignorado (NA)
ESCMAE	Faixas de anos de escolaridade da mãe	Nenhum, 1-3, 4-7, 8-11, 12+, Ignorado (NA)
ESTCIVMAE	Estado civil da mãe	Casada, Solteira, Separada judicialmente, Viúva, União consensual, União estável, Ignorado (NA)
QTDFILMORT	Número de perdas fetais/ abortos anteriores	numérico
QTDFILVIVO	Número de nascidos vivos anteriores	numérico
QTDGESTANT	Número de gestações anteriores	numérico
PARIDADE	Número de partos anteriores	numérico
PARTO	Tipo de parto	Cesáreo,Vaginal, Ignorado (NA)
GRAVIDEZ	Tipo de gravidez	única, dupla, tripla ou +, Ignorado (NA)
CONSULTAS	Faixas de número de consultas de pré-natal	Nenhuma, 1-3, 4-6, 7 ou +, Ignorado (NA)
CONSPRENAT	Número de consultas de pré-natal	numérico

6.1.2. Explorando a variável peso ao nascer por meio do box plot para investigar existência de valores discrepantes segundo ano de nascimento.

Analyses > Exploration > Descriptives > Variables: PESO > Split by: ano_nasc > Plots > Box plot (Ver Figura E6.1.2)

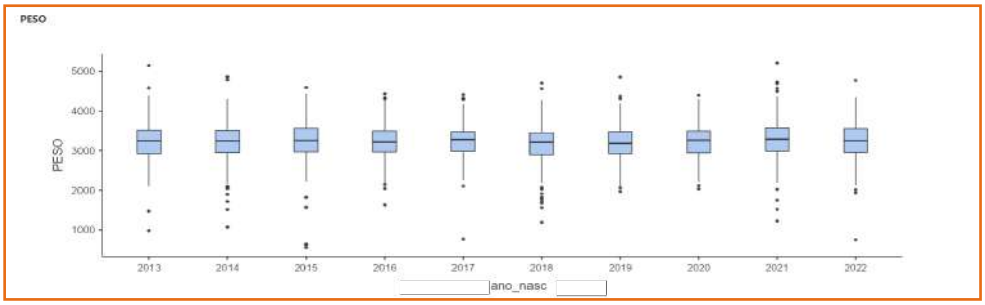


Figura E6.1.2. Boxplot da variável peso ao nascer segundo ano de nascimento

6.1.3. Explorando os dados de peso ao nascer segundo sexo por meio do gráfico de densidade.

Analyses > Exploration > Descriptives > Variables: PESO > Split by: sexo > Plots > Density (Ver Figura E6.1.3)

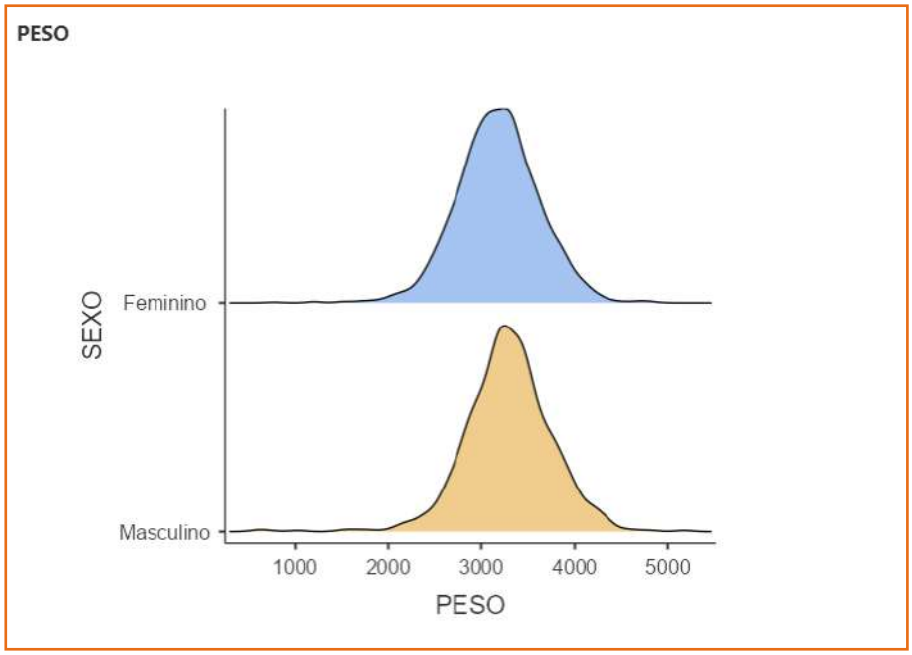


Figura E6.1.3. Gráfico de distribuição dos recém-nascidos segundo peso e sexo

6.1.4. Obtendo o gráfico de dispersão das variáveis de idade gestacional e peso ao nascer.

Analyses > Exploration > Scatter plot > X-Axis: SEMAGESTAC > Y-Axis: PESO > Regression Line: None > Marginals: None (Ver Figura E6.1.4)

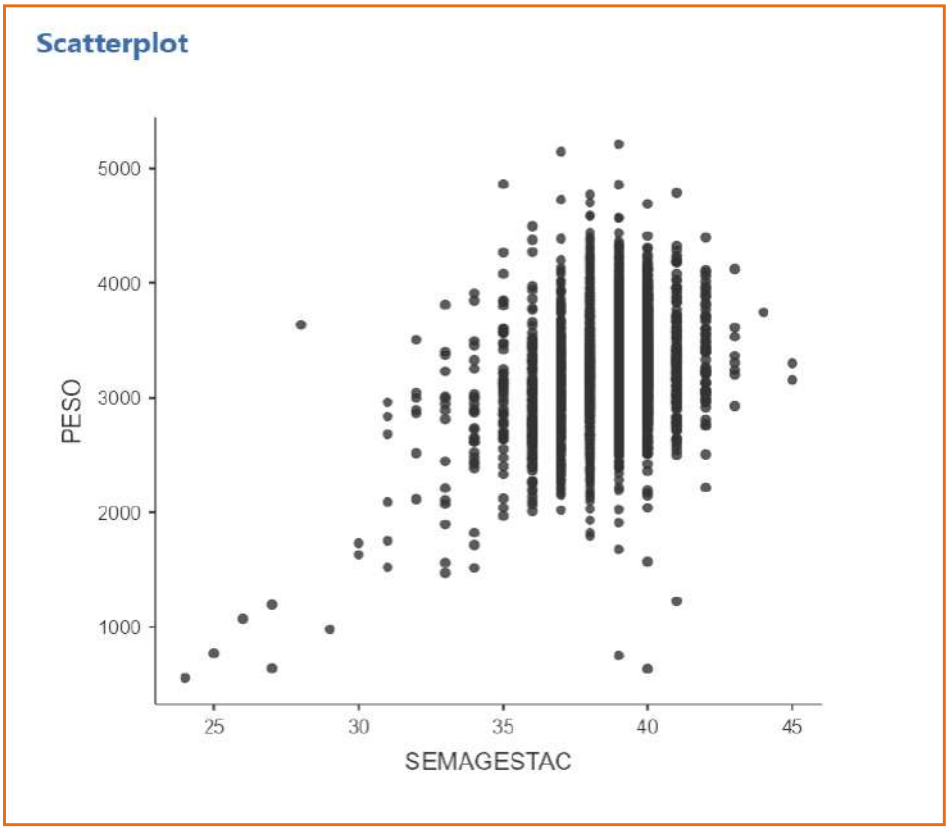


Figura E6.1.4. Gráfico de dispersão das variáveis de idade gestacional e peso ao nascer

6.1.5. Calculando o coeficiente de correlação linear de Pearson entre as variáveis peso ao nascer e idade gestacional.

Analyses > Regression > Correlation Matrix > Variáveis: SEMAGESTAC, PESO > Correlation Coefficients: Pearson > Additional Options: Report significance > Hypothesis: Correlated (Ver Figura E6.1.5)

Correlation Matrix

Correlation Matrix

		SEMAGESTAC	PESO
SEMAGESTAC	Pearson's r	—	
	df	—	
	p-value	—	
PESO	Pearson's r	0.299	—
	df	3159	—
	p-value	< .001	—

Figura E6.1.5. Resultados da correlação linear de Pearson para as variáveis idade gestacional e peso ao nascer

Observação: Como o valor de p é pequeno (p < 0,001) pode-se afirmar que existe evidência estatística de que os dados são provenientes de população onde existe correlação linear entre as variáveis.

6.1.6. Recalculando o Coeficiente de Correlação Linear de Pearson após remover possíveis outliers.

Criando variável de identificação de possíveis outliers considerando peso ao nascer acima de 3000 gramas e semana gestacional abaixo de 30 semanas; e peso ao nascer abaixo de 1000 gramas e semana gestacional acima de 35 semanas (Figura E6.1.6.1).

Variables>Transform>flag_outliers_peso_idade>Source Variable: PESO>-Create New Transform>

condição 1 (fórmula 1): `if $source > 3000 and SEMAGESTAC < 30 use 1`

condição 2 (fórmula 2): `if $source < 1000 and SEMAGESTAC > 35  use 1`

condição 3 (fórmula 3): `else use 0`

Criando filtro para remover possíveis outliers:

Data>Filters> fórmula: `flag_outlier_peso_idade != 1` (Deixar active)

Figura E6.1.6.1. Criação de filtro para remover possíveis outliers

Recalculando o Coeficiente de Correlação de Pearson sem as observações filtradas (Ver Figura E6.1.6.2):

Analyses > Regression > Correlation Matrix > Variáveis: SEMAGESTAC, PESO > Correlation Coefficients: Pearson > Additional Options: Report significance > Hypothesis: Correlated

Correlation Matrix

Correlation Matrix

		PESO	SEMAGESTAC
PESO	Pearson's r	—	
	df	—	
	p-value	—	
SEMAGESTAC	Pearson's r	0.308	—
	df	3157	—
	p-value	< .001	—

Figura E6.1.6.2. Resultados do Coeficiente de Correlação de Pearson para as variáveis peso ao nascer e idade gestacional com filtro para as observações que são possíveis outliers

Obtendo o gráfico de dispersão das variáveis de idade gestacional e peso ao nascer com o filtro removendo possíveis outliers.

Analyses > Exploration > Scatter plot > X-Axis: SEMAGESTAC > Y-Axis: PESO > Regression Line: None > Marginals: None (Ver Figura E6.1.6.3)

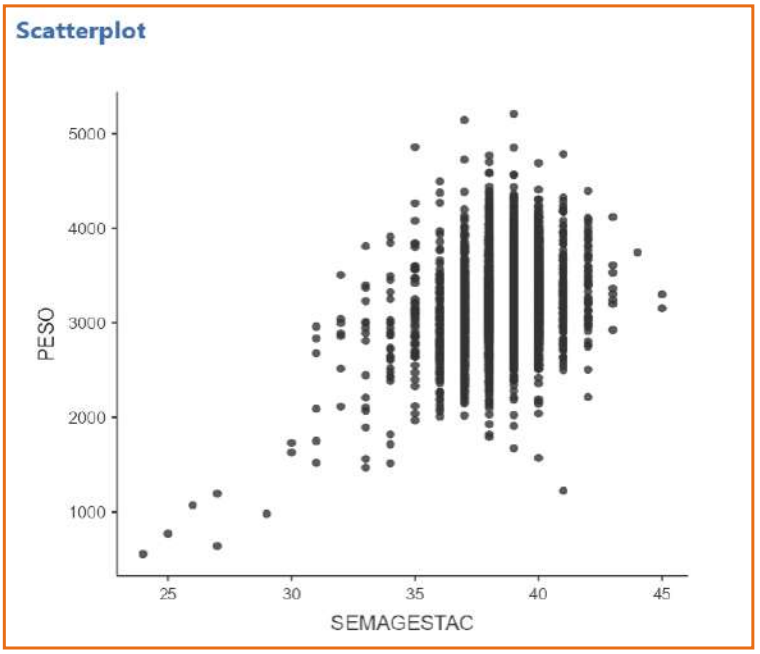


Figura E6.1.6.3. Gráfico de dispersão das variáveis de idade gestacional e peso ao nascer com fitro para remover possíveis outliers

Observação: Não se esqueça de desativar o filtro após o exercício. Para desativar, basta seguir o caminho (Ver Figura E6.1.6.4):

Data > Filters > Selecionar o filtro criado > Mudar para inactive

Figura E6.1.6.4. Desativar filtro

7. ANÁLISE DE VARIÁVEIS QUALITATIVAS

No módulo Frequencies pode-se encontrar vários modelos de análise de dados qualitativos. Nesse material será focado no subtópico Tabelas de Contingência, o teste de associação pelo Qui-quadrado de Pearson para amostras independentes (Figura 54).

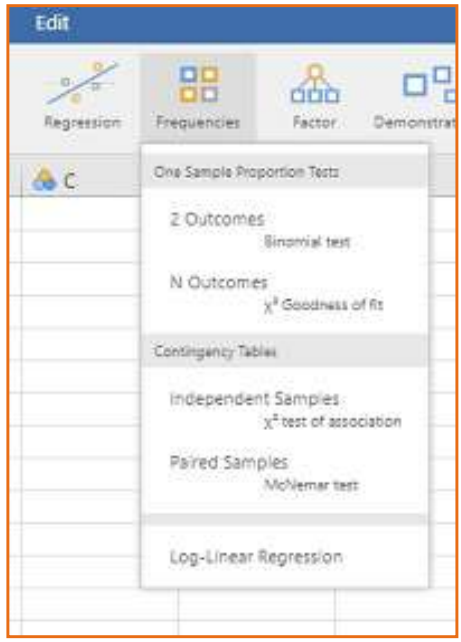


Figura 54. Teste de associação para amostras

O subtópico "Contingency Tables" será exemplificado com dados do estudo que inclui 118 pacientes psiquiátricas (Quadro E3.3.7, banco fem.omv). Selecionar na linha (rows) ans (ansiedade) e na coluna (Columns) vida para fazer a tabela de contingência (Figura 55).

Selecionar χ^2

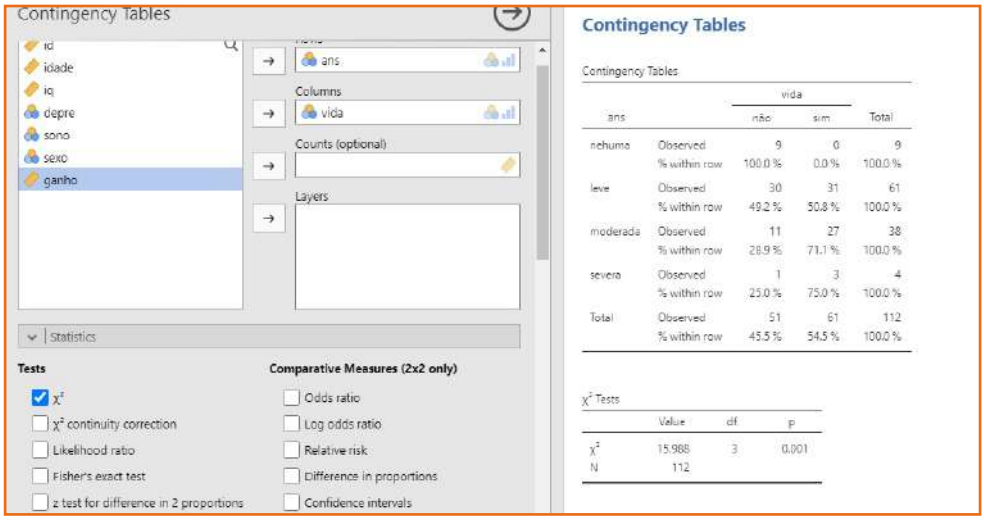


Figura 55. Módulo Contingency Tables; subtópico Tests

Selecionar porcentagem na linha, no menu Cells e, em Counts, selecionar Observed counts (Figura 56).

Em desenhos de corte transversal, a variável que deve ser fixada para receber o 100% é a variável (ansiedade) que "explica" o desfecho. A variável desfecho é vida (pensamento suicida).

Em Ordinal selecione Gamma para calcular o coeficiente de associação que indicará a força da associação.

Em Hypothesis (hipótese alternativa) selecione teste bicaudal, Group1 Group 2

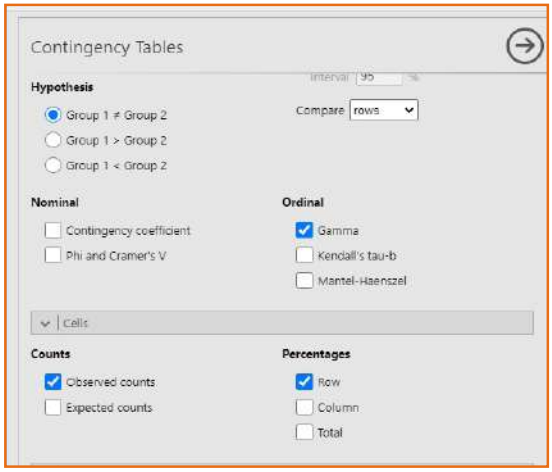


Figura 56. Módulo Contingency Tables; subtópico Hyphothesis

Em Plots selecione Bar chart, em Y selecione “Percentage within rows” porque dada a condição de depressão pode-se verificar a porcentagem de participantes que tiveram pensamento suicida.

O tipo de gráfico é o Stacked com categorias da variável ans no eixo X (Figura 57).

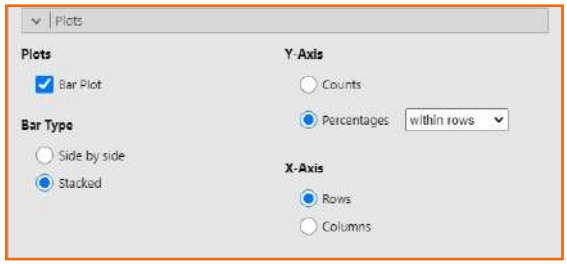


Figura 57. Módulo Contingency Tables; subtópico Plots

Observa-se que a porcentagem de pacientes que tiveram pensamento suicida aumenta com o grau de ansiedade (Figura 58).

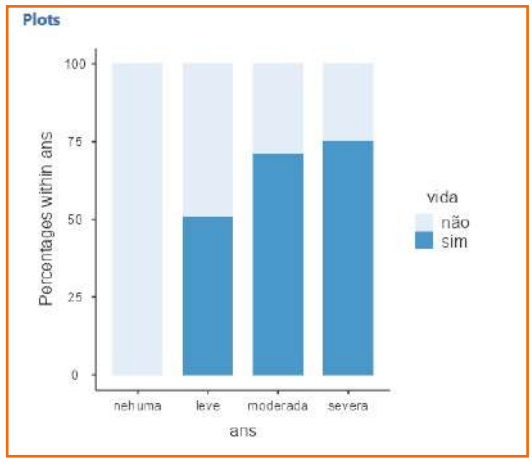
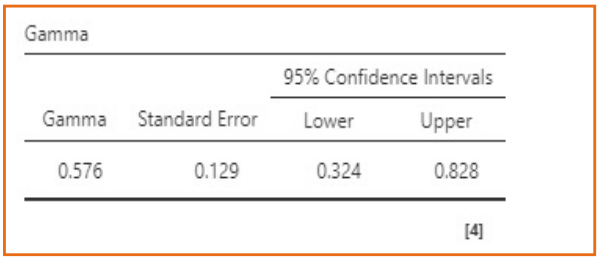


Figura 58. Distribuição de pacientes segundo grau de ansiedade e pensamento suicida

Considerando-se o resultado do teste (Figura 59), o gráfico (Figura 58), as porcentagens apresentadas na tabela de contingência e o coeficiente de associação Gamma, pode-se afirmar que existe associação entre as variáveis vida (pensamento suicida) e ans (graus de ansiedade) (p=0,001) e Coeficiente Gamma=0,576, que indica força de associação considerável.



Gamma

		95% Confidence Intervals	
Gamma	Standard Error	Lower	Upper
0.576	0.129	0.324	0.828

[4]

Figura 59. Resultados do teste Gamma

7.1. Exercícios

Considere o banco de dados "dados_exercicio_sinasc_jamovi.csv" (Bergamaschi, 2024). Também é possível reproduzir o exercício de geração desta base de dados seguindo o tópico 9 "Saiba Mais" da apostila. Banco proveniente de SINASC/DATASUS (https://datasus.saude.gov.br) com 3.616 observações e 21 variáveis, referente ao Município de Osvaldo Cruz, São Paulo, no período de 2012-2022. Extraído em 12 de dezembro de 2024. O Quadro E7.1 apresenta o dicionário das variáveis.

Quadro E7.1. Dicionário de variáveis contendo o nome original, o nome da variável, a definição e os códigos das categorias

Nome original	Definição	Códigos
DTNASC	Data de nascimento	dd/mm/aa
LOCNASC	Local de nascimento	Hospital, Outro estabelecimento de saúde, Domicílio, Outros, Ignorado (NA)
SEXO	Sexo do recém-nascido	Masculino, Feminino, Ignorado (NA)
PESO	Peso ao nascer (gramas)	numérico
APGAR1	APGAR no primeiro minuto	numérico, de 0 a 10
APGAR5	APGAR no quinto minuto	numérico, de 0 a 10
IDANOMAL	Anormalidade congênita	Sim, Não, Ignorado (NA)
SEMAGESTAC	Semanas de gestação	numérico
GESTACAO	Faixas de idade gestacional (semanas)	22-27, 28-31, 32-36, 37-41, 42+
IDADEMAE	Idade da mãe (anos)	numérico

continua...

Nome original	Definição	Códigos
RACACORMAE	Raça/cor da mãe	Branca, Preta, Amarela, Parda, Indígena, Ignorado (NA)
ESCMAE	Faixas de anos de escolaridade da mãe	Nenhum, 1-3, 4-7, 8-11, 12+, Ignorado (NA)
ESTCIVMAE	Estado civil da mãe	Casada, Solteira, Separada judicialmente, Viúva, União consensual, União estável, Ignorado (NA)
QTDFILMORT	Número de perdas fetais/ abortos anteriores	numérico
QTDFILVIVO	Número de nascidos vivos anteriores	numérico
QTDGESTANT	Número de gestações anteriores	numérico
PARIDADE	Número de partos anteriores	numérico
PARTO	Tipo de parto	Cesáreo,Vaginal, Ignorado (NA)
GRAVIDEZ	Tipo de gravidez	única, dupla, tripla ou +, Ignorado (NA)
CONSULTAS	Faixas de número de consultas de pré-natal	Nenhuma, 1-3, 4-6, 7 ou +, Ignorado (NA)
CONSPRENAT	Número de consultas de pré-natal	numérico

7.1.1. Investigar a existência de associação entre as variáveis peso ao nascer dicotomizada (exercício 6.5.1.1) e semanas de gestação categorizadas em pré-termo e não pré-termo.

Inicialmente, deve-se filtrar os dados de peso ao nascer somente para partos únicos (GRAVIDEZ=='única') e semanas de gestação diferente de ignorado (CONSULTAS!=NA).

Criação de filtros:

Data > Filters > Add new filter (+) > Fórmula: `GRAVIDEZ == 'única'`

Data > Filters > Add new filter (+) > Fórmula: `SEMAGESTAC != NA`

Criação da variável semanas de gestação dicotomizada “pre_termo”:

Variables > Transform > Variable: pre_termo > Source Variable: SEMAGESTAC > Create new transform > Fórmula: `if $source < 37 use 'pre_termo' else use 'nao_pre_termo'` (Figura 7.1.1.1)

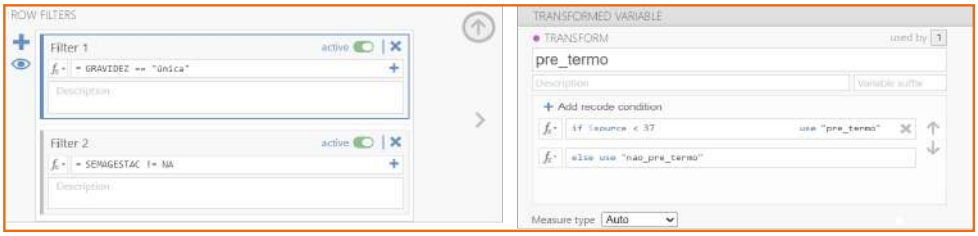


Figura 7.1.1.1. Criação de filtros e da nova variável

Análise:

Analyses > Frequencies > Contingency Tables > Independent samples x^2 test of association > Rows: pre_termo > Columns: baixo_peso > Tests: selecionar x^2 > Hypothesis: selecionar Group 1 Group 2 > Nominal: selecionar Contingency coefficient > Percentages: selecionar Row > Plots: selecionar Bar Plot, selecionar Bar Type Stacked, selecionar Y-Axis Percentages within rows, X-Axis selecionar Rows

A Figura 7.1.1.2 apresenta os resultados da análise.

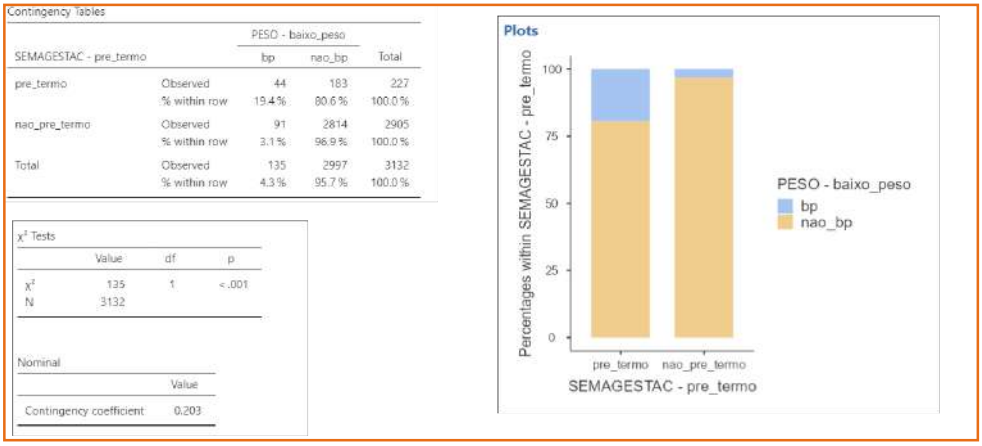


Figura 7.1.1.2. Resultados da análise

8. MODELOS DE REGRESSÃO

No módulo Regression 

pode-se executar modelos de regressão (Figura 47).

Utilizar os dados do Quadro 4.

Quadro 4. Rol de pacientes segundo características pressão arterial e idade

Indivíduo	PAS (Y)	Idade (X)		Indivíduo	PAS (Y)	Idade (X)
1	144	39		16	130	48
2	220	47		17	135	45
3	138	45		18	114	17
4	145	47		19	116	20
5	162	65		20	124	19
6	142	46		21	136	36
7	170	67		22	142	50
8	124	42		23	120	39
9	158	67		24	120	21
10	154	56		25	160	44
11	162	64		26	158	53
12	150	56		27	144	63
13	140	59		28	130	29
14	110	34		29	125	25
15	128	42		30	175	69

Fonte: Kleinbaum (1988).

Ativar o Módulo Regression e selecionar Linear Regression (Figura 60).

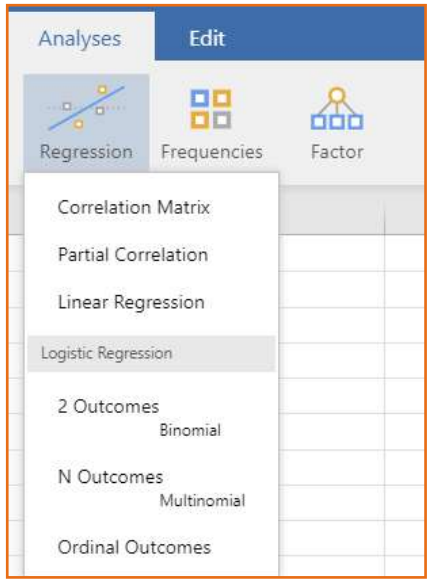


Figura 60. Módulo Regression

8.1. Regressão linear simples

Verificação de normalidade, uma das pressuposições para o modelo de regressão linear simples. Para tanto, utilizou-se o módulo Exploração. Observa-se que a distribuição da pressão arterial não segue distribuição normal (teste de Shapiro-Wilk; p=0,014) (Figura 61). Excluindo-se o par de valores (x=47; y=220), observa-se que a variável pressão arterial passa a ter distribuição normal (Shapiro-Wilk, p=0,642 (Figura 62).

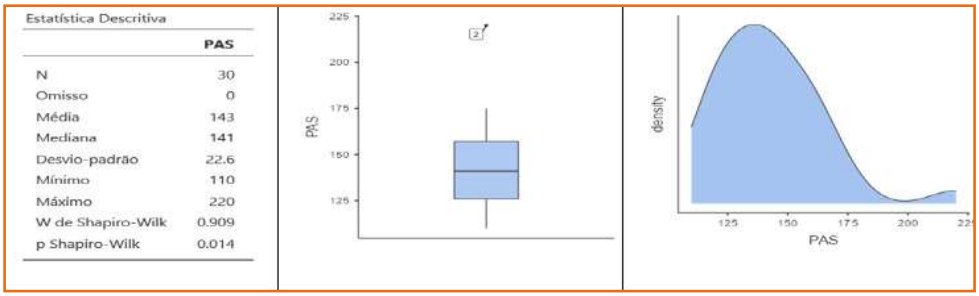


Figura 61. Estatísticas descritivas, box plot e distribuição da Pressão Arterial (PAS)

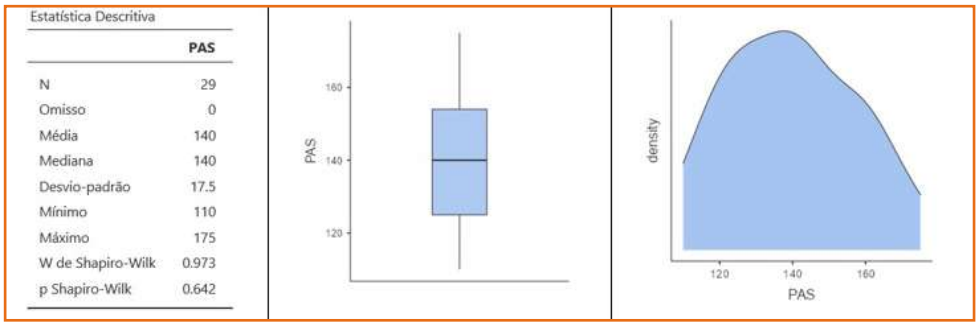


Figura 62. Estatísticas descritivas, box plot e distribuição da Pressão Arterial (PAS) após retirar o par de valores (47,220)

Gráfico de dispersão entre as variáveis utilizando-se o Módulo Exploração (Figura 63).

Por meio do gráfico observa-se que a relação entre as variáveis é linear. Isso indica que é adequado fazer o modelo de regressão linear simples para explicar a variabilidade da pressão arterial sistólica em função da idade (Figura 63).

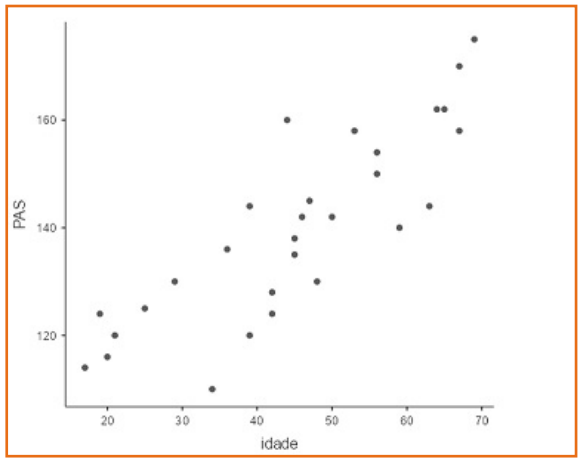


Figura 63. Diagrama de dispersão entre Pressão arterial sistólica

Após a ativação de "Linear regression", na tela que aparece, selecionar a variável dependente ("Dependent Variable") e a variável independente de interesse principal (em "Covariates" para quantitativas ou em "Factors para qualitativas"). No exemplo, escolher a variável pressão arterial (mmHg) como dependente e idade (anos) como independente (Figura 64).

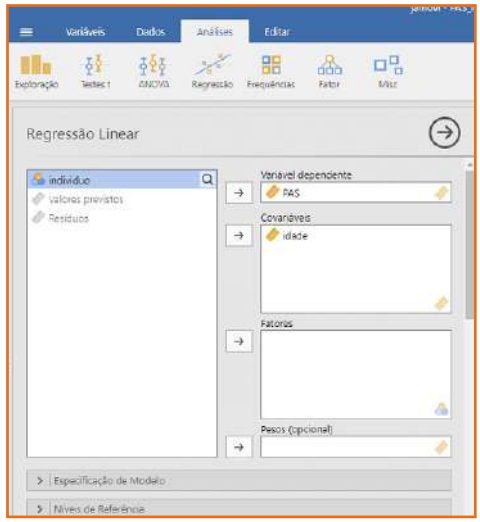


Figura 64. Entrada de variáveis para a realização da regressão linear

Na Figura 65, aba Coeficientes do Modelo, pode-se verificar por meio da ANOVA que idade é importante para explicar a pressão arterial ($p<0,001$).

O coeficiente da idade é diferente de zero (teste T, p<0,001). Os coeficientes do modelo permitem escrever a equação do modelo: $\hat{y} = 97,1 + 0,95\ (idade)$. O coeficiente angular indica que para cada ano a mais de idade, a pressão arterial sistólica aumenta em média, 0,95 mmHg. O modelo está bem ajustado aos dados (p<0,001) e o coeficiente de explicação = 0,712 indicando que idade explica 71,2% da variabilidade da pressão arterial sistólica.

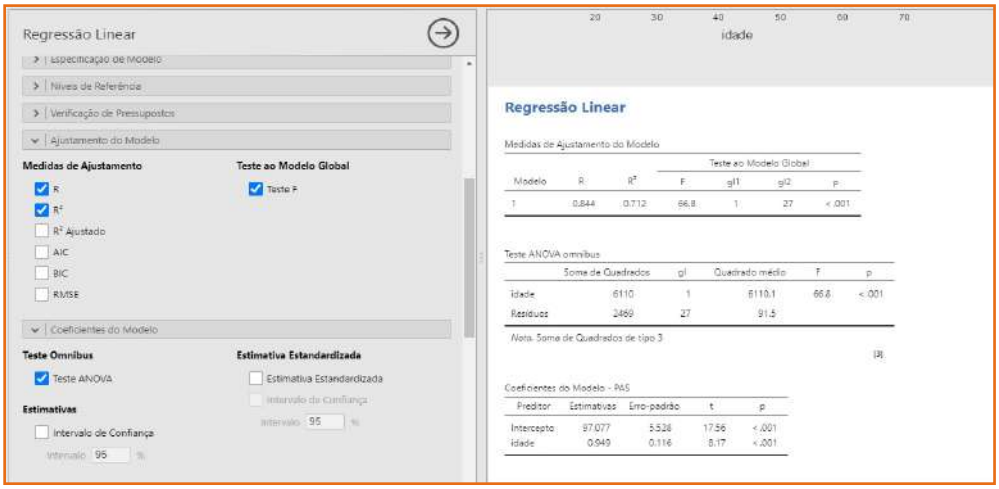


Figura 65. Saída das abas que compõem o Módulo Regress

A função linear ajustada aos dados pode ser vista na Figura 66.

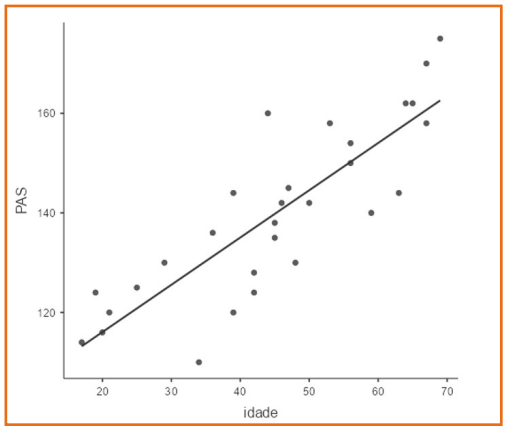


Figura 66. Reta de regressão ajustada aos dados de pressão arterial sistólica em função da idade

Diagnóstico do modelo

Na aba Gravar, solicitar o cálculo dos valores preditos pelo modelo e dos resíduos ($y_{observado}$-$y_{preditos}$). Esses valores são necessários para o diagnóstico do modelo por meio dos gráficos e teste de normalidade dos resíduos (Figura 67).

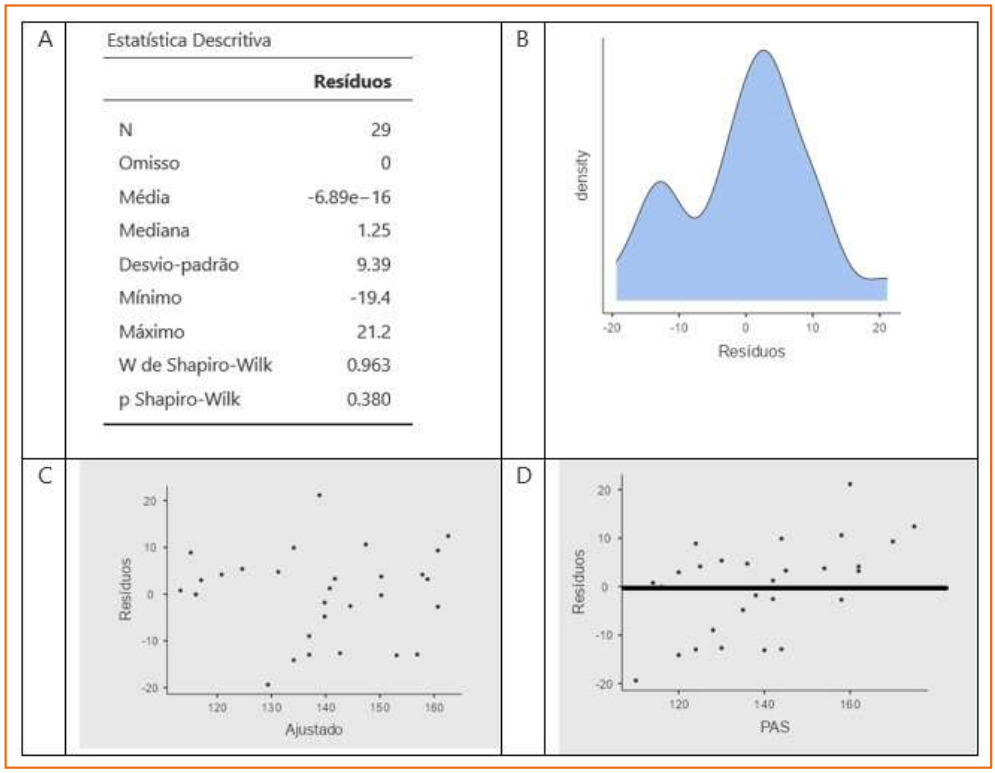


Figura 67. Resumo dos dados dos resíduos, teste de normalidade dos resíduos, gráficos de resíduos

8.2. Exercícios – Base de Dados

Considere a base de dados sobre poluição atmosférica de 41 cidades nos Estados Unidos, coletados por Sokal e Rohlf (1981). A variável dependente é a concentração anual de dióxido de enxofre (SO_2) em microgramas por metro cúbico. Os dados de SO_2 correspondem à média dos anos 1969 a 1971. Os valores de 6 variáveis explanatórias são também apresentados, duas das quais relacionadas a ecologia humana e quatro com o clima.

Considere o conteúdo dos Quadros E8.2.1 e E8.2.2 para inserir adequadamente os dados da base de poluição atmosférica no **Jamovi.**

Quadro E8.2.1. Dicionário de variáveis da base de dados sobre poluição atmosférica

Nome da variável	Descrição da varável	Nome da variável no banco de dados
Cidade	cidade onde os dados foram coletados	*cidade*
id	número de identificação da cidade	*id*
SO_2	dióxido de enxofre	*so2*
Temp	temperatura média anual em ^{0}F (Fahrenheit)	*temp*
Manufatura	número de fábricas com 20 ou mais trabalhadores	*manufa*
Pop	tamanho da população (censo de 1970) em milhares	*pop*
Vento	velocidade média anual em milhas por hora	*vento*
Precipitação	precipitação média anual (polegadas)	*precipita*
Dias	número médio anual de dias de precipitação	*dias*

Fonte: Adaptado de Hand et al. (1994).

Quadro E8.2.2. Conjunto de dados sobre poluição atmosférica

cidade	*id*	*so2*	*temp*	*manufa*	*pop*	*vento*	*precipita*	*dias*
Phoenix	1	10	70.3	213	582	6.0	7.05	36
Little Rock	2	13	61.0	91	132	8.2	48.52	100
San Francisco	3	12	56.7	453	716	8.7	20.06	67
Denver	4	17	51.9	454	515	9.0	12.95	86
Hartford	5	56	49.1	412	158	9.0	43.37	127
Wilmington	6	36	54.0	80	80	9.0	40.25	114
Washington	7	29	57.3	434	757	9.3	38.89	111
Jackson	8	14	68.4	136	529	8.8	54.47	116
Miami	9	10	75.5	207	335	9.0	59.80	128
Atlanta	10	24	61.5	368	497	9.1	48.34	115
Chicago	11	110	50.6	3344	3369	10.4	34.44	122
Indianapolis	12	28	52.3	361	746	9.7	38.74	121
Des Moines	13	17	49.0	104	201	11.2	30.85	103

continua...

cidade	id	so2	temp	manufa	pop	vento	precipita	dias
Wichita	14	8	56.6	125	277	12.7	30.85	82
Louisville	15	30	55.6	291	593	8.3	43.11	123
New Orleans	16	9	68.3	204	361	8.4	56.77	113
Baltimore	17	47	55.0	625	905	9.6	41.31	111
Detroit	18	35	49.9	1064	1513	10.1	30.96	129
Minneapolis	19	29	43.5	699	744	10.6	25.94	137
Kansas	20	14	54.5	381	507	10.0	37.00	99
St Louis	21	56	55.9	775	622	9.5	35.89	105
Omaha	22	14	51.5	181	347	10.9	30.18	98
Albuquerque	23	11	56.8	46	244	8.9	7.77	58
Albany	24	46	47.6	44	116	8.8	33.36	135
Buffalo	25	11	47.1	391	463	12.4	36.11	166
Cincinnati	26	23	54.0	462	453	7.1	39.04	132
Cleveland	27	65	49.7	1007	751	10.9	34.99	155
Columbia	28	26	51.5	266	540	8.6	37.01	134
Philadelphia	29	69	54.6	1692	1950	9.6	39.93	115
Pittsburgh	30	61	50.4	347	520	9.4	36.22	147
Providence	31	94	50.0	343	179	10.6	42.75	125
Memphis	32	10	61.6	337	624	9.2	49.10	105
Nashville	33	18	59.4	275	448	7.9	46.00	119
Dallas	34	9	66.2	641	844	10.9	35.94	78
Houston	35	10	68.9	721	1233	10.8	48.19	103
Salt Lake City	36	28	51.0	137	176	8.7	15.17	89
Norfolk	37	31	59.3	96	308	10.6	44.68	116
Richmond	38	26	57.8	197	299	7.6	42.59	115
Seattle	39	29	51.1	379	531	9.4	38.79	164
Charleston	40	31	55.2	35	71	6.5	40.75	148
Milwaukee	41	16	45.7	569	717	11.8	29.07	123

8.2.1. Exercício

A) Investigar a distribuição da variável desfecho “so2”. Caso não apresentar distribuição normal, utilizar a transformação logarítmica (função ln) e testar se a variável transformada segue uma distribuição normal. Verifique também a presença de possíveis valores discrepantes.

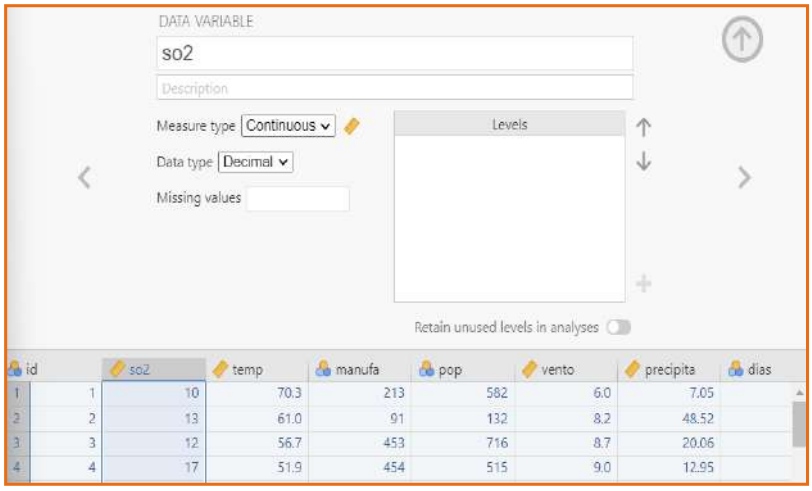
B) Se as pressuposições A for satisfeita, e, assumindo-se que existe homogeneidade de variâncias, investigar se a variável manufa contribui para explicar a variabilidade do nível de so2 por meio de uma regressão linear simples. Verificar o valor do coeficiente de determinação do mo-

delo e se os resíduos apresentam distribuição normal e são aleatórios em torno do 0.

C) Apresentar o gráfico de dispersão com a reta de regressão ajustada.

Resolução:

Abra o **Jamovi** e copie e cole os dados do Quadro E8.2.2 para a planilha Data. Note que devem ser copiados apenas os valores numéricos, sem o cabeçalho, que devem ser incluídos após a inserção dos dados. Observe que ao copiar os dados, o **Jamovi** não assume automaticamente o tipo correto de todas as variáveis; isso deve ser feito manualmente. Por exemplo, a variável so2 deve ser declarada como Measure Type "Continuous" e Data Type "Decimal" para diferenciar do measure type "Continuous" e Data Type "Integer", que seria referente a uma variável numérica discreta (contagem) (Figura 8.2.1.1).



	id	so2	temp	manufa	pop	vento	precipita	dias
1	1	10	70.3	213	582	6.0	7.05	
2	2	13	61.0	91	132	8.2	48.52	
3	3	12	56.7	453	716	8.7	20.06	
4	4	17	51.9	454	515	9.0	12.95	

Figura 8.2.1.1. Exemplo para atribuição do tipo de medida e de dado da variável so2

A) Analyses > Exploration > Descriptives > Variables: so2 > Statistics: selecionar Normality Shapiro-Wilk > Plots: selecionar Box plot

Os resultados da seleção podem ser observados na Figura 8.2.1.2.

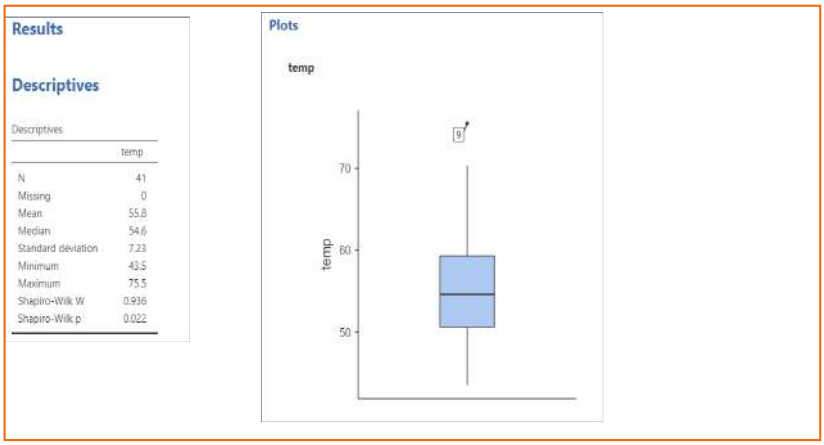


Figura 8.2.1.2. Resultados da seleção

Verifica-se a partir do teste de normalidade Shapiro-Wilk que so2 não segue distribuição normal (p-valor = 0,022) (para seguir distribuição normal, o valor de p teria que ser maior que 0,05).

Gerando a nova variável so2 com transformação ln (logaritmo neperiano):

Variables > Transform > Transformed Variable: so2_ln > Source Variable: so2 > using transform: Create New Transform > Fórmula: `ln($source)`

Testando se a variável transformada segue distribuição normal:

Analyses > Exploration > Descriptives > Variables: so2_ln > Statistics: selecionar Normality Shapiro-Wilk

Os resultados da seleção podem ser observados na Figura 8.2.1.3.

Descriptives

Descriptives

	so2_ln
N	41
Missing	0
Mean	3.15
Median	3.26
Standard deviation	0.702
Minimum	2.08
Maximum	4.70
Shapiro-Wilk W	0.954
Shapiro-Wilk p	0.097

Figura 8.2.1.3. Resultados do teste de normalidade para a variável so2 transformada

B) Analyses > Regression > Linear Regression > Dependent Variable: so2_ln > Covariates: manufa > Assumption Checks: selecionar Normality test e Residual plots

Os resultados da análise podem ser observados na Figura 8.2.1.4.

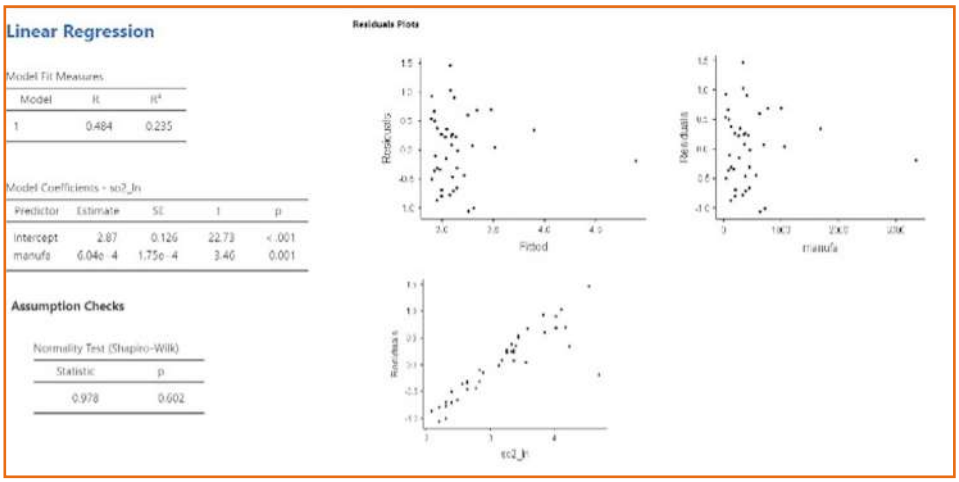


Figura 8.2.1.4 Resultados da regressão linear simples

C) Analyses > Descriptives > Scatterplot > X-Axis: so2_ln > Y-Axis: manufa > Regression Line: Linear

Os resultados da análise podem ser observados na Figura 8.2.1.5.

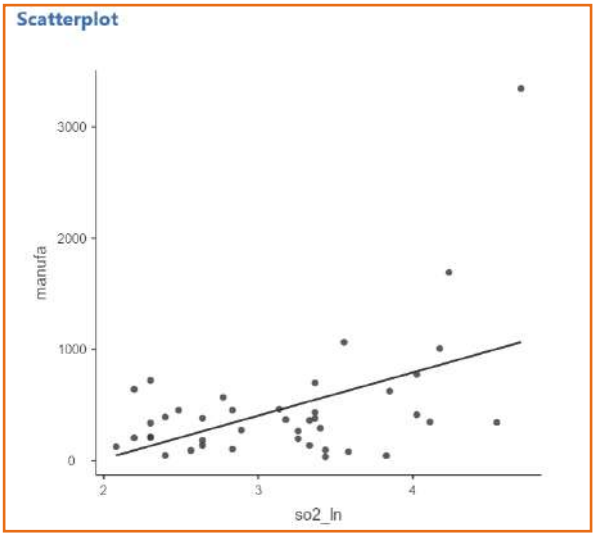


Figura 8.2.1.5. Gráfico de dispersão e reta estimada

Observa-se que o modelo não está bem ajustado, informado pelo valor de R^2 = 0,235, indicando que 23,5% da variabilidade de so2 é explicada pelo número de fábricas com 20 ou mais empregados. É possível que a relação entre as variáveis não seja linear e isso pode ser visto pelo gráfico de dispersão (Figura 8.2.1.5) que sugere uma relação não-linear entre as variáveis.

O gráfico de dispersão com regression line smooth (Figura 8.2.1.6) reforça a relação não-linear entre as variáveis.

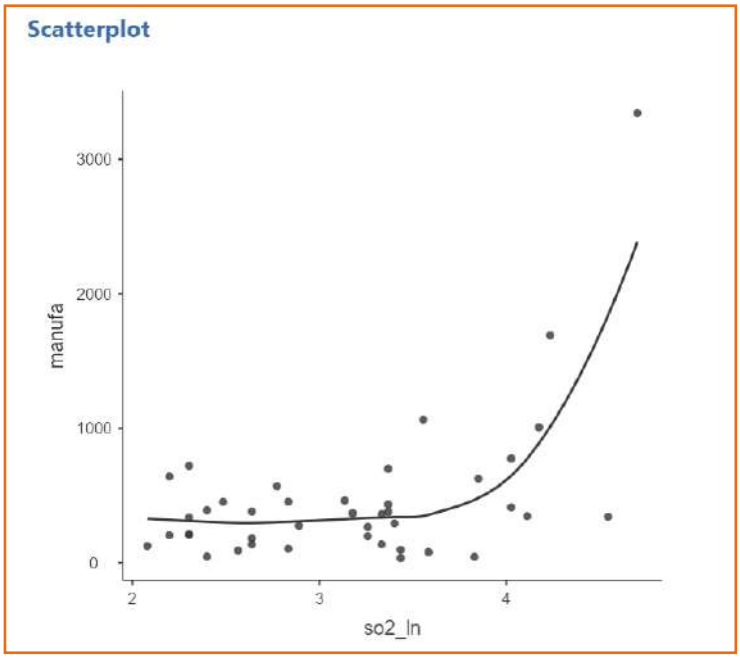


Figura 8.2.1.6. Gráfico de dispersão com opção de Regression Line: Smooth

9. SAIBA MAIS

É possível extrair dados diretamente do DATASUS por meio do TABNET (para consultas com até duas variáveis), procedendo da seguinte forma:

Na página oficial (Figura 68), escolher a informação de interesse, por exemplo, “Nascidos Vivos - desde 1994” (Figura 69).

Figura 68. Página oficial do TABNET

Fonte: DATASUS, [ca. 2024].

Figura 69. Opções de informação do TABNET

Fonte: DATASUS, [ca. 2024].

Em seguida, escolher as opções desejadas (Figura 70).

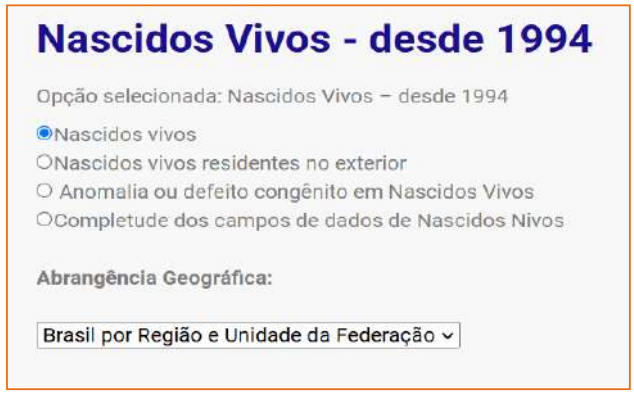


Figura 70. Opções para o exemplo de informações de Nascidos Vivos

Fonte: DATASUS, [ca. 2024].

E as variáveis de interesse (uma na linha e outra na coluna, ou apenas uma na linha ou na coluna) (Figura 71).

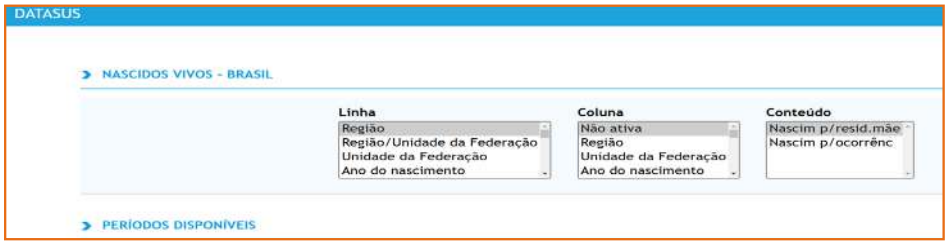


Figura 71. Seleção de variáveis de interesse

Fonte: DATASUS, [ca. 2024].

Selecionar “Mostra” (Figura 72).

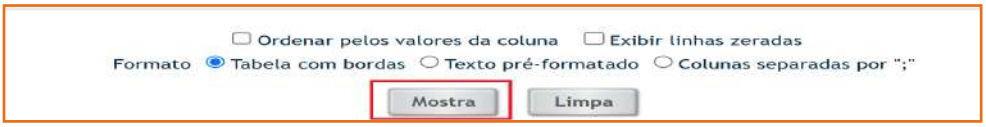


Figura 72. Seletor para gerar a extração desejada

Fonte: DATASUS, [ca. 2024].

Também é possível extrair dados diretamente do DATASUS por meio do TABWIN (para consultas com mais variáveis), procedendo da seguinte forma (Figura 73).

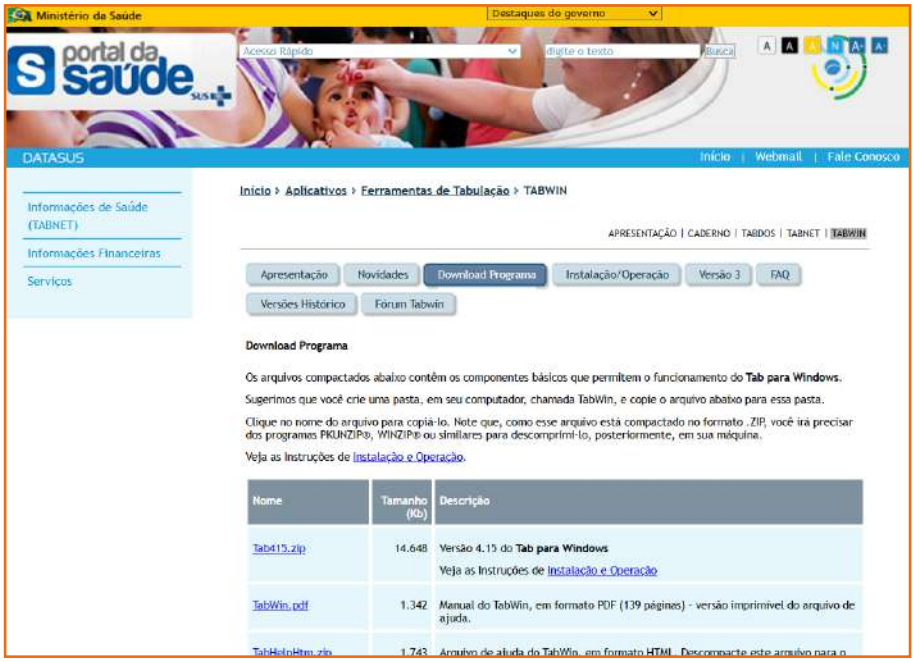


Figura 73. Página oficial de download do programa TABWIN

Fonte: DATASUS (2008).

É necessário instalar localmente os arquivos de dados brutos para acessá-los pelo programa TABWIN previamente instalado. Os arquivos de dados brutos referentes às análises desejadas podem ser encontrados na página oficial “Transferência de Arquivos” do DATASUS (Figura 74).

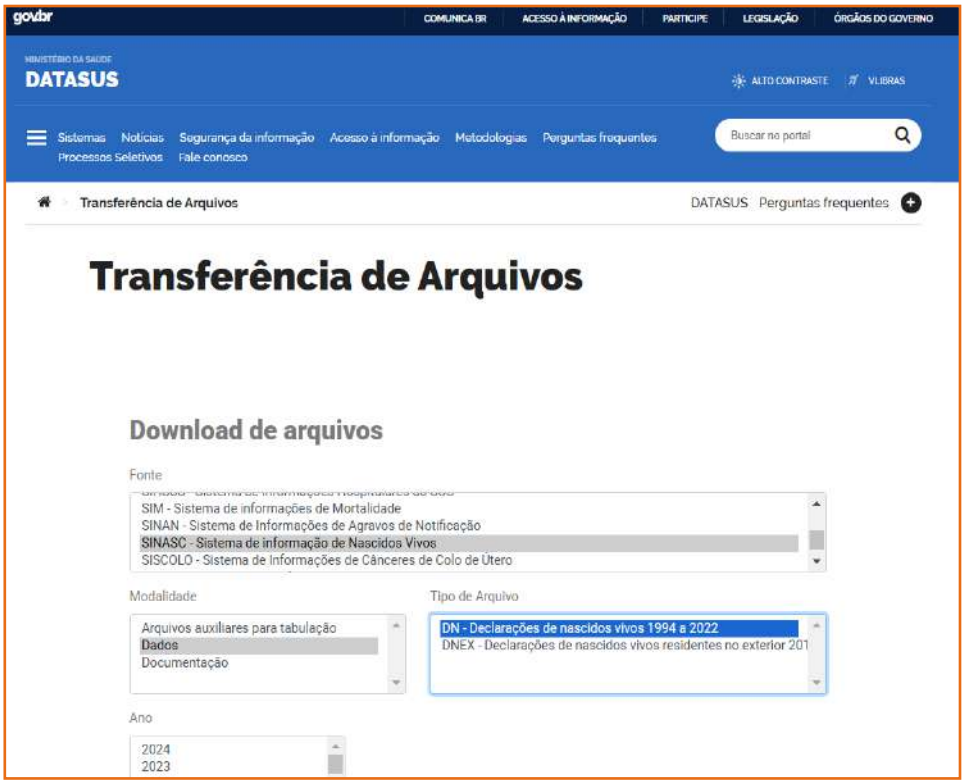


Figura 74. Página oficial de downloads de arquivos de dados brutos dos sistemas de informação do DATASUS

Fonte: DATASUS [2024].

Alternativamente, é possível obter os dados através do pacote estatístico R (versão 4.4.2) (https://cran.r-project.org/bin/windows/base/) por meio do pacote "microdatasus" (versão 2.3.1), conforme ilustrado nos comandos a seguir que foram empregados para obter a base de dados de exemplo do tópico 2.2:

```
#Instalando pacotes
remotes::install_github("rfsaldanha/microdatasus")
install.packages(dplyr, dep=T)

#Carregando pacotes
library(microdatasus)
library(dplyr)
```

```
#Extraindo os dados brutos do DATASUS
dados <- microdatasus::fetch_datasus(year_start = 2012,
year_end = 2022, uf = "SP", information_system = "SINASC")
dados <- microdatasus::process_sinasc(dados)

#Filtrando os dados para o município de Osvaldo Cruz
dados_mun <- dplyr::filter(dados, CODMUNNASC==353460)

#Seleção de variáveis de interesse
dados_mun <- dplyr::select(dados_mun,DTNASC,LOCNASC, IDA-
DEMAE,ESCMAE,ESTCIVMAE, QTDFILMORT, QTDFILVIVO,QTDGESTANT,
PARIDADE, GESTACAO,PARTO, GRAVIDEZ, IDANOMAL, CONSULTAS,
APGAR1, APGAR5, RACACORMAE, PESO, SEXO, SEMAGESTAC, CONS-
PRENAT)

#Exportando os dados para formato .csv
#(write.csv2() considerando variáveis de texto em PT-BR
#file.choose(new=TRUE) para escolher manualmente o diretório
#argumento "new=TRUE" para criação de novo arquivo
write.csv2(dados_mun, (file.choose(new=TRUE)))
```

10. REFERÊNCIAS

BERGAMASCHI, D. P. **Banco de Dados**. 2024. Disponível em: https://www.fsp.usp.br/denisepb/index.php/banco-de-dados/ Acesso em: 12 dez. 2024.

DATASUS. **Tabnet**. Brasília, Ministério da Saúde: [ca. 2024]. Disponível em: https://datasus.saude.gov.br/informacoes-de-saude-tabnet/ Acesso em: 12 dez. 2024.

DATASUS. **Tabwin**. Brasília, Ministério da Saúde: 2008. Disponível em: http://siab.datasus.gov.br/DATASUS/index.php?area=060805&item=3 Acesso em: 12 dez. 2024.

DATASUS. **Transferência de arquivos**. Brasília, Ministério da Saúde: [2024]. Disponível em: https://datasus.saude.gov.br/transferencia-de-arquivos/# Acesso em: 12 dez. 2024.

HAND, D. J.; DALY, F.; LUNN, A. D.; MCCONWAY, K. J.; OSTROWSKI, E. **A handbook of small data sets.** London: Chapman and Hall, 1994.

JAMOVI. **User guide oficial do Jamovi.** [S.l., ca. 2024]. Disponível em https://www.jamovi.org/user-manual.html Acesso em: 12 dez. 2024.

KLEINBAUM, D. G., et al. **Applied regression analysis and other multivariable methods**. Belmont, CA: Duxbury press, 1988.

RABE-HESKETH, S.; EVERITT, B. **A handbook of Statistical Analyses using Stata**. London: Chapman & Hall/RC, 2004.

ROSNER, B. **Fundamentos de Bioestatística.** Tradução Noveritis do Brasil; revisão técnica; Magda Carvalho Pires. São Paulo: Cengage Learning, 2016.

SISVAN. **Notas Técnicas.** Brasília, [ca. 2024]. Disponível em: http://tabnet.https://datasus.saude.gov.br/cgi-win/SISVAN/CNV/notas_sisvan.html Acesso em: 12 dez. 2024.