



# Bioestatística quantitativa aplicada

Edison Capp  
Otto Henrique Nienov  
Organizadores

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

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

# Bioestatística Quantitativa Aplicada

Porto Alegre 2020  
UFRGS

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

260p.

ISBN: 978-65-86232-43-1

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

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

NLM: WA950

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

Endereço:

PPG em Ciências da Saúde: Ginecologia e Obstetrícia

FAMED – UFRGS

Rua Ramiro Barcellos, 2400/2º andar

CEP 900035-003 – Porto Alegre – RS

Telefone: +55 51 3308 5607

E-mail: ppggo@ufrgs.br

Editoração e diagramação: Edison Capp

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

Edison Capp  
Otto Henrique Nienov  
Organizadores

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

## 7 Desfechos quantitativos: amostras independentes

*Caroline Darski*

*Cristiana Kuhl*

*Paula Barros Terraciano*

*Otto Henrique Nienov*

### Comparação de amostras

Pensemos que um fabricante de lâmpadas para microscópio óptico queira trocar um dos componentes na expectativa de obter maior durabilidade e potência. Antes de iniciar a produção em grandes quantidades, é preciso que o fabricante tenha informações suficientes que demonstrem que o novo componente realmente fará com que a lâmpada apresente maior durabilidade e potência. Pensando nisso, o fabricante reuniu, aleatoriamente, 15 unidades da lâmpada produzidas com o componente antigo (lâmpada A) e 15 unidades da lâmpada produzidas com o novo componente (lâmpada B). Em seguida, registrou a durabilidade e a potência de cada uma delas em uma planilha e calculou as respectivas médias. Comparando os dois modelos de lâmpadas, o fabricante identificou que a lâmpada B apresentou maiores durabilidade (38.500 horas *versus* 38.000 horas, respectivamente) e potência (30W *versus* 25W, respectivamente) médias do que as registradas para a lâmpada A. Essa informação é suficiente para concluirmos que a nova lâmpada tem melhor desempenho?

A resposta é não! Lembre-se que o experimento demonstrou que, somente naquela amostra, a lâmpada B apresentou um melhor desempenho do que a lâmpada A. E, se fossemos repetir o experimento em uma nova amostra aleatória, com outras 15 unidades de cada lâmpada, será que obteríamos o mesmo resultado? Como poderíamos verificar se realmente existe diferença significativa entre a durabilidade e a potência dos dois modelos de lâmpadas?

Através dos testes de hipóteses, que nos fornecem subsídios para verificar se uma hipótese é verdadeira ou não. No exemplo do fabricante, seria necessário utilizar um teste de hipóteses para testar se existem diferenças significativas na durabilidade e na potência entre os dois modelos de lâmpadas. Se o fabricante assumir que a duração e a potência médias das lâmpadas apresentam uma distribuição normal, utilizaríamos testes paramétricos; se o fabricante verificar que a duração e a potência médias das lâmpadas apresentam uma distribuição não normal, empregariamos testes não paramétricos e expressariamos os valores de durabilidade e potência das lâmpadas em mediana. Note que no exemplo, estamos nos referindo às variáveis quantitativas. Por isso, antes de falarmos sobre teste de normalidade e testes paramétricos e não paramétricos, vamos lembrar o que são variáveis quantitativas. As variáveis quantitativas são aquelas que podem ser mensuradas através de escalas quantitativas, isto é, escalas que têm unidades de medida. Podem ser classificadas em discretas (quando assumem apenas valores pertencentes a um conjunto finito ou enumerável de números) ou contínuas (quando os possíveis valores estão dentro de um intervalo, aberto ou fechado, dos números reais).

## Testes paramétricos e não paramétricos

A principal diferença entre os testes paramétricos e não paramétricos está na distribuição da variável de estudo. Quando é possível assumir a normalidade de distribuição da variável, podem ser utilizados testes paramétricos. Caso contrário, pode-se transformar a variável de modo a obter a normalidade (logaritmo, raiz quadrada, exponencial, etc.). No SPSS, na opção "Calcular variável..." do menu "Transformar", é possível obter o logaritmo decimal ("Log10") e exponencial ("Exp") através da função "Aritmética". Se, ainda assim, a distribuição obtida não for normal, utilizam-se testes não paramétricos.

A escolha adequada do teste (paramétrico ou não paramétrico) a ser utilizado é de fundamental importância, pois conclusões erradas podem ser obtidas se optarmos pelo teste inadequado. Além disso, a apresentação dos dados deve estar de

acordo com a suposição de distribuição da variável: se distribuição normal, utilizamos média e desvio padrão; se distribuição não normal, expressamos os dados em mediana e quartis ou valores mínimo e máximo. Na comparação de grupos, além da distribuição da variável estudada (em cada um dos grupos), a escolha do teste estatístico também dependerá do número de grupos a serem comparados e do tipo de amostras (independentes ou pareadas). Amostras (ou grupos) são consideradas independentes quando não há relação entre os sujeitos, ou seja, quando se caracterizam por indivíduos diferentes e não pareados. Já amostras (ou grupos) pareados correspondem àqueles que estão relacionados, ou seja, correspondem aos mesmos indivíduos ou a um controle emparelhado (Figura 1).

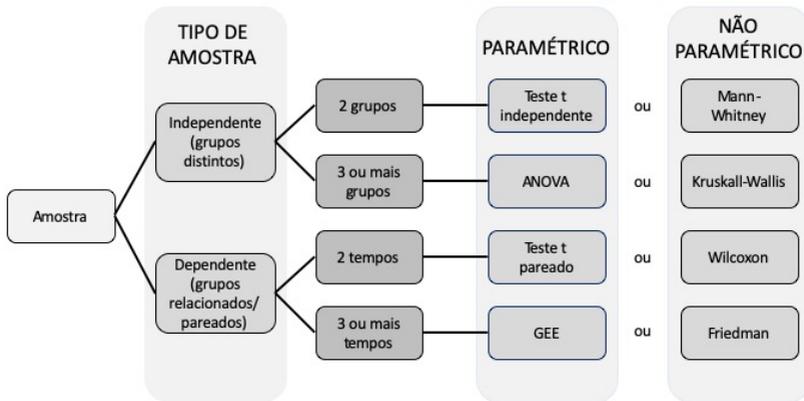


Figura 1. Testes paramétricos e não paramétricos para amostras independentes e pareadas (Mancuso e cols., 2017).

Os testes paramétricos são utilizados para comparar médias, supondo-se a normalidade da variável estudada dentro de cada grupo. Assim, devemos primeiramente verificar a hipótese de normalidade da distribuição, para então decidir qual análise estatística utilizar na comparação dos grupos. Os testes não paramétricos não fazem suposições sobre os parâmetros da distribuição, ou seja, não assumem uma distribuição normal da variável de estudo. São os mais apropriados quando não se conhece a distribuição dos dados na população.

Desta forma, os testes não paramétricos são menos exigentes, pois dispensam a suposição de normalidade e, logo, são mais eficientes do que os testes paramétricos quando os dados não possuem distribuição normal. E, quando normal, sua eficiência é levemente inferior. Também são úteis quando, na comparação, a distribuição é normal em alguns grupos e assimétrica em outros. Por isso, são testes de aplicação mais ampla do que os testes paramétricos. Geralmente, as probabilidades resultantes são exatas, com exceção de casos onde as amostras são grandes. Além disso, são mais fáceis de aplicar e exigem menor número de comandos.

Uma das grandes vantagens dos testes paramétricos é a possibilidade de testar interações, com fatores fixos, aleatórios e covariáveis. Em contrapartida, a obtenção, utilização e interpretação das tabelas do arquivo de saída são, em geral, mais complexas. Os testes não paramétricos não permitem testar interações, o que restringe seu uso em modelos mais complexos e, também, não levam em consideração a magnitude dos dados, pois, na maior parte das vezes, os dados são transformados em ordens ou sinais, ocasionando muitas vezes um “desperdício” de informações. Quando todas as exigências do modelo estatístico são satisfeitas, o teste paramétrico é mais poderoso. Portanto, para se obter a mesma eficiência no teste não paramétrico é necessário um tamanho amostral maior do que seria necessário com um teste paramétrico.

## Definindo a normalidade

A distribuição de probabilidades de uma variável é uma função que descreve a chance com que uma variável pode assumir certos valores ao longo de um subconjunto de possíveis valores. A normalidade de distribuição da variável pode ser verificada através da: 1) inspeção visual utilizando um histograma ou um gráfico quantil-quantil (Q-Q Plot). O Q-Q Plot é um gráfico de dispersão dos quantis (pontos estabelecidos em intervalos regulares que dividem os dados ordenados em “q” subconjuntos de dados de dimensão essencialmente igual) calculados na amostra (quantis amostrais) contra os quantis da distribuição normal (quantis teóricos), sendo útil para verificar a adequação de distribuição de frequência dos dados à

uma distribuição de probabilidades. Quanto mais próximos os pontos estão de uma linha reta, mais próximos estão da distribuição normal; 2) comparação entre os valores de média e mediana e; 3) observação dos valores de assimetria e curtose. Lembre-se que a assimetria é uma medida da distorção da distribuição de probabilidade de uma variável aleatória sobre sua média e, a curtose, informa a altura da distribuição dos dados em relação a uma curva de Gauss. Se ambas forem zero (0), os dados são perfeitamente simétricos. No entanto, essas formas de avaliar a distribuição são consideradas subjetivas.

Para testar a distribuição normal dos dados e evitar a subjetividade, diversos testes estatísticos podem ser utilizados, entre eles o teste de Kolmogorov-Smirnov e o teste de Shapiro-Wilk. Assumimos como  $H_0$  que a distribuição da variável não é assimétrica, enquanto que a  $H_A$  assume que a distribuição da variável é assimétrica. O teste de Shapiro-Wilk é considerado como de melhor potência, mas o SPSS nos fornece ambos os testes. Se a significância do teste for superior a 0,050 ( $p > 0,050$ ), as variáveis seguem uma distribuição normal e, logo, aplicam-se testes paramétricos e utiliza-se média e desvio padrão. Se a significância do teste for inferior ou igual a 0,050 ( $p \leq 0,050$ ), as variáveis seguem uma distribuição não normal e, por conseguinte, expressa-se em mediana e percentis ou valores mínimo e máximo e utiliza-se testes não paramétricos.

**Atenção!** É recomendada a realização do teste de normalidade somente quando, na variável estudada, o tamanho amostral de cada grupo for maior que 15 sujeitos. Caso contrário, deve-se proceder com análises não-paramétricas.

## Teste de normalidade para uma amostra

Mesmo quando queremos caracterizar uma amostra, sem considerar grupos, é necessário verificar a normalidade de distribuição das variáveis para aquela amostra. Para exemplificar, vamos verificar a normalidade das variáveis "Idade" e "Massa corporal" do "Banco de dados 2.sav" (disponível em <https://bit.ly/bancosdedados>). O banco de dados apresenta dados de sujeitos com obesidade. Lembre-se: é importante conhecer o banco de dados. Tome algum tempo para examiná-lo e conhecer as variáveis.

No menu "Analisar", "Estatísticas descritivas, clique em "Explorar". Na janela "Explorar", selecione, na tela da esquerda, as variáveis "Idade" e "Massa corporal" e as insira no campo "Lista dependente" com auxílio do botão da seta. Clique em "Estatísticas..." e marque as opções "Descritivos" e "Percentis". Clique em "Continuar". Clique em "Diagramas..." e selecione as opções "Histograma" e "Diagramas de normalidade com testes". Clique em "Continuar". Importante! Caso haja valores omissos, clique em "Opções" e selecione a opção "Excluir casos por par". Quando incluímos, na "Lista dependente", mais de uma variável e uma destas apresenta *missing value*, a normalidade será testada sem incluir os casos omissos em ambas as variáveis, mesmo nas que não apresentam dado faltante. Ou seja, o SPSS considera o valor omissos em todas as variáveis de estudo. Clique em "Continuar". Em seguida, clique em "Ok" ou "Colar".

Antes de verificarmos os testes de normalidade, no quadro das análises descritivas (*Descriptives*), compare os valores de média e mediana e observe os valores de assimetria e curtose. Também verifique os histogramas e a curva de normalidade. É possível afirmar que essas variáveis não possuem uma distribuição normal? Talvez sim, mas lembre-se que essas formas de verificar a normalidade são subjetivas. Para evitar essa subjetividade, vamos checar os valores dos testes de normalidade.

Conforme comentado anteriormente, o SPSS realiza dois testes de normalidade: o de Kolmogorov-Smirnov e o de Shapiro-Wilk, sendo este último considerado mais poderoso. Portanto, no quadro dos testes (*Tests of Normality*), vamos utilizar os valores do teste de Shapiro-Wilk. Lembre-se que, se a significância do teste for superior a 0,050, as variáveis de estudo apresentam uma distribuição normal e, se a significância do teste for inferior ou igual a 0,050, as variáveis de estudo seguem uma distribuição não normal. Neste caso, ambas as variáveis ("Idade" e "Massa corporal") apresentam distribuição não normal ( $p < 0,001$ ). Logo, podemos afirmar que, a mediana de idade foi de 37 (31-45) anos e a mediana da massa corporal foi de 111,1 (100,5-126,4) kg, onde temos "mediana (percentis 25-75)".

## Teste de normalidade para duas ou mais amostras

Quando comparamos grupos, como no exemplo do fabricante de lâmpadas para microscópio óptico, deve-se testar a normalidade em ambos os grupos. Para exemplificar, vamos verificar o comportamento da variável "Idade" entre as categorias da variável "Sexo" no "Banco de dados 2.sav". No menu "Analisar", "Estatísticas descritivas", clique em "Explorar...". Lembre-se que é possível redefinir as opções de teste clicando em "Redefinir". Em seguida, na janela "Explorar", selecione na tela da esquerda e insira a variável dependente ("Idade") na "Lista dependente", e a variável independente ("Sexo", que corresponde aos grupos) na "Lista de fator" com auxílio do botão da seta. Clique em "Estatísticas..." e marque as opções "Descritivos" e "Percentis". Clique em "Continuar". Clique em "Diagramas..." e selecione as opções "Histograma" e "Diagramas de normalidade com testes". Clique em "Continuar". Em "Opções", selecione "Excluir casos por par" e clique em "Continuar". Por fim, clique em "Ok" ou "Colar".

No arquivo de saída, temos as tabelas com as análises descritivas (*Descriptives*), os percentis (*Percentiles*), o teste de normalidade (*Tests of Normality*) e os gráficos (*Histogram*) para cada grupo, ou seja, para o sexo masculino e para o sexo feminino. Observando os valores de teste de Shapiro-Wilk, conclui-se que em ambos os grupos a variável apresenta distribuição não normal, com medianas de idade no sexo masculino de 37,5 (31,0-44,0) anos e no sexo feminino de 37,0 (31,0-45,0) anos.

Se fôssemos aplicar um teste estatístico para avaliar se as medianas de idade diferem entre os sexos, o teste de escolha seria não paramétrico neste exemplo. O teste paramétrico seria aplicado se em ambos os grupos a variável apresentasse distribuição normal, ou seja, se a significância do teste for superior a 0,050 e, logo, expressaríamos a idade em média e desvio padrão. Caso a variável apresentasse distribuição normal em um grupo e não normal noutro, consideraríamos teste não paramétrico, expressando, em ambos os grupos, as medidas em mediana e quartis ou valores mínimo e máximo.

## Testes para amostras independentes

Neste item, são abordados os testes paramétricos e não paramétricos para amostras independentes. Os testes indicados para amostras relacionadas serão abordados no capítulo 8.

As amostras (ou grupos) são consideradas independentes quando não há relação entre os sujeitos, ou seja, quando se caracterizam por indivíduos diferentes (grupos distintos) e não pareados.

**Atenção!** Antes de decidir qual teste utilizar (paramétrico ou não paramétrico) é necessário verificar a normalidade de distribuição da variável pelo teste de Shapiro-Wilk.

### *Teste t de Student*

O teste t de Student para amostras independentes compara as médias de uma variável quantitativa paramétrica (variável dependente) em dois grupos independentes (variável categórica independente, também chamada de fator, que apresenta duas categorias). Ou seja, quando os valores observados em um grupo não influenciam e nem se relacionam com os valores observados noutro grupo e apresentam distribuição normal. A  $H_0$  do teste t de Student para amostras independentes assume que as médias dos dois grupos são iguais, enquanto que a  $H_A$  assume que as médias dos dois grupos são diferentes.

Uma das suposições do teste t de Student para amostras independentes é de que as variâncias dos dois grupos sejam iguais. Quando comparamos médias, primeiramente devemos nos preocupar com a variabilidade encontrada em cada grupo. É preciso levar em conta se os grupos possuem ou não variâncias iguais, ou seja, se os grupos são homogêneos quanto à variabilidade. Para isso, o SPSS realiza o teste de Levene que compara as variâncias e, com base no resultado deste teste, escolhe-se o resultado do teste t supondo variâncias iguais ou variâncias diferentes. Em outras palavras, a  $H_0$  assume que as variâncias são iguais (homocedasticidade) e, a  $H_A$ , assume que pelo menos uma das variâncias se difere (heteroscedasticidade).

Na interpretação, se o teste de Levene for significativo, ou seja, se a significância do teste for inferior a 0,050 ( $p < 0,050$ ), decidimos que os grupos possuem variâncias diferentes. Assim, devemos considerar o segundo teste t (*Equal variances not assumed*) como válido (localizado na segunda linha da coluna *Sig.(2-tailed)*). Se o teste de Levene não for significativo, ou seja, se a significância do teste for superior ou igual a 0,050 ( $p \geq 0,050$ ), assumimos que não há diferença significativa entre a homogeneidade dos dois grupos. Assim consideramos o primeiro teste t (*Equal variances assumed*) como válido (localizado na primeira linha da coluna *Sig.(2-tailed)*).

Para exemplificar, no “Banco de dados 3 casos novos.sav” (disponível em <https://bit.ly/bancosdedados>), comparamos as médias de “Massa corporal” entre as categorias da variável “Sexo”. Lembre-se: antes de optarmos pelo teste, paramétrico ou não paramétrico, é preciso testar a normalidade de distribuição da variável entre os dois grupos pelo teste de Shapiro-Wilk. Neste caso, ambos os grupos apresentaram significância do teste superior a 0,050. Portanto, a variável de estudo (“Massa”) segue uma distribuição normal e podemos utilizar o teste t de Student para amostras independentes, visto que há apenas dois grupos a serem comparados.

Para executar o teste, no menu “Analisar”, “Comparar médias”, clique em “Teste T de amostras independentes...”. Na caixa de diálogo “Teste T de amostras independentes”, localize e selecione a variável “Massa corporal” na tela da esquerda, e a insira em “Variáveis de teste”. Em seguida, localize e selecione a variável “Sexo” na mesma tela, e a insira em “Variável de agrupamento”. Perceba que é necessário informar os códigos que definem os grupos (“? ?”). Para isso, clique em “Definir grupos...”. Na janela “Definir grupos”, defina as categorias correspondentes. Selecione os grupos que queremos comparar clicando em “Usar valores especificados” e indique os códigos numéricos “1” para “Sexo masculino” e “2” para “Sexo feminino”. Clique em “Continuar”. Em seguida, clique em “Ok” ou “Colar”.

No arquivo de saída, no primeiro quadro (*Group Statistics*), encontram-se as análises descritivas da variável “Massa corporal” para os dois grupos, com a frequência absoluta (N), a média

(Mean), o desvio padrão (Std. Deviation) e o erro padrão da média (Std. Error Mean). Observe que a média do sexo masculino (134,43) é maior que a média do sexo feminino (107,76). Será que a média do sexo masculino difere significativamente da média do sexo feminino?

Para responder essa pergunta, precisamos verificar o p-valor do teste t de Student para amostras independentes. Lembre-se que, primeiro, precisamos checar o teste de Levene para homogeneidade, verificando se a variabilidade entre cada grupo é semelhante ou não. Se o teste de Levene for significativo ( $p < 0,050$ ) decidimos que os grupos possuem variâncias diferentes e optamos pelo valor do teste t não assumindo igualdade de variâncias (segunda linha) como válido. Se o teste de Levene não for significativo ( $p \geq 0,050$ ) assumimos que não há diferença significativa entre a homogeneidade dos dois grupos e optamos pelo valor do teste t assumindo igualdade de variâncias (primeira linha) como válido.

No exemplo, o teste de Levene apresentou um p-valor de 0,353, ou seja, assumimos que não há diferença significativa entre a homogeneidade dos dois grupos e optamos pelo p-valor do teste t da primeira linha, que corresponde a  $p < 0,001$ . Assim, a média de massa corporal do sexo masculino ( $134,43 \pm 17,56$  kg) foi estatisticamente diferente da média de massa corporal do sexo feminino ( $107,76 \pm 11,60$  kg;  $p < 0,001$ ). Também podemos apresentar os resultados em uma tabela (tabela 1).

Tabela 1. Comparação das medidas de massa corporal entre sujeitos obesos do sexo masculino e feminino. Os dados são representados por média  $\pm$  desvio padrão. As unidades de medida e número de pacientes são apresentadas com as variáveis correspondentes.

	Sexo masculino (n = 9)	Sexo feminino (n = 21)	P-valor
Massa corporal (em kg)	134,43 $\pm$ 17,56	107,76 $\pm$ 11,60	<0,001*

n = número absoluto.

\*Teste t de Student independente. Significância estatística considerada de  $p \leq 0,050$ .

**Atenção!** Apesar da distribuição normal pelo teste de Shapiro-Wilk, temos nove sujeitos no sexo masculino. Neste caso, recomenda-se proceder com análise não paramétrica, pois o tamanho amostral de um dos grupos é menor que 15 sujeitos.

### **Teste U de Mann-Whitney**

O teste U de Mann-Whitney é um teste não paramétrico para a comparação de dois grupos independentes, substituindo o teste t de Student independente. A constatação de diferença pelo teste indica que as duas amostras diferem na variável de estudo. Para exemplificar, no "Banco de dados 3 casos novos.sav" (disponível em <https://bit.ly/bancosdedados>), vamos comparar as medianas de "HDL-Colesterol" entre as duas categorias da variável "Sexo". Lembre-se: antes de optarmos pelo teste, paramétrico ou não paramétrico, é preciso testar a normalidade de distribuição da variável entre os dois grupos pelo teste de Shapiro-Wilk. Neste caso, a variável "HDL-Colesterol" tem distribuição não normal no sexo feminino (significância do teste inferior ou igual a 0,050) e normal no sexo masculino (significância do teste superior a 0,050). Portanto, será utilizado o teste U de Mann-Whitney, visto que há dois grupos independentes e em um deles a variável possui distribuição não normal.

No menu "Analisar", "Testes não paramétricos", clique em "Amostras independentes...". Na janela "Testes Não Paramétricos: Duas ou Mais Amostras Independentes", na aba "Objetivo", a opção padrão do SPSS é "Compara automaticamente distribuições entre grupos". Clique na aba "Campos" e, selecione e insira a(s) variável(is) a ser(em) testada(s) (que devem estar declaradas como escalar) no campo "Testar campos" com o auxílio do botão da seta. Em seguida, selecione e insira, com o auxílio do botão da seta, a variável categórica (que deve ser declarada como numérica) em "Grupos".

Como mencionado, o teste não paramétrico é muito simples. Você define se as amostras são independentes ou pareadas e o restante (a escolha de qual teste utilizar) é automaticamente feito pelo SPSS. Para isso, na aba "Configurações", você deve marcar a opção "Escolhe automaticamente os testes com base nos dados". Desta forma, quando se deseja comparar grupos independentes, o

SPSS irá escolher automaticamente se fará o teste para dois grupos (teste U de Mann-Whitney) ou para três ou mais grupos (teste de Kruskal-Wallis), baseado no número de categorias da variável informada no campo “Grupos” da aba “Campos”.

**Atenção!** É necessário classificar corretamente as variáveis quanto ao seu tipo (escalar, ordinal ou nominal) para o teste ser aplicado adequadamente. Reveja os conceitos no capítulo 1. Você também pode definir o teste, selecionando a opção “Mann-Whitney U (2 amostras)” em “Customizar testes”. Por este caminho, é possível alterar o nível de significância ( $\alpha$ ) em “Opções de teste”. Por fim, clique em “Executar” ou “Colar” para obter o resultado do teste estatístico.

O teste utilizado é informado no arquivo de saída, junto com a hipótese de nulidade, o p-valor e a decisão de teste (padrão baseado em uma significância de 5%). Perceba que não obtemos a análise descritiva, somente o resultado do teste de hipótese (*Hypothesis Test Summary*). A mediana e percentis ou valores mínimo e máximo serão extraídos das tabelas com as análises descritivas (*Descriptives*) e percentis (*Percentiles*) solicitados no teste de normalidade. Conforme é indicado pelo teste de hipótese, rejeita-se  $H_0$  de que a distribuição dos níveis séricos de HDL-Colesterol seja igual entre as categorias da variável “Sexo”. Ou seja, com p-valor igual a 0,009, há diferença significativa nos níveis séricos de HDL-Colesterol entre os sexos masculino e feminino. A mediana para o sexo masculino foi de 35,0 (34,0-42) mg/dL e para o sexo feminino foi de 47,0 (44,0-52,0) mg/dL. Podemos apresentar os dados em uma tabela (Tabela 2).

Tabela 2. Comparação das medidas de HDL-Colesterol entre sujeitos obesos do sexo masculino e feminino. Os dados são representados por mediana e percentis (25-75). As unidades de medida e número de pacientes são apresentadas com as variáveis correspondentes.

	Sexo masculino (n = 9)	Sexo feminino (n = 21)	P-valor
HDL-Colesterol (mg/dL)	35,0 (34,0-42)	47,0 (44,0-52,0)	0,009*

HDL, lipoproteína de alta densidade; n = número absoluto.

\*Teste U de Mann-Whitney. Significância estatística considerada de  $p \leq 0,050$ .

Ao darmos um duplo clique no teste de hipóteses no arquivo de saída, uma nova janela de “Visualizador de modelo” mostrará um gráfico contendo a frequência da variável de cada grupo.

### **Análise de variância com um fator – ANOVA (one-way)**

Quando comparamos grupos, além de considerar a distribuição e a relação entre os grupos, também devemos levar em conta o número de grupos ou fatores a serem comparados. Quando a análise envolve comparar três ou mais grupos, supondo a normalidade dos dados pelo teste de Shapiro-Wilk, podemos realizar uma análise de variância (ANOVA). A ANOVA é uma generalização do teste t de Student, onde as variáveis categóricas devem ser selecionadas como fatores do modelo. Cada nível do fator pode ter um efeito diferente sobre a variável dependente. Além de considerarmos a normalidade e o número de grupos a serem comparados (três ou mais), a ANOVA possui outros dois pré-requisitos: a homogeneidade (ou homocedasticidade) das variâncias (pelo teste de Levene) e a independência das observações (definida pelo delineamento do estudo).

O objetivo da ANOVA é verificar se há diferença significativa entre pelo menos duas médias dos grupos. Para isso, a ANOVA considera como  $H_0$  que não há diferença entre as médias dos grupos ( $\mu_A = \mu_B = \dots = \mu?$ ) e, como  $H_A$ , que há diferença entre pelo menos duas das médias dos grupos ( $\mu_A \neq \mu?$ ). Desta forma, se não rejeitamos a  $H_0$  ( $p > 0,050$ ), concluímos a análise de que não há evidências significativas de diferença entre as médias dos grupos. Por outro lado, se rejeitamos a  $H_0$  ( $p \leq 0,050$ ), devemos prosseguir com a análise, comparando os grupos dois a dois (testes *post hoc*), para identificar quais médias são significativamente diferentes, já que pelo menos duas das médias dos grupos diferem entre si.

A ANOVA também produz um valor chamado F (*F-statistics* ou *F-ratio*). Esse valor de F compara a quantidade de variância sistemática dos dados com a quantidade de variância não sistemática. Em outras palavras, o valor de F indica a razão entre o modelo e seu erro. O valor de F vai nos dizer o quanto as médias desses grupos não são iguais. Para a comparação de médias entre os grupos, um valor de F elevado significa que há alguma diferença

entre esses grupos capaz de ser expressa adequadamente por meio de um modelo de regressão. No entanto, o valor de F só nos indica se houve diferença global entre as médias, mas não nos informa quais pares de grupos possuem médias significativamente diferentes. Para isso, precisamos dos testes *post hoc*.

Como exemplo, vamos comparar as médias da variável "Estatura" entre os níveis da variável "Circunferência da cintura recodificada", do "Banco de dados 3 Resolvido.sav" (disponível em <https://bit.ly/bancosdedados>). A variável "Circunferência da cintura recodificada" apresenta quatro categorias: "1" = até 104,00 cm; "2" = 104,01-111,00 cm; "3" = 111,01-120,00 cm; e "4" = mais de 120,01 cm e, a variável "Estatura" apresenta distribuição normal ( $p > 0,050$ ) em todas as quatro categorias. Lembre-se: antes de optarmos pelo teste, paramétrico ou não paramétrico, é preciso testar a normalidade de distribuição da variável entre os grupos pelo teste de Shapiro-Wilk.

Existem dois caminhos para a ANOVA one-way no SPSS. No primeiro, no menu "Analisar", "Comparar médias", clique em "ANOVA a um fator...". Na janela "ANOVA a um fator", selecione, na tela da esquerda, a variável "Estatura" como "Variável dependente" e a variável "Circunferência da cintura recodificada" em "Fator", com o auxílio do botão da seta. Clique em "Post Hoc" (comparação par a par). Na opção "Variâncias iguais presumidas", selecione "Tukey", "Bonferroni", "S-N-K" e "Duncan". Todos esses testes *post hoc* supõem variâncias iguais, mas sugere-se o teste de Tukey, por ser mais rigoroso, ou de Bonferroni, por ser mais conservador. Clique em "Continuar". Clique em "Opções" e marque, em "Estatísticas", as opções "Descritivo" e "Homogeneidade do teste de variância", e selecione a opção de "Diagrama de médias". Clique em "Continuar" e, em seguida, em "Ok" ou "Colar".

No arquivo de saída, na primeira tabela (*Descriptives*), temos a análise descritiva da variável dependente ("Estatura") pelos níveis do fator ("Circunferência da cintura recodificada"). São apresentados: frequência absoluta (N), média (*Mean*), desvio padrão (*Std. Deviation*), erro padrão (*Std. Error*), intervalo de confiança de 95% para a média (*95% Confidence Interval for Mean*) e os valores mínimo (*Minimum*) e máximo (*Maximum*) da

variável dependente para cada nível do fator. Em seguida, temos o quadro do teste de Levene para homogeneidade de variância (*Test of Homogeneity of Variances*). Podemos verificar que, no exemplo, o teste de Levene foi significativo ( $p = 0,001$ ), ou seja, o pré-requisito de homogeneidade não foi satisfeito (os grupos possuem variâncias diferentes) e a análise não pode prosseguir.

Apesar do teste de Shapiro-wilk ter resultado em um p-valor não significativo (o que indica a normalidade de distribuição dos dados), o teste de Levene foi significativo, ou seja, menor que 5%. Neste caso, teríamos que utilizar outro método estatístico para analisar nossos dados. Uma alternativa é utilizar testes não-paramétricos, como o teste de Kruskal-Wallis, que ainda será abordado neste capítulo.

Mas, suponhamos que o teste de Levene tenha apresentado p-valor não significativo ( $p \geq 0,050$ ), ou seja, os dados têm homogeneidade de variância (pressuposto da homogeneidade). Uma vez atendido o pressuposto, podemos seguir com a análise da ANOVA *one-way*. Ao observar a tabela da ANOVA, verificamos que o p-valor do teste foi significativo, ou seja, existe diferença de estatura entre os níveis da circunferência da cintura ( $p < 0,001$ ) e, o valor de F elevado (53,493), indica que há alguma diferença entre esses grupos. No entanto, precisamos verificar quais pares de grupos possuem médias significativamente diferentes. Para isso, devemos analisar os testes *post hoc*.

O quadro seguinte (*Multiple Comparisons*) apresenta os testes de Tukey HSD e Bonferroni com a comparação par a par das categorias (níveis) do fator para a variável dependente. Como mencionado, dentre os testes *post hoc*, sugere-se o teste de Tukey por ser o mais rigoroso. Por isso, iremos interpretar os dados para o teste *post hoc* de Tukey. A interpretação para os outros testes é similar. Além disso, no quadro, temos a comparação par a par para cada nível fixo do fator ((I) Circunferência da cintura recodificada) com os outros níveis ((J) Circunferência da cintura recodificada), a diferença entre as médias da variável dependente ("Estatura") para cada par de nível comparado (*Mean Difference (I-J)*), o erro padrão (*Std. Error*), o nível de significância (Sig.) e o intervalo de confiança (95% *Confidence Interval*).

Analisando estes resultados, podemos concluir, com 95% de confiança, que as categorias 2 e 3 não diferem em média na estatura ( $p = 0,718$ ). A categoria 1 difere significativamente da categoria 2 com  $p = 0,005$  e, das categorias 3 e 4, com  $p < 0,001$ . As categorias 2 e 3 se diferem significativamente da categoria 4 ( $p < 0,001$ ) e, esta última, se diferencia de todas ( $p < 0,001$ ). Uma forma de simplificar a representação destas informações é fazer uma tabela utilizando um sistema de letras, onde letras distintas representam médias significativamente diferentes (Tabela 3). Ou seja, primeiro colocamos os níveis (1, 2, 3 e 4) alinhados por ordem crescente em relação às médias. Em seguida, traduzimos os resultados da tabela de comparação par a par. Para o exemplo, a categoria 1, por ser a de menor média, recebe a letra "a" e as categorias que também não se diferenciam significativamente dela recebem a letra "a" (a categoria 1 se difere de todas); as categorias 2 e 3 são as únicas que não se diferem entre si, por isso recebem a mesma letra "b"; a categoria 4 se difere de todas as categorias, por isso recebe a letra "c".

Tabela 3. Comparação das medidas de estatura entre as categorias de circunferência da cintura de sujeitos obesos. Os dados são representados por média e desvio padrão. As unidades de medida e número de pacientes são apresentadas com as variáveis correspondentes.

Níveis de circunferência da cintura	Frequência absoluta (n)	Média (em metros)	Desvio padrão
1 (até 104,00 cm)	126	1,60 <sup>a</sup>	0,06
2 (104,01 a 111,00 cm)	110	1,63 <sup>b</sup>	0,06
3 (111,01 a 120,00 cm)	124	1,64 <sup>b</sup>	0,08
4 (mais de 120,01 cm)	102	1,72 <sup>c</sup>	0,09
Total	462	1,64	0,08

Letras distintas representam média de estatura entre os níveis de circunferência da cintura significativamente diferentes.

n = número absoluto.

\*Teste ANOVA *one-way*. Significância estatística considerada de  $p \leq 0,050$ ;

Os resultados também podem ser apresentados por gráficos. A figura 2 apresenta a comparação das médias da estatura entre os quatro níveis de circunferência da cintura, possibilitando a análise visual dos resultados descritos na tabela. Note que é possível perceber que a categoria 1 se difere de todas; as categorias 2 e 3 são as únicas que não diferem entre si; e a categoria 4 se difere de todas as categorias.

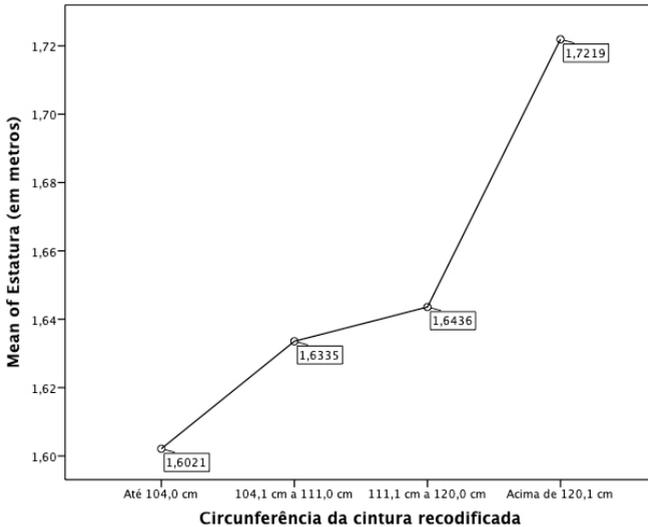


Figura 2. Comparação das médias da estatura entre os quatro níveis de circunferência da cintura.

Na segunda opção, além da ANOVA *one-way*, podemos realizar a ANOVA *two-way* (acrescentando dois fatores) e a ANCOVA (adicionando uma covariável). Estes testes serão abordados a seguir neste capítulo.

No menu "Analisar", "Modelo linear geral", clique em "Com uma única variável...". Na caixa de diálogo "Univariável", selecione a variável "Estatura" como "Variável dependente" e a variável "Circunferência da cintura recodificada" em "Fator(es) fixo(s)". Clique em "Post Hoc" e insira a variável "Fator(es)" em "Testes *post hoc* para". Na opção "Variâncias iguais presumidas" selecione "Tukey", "Bonferroni", "S-N-K" e "Duncan". Clique

em "Continuar". Em "Opções", marque as opções "Estatísticas descritivas", "Estimativas de tamanho de efeito", "Potência observada" e "Testes de homogeneidade". Clique em "Continuar" e, em seguida, em "Ok" ou "Colar".

Como visto da primeira maneira, nas primeiras tabelas temos a análise descritiva da variável dependente ("Estatura") pelos níveis do fator ("Circunferência da cintura recodificada") e, em seguida, temos o teste de Levene para homogeneidade de variância. Como podemos verificar, o pré-requisito de homogeneidade não foi satisfeito (os grupos possuem variâncias diferentes) e a análise não pode prosseguir. Se o requisito fosse cumprido, seguiríamos os mesmos passos de interpretação dos dados vistos anteriormente.

Desta forma, vamos apresentar as diferenças entre os dois caminhos de teste de ANOVA. A primeira diferença é que obtemos um coeficiente de correlação (*Partial Eta Square*). O coeficiente de correlação é uma medida de tamanho de efeito, que pode ser interpretado como o percentual da variância total da variável dependente explicado pela variação entre as categorias formadas pelas variáveis independentes. Neste exemplo, 25,9% da variância total da "Estatura" pode ser explicada pelos níveis do fator "Circunferência da cintura recodificada". A segunda diferença corresponde ao poder do teste (*Observed Power*). Relacionado ao nível de significância, o poder do teste indica a probabilidade de rejeitar corretamente a hipótese de nulidade. Assim, quanto maior a potência do teste, maior o grau de confiança nos resultados. Neste exemplo, essa estatística assume o valor máximo (igual a 1).

Outra diferença, é que não são construídos os gráficos. É necessário criá-los no menu "Gráficos", "Caixas de diálogo legendas", "Linha...". Na janela "Gráficos de linha", selecione as opções "Simples" e "Resumos para grupos de casos". Clique em "Definir". Na janela seguinte, clique em "Outra estatística (por exemplo, média)" e selecione, na tela da esquerda, a variável "Estatura" e a insira em "Variável". Clique em "Alterar estatística..." e indique "Média dos valores". Clique em "Continuar". Insira a variável "Circunferência da cintura recodificada" em "Eixo da categoria". Em seguida, clique em "Ok" ou "Colar". Assim, no arquivo de saída, teremos o mesmo gráfico apresentado na figura 2.

## **Análise de variância com dois fatores – ANOVA two-way**

A ANOVA two-way trabalha com dois fatores, ou seja, duas variáveis explicativas categóricas. O objetivo desta análise é verificar se o comportamento dos fatores em conjunto (interação) pode interferir na variável de estudo que, assim como na ANOVA one-way, deve ser quantitativa paramétrica. Desta forma, podemos analisar os efeitos principais dos dois fatores de forma independente (isoladamente) e o efeito da interação dos mesmos (em conjunto). O efeito principal corresponde ao efeito direto de uma variável independente sobre a variável dependente e o efeito de interação reflete o efeito conjunto de duas ou mais variáveis independentes sobre a variável dependente. Os fatores podem ser de efeitos fixos (são variáveis onde todas as categorias estão representadas) ou de efeitos aleatórios (são as variáveis em que apenas algumas das categorias, em relação ao todo, são selecionadas aleatoriamente).

Para entender as interações, vamos observar o exemplo ilustrado por Guimarães e Hirakata (2012), utilizando os fatores (variáveis categóricas), grupo e tempo, para explicar o comportamento da variável dependente ansiedade, apresentada por um escore (Figura 3). Na figura A1, não há efeito dos fatores, grupo e tempo, sobre a variável ansiedade, ou seja, não há relação entre os fatores e a variável dependente. Na figura A2, não há interação, mas há um efeito do grupo, mostrando que o Tempo 2 possui um escore maior que os demais tempos e separadamente certificamos que os grupos possuem médias diferentes, sendo que o Grupo A é sempre maior que o Grupo B. Na figura A3, temos o efeito (diferença) da variável grupo, mas não há efeito no tempo, diferente da figura A4 em que há efeito do tempo, mas não do grupo. Nas duas últimas figuras, observamos a dependência de uma variável em relação à outra. Na figura A5, há divergência dos grupos em relação ao terceiro tempo (interação) e, na figura A6, há um cruzamento entre os grupos, observando que no Tempo 1 o grupo B possui maiores médias de escores de ansiedade, se igualando no tempo 2 e reduzindo os valores médios no Tempo 3 (interação).

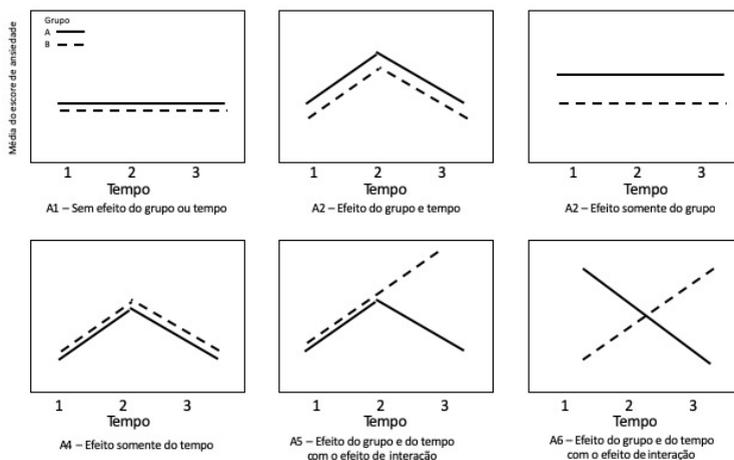


Figura 3. Interação entre os fatores, grupo e tempo, para explicar o comportamento da variável dependente “escore de ansiedade”. (Guimarães e Hirakata, 2012).

Assim como a ANOVA *one-way*, a ANOVA *two-way* também apresenta pré-requisitos para o teste. São estes: 1) o número de grupos a serem comparados (três ou mais); 2) a normalidade de distribuição dos dados (pelo teste de Shapiro-wilk); 3) a homogeneidade (ou homocedasticidade) das variâncias (pelo teste de Levene) e; 4) a independência das observações (definida pelo delineamento do estudo). A ANOVA de três ou mais vias segue os mesmos pré-requisitos, comandos e interpretações da ANOVA *two-way*.

Para exemplificar, supondo normalidade dos dados, vamos verificar a relação entre as médias da variável “Estatura” (variável dependente) entre os níveis dos fatores “Circunferência da cintura recodificada” e “Classificação IPAQ” (ambos fatores de efeito fixo), no “Banco de dados 3 Resolvido.sav” (disponível em <https://bit.ly/bancosdedados>). Vamos seguir os mesmos passos vistos na segunda forma para realizar a ANOVA *one-way*, mas incluindo dois fatores. Lembre-se: antes de optarmos pelo teste, paramétrico ou não paramétrico, é preciso testar a normalidade de distribuição da variável entre os grupos pelo teste de Shapiro-Wilk. Neste exemplo, apesar da normalidade da variável “Estatura” entre

os grupos da variável "Circunferência da cintura recodificada", temos uma distribuição não normal da variável entre os grupos da "Classificação IPAQ". Portanto, não poderíamos realizar a ANOVA two-way. Por isso, para o exemplo, iremos supor que o pressuposto de normalidade tenha sido cumprido.

No menu "Analisar", "Modelo linear geral", clique em "Com uma única variável...". Na janela "Univariável", selecione, na tela da esquerda, a variável "Estatura" como "Variável dependente" e, agora, vamos adicionar dois fatores ao invés de um, no caso, as variáveis "Circunferência da cintura recodificada" e "Classificação IPAQ" em "Fator(es) fixo(s)". Clique em "Diagramas" e insira no "Eixo horizontal" a variável "CCREC" e em "Linhas separadas" a variável "IPAQ". Clique em "Adicionar". Perceba que está sinalizado no quadro "Diagrama" a interação entre os dois fatores ("CCREC\*IPAQ"). Em seguida, clique em "Continuar". Em "Post Hoc", selecione os dois "Fator(es)" e insira-os em "Testes post hoc para". Na opção "Variâncias iguais presumidas" selecione "Tukey", "Bonferroni", "S-N-K" e "Duncan". Clique em "Continuar". Clique em "Opções", localize e selecione a variável da interação ("CCREC\*IPAQ") e a insira dentro do quadro "Exibir médias para". Marque as opções "Estatísticas descritivas", "Estimativas de tamanho de efeito", "Potência observada" e "Testes de homogeneidade" em "Exibir". Clique em "Continuar" e, em seguida, em "Colar", para gerar o arquivo de sintaxe.

**Atenção!** Será necessário adicionar alguns comandos (grifados no Quadro 1) ao algoritmo manualmente no arquivo de sintaxe para obtermos a análise de *post hoc*. Neste comando, os níveis do primeiro fator serão fixos e, o segundo fator com um nível fixo é comparado par a par com os outros níveis, e vice-versa, semelhante ao descrito no quadro da ANOVA one-way (*Multiple Comparisons*), para verificar a diferença das médias de estatura. Após adicionar o comando, selecione todo o algoritmo e execute-o.

Quadro 1. Comando para a análise *post hoc*, fixando os níveis de um dos fatores e comparando par a par os níveis de outro fator.

```
UNIANOVA ESTATURA BY CCREC IPAQ
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/POSTHOC=CCREC IPAQ(SNK TUKEY DUNCAN BONFERRONI)
/PLOT=PROFILE(CCREC*IPAQ)
/EMMEANS=TABLES(CCREC*IPAQ) COMPARE (CCREC) ADJ(BONFERRONI)
/EMMEANS=TABLES(CCREC*IPAQ) COMPARE (IPAQ) ADJ(BONFERRONI)
/PRINT=OPOWER ETASQ HOMOGENEITY DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/DESGIN=CCREC IPAQ CCREC*IPAQ.
```

No arquivo de saída, no primeiro quadro (*Between-Subjects Factors*), são apresentados os fatores, os níveis (categorias) de cada fator e a frequência absoluta (N) de cada nível. No quadro seguinte (*Descriptive Statistics*), temos a análise descritiva (média e desvio padrão) da variável dependente ("Estatura") para cada nível dos fatores ("Circunferência da cintura recodificada" e "Classificação IPAQ").

Em seguida, temos o teste de Levene para homogeneidade de variância (*Levene's Test of Equality of Error Variances*). Podemos verificar que, no exemplo, o teste de Levene foi significativo ( $p = 0,010$ ), ou seja, o pré-requisito de homogeneidade não foi satisfeito (os grupos possuem variâncias diferentes) e a análise não pode prosseguir.

Mas, imaginemos que o pré-requisito de homogeneidade tenha sido satisfeito (os grupos não possuem variâncias diferentes) e a análise pode prosseguir. No quadro seguinte (*Tests of Between-Subjects Effects*), são testados o efeito da interação (em conjunto) e os efeitos principais (separadamente). A leitura deve ser realizada de baixo para cima, ou seja, devemos primeiramente avaliar o efeito da interação e posteriormente, se não houver significância estatística na interação, os efeitos principais de cada fator sobre a

variável dependente. Com p-valor igual a 0,345, a interação não foi significativa, afirmado pelo valor de  $F(1,116)$ . Neste caso, trabalhamos os efeitos principais, como na ANOVA *one-way*, olhando para os testes *post hoc* (que se encontram ao final da análise).

Também podemos observar o coeficiente de correlação (*Partial Eta Square*) e o poder do teste (*Observed Power*). Neste exemplo, 2,9% da variância total na estatura pode ser explicado pela interação dos níveis (categorias) da variável circunferência da cintura e IPAQ e essa estatística assume o valor de 0,649 (grau de confiança nos resultados). Se o p-valor mostrasse uma interação significativa, isso significaria que há um efeito conjunto dos fatores ("Circunferência da cintura" e "IPAQ") com a variável dependente ("Estatura"). Sendo assim, não é correto analisar separadamente os efeitos principais dos fatores, somente a interação.

Também é apresentado um quadro (*Estimates*) com os níveis dos fatores e as médias (*Mean*), erro padrão (*Std. Error*) e intervalo de confiança de 95% (*95% Confidence Interval*) da variável dependente para cada nível.

Vamos supor que a interação entre os fatores tenha sido significativa e, assim, para completar a análise, precisamos analisar o teste de *post hoc* de Bonferroni (*Pairwise Comparisons*). O raciocínio é o mesmo visto na interpretação do teste de *post hoc* de Tukey visto na ANOVA *one-way*. O objetivo é verificar quais categorias (níveis) diferem significativamente entre si na comparação par a par. Para essa análise ocorrer, foi necessário inserir os comandos no algoritmo do arquivo de sintaxe.

Em um primeiro quadro, há a comparação média das estaturas entre os níveis (categorias) da circunferência da cintura para cada nível do IPAQ (variável fixa). Para interpretar os resultados, vamos somente olhar para os indivíduos classificados como sedentários no IPAQ. Conforme a comparação par a par do teste de *post hoc* de Bonferroni, nos sujeitos sedentários, a categoria 4 (acima de 120,01 cm) da circunferência da cintura se diferenciou significativamente, na estatura média, em comparação a todas as outras categorias ( $p < 0,001$ ). Essa interpretação deve ser feita para cada conjunto de comparação.

No segundo quadro, temos a comparação contrária, ou seja, a comparação média das estaturas entre as categorias (níveis) do IPAQ para cada categoria (nível) da circunferência da cintura (variável fixa). Para interpretar os resultados, vamos observar a última categoria da circunferência da cintura. Conforme a comparação par a par do teste de *post hoc* de Bonferroni, nos sujeitos com mais de 120,01 cm de circunferência da cintura (categoria 4), os sujeitos insuficientemente ativos A se diferenciam significativamente, na estatura média, em comparação aos sujeitos ativos ( $p = 0,031$ ). Essa mesma interpretação deve ser feita para cada conjunto de comparação. Da mesma forma vista na ANOVA *one-way*, podemos sumarizar os resultados em uma tabela (Tabela 4) ou gráfico (Figura 4).

Tabela 4. Interação entre fatores, circunferência da cintura e IPAQ, para explicar o comportamento da variável dependente estatura. Os dados são representados por média e desvio padrão. As unidades de medida e número de pacientes são apresentadas com as variáveis correspondentes.

Categories CC	1 (até 104,0 cm)	2 (104,01-111,0 cm)	3 (111,01-120,0 cm)	4 (mais de 120,01 cm)
Classificação IPAQ	Média ± DP	Média ± DP	Média ± DP	Média ± DP
Sedentário (n = 60)	1,57 ± 0,05 <sup>aA</sup>	1,60 ± 0,05 <sup>aA</sup>	1,64 ± 0,10 <sup>bA</sup>	1,73 ± 0,10 <sup>bAB</sup>
Insuf. B (n = 140)	1,62 ± 0,07 <sup>aA</sup>	1,64 ± 0,06 <sup>aA</sup>	1,65 ± 0,07 <sup>aA</sup>	1,71 ± 0,09 <sup>bAB</sup>
Insuf. A (n = 89)	1,60 ± 0,04 <sup>aA</sup>	1,64 ± 0,07 <sup>abA</sup>	1,66 ± 0,07 <sup>bA</sup>	1,77 ± 0,06 <sup>cA</sup>
Ativo (n = 157)	1,59 ± 0,06 <sup>aA</sup>	1,63 ± 0,07 <sup>aA</sup>	1,63 ± 0,07 <sup>aA</sup>	1,70 ± 0,09 <sup>bB</sup>
Muito ativo (n = 16)	1,57 ± 0,07 <sup>aA</sup>	1,66 ± 0,08 <sup>abA</sup>	1,64 ± 0,03 <sup>abA</sup>	1,73 ± 0,05 <sup>bAB</sup>

Letras minúsculas distintas representam médias significativamente diferentes entre categorias de circunferência da cintura, fixando IPAQ (→). Letras maiúsculas distintas representam médias significativamente diferentes entre IPAQ, fixando categorias de circunferência da cintura (↓). CC, circunferência da cintura; IPAQ, questionário internacional de atividade física; DP, desvio padrão, n, número absoluto

\*Teste ANOVA two-way. Significância estatística considerada de  $p \leq 0,050$ .

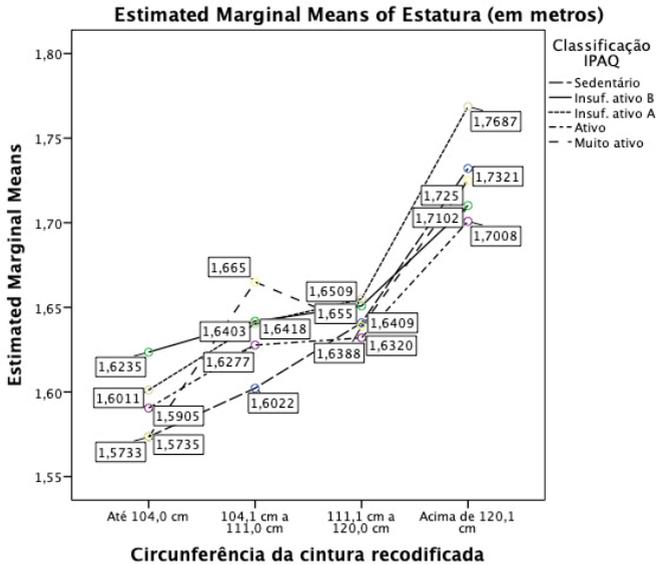


Figura 4. Interação entre fatores, circunferência da cintura e IPAQ, para explicar o comportamento da variável dependente estatura.

**Atenção!** Lembre-se que, como o p-valor foi igual a 0,345, a interação não foi significativa. Desta forma, devemos ver os efeitos principais (isoladamente), como na ANOVA one-way, para cada fator, olhando para os testes *post hoc* de Tukey (que se encontram ao final da análise).

Os quadros apresentados (*Multiple comparisons*) são idênticos a ANOVA one-way: temos a comparação par a par para cada nível fixo do fator ((I) Circunferência da cintura recodificada ou (I) Classificação IPAQ) com os outros níveis ((J) Circunferência da cintura recodificada ou (J) Classificação IPAQ), a diferença entre as médias da variável dependente ("Estatura") para cada par de nível comparado (*Mean Difference (I-J)*), o erro padrão (*Std. Error*), o nível de significância (*Sig.*) e o intervalo de confiança (95% *Confidence Interval*).

Sendo assim, na interpretação referente aos níveis de circunferência da cintura, podemos dizer que, com 95% de confiança, as categorias 2 e 3 são as únicas que não diferem em média para a estatura ( $p = 0,714$ ). Com relação as categorias do

IPAQ, podemos concluir, com 95% de confiança, que nenhuma categoria do IPAQ difere em média para a estatura ( $p > 0,050$ ).

### **Análise de covariância – ANCOVA**

A análise de covariância (ANCOVA) tem o mesmo objetivo da ANOVA, que é testar os efeitos principais e a interação de variáveis independentes (categóricas) sobre a variável dependente (contínua). No entanto, na ANCOVA, podemos controlar estes efeitos para outras variáveis independentes contínuas, as quais chamamos de covariáveis, ou ainda, variável auxiliar ou concomitante. A covariável complementa o controle local e, na grande maioria das situações, simplesmente o substitui. Desta forma, a ANCOVA permite que se faça um “ajuste” do efeito de uma variável resposta que sofreu influência de uma variável ou uma causa de variação não controlada. Obrigatoriamente, a covariável necessita estar correlacionada com a variável resposta para que se possa fazer uso de tal análise. Quando a ANCOVA é realizada com mais de uma covariável, chama-se de análise de ANCOVA<sup>2</sup>.

Para exemplificar, no “Banco de dados 4.sav” (disponível em <https://bit.ly/bancosdedados>), vamos verificar se há diferença na massa corporal após a cirurgia bariátrica (“Massa corporal 3”) entre os sexos (“Sexo”), supondo a normalidade de distribuição dos dados, e utilizando a massa corporal basal (“Massa corporal 1”) como controle (covariável). Ou seja, vamos corrigir essa comparação pela massa corporal basal dos indivíduos, para saber se este não é um fator confundidor. Iremos assim, verificar se as variáveis sexo (“Sexo”) e massa corporal basal (“Massa corporal 1”) possuem efeito na variável dependente massa corporal final (“Massa corporal 3”). Seguiremos o mesmo caminho de análise visto na ANOVA *one-way*, mas acrescentando uma covariável. Lembre-se: antes de optar pelo teste, paramétrico ou não paramétrico, é preciso testar a normalidade de distribuição da variável entre os grupos pelo teste de Shapiro-Wilk.

Este banco de dados contém a avaliação de sujeitos com obesidade no momento pré-operatório (1) e em dois momentos após a cirurgia bariátrica (2 e 3). Portanto, temos três medidas distintas, uma para cada ponto no tempo (pareadas), para as

variáveis massa corporal, circunferência da cintura, medidas pressóricas, hipertensão e atividade física avaliada pelo IPAQ versão curta. Além disso, temos descritos sexo, técnica cirúrgica e estatura dos sujeitos. Estes sujeitos também foram avaliados quanto a alterações de sensibilidade nos membros inferiores (neuropatia periférica) através do MNSI, por dois examinadores. Lembre-se: é importante conhecer o banco de dados. Tome algum tempo para examiná-lo e conhecer as variáveis.

Primeiro, temos que realizar uma ANOVA *one-way* para nos certificarmos que a variável independente ("Sexo") não tenha efeito sobre a covariável ("Massa corporal 1"). Para isso, o p-valor do teste deve ser não significativo, ou seja, maior que 0,050. Porém, no exemplo, há efeito do sexo sobre a massa corporal basal ( $p < 0,001$ ), evidenciado pelo valor de F elevado (59,458). Vamos supor que não haja efeito, para podermos prosseguir com a ANCOVA.

No menu "Analisar", "Modelo linear geral", clique em "Com uma única variável...". Na janela "Univariável", selecione, na tela da esquerda, a variável "Massa corporal 3" como "Variável dependente", a variável "Sexo" como "Fator(es) fixo(s)" e a variável "Massa corporal 1" como "Covariável". Em "Modelo", é necessário que a opção "Fatorial completo" esteja selecionada (padrão). Clique em "Opções" e, selecione e insira a variável "Sexo" para "Exibir médias para". Clique em "Comparar efeitos principais" e selecione na aba "Ajuste do intervalo de confiança" o teste de "Bonferroni". Marque as opções: "Estatísticas descritivas", "Estimativas de tamanho de efeito", "Potência observada" e "Testes de homogeneidade" em "Exibir". Clique em "Continuar", e em seguida, clique em "Ok" ou "Colar".

No arquivo de saída, no primeiro quadro (*Descriptive Statistics*) observamos a análise descritiva da variável dependente ("Massa corporal 3") discriminada pelas categorias da variável "Sexo", com as médias (*Mean*), desvio padrão (*Std. Deviation*) e frequência absoluta (N). Assim como na ANOVA, temos o teste de Levene para homogeneidade (*Levene's Test of Equality of Error Variances*). Com  $p = 0,066$ , aceitamos a hipótese de nulidade de homogeneidade das variâncias entre as categorias de sexo.

Nos próximos quadros, temos a análise de covariância (*Tests of Between-Subjects Effects*) e o teste de *post hoc* de Bonferroni (*Pairwise Comparisons*). Com  $p = 0,560$ , não há diferença significativa entre os sexos em relação à massa corporal final, ajustando a análise pela massa corporal inicial ( $p < 0,001$ ). Além disso, no quadro "*Estimates*", são apresentadas novas médias controladas, ou seja, as médias estimadas para os sexos e corrigidas pela massa corporal inicial (fixada em 112,99). Perceba que as novas médias de massa final (74,44 kg para o sexo masculino e 73,19 kg para o sexo feminino) diferem para as médias "reais" sem o controle da massa corporal inicial (85,26 kg para o sexo masculino e 71,15 para o sexo feminino).

Para verificar a interação dos fatores, precisamos desfazer as ações realizadas em "Opções" e, em "Modelo", marcar a opção "Personalizado", onde deve-se solicitar o efeito principal do fator ("Sexo") e da covariável ("Massa corporal 1"), e a interação ("Massa corporal 1\*Sexo"). Para isso, basta selecionar ambas as variáveis em "Fatores e covariáveis", marcar o "Tipo: interação" e, com auxílio do botão da seta, inserir no "Modelo". Clique em "Continuar", e em seguida, clique em "Ok" ou "Colar".

No arquivo de saída, é possível observar que não há interação entre sexo e a massa corporal basal ( $p = 0,054$ ). Ou seja, as categorias apresentam o mesmo comportamento. Se houvesse interação ( $p \leq 0,050$ ), as categorias (sexos masculino e feminino) apresentariam comportamentos de massa corporal basal diferentes, por exemplo, um seria o inverso do outro.

Podemos demonstrar os resultados da ANCOVA através de um gráfico (Figura 5). Para construção do gráfico, clique em "Gráficos", "Caixas de diálogo legadas" e selecione a opção "Dispersão/ponto...". Clique em "Dispersão simples" e em "Definir". Insira a variável "Massa corporal 3" no "Eixo Y" e "Massa corporal 1" no "Eixo X". Selecione a variável "Sexo" e insira em "Colunas". Clique em "Ok" ou "Colar".

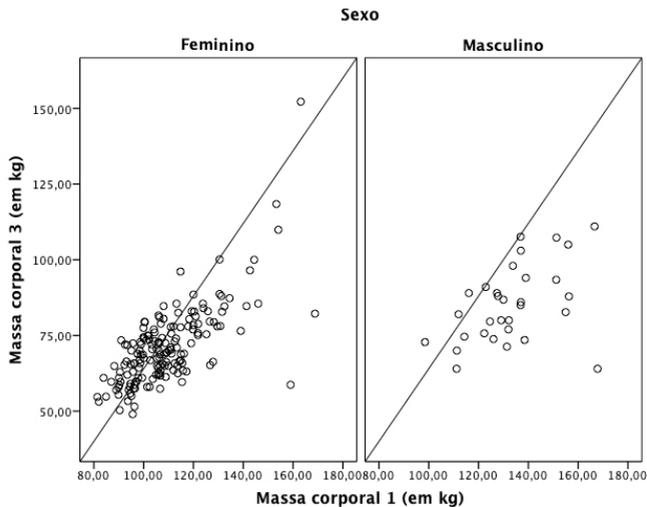


Figura 5. Gráfico de dispersão para a massa corporal basal versus a massa corporal final, discriminado pelo sexo.

O gráfico obtido é do tipo “scatter plot” (gráfico de dispersão) para a massa corporal basal versus a massa corporal final, discriminado pelo sexo. A linha de tendência foi desenhada com um duplo clique na imagem do gráfico no arquivo de saída. Com o botão direito, clique sobre o gráfico na janela “Editor de gráfico” e selecione “Adicionar linha de referência da equação”. Analisando, então, a associação entre a massa corporal final e basal discriminado por sexo, observamos que em ambos os sexos o comportamento é o mesmo. Assim, para a ANCOVA, não há interação dos fatores.

### Teste de Kruskal-Wallis

Quando os pré-requisitos necessários para o teste paramétrico não podem ser cumpridos para a variável de estudo, não se pode confiar no resultado de uma análise de variância, pois a probabilidade de se cometer um erro do tipo I afasta-se perceptivelmente de  $\alpha$ . O teste de Kruskal-Wallis substitui a ANOVA one-way quando isto ocorre. Juntamente com o teste U de Mann-Whitney, o teste de Kruskal-Wallis é um dos testes não paramétricos mais utilizados. É utilizado para comparação de variáveis ordinais ou contínuas (intervalares) entre três ou mais grupos independentes. Quando constatada uma diferença significativa, temos indicação

de que pelo menos duas amostras diferem na variável de interesse. Para verificar qual dos pares se diferenciam, o teste de Kruskal-Wallis apresenta uma análise de *post hoc* de Dunn.

Para exemplificar, no "Banco de dados 3 casos novos.sav" (disponível em <https://bit.ly/bancosdedados>), vamos comparar as medianas da variável "Massa corporal" nos quatro grupos da variável "Classificação IPAQ". Para tanto, será utilizado o teste de Kruskal-Wallis para amostras independentes, visto que há quatro grupos a serem comparados e a variável possui distribuição não normal em um dos grupos. Lembre-se: antes de optarmos pelo teste, paramétrico ou não paramétrico, é preciso testar a normalidade de distribuição da variável entre os grupos pelo teste de Shapiro-Wilk.

O teste é executado da mesma forma que se realiza o teste U de Mann-Whitney. No menu "Analisar", "Testes não paramétricos", clique em "Amostras independentes...". Na aba "Objetivo", consta como padrão "Compara automaticamente distribuições entre grupos". Na aba seguinte ("Campos"), localize e insira, na tela da esquerda, a(s) variável(is) a ser(em) testada(s) (que deve estar declarada como escalar) no campo "Testar campos" e a variável categórica (que deve ser declarada como numérica) em "Grupos", com o auxílio do botão da seta.

Como mencionado no teste U de Mann-Whitney, a comparação não paramétrica é muito simples. Você define se as amostras são independentes ou pareadas e o restante (a escolha de qual teste utilizar) é automaticamente feito pelo SPSS. Para isso, na aba "Configurações", você deve marcar a opção "Escolhe automaticamente os testes com base nos dados". Desta forma, quando se deseja comparar grupos independentes, o SPSS irá escolher automaticamente se fará o teste para dois grupos (teste U de Mann-Whitney) ou para três ou mais grupos (teste de Kruskal-Wallis), baseado no número de categorias da variável informada no campo "Grupos" da aba "Campos".

**Atenção!** É necessário classificar corretamente as variáveis quanto ao seu tipo (escalar, ordinal ou nominal) para o teste ser aplicado adequadamente. Reveja os conceitos no capítulo 1. Você também pode definir o teste manualmente, selecionando a opção

“Kruskal-Wallis 1-way ANOVA (amostras k)” em “Customizar testes”. Por este caminho, é possível alterar o nível de significância ( $\alpha$ ) em “Opções de teste”. Por fim, clique em “Executar” ou “Colar” para obter o resultado do teste estatístico.

O teste utilizado é informado no arquivo de saída, junto com a hipótese de nulidade, o p-valor e a decisão do teste (baseada em uma significância de 5%). Conforme informado pelo teste de hipótese (*Hypothesis Test Summary*), com  $p = 0,346$ , deve-se reter  $H_0$  de que a distribuição de massa corporal entre as categorias do IPAQ é a mesma. Ou seja, podemos afirmar que não há diferença significativa na massa corporal entre as categorias de atividade física avaliadas pelo IPAQ ( $p = 0,346$ ). Podemos, ainda, descrever os dados em uma tabela (Tabela 5).

Tabela 5. Comparação da massa corporal entre as categorias do IPAQ. Os dados são representados por mediana e quartis (25-75). As unidades de medida e número de pacientes são apresentadas com as variáveis correspondentes.

Categorias do IPAQ	Sedentário (n = 4)	Insuf. Ativo B (n = 8)	Insuf. Ativo A (n = 8)	Ativo (n = 10)	P-valor
Massa corporal (em kg)	126,0 (120,5-131,5)	1120,0 (100,8-131,0)	108,9 (99,5-118,0)	110,5 (101,0-119,0)	0,346*

IPAQ, questionário internacional de atividade física; n, número absoluto.

\*Teste de Kruskal-Wallis. Significância estatística considerada de  $p \leq 0,050$ .

Supondo que o teste de Kruskal-Wallis fosse significativo, isso indicaria que pelo menos duas categorias possuem distribuições diferentes. Mas, quais são as categorias? Para obter esse resultado, deve-se clicar duas vezes sobre o resultado da hipótese (*Hypothesis Test Summary*), no arquivo de saída. Na janela “Visualizador do modelo”, no campo inferior direito, em “Visualização”, selecione a opção “Comparações por pares”. O novo resultado apresentado, informa uma tabela com as comparações par a par e um gráfico que indicará as diferenças entre os pares. A comparação par a par corresponde ao teste de *post hoc* de Dunn (ou Dunn-Bonferroni). Na tabela, na primeira coluna, estará a comparação par a par e, nas

duas últimas colunas, o p-valor e p-valor ajustado (este segundo deve ser levado em consideração). Em algumas situações, é possível que o teste global indique que há diferença entre os grupos, porém, na comparação par a par, nenhuma seja significativa, já que esses são testes independentes.

## Referências

Callegari-Jacques SM. Bioestatística: princípios e aplicações. Porto Alegre: ArtMed, 2011. 255p.

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

Guimarães L.S.P, Hirakata V.N. Uso do Modelo de Equações de Estimções Generalizadas na análise de dados longitudinais. Revista HCPA. 2012;32(4):503-511.

Mancuso ACB, Nune LN, Guimarães LSP, Camey AS, Hirakata VN. Métodos Estatísticos para Desfechos Quantitativos utilizando o SPSS v.18, 2017. Disponível em <https://sites.google.com/a/hcpa.edu.br/bioestatcursos/home/ii>. Acessado em 13/07/2020.

## Exercícios sugeridos

1. No “Banco de dados 1 Resolvido.sav”, a idade é diferente entre os alunos do nível mestrado e doutorado?
2. No “Banco de dados 3 Resolvido.sav”, há diferenças nas medidas pressóricas (PAS, PAD e PAM) e lipídicas (colesterol total, triglicerídeos, HDL-Colesterol e LDL-Colesterol) entre os sexos? Construa a tabela apresentando os dados da análise descritiva, o teste estatístico utilizado e o p-valor encontrado e, interprete os resultados para um nível de significância de 5%.
3. No “Banco de dados 3 casos novos.sav”, há diferenças nas medidas pressóricas (PAS e PAD) e lipídicas (colesterol total, triglicerídeos e HDL-Colesterol) entre os sexos? Construa a tabela apresentando os dados da análise descritiva, o teste estatístico utilizado e o p-valor encontrado e, interprete os resultados para um nível de significância de 5%.