Visualização de Dados do Inmetro

Manual do Usuário

André Miyazawa, Werickson Fortunato de Carvalho Rocha

Index

1

1 Introdução e Inicialização 4
1.1 Sobre o aplicativo 4
1.2 Salvar imagens 4
1.3 Layout 5

2

2 Gráficos Básicos 7
2.1 Inserção dos Dados 7
2.1.1 Importar os dados 7
2.1.2 Inserir os dados na planilha 9
2.1.3 Resumo dos dados 9
2.2 Gráficos 2D 10
2.2.1 Histograma 10
2.2.2 Box Plot 14
2.2.3 Violino 15
2.2.4 Gráfico de pontos 16
2.2.5 Gráfico de Densidade 20
2.2.6 Gráfico de Erro 21
2.3 Gráficos 3D 22
2.3.1 Histograma 3D 23
2.3.2 Gráfico de Densidade 24
2.3.3 Gráfico de pontos 26
2.3.4 Gráfico de Barras 30

3

3 Testes Estatísticos 32
3.1 Inserção dos Dados 32
3.1.1 Instruções de inserção de cada tipo de dado 32
3.1.2 Exemplos 38
3.1.3 Inserir os dados na planilha 40
3.2 Avaliando os Dados 41
3.2.1 Distribuição de dados normais 42
3.2.2 Homogeneidade das variâncias 46
3.2.3 Transformando os dados para a normalidade 47
3.2.4 Avaliando a esfericidade 50
3.3 Comparando duas médias 51
3.3.1 Teste T 51
3.3.2 Teste de Wilcoxon 54
3.3.3 Teste do Sinal (Sign Test) 57
3.4 Comparando múltiplas medias 59
3.4.1 ANOVA 59
3.4.2 ANOVA Medidas Repetidas 60
3.4.3 ANOVA Medidas Misturadas 61
3.4.4 ANCOVA 63
3.4.5 MANOVA 65
3.4.5 Teste de Friedman 70
3.4.5 Teste de Kruskal-Wallis 68

4

4 Gráficos em Mesh 72
4.1 Inserção dos Dados 72
4.1.1 Importar os dados 72
4.2 Construir Gráficos em Mesh 74

5

5 Referências 75

1 Introdução e Inicialização

1.1 Sobre o aplicativo:

Este aplicativo tem função de visualizar, comparar diversos dados e realizar testes estatísticos com múltiplas variáveis, mostrando observações e predições. O aplicativo aceita dados em arquivos ".xlsx", e possui três funcionalidades.

A primeira o usuário envia uma tabela, e consegue projetar diversos gráficos 2D e 3D, com diversas configurações, além de uma tabela mostrando dados estatísticos importantes.

A segunda funcionalidade são os testes estatísticos, em que o usuário insere a tabela desejada, e o software calcula cada um dos testes. São testes que podem ser comparados com 2 ou mais grupos.

A terceira, por fim é uma tabela específica para projetar gráficos Mesh, onde o usuário insere um arquivo ".xlsx" com diversas planilhas, e o programa mostra os gráficos.

Por fim, todos os gráficos podem ser baixados em um arquivo .png.

1.2 Salvar imagens:

Todos os gráficos possuem pelo menos um meio para salvar os gráficos em uma imagem .png

Para gráficos no modelo **plotly**, deve-se clicar no botão "**Download plot as a png**" como no exemplo abaixo:

Já para gráficos que **não utilizam o plotly** como ferramenta de visualização, terá um botão embaixo do gráfico escrito "**Baixe o gráfico em png**:" como no exemplo abaixo:

1.3 Layout:

Alguns gráficos visuais possuem uma aba, geralmente embaixo das configurações de Controle de opções, chamada **Controle de Layout**. Nesta

aba o usuário pode personalizar um gráfico específico modificando as legendas, as cores e os eixos.

2 Gráficos Básicos

A barra de navegação de dados unidimensionais serve para você inserir um ou mais dados unidimensionais, podendo verificar gráficos, distribuições, testes e outros aspectos.

Para os gráficos básicos, você deve inserir os dados como mostra na página "*2.1.1 Inserção dos Dados*". Com os dados já inseridos, a barra de navegação possui quatro funcionalidades, mostrar um resumo com algumas informações sobre os dados inseridos, um painel contendo gráficos 2D, um painel contendo gráficos 3D e um painel checando as propriedades do gráfico.

Existem 6 tipos de gráficos 2D: **Histogramas**, **Box Plot**, **Violino**, **Gráfico de pontos**, **Gráfico de Densidade** e **Gráfico de Erros**, e existem 5 tipos de gráficos 3D: **Histogramas 3D**, **Gráfico em Superfície**, **Gráfico de pontos 3D**, **Gráfico de Densidade 3D** e **Gráfico de Barras** 3D.

Na parte de checagem de dados, existe a checagem se os dados seguem uma **distribuição normal**, a **homogeneidade das variações**, **avaliando a esfericidade** e **transformando os dados** recebidos para uma distribuição normal.

2.1 Inserção dos Dados

2.1.1 Importar os dados:

Existem duas opções para a **importação** dos dados pelo usuário, dados de exemplo que já vem baixados com o aplicativo e inserir dados de uma planilha Excel.

Para modificar o título, o usuário deve escrever um título abaixo do texto "Digite o título".

Para os dados exemplares você deve selecionar o painel:

Gráficos Básicos -> **Carregue seus dados** -> **Importe seus dados** -> **Usar um arquivo exemplo**

Existem 6 opções de exemplos: "**Obs de Laboratório**" que são dados inteiros e aleatórios, cada um com a aleatoriedade começando de um certo valor e terminando em outro.

"**Distribuições aleatórias**" são diversas distribuições com dados aleatórios, possui a distribuição normal, de Poisson, Binomial com 50% para cada. Exponencial e de Chi-Quadrado.

"**Votações na Florida**" mostra o número de pessoas que votaram para cada um dos candidatos a presidente no ano de 2000. Cada uma das observações é de uma determinada cidade da Flórida.

"**Construções**", que mostra o número de construções de cada uma das quatro regiões do Centro-Oeste dos Estados Unidos, para cada um dos meses, desde janeiro de 2018 até setembro de 2018.

"**Temperatura Lincoln**", que mostra 3 categorias. A média das temperaturas mínimas, a média das temperaturas médias, e a média das temperaturas máximas de cada mês.

"**demografia – Midwest**" mostra a demografia étnica dos povos no centro-oeste dos Estados Unidos.

Para inserir uma planilha Excel com os dados você deve selecionar o painel:

Dados Unidimensionais -> **Carregue seus dados** -> **Importe seus dados** -> **Importar um arquivo .xlsx**

Os dados devem estar neste formato:

	A	B	C	D	E	F
1	Arvore 1	Arvore 2	Arvore 3	Arvore 4	Arvore 5	Arvore 6
2	17	15	21	14	17	20
3	18	13	24	16	16	19
4	22	17	22	19	18	19
5	19	19	19	24	19	20
6	18	18	18	13	18	21
7	21	17	18	18	16	21
8	20	16	21	16	17	19
9	20	19	20	23	18	20

Com a primeira linha contendo o nome de cada uma das colunas e as demais linhas contendo os dados de cada uma das colunas. O nome das colunas é essencial.

2.1.2 Inserir os dados na planilha:

Para inserir uma planilha Excel com os dados você deve selecionar o painel:

Gráficos Básicos -> Carregue seus dados -> Digite seus dados na planilha

Os dados devem estar neste formato:

Arvore 1	Arvore 2	Arvore 3	Arvore 4	Arvore 5	Arvore 6	V7	
17	15	21	14	17	20		
18	13	24	16	16	19		
22	17	22	19	18	19		
19	19	19	24	19	20		
18	18	18	13	18	21		
21	17	18	18	16	21		
20	16	21	16	17	19		
20	19	20	23	18	20		
18	14	17	20	18	1		

De novo, a primeira linha é o nome das colunas e as demais são os seus respectivos dados.

Opções de formatação como **copiar e colar** dados excel funcionam na planilha, além de outras funcionalidades como selecionando um conjunto retangular de células e arrastando no canto inferior direito, fazendo uma segunda seleção copiando a primeira seleção e colando na segunda.

2.1.3 Resumo dos dados:

Após carregar os dados irá surgir um painel escrito “Informações gerais dos dados” na aba de Dados Unidimensionais, e neste painel contém uma tabela com os dados carregados e uma outra tabela abaixo com as informações dos dados, como a **média**, **desvio padrão**, **valores mínimos e máximos**, a **variância**, e o **primeiro e terceiro quadrante**.

A **segunda tabela** pode ser transposta na barra de controle de opções, ao clicar em: "**Transcrever tabela de dados estatísticos**"

Dados estatísticos:

how 10 entries Search:

	Média	Desvio Padrão	Mínimo	1º Quadrante	Mediana	3º Quadrante	Máximo	Variância
Lab 1	52.72	14.84	27	40	53	65.75	77	220.2256
Lab 2	39.573	6.634	30	33	39	46	50	44.009956
Lab 3	55.953	20.379	20	39.25	56.5	74.5	89	415.303641
Lab 4	45.16	26.087	1	23	45	61	97	680.531569
Lab 5	61.573	6.768	1	58	62	66	70	45.805824

howing 1 to 5 of 5 entries Previous 1 Next

2.2 Gráficos 2D

Os gráficos 2D podem ser encontrados na aba de Dados Unidimensionais, após carregar os dados no aplicativo.

2.2.1 Histograma:

O painel Histograma possui dois tipos de histogramas, o primeiro está em **formato de barras**, da frequência sobre os dados numéricos. No controle de opções o usuário pode modificar o gráfico para um histograma empilhado e/ou com espaçamento entre as barras.

Histograma sem modificações:

Histograma empilhado:

Histograma com espaçamento:

Mais modificações:

O usuário pode modificar a **quantidade de bandas**, ou seja, aumentar o número de barras representativas do gráfico correspondente ao seu tamanho, desativando o intervalo padrão do histograma; e também a opacidade das barras.

O intervalo padrão do histograma é calculado utilizando a regra de Sturges

O tamanho do espaçamento pode ser modificado caso selecione ativar os espaçamentos.

Um **histograma de densidade** pode ser inserido, podendo modificar a escala do histograma, a opacidade do histograma e a opacidade da área do histograma.

No segundo tipo de histograma, mostra um gráfico contendo um **histograma único para cada uma das colunas**, cada um dos histogramas empilhados.

Histograma com as configurações padrões

Mais modificações:

Na configuração do gráfico, o usuário pode escolher em **mostrar os quartis** e selecionar sua quantidade, modificar a largura de banda padrão do histograma, e modificar a sua altura.

Na configuração dos pontos, o usuário pode escolher em **mostrar os pontos**, e assim poderá escolher seu tamanho, opacidade, formato (linhas, pontos ou cruz) e sua posição. Entre as posições ele poderá escolher a "identidade", em que todos os pontos vão estar no y = 0. "dentro do gráfico", com os pontos dentro do gráfico, "espalhados", com os pontos espalhados, e "abaixo do gráfico" com os pontos abaixo do gráfico.

Na configuração das cores do histograma, o usuário pode escolher **diversas paletas de cores** para o histograma e o sentido da paleta de cores (reverso ou não).

Histograma apenas mostrando a mediana, largura de banda reduzida, altura aumentada, pontos em linhas com opacidade baixa, tamanho aumentado na posição de identidade. Paleta de cores D reversa:

Mesmo histograma, agora com os pontos em formato de cruz, opacidade em 70%, e dentro do gráfico. Paleta de cores C:

Este gráfico pode demorar mais para ser carregado.

2.2.2 Box Plot:

O painel Box Plot mostra um gráfico Box Plot, e no eixo X contem as categorias e no eixo Y contendo os Dados.

Gráfico Box Plot sem modificações:

Mais modificações:

O usuário pode escolher o **algoritmo do cálculo dos quartis** (Linear, Inclusivo e Exclusivo), e se deseja mostrar uma linha tracejada que é a **média** de cada um dos dados.

Escolhendo o método “linear”, o programa irá utilizar a interpolação linear, já nos métodos “inclusivo” e “exclusivo” utiliza a mediana para dividir os dados em dois. Caso a quantidade de dados seja ímpar, no inclusivo não inclui a mediana em nenhuma das metades, já no metodo exlusivo inclui a mediana em ambas as metades. E assim o primeiro quarteto é a mediana da primeira metade, e o segundo quarteto é a mediana da segunda metade.

Na configuração dos pontos o usuário pode **mostrar os pontos** aumentar ou reduzir a distância entre os pontos, a posição dos pontos conforme cada uma das colunas, o tamanho dos pontos, a largura das linhas dos pontos e a opacidade dos pontos.

Gráfico com algoritmo de cálculo linear, média visivel, com 0.13 de distância entre os pontos, -1.2 da posição dos pontos, 2 de tamanho, 0 de largura das bordas dos pontos, 60% de opacidade com formato de circulo:

2.2.3 Violino:

O gráfico em violino mostra um gráfico parecido com o gráfico Box Plot porém com as curvas no entorno do gráfico representando a função de densidade. No eixo X contem as categorias e no eixo Y contendo os Dados.

Gráfico em Violino sem modificações:

Mais modificações:

O usuário pode escolher o **algoritmo do cálculo dos quartis** (Linear, Inclusivo e Exclusivo), e se deseja mostrar uma linha tracejada que é a **média** de cada um dos dados. Ele também pode **modificar as curvas de densidade** do gráfico ao desativar a largura de banda padrão.

Escolhendo o método "linear", o programa irá utilizar a interpolação linear, já nos métodos "inclusivo" e "exclusivo" utiliza a mediana para dividir os dados em dois. Caso a quantidade de dados seja ímpar, no inclusivo não inclui a mediana em nenhuma das metades, já no metodo exlusivo inclui a mediana em ambas as metades. E assim o primeiro quarteto é a mediana da primeira metade, e o segundo quarteto é a mediana da segunda metade.

Na configuração dos pontos o usuário pode **mostrar os pontos** aumentar ou reduzir a distância entre os pontos, a posição dos pontos conforme cada uma das colunas, o tamanho dos pontos, a largura das linhas dos pontos e a opacidade dos pontos.

Gráfico com algoritmo de cálculo inclusivo, média visivel, largura de banda não padrão (1), com 0.2 de distância entre os pontos, -1.2 da posição dos pontos, 3 de tamanho, 0 de largura das bordas dos pontos, 60% de opacidade com formato de cruz:

2.2.4 Gráfico de pontos:

O gráfico de pontos, geralmente quando são dados parecidos, com várias observações e com frequências altas, representam os dados e suas frequências.

O painel do gráfico de pontos possui dois tipos de gráficos, o gráfico de pontos simples e o gráfico de pontos em formato bee swarm.

Gráfico de pontos Simples:

O gráfico de pontos simples possui os dados no eixo X e as frequências dos dados no eixo Y.

Gráfico de pontos simples sem modificações:

Mais modificações:

Nas configurações dos pontos, o usuário pode escolher o **formato dos pontos**, mostrar **linhas nas bordas** dos pontos, o **tamanho** dos pontos, a **opacidade** dos pontos e o intervalo entre os valores (tamanho das bandas).

Nas configurações das linhas, o usuario ao ativar as linhas entre os pontos, e também pode escolher a opacidade das linhas.

Nas configurações das elipses, ao **mostrar as elipses** pode escolher em mostrar a área da elipse e a sua opacidade, e o intervalo de confiança CI da elipse.

A elipse é calculada sobrepondo os contornos de probabilidade normal sobre um gráfico de dispersão de dados.

Gráfico com pontos com opacidade 0 e intervalo entre valores 1, com linhas com 70% de opacidade e uma elipse com 10% de opacidade e com um intervalo de confiança de 95%:

Gráfico com os pontos em formato de circulo, com linhas nos pontos, tamanho 10, opacidade em 80% e intervalo entre valores 1. Sem linhas e com uma elipse com 15% de opacidade e intervalo de confiança de 90%:

Gráfico de pontos em bee swarm:

O gráfico de pontos simples possui a dificuldade de visualização quando os pontos, de categorias diferentes, possuem mesma frequência e mesmos valores, ou até na mesma categoria, com frequências e valores muito parecidos.

Este problema pode ser concertado alterando a opacidade, o formato dos pontos, adicionando as bordas dos pontos ou até adicionando linhas entre os pontos. Ou também pode ser concertado utilizando o gráfico em bee swarm.

Gráfico em bee swarm modificando apenas o espaçamento entre os pontos para 10:

Mais modificações:

Nas configurações do gráfico o usuário pode modificar o **tamanho** (escala) do gráfico, o **espaçamento entre os pontos**, se deseja que o gráfico fique de lado, o método de espalhamento dos pontos, e a prioridade de encaixe dos pontos.

Nas configurações dos pontos, o usuário pode modificar o formato dos pontos (circulos, quadrados, triângulos ou formas diversas).

Gráfico com tamanho 1, espaçamento entre os pontos 4, espaçamento entre os gráficos 0.1, com o método “swarm” e com a prioridade “crescente”:

2.2.5 Gráfico de Densidade:

O gráfico de densidade faz um cálculo de densidade, com o eixo X sendo os dados e o eixo Y a densidade dos dados.

Gráfico de densidade sem modificações:

Mais modificações:

As configurações do gráfico de densidade permitem a **adição da área** ou das **linhas** do gráfico de densidade, e a **escolha de uma estimativa de densidade kernel** (EDK) dentre gaussiana, retangular, triangular, Epanechnikov, biweight, cosseno e optcosseno, sendo o padrão gaussiana.

A estimativa de densidade kernel é uma forma não-paramétrica para estimar a Função densidade de probabilidade (FDP) de uma variável aleatória. Para mais informações:

https://pt.wikipedia.org/wiki/Estimativa_de_densidade_kernel

Gráfico mostrando a área, porém não as linhas, com a estimativa de densidade de kernel igual a “Retangular”:

2.2.6 Gráfico de Erro:

O gráfico de erro é um gráfico que o eixo X são as colunas dos dados, e o eixo Y são as médias de cada um dos dados e os seus respectivos erros.

A linha de erro é realizada com o **desvio padrão**, e o usuário consegue modificar no controle de opções.

Gráfico de erro sem modificações:

Mais modificações:

Nas configurações do gráfico de erro, o usuário pode **remover a linha de erro**, modificar a opacidade da linha de erro e **alterar o algoritmo para o cálculo de erro** (Desvio padrão, Erro padrão, ou selecionar um intervalo de confiança).

O Desvio Padrão é uma medida que indica a **dispersão dos dados** dentro de uma amostra com relação à média calculada pela raiz quadrada do somatório de todas as observações subtraídas pela média das observações ao quadrado, dividido pelo número de observações.

O erro padrão é uma medida de **variação de uma média amostral** em relação à média da população calculada dividindo o desvio padrão pela raiz quadrada do número de observações.

Para calcular o intervalo de confiança basta multiplicar o erro padrão pelo percentil associado ao nível de significância observado em uma distribuição normal padrão, ou seja, que apresenta média 0 e desvio-padrão igual a 1.

Nas configurações das barras, o usuário pode **adicionar uma barra** e alterar a sua opacidade.

Nas configurações dos pontos o usuário pode modificar o **formato dos pontos** (Círculos, Quadrados, Triângulos ou formas diversas), adicionar uma linha de borda nos pontos, alterar o tamanho dos pontos e a sua opacidade.

Gráfico com linha de erro, opacidade em 100%, algoritmo utilizado: desvio padrão, com barras com 60% de opacidade, pontos no formato de círculos, com linhas de borda nos pontos, tamanho 16 e opacidade 100%

2.3 Gráficos 3D

Os gráficos 3D podem ser encontrados na aba Dados Unidimensionais após carregar os dados no aplicativo.

Os gráficos 3D podem ser controlados pelo usuário, podendo rotacionar utilizando o botão esquerdo do mouse, e ser deslocado utilizando o botão direito do mouse. No canto superior direito, existem duas opções de rotação, **turntable rotation** que o usuário pode rotacionar o gráfico em relação à um dos eixos (eixo z), e **orbital rotation** que o usuário pode fazer uma rotação orbital.

2.3.1 Histograma 3D:

O histograma 3D, possui este formato de linhas representando os dados sobre sua frequência, podendo modificar o intervalo de frequências nos controles de opções.

Histograma 3D sem modificações:

Mais modificações:

Na configuração do gráfico, o usuário pode modificar a **quantidade de intervalos de frequência**, nas configurações de linha o usuário pode modificar o comprimento das linhas e a sua opacidade, e nas configurações de pontos o usuário pode determinar o tamanho dos pontos caso deseje mostrá-los.

Gráfico com comprimento de linha 5, opacidade em 90% e com pontos no tamanho 3.

2.3.2 Gráfico de Densidade:

O gráfico de densidade faz um cálculo de densidade, com o eixo X sendo os dados e o eixo Z a densidade dos dados, com as classificações no eixo Y.

Gráfico de densidade sem modificações:

Mais modificações:

As configurações do gráfico de densidade **permitem modificação da largura das linhas** ou a **escolha de uma estimativa de densidade kernel** (EDK) dentre gaussiana, retangular, triangular, Epanechnikov, biweight, cosseno e optcosseno, sendo o padrão gaussiana.

A estimativa de densidade kernel é uma forma não-paramétrica para estimar a Função densidade de probabilidade (FDP) de uma variável aleatória. Para mais informações:

https://pt.wikipedia.org/wiki/Estimativa_de_densidade_kernel

Gráfico com as linhas com largura 5, com a estimativa de densidade de kernel igual a “Epanechnikov”:

2.3.3 Gráfico de pontos:

O gráfico de pontos, geralmente quando são dados parecidos, com várias observações e com frequências altas, representam os dados e suas frequências.

O painel do gráfico de pontos possui dois tipos de gráficos, o gráfico de pontos simples e o gráfico de pontos em formato bee swarm.

Gráfico de pontos 3D Simples:

O gráfico de pontos simples possui os dados no eixo X, as frequências dos dados no eixo Z, e cada uma das classificações no eixo Y.

Gráfico de pontos 3D simples sem modificações:

Mais modificações:

Nas configurações dos pontos, o usuário pode escolher o **tamanho dos pontos**, a **opacidade** dos pontos, adicionar uma **linhas de borda** preta nos pontos e o **intervalo entre os valores** (tamanho das bandas).

Nas configurações das elipses, ao **mostrar as elipses** pode escolher em mostrar a área da elipse e a sua opacidade, o contorno da elipse (contorno preto), sua largura e sua opacidade e o intervalo de confiança CI da elipse.

A elipse é calculada sobrepondo os contornos de probabilidade normal sobre um gráfico de dispersão de dados.

Gráfico com pontos com opacidade em 100% e intervalo entre valores 1, com linha de borda, e uma elipse sem mostrar o contorno da elipse e mostrando a área com 75% de opacidade e com um intervalo de confiança de 95%:

Gráfico com pontos com opacidade em 100% e intervalo entre valores 1, sem linha de borda, e uma elipse mostrando o contorno da elipse, com opacidade em 100% e com contorno igual a 2,5 e mostrando a área com 20% de opacidade e com um intervalo de confiança de 87%:

Gráfico de pontos 3D em Bee swarm:

O gráfico de pontos simples possui a dificuldade de visualização quando os pontos, de categorias diferentes, possuem mesma frequência e mesmos valores, ou até na mesma categoria, com frequências e valores muito parecidos.

Este problema pode ser concertado alterando a opacidade, o formato dos pontos, adicionando as bordas dos pontos ou até adicionando linhas entre os pontos. Ou também pode ser concertado utilizando o gráfico em bee swarm.

Gráfico de bee swarm sem modificações:

Mais modificações:

Nas configurações do gráfico o usuário pode modificar o **espaçamento entre os pontos**, se deseja que o gráfico fique **de lado**, o método de espalhamento dos pontos, e a prioridade de encaixe dos pontos.

Nas configurações dos pontos o usuário pode **adicionar linhas** nos pontos, escolher o **tamanho dos pontos** e escolher a **opacidade** deles.

Gráfico com 10 de espaçamento entre os pontos, de lado, com método Swarm na prioridade Crescente, pontos com bordas pretas, com tamanho igual a 7 e opacidade em 100%

2.3.4 Gráfico de Barras:

O gráfico de barras é um gráfico de histograma em barras com os dados no eixo X, as Frequências no eixo Y e as classificações no eixo Z. O gráfico de barras utiliza o mesmo algoritmo que os histogramas.

Gráfico de barras sem modificações:

Mais modificações:

Nas configurações dos algoritmos pode-se escolher o **algoritmo para o cálculo do número de intervalos**, pode se escolher entre o “Sturges” (padrão), “Scott”, e o “Freedman-Diaconis”.

Para saber mais sobre estes algoritmos (Contém os 3 algoritmos na página):

https://pt.wikipedia.org/wiki/Histograma#Escolha_do_n%C3%BAmero_de_barras_pela_f%C3%B3rmula_de_Sturges

Nas configurações das barras o usuário pode modificar o **espaçamento** entre as barras e a **opacidade** das barras.

Gráfico utilizando o algoritmo de Scott, com 1 de espaçamento entre as barras e com opacidade em 60%:

3 Testes Estatísticos

Aqui vão estar diversos testes estatísticos multivariados. Os testes vão mostrar gráficos personalizados e específicos para cada um dos testes, e tabela de verificação de hipóteses.

3.1 Inserção dos Dados.

Cada um dos testes requer um certo tipo de input pelo usuário, assim cada um dos testes vai ter uma introdução de como inserir cada um dos dados.

3.1.1 Instruções de inserção de cada tipo de dado

- Uma Variável: O tipo “Uma variável” requer apenas uma coluna de dados numéricos, sem limite do número de linhas, como mostra o exemplo:

Desfazer | Área de Trai

T26

	A	B
1	Variável	
2	37.71826	
3	40.50976	
4	49.86006	
5	93.16806	
6	96.7968	
7	20.86402	
8	30.05874	
9	34.09552	
10	10.20932	
11	33.64309	
12	27.23616	
13	48.30598	
14	28.16232	
15	10.10477	
16	49.30855	
17	47.45358	
18	77.69113	
19	29.00689	
20	85.93355	

O número mínimo de linhas para cada uma das variáveis deve ser no mínimo 2.

Lembrando que o nome da(s) variável(is) deve(m) estar na primeira linha.

- **Duas Médias:** O tipo “Duas Médias” requer duas colunas de dados numéricos, sem limite do número de linhas, como mostra o exemplo:

O1

	A	B
1	Variável 1	Variável 2
2	57.731988	38.81103
3	57.953556	78.96223
4	99.156333	76.86954
5	69.974342	30.4395
6	29.65567	30.01733
7	85.098499	61.19835
8	15.748553	94.10047
9	44.510894	21.57128
10	37.661523	95.18191
11	6.9779419	5.522143
12	71.367954	99.03103
13	75.133143	33.21646
14	46.326051	89.94005
15	41.802432	88.9385
16	43.366282	60.31299
17	62.801563	97.75909
18	73.926776	49.91658
19	97.450259	76.04673
20	59.016614	21.81934
21	67.953455	17.86627
22	9.3595788	89.58029
23	7.4247716	63.21678
24	55.390171	
25	72.791229	
26	22.939578	
27	35.034029	

Pode-se notar que o número de linhas da Variável 1é diferente do número de linhas da Variável 2, assim não é necessariamente obrigatório que eles possuam a mesma quantidade de observações, apesar que, para testes pareados, existe a obrigatoriedade de ambas as variáveis possuírem a mesma quantidade de observações. Além disso, o número mínimo de linhas para cada uma das variáveis deve ser no mínimo 2.

Lembrando que o nome da(s) variável(is) deve(m) estar na primeira linha.

- **ANOVA:** O tipo “ANOVA” requer duas colunas de dados, sem limite do número de linhas, sendo a primeira coluna a variável dependente (numérica), e a segunda coluna a variável independente (grupo), como mostra o exemplo:

	A	B
1	Variável Dependente	Grupos
2	90.5648243	Grupo 1
3	73.07130966	Grupo 1
4	48.72972481	Grupo 1
5	62.80448895	Grupo 1
6	86.78307247	Grupo 1
7	7.353382348	Grupo 1
8	50.0966085	Grupo 1
9	24.58780318	Grupo 1
10	37.19413448	Grupo 1
11	14.83209503	Grupo 2
12	39.21461474	Grupo 2
13	26.65130701	Grupo 2
14	65.64404868	Grupo 2
15	95.55375106	Grupo 2
16	68.14877026	Grupo 2
17	66.06237753	Grupo 2
18	43.26391409	Grupo 3
19	52.20261993	Grupo 3
20	53.98819757	Grupo 3
21	10.70544363	Grupo 3
22	73.03258886	Grupo 3
23	50.04766837	Grupo 3
24	48.62337572	Grupo 3
25	56.03990647	Grupo 3
26	75.95040875	Grupo 3
27	66.56036004	Grupo 3

Pode-se notar que os tamanhos dos grupos são diferentes (Grupo 1 possui 9 observações, Grupo 2 possui 7 observações e Grupo 3 possui 10 observações), assim não é necessariamente obrigatório que eles possuam a mesma quantidade de observações, apesar que, para ser considerado um grupo, o grupo deve conter no mínimo 2 observações. Além disso, o número mínimo de linhas para cada uma das variáveis deve ser no mínimo 2. Por fim, o número de linhas de ambas as variáveis independente e dependente devem ser o mesmo.

Lembrando que o nome da(s) variável(is) deve(m) estar na primeira linha. Dúvidas sobre variável dependente e independente: https://cuboup.com/conteudo/variaveis-independentes-e-dependentes/

- ANOVA – Medidas Repetidas: O tipo "ANOVA – Medidas Repetidas" requer três colunas de dados, sem limite do número de linhas, sendo a primeira coluna a variável dependente (numérica), a segunda coluna é a variável independente (grupo), e a terceira o identificador numérico do caso variável, como mostra o exemplo:

	A	B	C
1	Variável Dependente	Grupos	Identificador
2	29.58651668	Grupo 1	1
3	87.13253114	Grupo 1	2
4	10.09829917	Grupo 1	3
5	51.39193403	Grupo 1	4
6	25.81401203	Grupo 1	5
7	27.17450711	Grupo 1	6
8	54.32859427	Grupo 1	7
9	74.28960048	Grupo 1	8
10	17.35068635	Grupo 1	9
11	7.380499941	Grupo 2	1
12	75.60794941	Grupo 2	2
13	54.25610063	Grupo 2	3
14	99.5698965	Grupo 2	4
15	63.34829213	Grupo 2	5
16	46.41443742	Grupo 2	6
17	38.53835903	Grupo 2	7
18	31.93785999	Grupo 3	1
19	9.132917287	Grupo 3	2
20	51.84542699	Grupo 3	3
21	68.17886404	Grupo 3	4
22	62.43934396	Grupo 3	5
23	84.0817685	Grupo 3	6
24	44.4840137	Grupo 3	7
25	14.86356973	Grupo 3	8
26	82.79629308	Grupo 3	9
27	81.77454083	Grupo 3	10
28			

Pode-se notar que os tamanhos dos grupos são diferentes (Grupo 1 possui 9 observações, Grupo 2 possui 7 observações e Grupo 3 possui 10 observações), assim não é necessariamente obrigatório que eles possuam a mesma quantidade de observações, apesar que, para ser considerado um grupo, o grupo deve conter no mínimo 2 observações. Além disso, o número mínimo de linhas para cada uma das variáveis deve ser no mínimo 2.

Por fim, o identificador deve ser um número (não necessariamente um inteiro, porém um identificador pequeno pode ajudar na visualização), indicando a observação correspondente. Cada grupo possui observações com identificação diferentes, e toda observação deverá ter uma variável dependente, um grupo pertencente e um identificador.

Lembrando que o nome da(s) variável(is) deve(m) estar na primeira linha. Dúvidas sobre variável dependente e independente: https://cuboup.com/conteudo/variaveis-independentes-e-dependentes/

ANOVA - Médias Misturadas: O tipo "ANOVA – Medidas Repetidas" requer quatro colunas de dados, sem limite do número de linhas, sendo a primeira coluna a variável dependente (numérica), a

segunda e a terceira coluna são as variáveis independentes (grupo), e a quarta o identificador numérico do caso variável, como mostra o exemplo:

Desfazer | Area de Transferência | Fonte

Q1

	A	B	C	D
1	Variável Dependente	Variável Independente 1	Variável Independente 2	Identificador
2	75.59377731	Grupo 1	T 1	1
3	13.17643956	Grupo 1	T 1	2
4	6.628054566	Grupo 1	T 1	3
5	43.58614068	Grupo 1	T 1	4
6	49.89442973	Grupo 1	T 1	5
7	71.09008785	Grupo 1	T 1	6
8	70.35651255	Grupo 1	T 1	7
9	24.68446559	Grupo 1	T 1	8
10	22.156602	Grupo 1	T 1	9
11	17.80769899	Grupo 2	T 1	1
12	43.96391181	Grupo 2	T 1	2
13	84.10497798	Grupo 2	T 1	3
14	54.05192647	Grupo 2	T 1	4
15	17.26911161	Grupo 2	T 1	5
16	32.70788223	Grupo 2	T 1	6
17	82.44548398	Grupo 2	T 2	7
18	58.6963355	Grupo 3	T 2	1
19	14.49705339	Grupo 3	T 2	2
20	78.26100442	Grupo 3	T 2	3
21	25.65245329	Grupo 3	T 2	4
22	14.07972378	Grupo 3	T 2	5
23	40.92811208	Grupo 3	T 2	6
24	70.32771152	Grupo 3	T 2	7
25	99.88997648	Grupo 3	T 2	8
26	8.045590038	Grupo 3	T 2	9
27	22.92366772	Grupo 3	T 2	10

Pode-se notar que os tamanhos dos grupos de ambas as variáveis independentes são diferentes (Grupo 1 possui 9 observações, Grupo 2 possui 7 observações e Grupo 3 possui 10 observações, assim como T1 possui 15 observações e T2 possui 11 observações), assim não é necessariamente obrigatório que eles possuam a mesma quantidade de observações, apesar que, para ser considerado um grupo, o grupo deve conter no mínimo 2 observações. Além disso, o número mínimo de linhas para cada uma das variáveis deve ser no mínimo 2.

Por fim, o identificador deve ser um número (não necessariamente um inteiro, porém um identificador pequeno pode ajudar na visualização), indicando a observação correspondente. Cada grupo possui observações com identificação diferentes, e toda observação deverá ter uma variável dependente, as duas variáveis independente e um identificador.

Lembrando que o nome da(s) variável(is) deve(m) estar na primeira linha. Dúvidas sobre variável dependente e independente: https://cuboup.com/conteudo/variaveis-independentes-e-dependentes/

ANCOVA: O tipo "ANCOVA" requer tres colunas de dados, sem limite do número de linhas, sendo a primeira coluna a variável dependente

(numérica), a segunda coluna a variável independente (grupo), e a terceira é a covariável (numérica), como mostra o exemplo:

Desfazer | Área de Transferência | Fonte

N1

	A	B	C
1	Variável Dependente	Grupos	Covariável
2	69.92486793	Grupo 1	0.035210247
3	9.808196989	Grupo 1	5.008243358
4	71.65429816	Grupo 1	2.657861343
5	37.79044372	Grupo 1	9.058531955
6	7.065035067	Grupo 1	9.688262565
7	95.58908743	Grupo 1	1.068320462
8	69.90679824	Grupo 1	0.327953002
9	72.12038557	Grupo 1	4.254098963
10	87.30024599	Grupo 1	4.023347295
11	98.34899255	Grupo 2	4.494794052
12	29.44167635	Grupo 2	8.186671325
13	22.86359305	Grupo 2	5.691172402
14	65.69845275	Grupo 2	4.20981896
15	86.24596159	Grupo 2	7.712280961
16	0.432850105	Grupo 2	3.282439007
17	18.6118256	Grupo 2	3.881082576
18	30.53316658	Grupo 3	9.871478717
19	20.80349477	Grupo 3	2.009767023
20	40.78660394	Grupo 3	1.10451294
21	44.88285479	Grupo 3	3.074118409
22	34.64070505	Grupo 3	7.184273269
23	0.891330906	Grupo 3	5.641146746
24	12.66996525	Grupo 3	3.882563625
25	67.27937109	Grupo 3	6.38212563
26	86.55588214	Grupo 3	4.530618595
27	6.114215412	Grupo 3	0.748557794

Pode-se notar que os tamanhos dos grupos são diferentes (Grupo 1 possui 9 observações, Grupo 2 possui 7 observações e Grupo 3 possui 10 observações), assim não é necessariamente obrigatório que eles possuam a mesma quantidade de observações, apesar que, para ser considerado um grupo, o grupo deve conter no mínimo 2 observações. Além disso, o número mínimo de linhas para cada uma das variáveis deve ser no mínimo 2.

Por fim, toda observação deverá ter uma variável dependente, a variável independente, e a covariável.

Lembrando que o nome da(s) variável(is) deve(m) estar na primeira linha. Dúvidas sobre variável dependente e independente: https://cuboup.com/conteudo/variaveis-independentes-e-dependentes/

MANOVA: O tipo "ANCOVA" requer, entre 3 a 5 colunas de dados, sem limite do número de linhas, sendo a primeira coluna até a penúltima coluna, variáveis dependentes (numérica), e a última coluna a variável independente (grupo), como mostra os exemplos:

	A	B	C	D	E	F
1	VD 1	VD 2	VD 3	VD 4	VD 5	Grupos
2	24	10	3	31	93	Grupo 1
3	59	97	56	20	7	Grupo 1
4	74	85	74	26	99	Grupo 1
5	76	43	2	50	2	Grupo 1
6	49	12	25	85	61	Grupo 1
7	9	98	78	74	27	Grupo 1
8	22	97	22	42	49	Grupo 1
9	61	15	34	27	25	Grupo 1
10	67	86	50	94	92	Grupo 1
11	1	16	3	63	59	Grupo 2
12	91	64	53	47	2	Grupo 2
13	24	73	37	7	80	Grupo 2
14	65	9	30	63	78	Grupo 2
15	53	97	69	47	18	Grupo 2
16	42	43	24	59	18	Grupo 2
17	96	13	69	58	80	Grupo 2
18	26	65	6	16	16	Grupo 3
19	83	3	46	35	76	Grupo 3
20	4	26	4	5	19	Grupo 3
21	34	70	35	74	4	Grupo 3
22	12	48	69	68	79	Grupo 3
23	64	92	24	6	28	Grupo 3
24	66	85	89	54	74	Grupo 3
25	53	61	17	13	40	Grupo 3
26	83	99	31	29	97	Grupo 3
27	42	92	41	62	0	Grupo 3

	A	B	C	D
1	VD 1	VD 2	VD 3	Grupos
2	20	59	80	Grupo 1
3	53	34	41	Grupo 1
4	5	0	65	Grupo 1
5	4	74	96	Grupo 1
6	81	62	88	Grupo 1
7	7	88	4	Grupo 1
8	10	92	38	Grupo 1
9	17	90	98	Grupo 1
10	67	43	13	Grupo 1
11	58	94	7	Grupo 2
12	7	1	50	Grupo 2
13	23	80	98	Grupo 2
14	23	60	38	Grupo 2
15	37	69	85	Grupo 2
16	71	72	62	Grupo 2
17	70	53	10	Grupo 2
18	30	27	77	Grupo 3
19	54	57	56	Grupo 3
20	82	81	83	Grupo 3
21	68	53	65	Grupo 3
22	25	6	40	Grupo 3
23	3	92	78	Grupo 3
24	4	2	1	Grupo 3
25	18	56	36	Grupo 3
26	77	4	61	Grupo 3
27	37	31	47	Grupo 3

Pode-se notar que os tamanhos dos grupos são diferentes (Grupo 1 possui 9 observações, Grupo 2 possui 7 observações e Grupo 3 possui 10 observações), assim não é necessariamente obrigatório que eles possuam a mesma quantidade de observações, apesar que, para ser considerado um grupo, o grupo deve conter no mínimo 2 observações. Além disso, o número mínimo de linhas para cada uma das variáveis deve ser no mínimo 2.

Por fim, toda coluna criada deverá ter o mesmo número de linhas, neste exemplo todas as colunas possuem 27 linhas (contando com a linha do título)

3.1.2 Exemplos:

Existem duas opções para a inserção dos dados pelo usuário, dados de exemplo que já vem baixados com o aplicativo e inserir dados de uma planilha Excel.

Para modificar o título, o usuário deve escrever um título abaixo do texto "Digite o título".

Para os dados exemplares você deve selecionar o painel:

Testes Estatísticos-> **Carregue seus dados** -> **Importe seus dados** -> **Usar um arquivo exemplo**

Exemplos do modo "Duas Médias":

"**Combustível**" são duas colunas, a primeira chamada "Velocity_Gas", que contêm a velocidade dos carros utilizando gasolina, e "Velocity_Kerosene" que contêm a velocidade dos carros utilizando querosene.

"**Ratos**" são duas colunas, a primeira com os pesos de 10 ratos antes de um tratamento, e a segunda com os pesos dos mesmos 10 ratos após o tratamento.

"**Peso entre gêneros**" são duas colunas, a primeira com 20 observações de pesos de pessoas do gênero feminino, e a segunda com 20 observações de pessoas do gênero masculino.

Exemplos do modo "ANOVA":

"**Combustível**" são duas colunas, a primeira chamada "Velocity.ratio", que contêm observações da velocidade de cada um dos carros (variável dependente), e a segunda é o "Fluido" que é a variável independente, com as opções "gás", "querosene", "Siptech_34_cSt", "Siptech_46_cSt" e "Siptech_90_cSt".

"**Crescimento de Plantas**" são duas colunas, a primeira chamada "weight", que contêm observações da velocidade de cada um dos pesos de diversas plantas (variável dependente), e a segunda é o "group" que é a variável independente, com as opções "ctrl", "trt1" e "trt2".

Exemplos do modo "ANOVA medidas repetidas ou misturadas":

"**Auto estima**", são três colunas, a primeira chamada "score", que contêm várias observações da velocidade de cada uma das 10 pessoas no teste (variável dependente), a secunda é o "time", que é a variável independente com as opções "t1", "t2" e "t3", e a terceira é a organização da repetição das variáveis. Podemos ver que é dividida em id's de 1 até 10, assim são 10 pessoas, repetindo observações 3 vezes.

Exemplos do modo “ANCOVA”:

“**Combustível**” são três colunas, a primeira chamada “Velocity.ratio”, que contêm observações da velocidade de cada um dos carros (variável dependente), “Err” que é o erro ou a covariável do problema, e “Fluido” que é a variável independente, com as opções “gás”, “querosene”, “Siptech_34_cSt”, “Siptech_46_cSt” e “Siptech_90_cSt”.

Para inserir uma planilha Excel com os dados você deve selecionar o painel:

Testes Estatísticos -> Carregue seus dados -> Importe seus dados -> Importar um arquivo .xlsx

Os dados devem estar neste formato:

	A	B	C	D	E	F	G
1	Sujeito	Gênero	Grau_Instr	n_Filhos	Idade	Salário	
2	1	M	Superior	1	31	8.2	
3	2	F	Superior	0	25	5.3	
4	3	M	Superior	0	33	9.4	
5	4	M	Ens Fundai	0	20	2.9	
6	5	F	Ensino Mé	0	23	3.7	
7	6	F	Superior	2	37	4.4	
8	7	M	Ens Fundai	0	38	4.7	
9	8	M	Superior	0	37	5.4	
10	9	M	Superior	1	34	5.8	
11	10	M	Ensino Mé	2	40	3.2	
12	11	F	Superior	0	41	6	

Com a primeira linha sendo os nomes das colunas, e as demais os seus respectivos valores.

Para saber o formato necessário para o usuário inserir os dados, leia o manual de cada teste específico.

3.1.3 Inserir os dados na planilha:

Para inserir uma planilha Excel com os dados você deve selecionar o painel:

Testes Estatísticos -> Carregue seus dados -> Digite seus dados na planilha

Os dados devem estar neste formato:

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	Sujeito	Gênero	Grau_Instrução	n_Filhos	Idade	Salário					
2	1	M	Superior	1	31	8.2					
3	2	F	Superior	0	25	5.3					
4	3	M	Superior	0	33	9.4					
5	4	M	Ens Fundamental	0	20	2.9					
6	5	F	Ensino Médio	0	23	3.7					
7	6	F	Superior	2	37	4.4					
8	7	M	Ens Fundamental	0	38	4.7					
9	8	M	Superior	0	37	5.4					
10	9	M	Superior	1	34	5.8					
11	10	M	Ensino Médio	2	40	3.2					
12	11	F	Superior	0	41	6					
13	12	F	Ensino Médio	3	46	6.9					
14	13	F	Ensino Médio	0	26	2					

De novo, a primeira linha é o nome das colunas e as demais são os seus respectivos dados.

Para saber o formato necessário para o usuário inserir os dados, leia o manual na página 3.1.1

Opções de formatação como **copiar e colar** dados excel funcionam na planilha, além de outras funcionalidades como selecionando um conjunto retangular de células e arrastando no canto inferior direito, fazendo uma segunda seleção copiando a primeira seleção e colando na segunda.

3.2 Avaliando os Dados

Nesta página você pode checar os dados inseridos para verificar as **estatísticas**, **normalidades**, **distribuições**, **testes** e **homogeneidades**, além de poder realizar a **transformação** de certos dados.

Esta página está localizada em:

Testes Estatísticos -> Avaliando os Dados

Inserção e importação dos dados:

A inserção deste teste requer duas colunas de dados, de preferência com um mesmo número de linhas, cada uma das colunas, com cada um dos dados representando uma observação. Por exemplo:

Título

Show 10 entries Search:

	Velocity_Gas	Velocity_Kerosene
1	1.16001	1.193986
2	1.160048	1.198517
3	1.158883	1.193928
4	1.158861	1.193736
5	1.158976	1.194131
6	1.166736	1.197274
7	1.165601	1.185119
8	1.166457	1.183752
9	1.165694	1.18652
10	1.167259	1.183375

3.2.1 Distribuição de dados normais:

Primeiramente, para verificar se os dados são normais podemos utilizar métodos visuais e estatísticos. Podemos ver no canto superior da tela 3 abas, “Gráfico de Densidade”, “Gráfico quantile-quantile” e “Tabela de verificação de normalidade”. Para todos os tipos de dados citados na página 3.1.1, existe a verificação de normalidade nas 3 abas.

Gráfico de densidade, um método visual para o usuário verificar a normalidade dos dados, assim o usuário pode julgar se a distribuição está no formato de sino (curva normal).

Mostrando exemplos de “Uma Variável”, “Duas Médias” e “ANOVA”:

O **gráfico Q-Q** (Gráfico Quantil – Quantil) que é um método visual para comparar com a distribuição normal, traçando seus quantis uns contra os outros. Uma linha de referência também é mostrada. No gráfico Q-Q, cada observação é plotada como um ponto único. Se os dados são normais os pontos formam uma linha reta.

Mostrando exemplos de “ANOVA”, “ANOVA – Médias Misturadas”, e “MANOVA”:

Por fim, como os testes visuais não são muito confiáveis, na "**Tabela de verificação de Normalidade**", existe os testes de normalidade para verificar a normalidade dos dados como o teste de **Shapiro-Wilk** e o teste de normalidade de **Kolmogorov-Smirnov (K-S)**.

Para ambos os testes, a hipótese nula é que a população está normalmente distribuída, assim, se o valor de p for menor que o nível alfa escolhido, a hipótese nula é rejeitada e há evidências de que os dados testados não são normalmente distribuídos

O teste de Shapiro-Wilk é amplamente recomendado para teste de normalidade e fornece melhores resultados do que o teste K-S. É baseado na correlação entre os dados e o valor normal correspondente.

Já o teste de normalidade de Kolmogorov-Smirnov é recomendado para amostras de dimensão igual ou maior do que 30. Para amostras com dimensões menores, é recomendado o teste de Shapiro-Wilk.

Mostrando exemplos de "Uma Variável", "Duas Médias", "ANOVA" e "MANOVA":

Teste de Shapiro Wilk

Show 10 entries Search:

	Estatística	p	Normalidade
W	0.9689	0.5091	Normal

Showing 1 to 1 of 1 entries Previous 1 Next

Teste de Kolmogorov-Smirnov

Show 10 entries Search:

	Estatística	p	Normalidade
D	1	0	Não normal

Showing 1 to 1 of 1 entries Previous 1 Next

Teste de Shapiro Wilk

Show 10 entries Search:

	Estatística	p	Normalidade
1	0.9385	0.2243	Normal
2	0.9863	0.9886	Normal

Showing 1 to 2 of 2 entries Previous 1 Next

Teste de Kolmogorov-Smirnov

Show 10 entries Search:

	Estatística	p	Normalidade
1	1	0	Não normal
2	1	0	Não normal

Showing 1 to 2 of 2 entries Previous 1 Next

Teste de Shapiro Wilk

Show 10 entries Search:

	Estatística	p	Normalidade
W	0.9661	0.4379	Normal

Showing 1 to 1 of 1 entries Previous 1 Next

Teste de Kolmogorov-Smirnov

Show 10 entries Search:

	Estatística	p	Normalidade
D	0.1829	0.2376	Normal

Showing 1 to 1 of 1 entries Previous 1 Next

Teste de Shapiro Wilk

Show 10 entries Search:

	Estatística	p	Normalidade
1	0.9879	0.2189	Normal
2	0.9895	0.323	Normal
3	0.9811	0.03676	Não normal
4	0.9722	0.003866	Não normal

Showing 1 to 4 of 4 entries Previous 1 Next

Teste de Kolmogorov-Smirnov

Show 10 entries Search:

	Estatística	p	Normalidade
1	0.1856	0.00006507	Não normal
2	0.2777	1.778e-10	Não normal
3	0.2271	3.827e-7	Não normal
4	0.3455	5.551e-16	Não normal

Showing 1 to 4 of 4 entries Previous 1 Next

Nas duas primeiras tabelas, é feito utilizando as variáveis, e nas outras duas os resíduos.

3.2.2 Homogeneidade das variâncias

Aqui vamos checar a homogeneidade das variâncias entre dois ou mais grupos utilizando diversos testes como o Teste F, o Teste de Bartlett, Teste de Levene e o Teste de Fligner-Killeen. Para isso os tipos de dados possíveis são os de ANOVA, ANCOVA e Duas Médias.

Em todos os testes, o usuário pode modificar o intervalo de confiança, entre 0 e 1.

O **Teste F** compara a variância de exatos dois grupos (Com os dados "Duas Médias" ele aceita, e em "**ANOVA**" e "**ANCOVA**" apenas se existir dois grupos), e requer que os dados estejam em uma distribuição normal.

Teste F para comparação de duas variáveis

Show 10 entries Search:

	F	Num_df	Denom_df	p
F	0.3512	24	29	0.01084

Showing 1 to 1 of 1 entries Previous 1 Next

Com um intervalo de confiança de 95 %:

0.1631 0.7788

O valor de p = **0.01084** , o que é **menor ou igual ao nivel de significância 0.05** . Assim sugere que **há diferênças significantes** entre as duas variâncias

O **Teste de Levene** compara a variância de dois ou mais grupos, e requer que os dados estejam em uma distribuição normal, porém ele é menos sensível a dados fora da normalidade.

Teste de Levene para comparação múltiplas variáveis

Show 10 entries Search:

	df1	df2	F	Sig
1	1	53	9.888	0.002726
2	1	53		

Showing 1 to 2 of 2 entries Previous 1 Next

O valor de p = **0.002726** , o que é **menor ou igual ao nivel de significância 0.05** . Assim sugere que **há diferências significantes** entre, pelo menos 2 variâncias das variáveis

O **Teste de Bartlett** compara a variância de dois ou mais grupos, e requer que os dados estejam em uma distribuição normal.

Teste de Bartlett para comparação múltiplas variáveis

Show 10 entries Search:

	F	df	p
Bartlett's K-squared	6.55152	1	0.0104795

Showing 1 to 1 of 1 entries Previous 1 Next

O valor de p = **0.01048** , o que é **menor ou igual ao nivel de significância 0.05** . Assim sugere que **há diferências significantes** entre, pelo menos 2 variâncias das variáveis

O **Teste de Fligner-Killeen** é um teste não paramétrico baseado em ranks, útil quando os dados não são normais.

Teste de Fligner-Killeen para comparação múltiplas variáveis

Show 10 entries Search:

	Chi_Quadrado	df	p
Fligner-Killeen:med chi-squared	9.495	1	0.002061

Showing 1 to 1 of 1 entries Previous 1 Next

O valor de p = **0.002061** , o que é **menor ou igual ao nivel de significância 0.05** . Assim sugere que **há diferências significantes** entre, pelo menos 2 variâncias das variáveis

3.2.3 Transformando os dados para a normalidade

Como diversos testes paramétricos precisam de dados normalizados como o Teste T e a ANOVA, em situações em que a normalidade não é encontrada, podemos considerar a transformação dos dados para a normalidade.

Para transformar os dados para a normalidade, primeiro devemos identificar para que lado a curva dos dados está (para esquerda ou para a

direita). O aplicativo mostra um gráfico de densidade para a verificação além do coeficiente de distorção.

Verificando a curva, o usuário pode escolher entre positivo (curva voltada para a direita), ou negativo (curva voltada para a esquerda), como mostra no exemplo:

Agora o usuário pode escolher entre 4 transformações:

- Raiz quadrada: Para curvas pouco tortas.
- Log: Para curvas bastante tortas.
- Inversa: Para curvas extremamente tortas.

No final, o usuário vai poder baixar os novos dados em um arquivo .xlsx.

3.2.4 Avaliando a esfericidade

O teste de Mauchly para a avaliação de esfericidade serve para mostrar que as diferenças de variações entre todas as combinações de condições relacionadas, são iguais, um dos requisitos fortes para a ANOVA – medidas repetidas e ANOVA – medidas misturadas.

O teste de Mauchly aceita os tipos de dados de "**ANOVA – medidas repetidas**" e "**ANOVA – medidas misturadas**", sem nenhuma restrição fora as obrigatórias. Dúvidas na inserção dos dados, entre na página **3.1.1**.

Primeiramente, o aplicativo computa uma ANOVA, sem correções e o teste de esfericidade de Mauchly.

Exemplos para ANOVA – medidas repetidas, e ANOVA – medidas misturadas respectivamente:

ANOVA

Show 10 entries Search:

	DFn	DFd	F	p	p<.05	ges
1	2	18	55.469	2.01e-8	*	0.829

Showing 1 to 1 of 1 entries Previous 1 Next

Teste de Esfericidade de Mauchly

Show 10 entries Search:

	Effect	W	p	p<.05
1	vi	0.551	0.092	

Showing 1 to 1 of 1 entries Previous 1 Next

ANOVA

Show 10 entries Search:

	Effect	DFn	DFd	F	p	ges
1	group	2	42	4.352	0.019	0.168
2	time	2	84	394.909	1.91e-43	0.179
3	group - time	4	84	110.188	1.38e-32	0.108

Showing 1 to 3 of 3 entries Previous 1 Next

Teste de Esfericidade de Mauchly

Show 10 entries Search:

	Effect	W	p	p<.05
1	group	0.884	0.079	
2	group - time	0.884	0.079	

Showing 1 to 2 of 2 entries Previous 1 Next

Em seguida, verificamos o valor de p do teste de esfericidade:

Resultados:

Pelo teste de esfericidade de mauchly, **p = 0.092** .

As variâncias das diferenças entre os grupo **são iguais** conforme o intervalo de confiança, assim podemos assumir a esfericidade.

Resultados:

Pelo teste de esfericidade de mauchly, **p = 0.079** .

As variâncias das diferenças entre os grupo **são iguais** conforme o intervalo de confiança, assim podemos assumir a esfericidade.

Caso a verificação de esfericidade foi violada, devemos utilizar uma das correções. No caso o aplicativo automaticamente roda ambas as correções de Greenhouse-Geisser e de Huynh-Feldt.

Dos dois métodos de correção de esfericidade, a correção de Huynh-Feldt é considerada a menos conservadora (superestimar o epsilon), enquanto a de Greenhouse-Geisser é considerada mais conservadora (subestimar o epsilon quando o epsilon está próximo de 1).

A recomendação geral é usar a correção Greenhouse-Geisser, principalmente quando épsilon < 0,75. Na situação em que o epsilon é maior que 0,75, alguns estatísticos recomendam usar a correção de Huynh-Feldt (Girden, 1992).

Correções de esfericidade

Correção Greenhouse-Geisser

Show 10 entries Search:

	Effect	DFn	DFd	F	p	ges
1	time	1.38	12.42	55.469	0.00000216	0.829

Showing 1 to 1 of 1 entries Previous 1 Next

Correção Huynh-Feldt

Show 10 entries Search:

	Effect	DFn	DFd	F	p	ges
1	time	1.55	13.94	55.469	6.03e-7	0.829

Showing 1 to 1 of 1 entries Previous 1 Next

Correções de esfericidade

Correção Greenhouse-Geisser

Show 10 entries Search:

	Effect	DFn	DFd	F	p	ges
1	group	2	42	4.352	0.019	0.168
2	time	1.79	75.24	394.909	3.48e-39	0.179
3	group - time	3.58	75.24	110.188	1.97e-29	0.108

Showing 1 to 3 of 3 entries Previous 1 Next

Correção Huynh-Feldt

Show 10 entries Search:

	Effect	DFn	DFd	F	p	ges
1	group	2	42	4.352	0.019	0.168
2	time	1.87	78.38	394.909	1.04e-40	0.179
3	group - time	3.73	78.38	110.188	1.46e-30	0.108

Showing 1 to 3 of 3 entries Previous 1 Next

3.3 Comparando duas médias

Inserção e importação dos dados:

A inserção deste teste requer o tipo de dado “Uma Variável” ou “Duas Médias”, cada um com sua configuração. Para saber mais sobre como inserir os dados, entre na página **3.1.1**.

3.3.1 Teste T:

O **Teste T de Student** ou somente **Teste T** é um teste de hipótese paramétrico que segue uma distribuição t de Student. utilizado para comparar se as duas médias são iguais. No aplicativo mostra 3 formas, a primeira com uma amostra de uma média, a segunda com amostras independentes, e a terceira com amostras pareadas.

One Way:

No Teste T, One Way, o software aceita ou uma coluna ou duas colunas e compara duas médias, com uma média arbitrária. Primeiro ele verifica os pré-requisitos como a normalidade e verificação de outliers.

Verificando Outliers

Não exstem outliers

Não exstem outliers

Por fim ele mostra a hipótese do Teste T para ambas as colunas:

Resultados

Teste T - Velocity_Gas

Show 10 entries

Search:

	p	estatística	df
Test T - Velocity_Gas	9.176e-63	1.174	24

Showing 1 to 1 of 1 entries Previous 1 Next

A área de efeito da variável **Velocity_Gas** , com mu = **5** é de: **NA**

Teste T - Velocity_Kerosene

Show 10 entries

Search:

	p	estatística	df
Test T - Velocity_Kerosene	6.066e-69	1.17	29

Showing 1 to 1 of 1 entries Previous 1 Next

A área de efeito da variável **Velocity_Kerosene** , com mu = **5** é de: **207.1**

Exemplo para “Uma Variável”:

Testando Normalidade

O valor de p utilizando o teste de Shapiro Wilk é de: **0.01018** (Estatísticamente não normal)

Recomenda-se utilizar o teste de Wilcoxon.

Verificando Outliers

Não exstem outliers

Resultados

Show 10 entries Search:

	p	estatística	df
Test T - Tamanho.da.Sépala	3.331e-129	5.843	149

Showing 1 to 1 of 1 entries Previous 1 Next

O valor de p para o teste T é: **3.331e-129**

Ou seja, a média dos dados é estatisticamente diferente de 0

A área de efeito da variável **Tamanho.da.Sépala** , com mu = **0** é de: **7.057**

Two Way e Pareado:

No Teste T, Two Way ou no teste Pareado, o software aceita somente duas colunas de dados (para o teste pareado, ambas as colunas devem possuir o mesmo número de colunas) , e compara as duas médias das duas colunas. Primeiro ele verifica os pré-requisitos como a normalidade, verificação de outliers, e se possuem uma variância padrão:

Testando Normalidade

Verificando Outliers

Os valores de p utilizando o teste de Shapiro Wilk é de:
Velocity_Gas - 0.03894 e **Velocity_Kerosene - 0.018**

Teste de homocedasticidade

Show 10 entries Search:

	Estimativa	p	Estatística
ratio of variances	0.216915	0.00164793	0.216915

Showing 1 to 1 of 1 entries Previous 1 Next

Por fim ele mostra a hipótese do Teste T:

Resultados:

Show 10 entries Search:

	p	estatística
Teste de Wilcoxon	1.45e-11	0

Showing 1 to 1 of 1 entries Previous 1 Next

A área de efeito entre as variáveis **Dados** , e **Classificação** é de: **-6.575**

3.3.2 Teste de Wilcoxon:

O **Teste de Wilcoxon** ou teste dos postos sinalizados de Wilcoxon é um teste de hipóteses não paramétrico utilizado quando se deseja comparar duas amostras relacionadas, amostras emparelhadas ou medidas repetidas em uma única amostra para avaliar se os postos médios populacionais diferem.

Ele substitui o **Teste T de Student** para amostras pareadas quando os dados não satisfazem as exigências. No aplicativo mostra 3 formas, a primeira com uma amostra de uma média, a segunda com amostras independentes, e a terceira com amostras pareadas.

One Way:

No Teste de Wilcoxon, One Way, o software aceita ou uma coluna ou duas colunas e compara duas médias, com uma média arbitrária. Primeiro ele verifica os pré-requisitos como a simetria ao redor da mediana, e a verificação de outliers.

Verificando Outliers

Não exstem outliers

Não exstem outliers

Por fim ele mostra a hipótese do Teste de Wilcoxon para ambas as colunas:

Resultados

Teste de Wilcoxon - Velocity_Gas

Show 10 entries

Search:

	p	estatística
Test de Wilcoxon - Velocity_Gas	5.96e-8	325

Showing 1 to 1 of 1 entries

Previous 1 Next

A área de efeito da variável **Velocity_Gas** , com mu = 0 é de: **0.8745**

Teste de Wilcoxon - Velocity_Kerosene

Show 10 entries

Search:

	p	estatística
Test de Wilcoxon - Velocity_Kerosene	1.86e-9	465

Showing 1 to 1 of 1 entries

Previous 1 Next

A área de efeito da variável **Velocity_Kerosene** , com mu = 0 é de: **0.8731**

Exemplo para “Uma Variável”:

Testando Simetria ao redor da mediana

Verificando Outliers

Não exstem outliers

Resultados

Teste de Wilcoxon - Tamanho.da.Sépala

Show 10 entries Search:

	p	estatística
Test de Wilcoxon - Tamanho.da.Sépala	2.26e-26	11320

Showing 1 to 1 of 1 entries Previous 1 Next

O valor de p para o teste de Wilcoxon é: **2.26e-26**

Ou seja, a média dos dados é estatisticamente diferente de 0

A área de efeito da variável **Tamanho.da.Sépala** , com mu = 0 é de: **0.8677**

Teste U de Mann-Whitney e o Teste de Wilcoxon Pareado:

No Teste de Mann-Whitney ou no teste de Wilcoxon Pareado, o software aceita somente duas colunas de dados, e compara as duas médias das duas colunas. Primeiro ele verifica os pré-requisitos como a verificação de outliers (Testa Simetria apenas no teste pareado):

Por fim ele mostra a hipótese do Teste:

Resultados:

Show 10 entries Search:

	p	estatística
1	0.586	408

Showing 1 to 1 of 1 entries Previous 1 Next

A área de efeito entre as variáveis **Velocity_Gas** , e **Velocity_Kerosene** , é de: **0.07521**

3.3.3 Teste do Sinal (Sign Test)

O **Teste do Sinal** é um método estatístico para testar diferenças consistentes entre pares de observações, tal como o peso dos sujeitos antes e depois do tratamento, e assim não aceita uma coluna de dados. É uma alternativa para os testes T e de Wilcoxon pareados, onde a distribuição dos dados não é normal nem simétrica, com o único requisito sendo manter as observações nos outliers.

O resultado da hipótese do Teste do Sinal:

Estatísticas

Show 10 entries Search:

	p	estatística	df
Test do Sinal	0.69	11	25

Showing 1 to 1 of 1 entries Previous 1 Next

3.4 Comparando múltiplas medias

3.4.1 ANOVA

A análise de variância compara médias de diferentes populações para verificar se essas populações possuem médias iguais ou não. Assim, essa técnica permite que vários grupos sejam comparados a um só tempo.

Em outras palavras, a análise de variância é utilizada quando se quer decidir se as diferenças amostrais observadas são reais (causadas por diferenças significativas nas populações observadas) ou casuais (decorrentes da mera variabilidade amostral).

Portanto, essa análise parte do pressuposto que o acaso só produz pequenos desvios, sendo as grandes diferenças geradas por causas reais.

Inserção e importação dos dados:

A inserção deste teste requer o tipo de dado "ANOVA", ou "ANCOVA", para saber mais como inserir os tipos de dado, entrar na página **3.1.1**.

Teste:

Primeiramente vamos realizar a verificação dos requisitos como o Teste de Normalidade, a Verificação de Outliers e a Verificação da Homogeneidade das Variâncias:

Verificando da Homogeneidade de Variância

Show 10 entries Search:

	df1	df2	statistic	p
1	4	130	5.967	0.0001933

Showing 1 to 1 of 1 entries Previous 1 Next

O valor de p do teste de Levene é: 0.0001933

Por fim, o software mostra o resultado da hipótese do Teste de ANOVA e a Tabela Post Hoc que compara cada uma das combinações 2 a 2 da variável independente, e sua significância de acordo com o valor de P:

Resultado do teste de ANOVA

Show 10 entries Search:

	F	p	p<.05	ges
1	4.846	0.016	*	0.264

Showing 1 to 1 of 1 entries Previous 1 Next

O valor de p do anova é: **0.016** . Ou seja, existe difereça significativa entre, pelo menos dois grupos.

Tabela Post Hoc

Show 10 entries Search:

	group1	group2	null.value	estimate	conf.low	conf.high	p.adj
1	gas	kerosene	0	-0.003841	-0.01375	0.00607	0.821
2	gas	Siptech_34_cSt	0	0.02987	0.01951	0.04022	6.66e-12
3	gas	Siptech_46_cSt	0	0.03789	0.0283	0.04747	3.2e-14
4	gas	Siptech_90_cSt	0	0.04924	0.03827	0.06022	6e-15
5	kerosene	Siptech_34_cSt	0	0.03371	0.0238	0.04362	5.33e-14
6	kerosene	Siptech_46_cSt	0	0.04173	0.03262	0.05083	4.66e-15
7	kerosene	Siptech_90_cSt	0	0.05309	0.04252	0.06365	3.55e-15
8	Siptech_34_cSt	Siptech_46_cSt	0	0.00802	-0.001563	0.0176	0.147
9	Siptech_34_cSt	Siptech_90_cSt	0	0.01938	0.0084	0.03036	0.0000294
10	Siptech_46_cSt	Siptech_90_cSt	0	0.01136	0.001101	0.02162	0.022

Showing 1 to 10 of 10 entries Previous 1 Next

3.4.2 ANOVA Medidas Repetidas

A ANOVA com medidas repetidas é igual a ANOVA, porém com medidas repetidas, dependendo das observações.

Inserção e importação dos dados:

A inserção deste teste requer o tipo de dado “ANOVA – Medidas Repetidas”. Para saber mais como inserir os tipos de dado, entrar na página **3.1.1**.

Primeiramente vamos realizar a verificação dos requisitos como o Teste de Normalidade, a Verificação de Outliers e a Verificação de Esfericidade:

Testando Normalidade

O valor p do teste de Shapiro-Wilk para estest dados são: 0.1892

Verificando Outliers

Esfericidade de Mauchly

Show 10 entries Search:

	Effect	W	p
1	vi	0.551	0.092

Showing 1 to 1 of 1 entries Previous 1 Next

Por fim, o software mostra o resultado da hipótese do Teste de ANOVA e a Tabela Post Hoc que compara cada uma das combinações 2 a 2 da variável independente, e sua significância de acordo com o valor de P:

Resultado do teste de ANOVA

Show 10 entries Search:

	Effect	DFn	DFd	F	p	p<.05	ges
1	vi	2	18	55.469	2.01e-8	*	0.829

Showing 1 to 1 of 1 entries Previous 1 Next

Tabela Post Hoc

Show 10 entries Search:

	group1	n2	statistic	df	p	p.adj
1	t1	10	-4.96761780255934	9	0.000772	0.002
2	t1	10	-13.2281484242824	9	3.34e-7	0.000001
3	t2	10	-4.8678156753501	9	0.000886	0.003

3.4.3 ANOVA Medidas Misturadas

A ANOVA com medidas misturadas é igual a ANOVA com medidas repetidas e ANOVA padrão. Possui dois diferentes tipos de fatores de variáveis, incluindo fatores entre grupos (que possui categorias independentes) e fatores com medidas repetidas.

Inserção e importação dos dados:

A inserção deste teste requer o tipo de dado "ANOVA – Medidas Misturadas". Para saber mais como inserir os tipos de dado, entrar na página **3.1.1**.

Primeiramente vamos realizar a verificação dos requisitos como o Teste de Normalidade, a Verificação de Outliers, Verificação de Homogeneidade das Variâncias e da Covariância:

Testando Normalidade

O valor p do teste de Shapiro-Wilk para estest dados são: 0.5207

Verificando Outliers

Verificando Homogeneidade de Variância

Show 10 entries Search:

	group	df1	df2	Estatística	p
1	grp1	2	42	0.08391	0.9197
2	grp2	2	42	0.1078	0.898
3	grp3	2	42	0.05175	0.9496

Showing 1 to 3 of 3 entries Previous 1 Next

Verificando Homogeneidade de Covariância

Show 10 entries Search:

	Estatística	p
1	3.352	0.1871

Showing 1 to 1 of 1 entries Previous 1 Next

Esfericidade de Mauchly

Show 10 entries Search:

	Effect	W	p
1	group	0.884	0.079
2	group - time	0.884	0.079

Showing 1 to 2 of 2 entries Previous 1 Next

Por fim, o software mostra o resultado da hipótese do Teste de ANOVA.

Resultado do teste de ANOVA

Show 10 entries Search:

	Effect	DFn	DFd	F	p	ges
1	group	2	42	4.352	0.019	0.168
2	time	2	84	394.909	1.91e-43	0.179
3	group - time	4	84	110.188	1.38e-32	0.108

Showing 1 to 3 of 3 entries Previous 1 Next

Em cima também existe um painel chamado "ANOVA – medidas misturadas – testes pareados", contendo os testes de ambos os grupos pareados, como a tabela Post Hoc, contendo os grupos e o valor de P para cada uma das possibilidades.

Tabela de ambos grupos pareados

Show 10 entries Search:

	group	Grupo 1	Grupo 2	p
1	grp1	t1	t2	0.789
2	grp1	t1	t3	0.335
3	grp1	t2	t3	0.484
4	grp2	t1	t2	0.767
5	grp2	t1	t3	0.071
6	grp2	t2	t3	0.127
7	grp3	t1	t2	0.000272
8	grp3	t1	t3	2.29e-8
9	grp3	t2	t3	0.00611

Showing 1 to 9 of 9 entries Previous 1 Next

Show 10 entries Search:

	time	Grupo 1	Grupo 2	p
1	t1	grp1	grp2	0.43
2	t1	grp1	grp3	0.895
3	t1	grp2	grp3	0.51
4	t2	grp1	grp2	0.435
5	t2	grp1	grp3	0.00212
6	t2	grp2	grp3	0.0169
7	t3	grp1	grp2	0.0937
8	t3	grp1	grp3	0.00000638
9	t3	grp2	grp3	0.00132

Showing 1 to 9 of 9 entries Previous 1 Next

3.4.4 ANCOVA

A ANCOVA (Análise de covariância) é um é um modelo linear geral que combina análise de variância e regressão.

A ANCOVA, parecida com a ANOVA, avalia se as médias de uma variável dependente (VD) são iguais entre os diferentes níveis de uma variável independente categórica (VI), muitas vezes chamada de tratamento, apesar de estatisticamente controlar os efeitos de outras variáveis contínuas que não são de interesse primário, conhecidas como covariáveis (CV) ou variáveis incômodas.

Inserção e importação dos dados:

A inserção deste teste requer o tipo de dado "ANCOVA". Para saber mais como inserir os tipos de dado, entrar na página **3.1.1**.

Primeiramente vamos realizar a verificação dos requisitos como o teste de Linearidade, Regressão de Homogeneidade, Homogeneidade das Variâncias e o Teste de Normalidade.

Teste de Linearidade

Regressão de Homogeniedade

Show 10 entries Search:

	Effect	DFn	DFd	F	p	p<.05	ges
1	vi	4	125	22.941	3.14e-14	*	0.423
2	cov	1	125	709.136	2.33e-53	*	0.85
3	vi:cov	4	125	24.715	4.36e-15	*	0.442

Showing 1 to 3 of 3 entries Previous 1 Next

O valor de **F (4 , 125) = 24.715** e o valor de P é: **4.36e-15**

Homogeniedade das Variâncias

Show 10 entries Search:

	F	Df1	Df2	p
Teste de Levene	8.671	4	130	0.000003097

Showing 1 to 1 of 1 entries Previous 1 Next

Teste de Normalidade

Show 10 entries Search:

	Estatística	p
Teste de Shapiro-Wilk	0.9871	0.2392

Showing 1 to 1 of 1 entries Previous 1 Next

Além disso, embaixo do gráfico tem uma tabela ANCOVA **removendo o efeito da covariável**. Após o ajuste o usuário consegue observar estatisticamente a diferença.

ANCOVA

Show 10 entries Search:

	Soma.de.quadrados	DF	F	p
Velocity.ratio	1.3859	1	408.6387	0
Fluido	0.0443	4	3.2659	0.0138
Residuos	0.4375	129		

Showing 1 to 3 of 3 entries Previous 1 Next

Por fim existe a tabela **PostHoc** que compara cada uma das combinações 2 a 2 da variável independente, e sua significância de acordo com o valor de P;

Tabela Posthoc:

Show 10 entries Search:

	Grupo 1	Grupo 2	df	Estatistica	p	p.adj	Significância
1	gas	kerosene	129	2.3375	0.021	0.2095	ns
2	gas	Siptech_34_cSt	129	2.0817	0.0393	0.3935	ns
3	gas	Siptech_46_cSt	129	2.6342	0.0095	0.0947	ns
4	gas	Siptech_90_cSt	129	0.7978	0.4264	1	ns
5	kerosene	Siptech_34_cSt	129	0.2362	0.8136	1	ns
6	kerosene	Siptech_46_cSt	129	0.8601	0.3913	1	ns
7	kerosene	Siptech_90_cSt	129	-0.6196	0.5366	1	ns
8	Siptech_34_cSt	Siptech_46_cSt	129	0.8873	0.3766	1	ns
9	Siptech_34_cSt	Siptech_90_cSt	129	-1.1184	0.2655	1	ns
10	Siptech_46_cSt	Siptech_90_cSt	129	-2.0744	0.04	0.4003	ns

Showing 1 to 10 of 10 entries Previous 1 Next

3.4.5 MANOVA

A análise multivariada da variância, parecido com a ANOVA, é um procedimento para comparação de médias amostrais, porém esta, médias amostrais multivariadas (duas ou mais variáveis). Como um procedimento multivariado, é usada quando há duas ou mais variáveis dependentes e é tipicamente seguida por testes de significância envolvendo variáveis dependentes individuais separadamente.

A análise multivariada da variância é uma forma generalizada da análise de variância univariada (ANOVA), embora, diferentemente desta, a análise multivariada da variância use a covariância entre variáveis dependentes ao testar a significância estatística das diferenças de média.

No aplicativo, o número máximo de variáveis dependentes é 4, para que o resultado continue o mais válido possível, e visualmente verificável.

Inserção e importação dos dados:

A inserção deste teste requer o tipo de dado “MANOVA”. Para saber mais como inserir os tipos de dado, entrar na página **3.1.1**.

Existe diversos requisitos para o uso da MANOVA corretamente, portanto vamos dividi-los em 2 tipos, univariados (comparando uma variável de cada vez), e multivariados.

Primeiramente, começando com os requisitos multivariados, vamos verificar os Outliers, a Normalidade Multivariada, Checando o Tamanho da Amostra, Identificando Multicolinearidade, e a Homogeneidade das Variâncias e da Covariancia:

Dados

Detectando Outliers Multivariados

Show 10 entries Search:

id	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	mahal.dist	is.outlier
No data available in table						

Showing 0 to 0 of 0 entries Previous Next

Checando Normalidade Multivariada

Show 10 entries Search:

	Estatística	p
1	0.9793	0.02342

Showing 1 to 1 of 1 entries Previous 1 Next

O valor de p utilizando o teste múltiplo de Shapiro Wilk é de: **0.02342** (Estatísticamente não normal)

Checando Tamanho da Amostra

Show 10 entries Search:

	Species	Freq
1	setosa	50
2	versicolor	50
3	virginica	50

Showing 1 to 3 of 3 entries Previous 1 Next

A suposição de tamanho da amostra adequada é satisfeita, como **50 > 4** .

Identificando Multicollinearidade

Show 10 entries Search:

	var1	var2	statistic	p
1	Sepal.Length	Sepal.Length	816400000	0
2	Sepal.Length	Sepal.Width	-1.44	0.152
3	Sepal.Length	Petal.Length	21.65	1.04e-47
4	Sepal.Length	Petal.Width	17.3	2.33e-37
5	Sepal.Width	Sepal.Length	-1.44	0.152
6	Sepal.Width	Sepal.Width		0
7	Sepal.Width	Petal.Length	-5.768	4.51e-8
8	Sepal.Width	Petal.Width	-4.786	0.00000407
9	Petal.Length	Sepal.Length	21.65	1.04e-47
10	Petal.Length	Sepal.Width	-5.768	4.51e-8

Showing 1 to 10 of 16 entries Previous 1 2 Next

Checando Homogeniedade das Covariâncias

Show 10 entries Search:

	Estatística	p	df	Método
1	140.9	3.352e-20	20	Box's M-test

Showing 1 to 1 of 1 entries Previous 1 Next

Checando Homogeniedade das Variâncias

Show 10 entries Search:

	Variáveis	df1	df2	Estatística	p
1	Petal.Length	2	147	19.48	3.129e-8
2	Petal.Width	2	147	19.89	2.261e-8
3	Sepal.Length	2	147	6.353	0.002259
4	Sepal.Width	2	147	0.5902	0.5555

Showing 1 to 4 of 4 entries Previous 1 Next

Por fim, neste mesmo painel está o resultado do teste de MANOVA com os valores de p:

Resultado do teste de MANOVA

```
                  Df Pillai approx F num Df den Df    Pr(>F)
var_independente   2 1.1919   53.466      8    290 < 2.2e-16 ***
Residuals        147
                                  ---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Para as verificações univariadas, vamos realizar os gráficos QQPlot para cada uma das variáveis dependentes de cada um dos grupos da variável independente, isto para cada uma das colunas inseridas. Em seguida foi realizado Testes de Shapiro Wilk pareados dos resíduos dos dados e a Checagem da Linearidade pareada.

Checando Normalidade

Testes pareados de Shapiro Wilk

Show 10 entries Search:

	Variável	Estatística	p
1	Petal.Length	0.8763	7.412e-10
2	Petal.Width	0.9018	1.68e-8
3	Sepal.Length	0.9761	0.01018
4	Sepal.Width	0.9849	0.1012

Showing 1 to 4 of 4 entries Previous 1 Next

Checando Linearidade

3.4.5 Teste de Kruskal-Wallis

O teste de Kruscal-Wallis é o teste não paramétrico utilizado na comparação de três ou mais amostras independentes. Ele nos indica se há diferença entre pelo menos dois deles. A aplicação do teste utiliza os valores numéricos transformados em postos e agrupados num só conjunto de dados. A comparação dos grupos é realizada por meio da média dos postos (posto médio).

O método consiste em primeiramente atribuir a cada valor observado, um posto, sempre atribuindo o menor posto ao menor valor e o maior posto ao maior valor.

Este ordenamento global, quando posteriormente adicionamos as ordens de cada coluna em separado, permite-nos obter o total das ordens para cada situação. Se existirem apenas diferenças aleatórias entre as situações, como é postulado na hipótese nula, é de se esperar que ordens altas e baixas se distribuam de forma aproximadamente equivalente pelas diferentes situações.

Mas se pelo contrário, houver uma preponderância de altos ou baixos resultados em qualquer uma das situações, é provável que tal fato reflita diferenças significativas devidas à variável independente.

O Teste de Kruskal-Wallis não funciona para a ANOVA de medidas repetidas, assim, para um teste não paramétrico deste tipo, utilizar o Teste de Friedman.

Inserção e importação dos dados:

A inserção deste teste requer o tipo de dado “ANOVA”, ou “ANCOVA”, para saber mais como inserir os tipos de dado, entrar na página **3.1.1**.

Uma das condições para realizar este teste, é a comparação de três ou mais amostras independentes e o tamanho mínimo de cada amostra deve ser de 6.

Teste:

Vamos primeiro realizar a Verificação de Outliers:

Teste de Kruskal-Wallis

Detectando Outliers

Em seguida vamos realizar o cálculo do teste de Kruskal Wallis

Calculo do Teste de Kruskal Wallis

Show 10 entries Search:

	statistic	df	p	method
1	100.9	4	6.45e-21	Kruskal-Wallis

Showing 1 to 1 of 1 entries Previous 1 Next

O valor de p do teste de Kruskal Wallis é: **6.45e-21** . Ou seja, existe difereça significativa entre, pelo menos dois grupos.

O valor da área de efeito do teste é: **0.7451**

Por fim vamos fazer a Tabela Post Hoc, com as múltiplas comparações entre pares, utilizando o Teste de Dunn e o Teste de Wilcoxon.

Comparado ao teste de Wilcoxon, o teste de Dunn leva em consideração os rankings usados por teste de Kruskal-Wallis. Também faz ajustes de laços.

Múltiplas comparações entre pares

Teste de Dunn

Show 10 entries Search:

	group1	group2	statistic	p
1	gas	kerosene	-0.1624	0.871
2	gas	Siptech_34_cSt	4.487	0.000007229
3	gas	Siptech_46_cSt	6.175	6.607e-10
4	gas	Siptech_90_cSt	7.289	3.129e-13
5	kerosene	Siptech_34_cSt	4.849	0.000001243
6	kerosene	Siptech_46_cSt	6.676	2.456e-11
7	kerosene	Siptech_90_cSt	7.727	1.101e-14
8	Siptech_34_cSt	Siptech_46_cSt	1.329	0.1839
9	Siptech_34_cSt	Siptech_90_cSt	3.058	0.002225
10	Siptech_46_cSt	Siptech_90_cSt	2.032	0.04216

Showing 1 to 10 of 10 entries Previous 1 Next

Teste de Wilcoxon

Show 10 entries Search:

	group1	group2	statistic	p
1	gas	kerosene	408	0.586
2	gas	Siptech_34_cSt	20	4.29e-11
3	gas	Siptech_46_cSt	0	3.85e-17
4	gas	Siptech_90_cSt	0	6.31e-13
5	kerosene	Siptech_34_cSt	34	4.28e-11
6	kerosene	Siptech_46_cSt	11	1.3e-16
7	kerosene	Siptech_90_cSt	0	4.24e-14
8	Siptech_34_cSt	Siptech_46_cSt	315	0.067
9	Siptech_34_cSt	Siptech_90_cSt	33	3.34e-8
10	Siptech_46_cSt	Siptech_90_cSt	164	0.000867

Showing 1 to 10 of 10 entries Previous 1 Next

3.4.5 Teste de Friedman

O teste de Friedman é o teste não-paramétrico utilizado para comparar dados amostrais vinculados, ou seja, quando o mesmo indivíduo é avaliado mais de uma vez. Uma alternativa para ANOVA medidas repetidas.

Este, por sua vez, não utiliza os dados numéricos diretamente, mas sim os postos ocupados por eles após a ordenação feita para cada grupo separadamente. Após a ordenação é testada a hipótese de igualdade da soma dos postos de cada grupo.

Inserção e importação dos dados:

A inserção deste teste requer o tipo de dado “ANOVA – Medidas Repetidas”. Para saber mais como inserir os tipos de dado, entrar na página **3.1.1**.

Teste:

Vamos primeiro realizar a Verificação de Outliers:

Teste de Friedman

Detectando Outliers

Em seguida vamos realizar o cálculo do teste de Kruskal Wallis

Calculo do Teste de friedman Wallis

Show 10 entries Search:

	statistic	df	p	method
1	18.2	2	0.0001117	Friedman test

Showing 1 to 1 of 1 entries Previous 1 Next

O valor de p do teste de Friedman é: **0.0001117** . Ou seja, existe difereça significativa entre, pelo menos dois grupos.

O valor da área de efeito do teste é: **0.91**

Por fim vamos fazer a Tabela Post Hoc, com as múltiplas comparações entre pares, utilizando o Teste de Wilcoxon e o Teste do Sinal.

Recomenda-se utilizar somente o Teste de Wilcoxon, porém, observe que também é possível realizar comparações pareadas usando o Sign Test, que pode carecem de poder para detectar diferenças em conjuntos de dados pareados. No entanto, é útil porque tem algumas suposições sobre as distribuições dos dados para comparar.

Múltiplas comparações entre pares

Teste de Wilcoxon

Show 10 entries Search:

	group1	group2	statistic	p
1	t1	t2	0	0.002
2	t1	t3	0	0.002
3	t2	t3	1	0.004

Showing 1 to 3 of 3 entries Previous 1 Next

Teste do Sinal

Show 10 entries Search:

	group1	group2	statistic	df	p
1	t1	t2	0	10	0.002
2	t1	t3	0	10	0.002
3	t2	t3	1	10	0.021

Showing 1 to 3 of 3 entries Previous 1 Next

4 Gráficos em Mesh

4.1 Inserção dos Dados

A barra de navegação dos gráficos em Mesh serve para você inserir um ou mais dados tridimensionais, e recebe um ou mais gráficos em Mesh tridimensional para a visualização e comparação dos dados.

4.1.1 Importar os dados:

Existem duas opções para a inserção dos dados pelo usuário, dados de exemplo que já vem baixados com o aplicativo e inserir dados de uma planilha Excel.

Para modificar o título, o usuário deve escrever um título abaixo do texto “Digite o título”.

Para os dados exemplares você deve selecionar o painel:

Gráficos em Mesh -> **Carregue seus dados** -> **Importe seus dados** -> **Usar um arquivo exemplo**

Existem 3 opções de exemplos: “gas”, que mostra 5 planilhas com diversos combustíveis, a respeito de outras duas variáveis, e “zoneA” e “zoneB”, que são uma lista de matrizes bem grandes.

Para inserir uma planilha Excel com os dados você deve selecionar o painel:

Dados Tridimensionais -> Carregue seus dados -> **Importe seus dados** -> **Importar um arquivo .xlsx**

Criando diversas planilhas, para cada planilha, a primeira linha são os nomes das colunas, e as demais os seus respectivos valores.

Em seguida você deve informar o número de planilhas que existem no seu arquivo .xlsx. Vale ressaltar que não é necessário que todas as planilhas tenham o mesmo número de linhas e colunas, como podem ver no exemplo abaixo:

	A	B	C	D	E
1	%Err (Ref.Vol) (%)	Ref. Vol. Flow (litre/s)	Fluid Kin. Viscosity (cSt)		
2	1.23	129.937452	10.32		
3	1.21	129.9202346	10.31		
4	1.23	129.9914902	10.36		
5	1.22	130.0470201	10.38		
6	1.25	129.9913216	10.42		
7	1.25	130.0443494	10.44		
8	1.20	89.76443802	10.65		
9	1.25	90.04478425	10.71		
10	1.28	89.98729684	10.69		
11	1.18	90.07862331	10.67		
12	1.22	90.08810392	10.65		
13	1.34	65.52886776	10.62		
14	1.28	65.51356611	10.61		
15	1.34	65.49131991	10.59		
16	1.32	65.48238396	10.58		
17	1.26	65.47099372	10.55		
18	1.38	39.37665209	10.55		
19	1.33	39.37155512	10.56		
20	1.36	39.36693941	10.56		
21	1.36	39.36451462	10.57		

Siptech46cSt | Siptech 35 cSt | Siptech 90 cSt | Kerosene 3 cSt

	A	B	C	D
1	%Err (Ref.Vol) (%)	Ref. Vol. Flow (litre/s)	Fluid Kin. Viscosity (cSt)	
2	1.35	5.946109086	2.66E+00	
3	1.39	5.939520346	2.65E+00	
4	1.40	5.938455765	2.66E+00	
5	1.42	5.937907784	2.66E+00	
6	1.38	5.937032138	2.65E+00	
7	1.42	5.938753981	2.65E+00	
8	1.30	11.28235841	2.66E+00	
9	1.31	11.28654057	2.65E+00	
10	1.29	11.28998565	2.66E+00	
11	1.29	11.28997376	2.65E+00	
12	1.31	11.29100499	2.65E+00	
13	1.29	11.29141684	2.65E+00	
14	1.19	22.40339738	2.65E+00	
15	1.26	22.41985057	2.65E+00	
16	1.22	22.418768	2.65E+00	
17	1.24	22.42015651	2.65E+00	
18	1.22	22.41718282	2.65E+00	
19	1.18	41.70443356	2.65E+00	
20	1.17	41.71003029	2.65E+00	
21	1.21	41.71003453	2.65E+00	
22	1.18	41.70913261	2.65E+00	

Siptech46cSt | Siptech 35 cSt | Siptech 90 cSt | **Kerosene 3 cSt**

Pronto Acessibilidade: tudo certo

Podemos observar 5 planilhas, sendo que cada uma contem 3 colunas, que serão os eixos x, y e z respectivamente, e o numero de linhas que é a quantidade de observações, esta sem limitações.

4.2 Construir Gráficos em Mesh

Os gráficos em Mesh são uma comparação de diversos Gráficos em Mesh de dados tridimensionais. A página está em Dados Tridimensionais após o usuário carregar os dados tridimensionais.

Exemplo de 3 gráficos em Mesh comparando de lado, um a um.

Dado os dados inseridos, o usuário pode escolher quais gráficos mesh ele deseja mostrar, podendo definir as dimensões que ele deseja mostrar o gráfico completo.

Vale ressaltar que as posições dos gráficos em Mesh vão ser reiniciadas se o usuário adicionar mais um gráfico em Mesh.

5 Referências

- Practical Statistics in R II - Alboukadel Kassambara
- Wikipédia Teste de Wilcoxon - https://pt.wikipedia.org/wiki/Teste_de_Wilcoxon
- Wikipédia Teste T de Student - https://pt.wikipedia.org/wiki/Teste_t_de_Student
- Base de Dados - https://www.rdocumentation.org/
- Wikipédia ANOVA – https://pt.wikipedia.org/wiki/An%C3%A1lise_de_vari%C3%A2ncia
- Wikipédia ANCOVA - https://pt.wikipedia.org/wiki/An%C3%A1lise_de_covari%C3%A2ncia
- Wikipédia MANOVA - https://pt.wikipedia.org/wiki/An%C3%A1lise_multivariada_da_vari%C3%A2ncia
- Link explicando as variáveis dependentes e independentes https://cuboup.com/conteudo/variaveis-independentes-e-dependentes/
- Análise de Variância em geral https://www.lampada.uerj.br/bioestatisticabasica/anova.html
- Correção de Esfericidade https://statistics.laerd.com/statistical-guides/sphericity-statistical-guide-2.php
- Correção de Esfericidade – Greenhouse-Geisser https://en.wikipedia.org/wiki/Greenhouse%E2%80%93Geisser_correction