



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

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

8 Desfechos quantitativos: amostras pareadas

*Fernanda Vargas Ferreira
Fernanda Dapper Machado
Edison Capp
Otto Henrique Nienov*

Neste capítulo, iremos abordar os testes paramétricos e não paramétricos indicados para amostras pareadas. 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.

Os testes para amostras pareadas (relacionadas ou emparelhadas), que são consideradas independentes quanto à relação entre os indivíduos, são utilizados quando se objetiva comparar variáveis que tenham alguma dependência, seja por serem do mesmo sujeito de pesquisa ou de um controle emparelhado (por exemplo, por sexo, idade, etc.).

Teste t pareado

É um teste análogo do teste t de Student, para amostras paramétricas relacionadas, ou seja, é utilizado quando a suposição de normalidade é aceita para ambos os grupos pareados. Alguns autores defendem que a normalidade das diferenças entre os pares também seja um pressuposto do teste. Por exemplo, quando o interesse é comparar duas medições (no tempo) em um único grupo, onde é realizado entre elas uma intervenção, a análise recomendada é o teste t pareado.

O teste t pareado avalia se as médias de duas medidas relacionadas são estatisticamente diferentes uma da outra. Neste caso, a hipótese de nulidade de um teste t para amostras pareadas é de que a média das diferenças entre as medidas relacionadas é igual a zero, ou seja, não há diferença entre as medidas.

Além da comparação de tempo, métodos também podem ser comparados pelo teste t pareado, quando estes estiverem na mesma unidade e atenderem aos pressupostos do teste. Por exemplo, o cálculo da medida de consumo energético utilizando o Questionário de Frequência Alimentar comparado à média do Recordatório de 24 horas.

Para exemplificar, no "Banco de dados 4.sav" (disponível em <https://bit.ly/bancosdedados>), vamos comparar as medidas de massa corporal avaliadas em dois momentos ("Massa corporal 1" e "Massa corporal 2") no sexo masculino. Para isso, é necessário dividir o arquivo, no menu "Dados", "Dividir arquivo", "Organizar saída por grupos" baseados pela variável "Sexo". Em seguida, precisamos verificar a distribuição de normalidade das variáveis pelo teste de Shapiro-Wilk, como visto no capítulo 7, no menu "Analisar", "Estatísticas descritivas", "Explorar...".

Com o arquivo dividido, observa-se uma distribuição normal (significância do teste de Shapiro-Wilk superior a 0,05) para o sexo masculino, sendo apropriado o uso do teste t pareado para comparar as médias de massa corporal e, para o sexo feminino, observa-se uma distribuição não-normal (significância do teste de Shapiro-Wilk inferior ou igual a 0,05). Logo, para comparar as medianas de massa corporal no sexo feminino, utilizaremos o teste de Wilcoxon, que será abordado a seguir.

Para realizar o teste t pareado, no menu "Analisar", "Comparar médias", clique em "Teste T de amostras em Pares". Na janela "Teste T de amostras em pares", insira em "variáveis em pares" as variáveis "Massa corporal 1" ("Variável1") e "Massa corporal 2" ("Variável2"). Em seguida, clique em "Ok" ou "Colar".

No arquivo de saída, encontramos os resultados para ambos os sexos. No entanto, considerando a distribuição não normal para o sexo feminino, iremos apenas interpretar o valor do teste para o sexo masculino. Na primeira tabela (*Paired Samples Statistics*), temos a análise descritiva, com média (*Mean*), frequência absoluta (*N*), desvio padrão (*Std. Deviation*) e erro padrão da média (*Std Error Mean*). Em seguida (*Paired Samples Correlations*), é apresentado o coeficiente de correlação de

Pearson ($r = 0,885$) e o p-valor ($p < 0,001$). Os testes de correlação serão abordados no capítulo 9.

E, finalmente, na terceira tabela (*Paired Samples Test*), encontra-se o resultado do teste t pareado. São apresentados média, desvio padrão, erro padrão da média e intervalos de confiança calculados a partir da diferença entre as medidas de massa corporal. A coluna "t" traz o cálculo da diferença média dividida pelo seu desvio-padrão e, em "df", temos o grau de liberdade do teste. O valor de p do teste t de amostras pareadas se encontra na última coluna da tabela (*Sig. (two-tailed)*).

Na interpretação, com p-valor inferior a 0,05, há diferença significativa entre as médias de massa corporal no sexo masculino, ou seja, a massa corporal após a cirurgia bariátrica ($96,6 \pm 14,3$ kg) diminuiu significativamente em comparação à massa corporal basal ($133,3 \pm 16,7$ kg, $p < 0,001$). Podemos apresentar o resultado em uma tabela (Tabela 1).

Tabela 1. Comparação da massa corporal antes e após a cirurgia bariátrica no sexo masculino. Os dados são representados por média \pm desvio padrão. As unidades de medida e número de pacientes são apresentadas com as variáveis correspondentes.

	Tempo 1 (n = 31)	Tempo 2 (n = 31)	P-valor*
Massa corporal (em kg)	$133,3 \pm 16,7$	$96,6 \pm 14,3$	$<0,001$

n, número absoluto.

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

Teste de Wilcoxon

O teste de Wilcoxon é uma opção de teste não paramétrico para variáveis contínuas, quando se deseja comparar duas amostras emparelhadas, sendo substituto do teste t pareado. A hipótese de nulidade deste teste é que a mediana das diferenças entre os pares é igual a zero. Este teste avalia a diferença entre a quantidade de pares com diferenças positivas e a quantidade de pares com diferenças negativas. O teste de Wilcoxon não é recomendado quando há muitos pares com diferença igual a

zero, pois estes acabam por reduzir o tamanho da amostra, sendo excluídos da análise.

Como as variáveis “Massa corporal 1” e “Massa corporal 2” apresentaram distribuição não normal no sexo feminino pelo teste de Shapiro-Wilk, utilizaremos o teste de Wilcoxon para avaliar a diferença entre as medianas de massa corporal avaliadas antes e após a cirurgia bariátrica. Para isso, no “Banco de dados 4.sav”, no menu “Analisar”, “Testes não paramétricos”, clique em “Amostras relacionadas”. Na janela “Testes Não Paramétricos: Duas ou Mais Amostras Relacionadas”, na aba “Objetivo”, a opção padrão do SPSS é “Compara automaticamente dados observados a hipotetizados”. Clique na aba “Campos”, selecione e insira as variáveis a serem testadas (“Massa corporal 1” e “Massa corporal 2”) em “Testar Campos”, com o auxílio do botão da seta. É importante destacar que somente variáveis classificadas como escalares poderão ser adicionadas ao teste.

Assim como mencionado nos testes para amostras independentes do capítulo 7, o teste não paramétrico é muito simples. Na aba “Configurações”, deve-se marcar a opção “Escolhe automaticamente os testes com base nos dados” e o SPSS irá escolher automaticamente se realizará o teste para dois grupos (teste de Wilcoxon) ou para três ou mais grupos (teste de Friedman) baseado no número de variáveis no campo “Testar campos”. Você também pode definir o teste, selecionando a opção “Wilcoxon (2 amostras)” em “Customizar testes”. Por este caminho, é possível alterar o nível de significância (α) em “Opções de teste”. Note que também temos as opções de testes para dados binários (dicotômicos): testes de McNemar e Q de Cochran, vistos no capítulo 6. Por fim, clique em “Executar” ou “Colar” para obter o resultado do teste estatístico.

O teste utilizado é informado na saída, bem como a hipótese de nulidade, o p-valor e a decisão baseada em uma significância de 5%. Novamente, teremos os resultados de teste para ambos os sexos, mas iremos considerar apenas o resultado para o sexo feminino. 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 diferença entre as medianas de massa corporal seja igual a zero (0). Ou seja, com p-valor inferior a 0,001, há diferença significativa nas medianas de massa corporal entre os tempos pré- e pós-operatório no sexo feminino. A mediana de massa corporal basal foi de 106,5 (98,0-115,7) kg e a do pós-operatório foi de 78,4 (71,2-87,2) kg. Podemos apresentar os dados em uma tabela (Tabela 2).

Tabela 2. Comparação da massa corporal antes e após a cirurgia bariátrica no sexo 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.

	Tempo 1 (n = 164)	Tempo 2 (n = 164)	P-valor
Massa corporal (em kg)	106,5 (98,0-115,7)	78,4 (71,2-87,2)	<0,001*

n, número absoluto.

*Teste de Wilcoxon. 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 histograma contendo os deltas (ou diferenças) da massa corporal, o percentual do delta, o número de diferenças negativas, positivas e iguais a zero e, apesar de não indicarem o tamanho de efeito, essas informações podem indicar a direção do efeito.

Análise de variância para medidas repetidas (ANOVA-MR)

A análise de variância para medidas repetidas (ANOVA-MR) é uma extensão da ANOVA abordada no capítulo anterior, que pode ser utilizada para análises longitudinais. Medidas repetidas é o termo utilizado quando o sujeito analisado participa de todas as condições do experimento. Na ANOVA-MR, a variável resposta é acompanhada ao longo do tempo, criando-se um modelo longitudinal com medidas repetidas.

Para este modelo, é necessário respeitar os seguintes pressupostos: a) independência, no sentido de que cada participante é independente, não interferindo no resultado do outro, assim como as variáveis não interferem umas nas outras; b) normalidade, avaliada em cada medida no tempo da variável pelo teste de Shapiro-Wilk; c) homocedasticidade, que sugere a igualdade de variância, indicada pelo teste de Levene; e d) esfericidade, medida pela esfericidade de Mauchly e correções. Este último pré-requisito assume que a relação entre os pares das condições experimentais nas quais as medidas foram realizadas é similar. A esfericidade requer que a variância das diferenças de todos os pares seja constante. Se $p \geq 0,05$, não se rejeita a H_0 de que as variâncias sejam iguais e, se $p < 0,05$, a esfericidade não pode ser assumida e, assim, existem diferenças significativas entre as variâncias.

Alguns estatísticos também defendem a conferência da normalidade dos resíduos em análises de ANOVA-MR. Os resíduos são automaticamente salvos no banco de dados como variáveis e, podemos fazer um teste de normalidade conforme vimos no capítulo 7.

Para exemplificar, no "Banco de dados 4.sav", vamos verificar se há diferença entre as medidas de massa corporal pré- e pós-operatórias ("Massa corporal 1", "Massa corporal 2" e "Massa corporal 3") em relação à atividade física basal ("Classificação do IPAQ 1"), supondo a normalidade dos dados. Para isso, não iremos considerar o banco de dados dividido pelo sexo. No menu "Dados", clique em "Dividir arquivo" e selecione "Analisar todos os casos, não criar grupos".

Para a análise de ANOVA-MR, procede-se da seguinte forma: No menu "Analisar", "Modelo Linear Geral", clique em "Medidas Repetidas". Na janela dos "Fator(es) de definição de medida...", em "Número de Níveis", informe o fator (por exemplo, "Massa") e o número de vezes que o fator se repete (neste caso, "3"). Clique em "Adicionar" e, em seguida, em "Definir". Na próxima janela ("Medidas repetidas"), aparecerão três pontos de interrogação ("?_1)", "?_2)", ("?_3)"), na tela "Variáveis dentro de assuntos (fator1)". Esses pontos devem ser substituídos pelas variáveis "Massa 1", "Massa

2" e "Massa 3", respectivamente. Em "Fator(es) entre assuntos", inserimos o fator "Classificação IPAQ 1". Em "Diagramas", no "Eixo Horizontal" vamos inserir "Massa" e, em "Linhas Separadas", o fator "IPAQ1". Clique em "Adicionar" e, então, em "Diagramas", teremos a interação "Massa*IPAQ1". Clique em "Continuar". Em "Post Hoc", selecione em "Fator(es)" o "IPAQ1" e insira em "Testes post hoc para". Na opção "Variâncias iguais presumidas", selecione "Tukey", "Bonferroni", "S-N-K" e "Duncan". Estes testes correspondem ao *post hoc*. Clique em "Continuar". Em seguida, clique em "Opções" e envie para "Exibir médias para" as variáveis ("IPAQ1" e "Massa") e a interação ("IPAQ1*Massa"). Marque a opção "Compartilhar efeitos principais" e selecione na aba "Ajuste do intervalo de confiança" o teste de "Bonferroni". Marque, em "Exibir", as opções "Estatísticas descritivas", "Estimativas de tamanho de efeito", "Potência observada" e "Testes de homogeneidade". Clique em "Continuar" e, por fim, em "Ok" ou "Colar".

No arquivo de saída, na primeira tabela (*Descriptive Statistics*), temos a análise descritiva das medidas repetidas de massa corporal para cada categoria do fator, com valores de média (*Mean*), desvio padrão (*Std. Deviation*) e absoluto (N). Em seguida, é apresentado o teste de Levene (*Levene's Test of Equality of Error Variances*). Se o teste de Levene for significativo ($p \leq 0,05$), os grupos possuem variâncias diferentes. Se o teste de Levene não for significativo ($p > 0,05$), assumimos igualdade de variâncias. Ou seja, não há diferença significativa entre a homogeneidade dos dois grupos. No exemplo, assumimos a igualdade.

Em seguida, temos o teste de esfericidade de Mauchly (*Mauchly's Test of Sphericity*), que nos fornece a estimativa de ϵ , que corresponde à medida de esfericidade. Quando há esfericidade, o valor de ϵ é igual a 1. Valores menores indicam desvio de esfericidade, como neste caso, em que temos um valor de 0,721. Este teste é altamente influenciado pelo tamanho da amostra (falha em demonstrar esfericidade em amostras pequenas e superestima em amostras grandes). Podemos verificar que nesse exemplo o teste de Mauchly foi significativo ($p < 0,001$), ou seja, existe diferença entre as variâncias. Quando isso ocorre, o próprio teste oferece um ajuste, que são as correções de Greenhouse-Geisser e de Huynh-Feldt. A correção de Greenhouse-Geisser

é mais rígida que a de Huynh-Feldt, sendo a preferencialmente utilizada. De um modo geral, recomenda-se o uso de Huynh-Feldt quando $\epsilon > 0,75$ e de Greenhouse-Geisser quando o $\epsilon < 0,75$.

Na ANOVA-MR são incluídos os resultados das análises uni e multivariadas que, em alguns casos, podem apresentar-se discordantes. Quando há esfericidade, os resultados da análise univariada têm mais poder e, do contrário, pode-se usar os resultados multivariados quando não há esfericidade. No quadro seguinte (*Tests of Between-Subjects Effects*), temos a análise de variância da ANOVA-MR, com os valores da esfericidade assumida (*Sphericity Assumed*) e corrigida (Greenhouse-Geisser e Huynh-Feldt) para a interação e o fator. Para ambos, são expressos o grau de liberdade (df), a média quadrada (*Mean Square*), o valor de F (F) e de significância (Sig.). Como visto no capítulo 7, também obtemos o coeficiente de correlação (*Partial Eta Square*), que é uma medida de tamanho de efeito, e o poder do teste (*Observed Power*), que indica a probabilidade de rejeitar corretamente a hipótese de nulidade. No exemplo, rejeita-se a H_0 do fator porque há pelo menos uma diferença entre as medidas repetidas ($p < 0,001$), mas para a interação aceitamos a H_0 de que não existe diferença ($p = 0,518$). Ou seja, na interação das massas corporais com o IPAQ, não houve diferença. Quando a interação é estatisticamente significativa, ela indica que o padrão de mudança entre os tempos é diferente entre os grupos.

Nos quadros seguintes (*Pairwise Comparisons*), são identificadas quais as medidas que diferem entre si, na avaliação *post-hoc*. As análises *post-hoc* só devem ser verificadas quando a ANOVA-MR tiver um valor de $p \leq 0,05$. Lembre-se: os testes *post-hoc* consistem na análise por pares dos grupos. A escolha do teste *post-hoc* é definida levando-se em consideração diversos fatores. Considerando os testes utilizados na análise, os testes de Bonferroni e Tukey são considerados bastante conservadores, porém, o teste de Tukey requer que os grupos tenham tamanhos iguais e permite análise por grupos. Os dois fazem a comparação por pares com intervalo de confiança 95%. O teste de Tukey é indicado quando há esfericidade. Quando tal pressuposto não é assumido, utiliza-se Bonferroni. Já os testes de Duncan e S-N-K são considerados liberais, porém também exigem grupos de

tamanhos iguais e não fornecem a comparação por pares com intervalo de confiança de 95%.

Sendo assim, como mencionado, os valores de p não foram significativos para a interação. No gráfico (Figura 1), é possível visualizar que as linhas da massa corporal seguem de forma semelhante nos cinco grupos da classificação do IPAQ. Em outras palavras, a massa corporal (efeito principal) diminui significativamente conforme o tempo ($p < 0,001$), porém, comparando os grupos (categorias do IPAQ), a massa corporal diminui da mesma forma (sem diferença estatisticamente significativa).

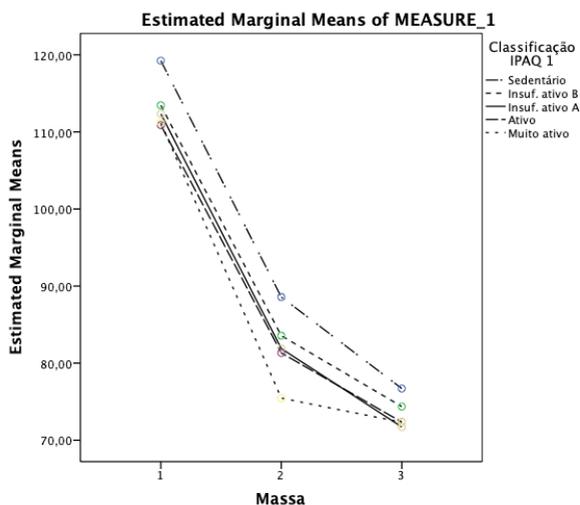


Figura 1. Comparação das medidas de massa corporal pré- e pós-operatórias em relação às categorias de atividade física basal avaliadas pelo IPAQ.

Uma amostra com muitos valores omissos limita o uso da ANOVA-MR, pois sujeitos com uma medida faltante são excluídos da análise. Ao se reportar os resultados de uma análise de ANOVA-MR, é importante incluir as médias, desvio padrão, valores de p para efeitos principais e interações.

Equações de estimativas generalizadas (GEE)

A análise de Equações de Estimativas Generalizadas (GEE) é utilizada para dados pareados e longitudinais. Esta análise tem sido considerada melhor que a ANOVA-MR, que possui pré-requisitos difíceis de serem cumpridos, além de outras limitações, tais como: o tempo, que é considerado uma variável categórica; os dados devem estar balanceados, ou seja, não deve existir nenhuma falha na coleta de dados de nenhum indivíduo e; as covariáveis ou os fatores envolvidos devem ser fixos, não podendo variar com o tempo.

A análise de GEE se baseia na metodologia dos modelos lineares generalizados, permitindo que sejam trabalhados desfechos categóricos ou quantitativos. As covariáveis ou fatores estudados podem ser avaliados novamente, mudando seus valores conforme o passar do tempo, acrescentando-se no modelo essa variação. Além disso, a variável tempo pode ser usada como uma covariável, ou seja, pode ser analisado o acréscimo de uma unidade de tempo no efeito do desfecho. Dessa forma, o método de análise de dados longitudinais pelo GEE possibilita a análise de desfechos contínuos, mesmo quando a variável não apresenta distribuição normal ou esfericidade.

A aplicação da GEE no SPSS exige uma reestruturação do banco de dados para um formato chamado de dados empilhados. No exemplo, iremos transpor as medidas de massa corporal dos casos, que se encontram nas colunas, para as linhas. Na execução da GEE, recomenda-se reformular o banco de dados separadamente para cada variável a ser analisada. Porém, é possível reformular e incluir várias variáveis para serem analisadas de uma só vez. Na reformulação, o SPSS multiplica cada caso de acordo com o número de tempos a serem avaliados, incluindo todas as avaliações em uma única variável.

Para exemplificar, no "Banco de dados 4.sav", será utilizada GEE para comparar as três medidas de massa corporal ("Massa corporal 1", "Massa corporal 2" e "Massa corporal 3") no sexo masculino, pois as variáveis apresentaram distribuição normal pelo teste de Shapiro-Wilk. Mas primeiro, é preciso reestruturar o banco de dados. Para reestruturar o banco de dados, no menu

“Dados”, clique em “Reestruturado”. Na janela “Assistente de reestruturação de dados”, haverá três opções de reestruturação: a primeira transforma um banco horizontal em vertical (opção que será utilizada), a segunda opção modifica um banco vertical em horizontal e a terceira troca de lugares colunas e linhas, transpondo o banco. Selecione a opção “Reestruturar variáveis selecionadas em casos” e clique em “Próximo”.

Na etapa 2 de 7, especifica-se o número de grupos de variáveis que serão reestruturados. Se somente um, selecione a primeira opção (neste caso) ou, se mais de um, a segunda opção e indique o número de grupos de variáveis a serem reestruturados. Clique em “Próximo”. Como visto no capítulo 3, no banco de dados, deve-se ter uma variável que identifique os casos da amostra, de preferência numérica, para organizar o banco de dados e facilitar a localização dos casos. Assim, em “Identificação de grupo de caso”, selecione “Usar variável selecionada” e indique a variável referente à identificação dos casos (“Identificação do paciente”). Em “Variável de destino”, selecione “trans 1” e insira as três medidas de massa corporal (“Massa 1”, “Massa 2” e “Massa 3”), que corresponderão ao grupo de variáveis. Caso seja reestruturado mais de um grupo de variáveis, alternamos para “trans 2” e selecionamos as medidas da variável a serem reestruturadas. Em seguida, repetimos o processo em “trans 3” e, assim, sucessivamente. No último campo da etapa 3 de 7, devem ser inseridas as variáveis fixas, que permanecem constantes no tempo. Neste caso, a variável “Sexo” será selecionada. Em seguida, clique em “Próximo”. Na etapa 4 de 7, indica-se o número de variáveis auxiliares que serão utilizadas. Neste caso, somente uma. Clique em “Próximo”.

Como teremos uma variável auxiliar, precisamos informar o tipo de valor de índice que será utilizado e, pode-se editar o nome e o rótulo dessa variável. Há duas escolhas para os valores da variável índice: 1) colocar uma sequência numérica (valores de índice 1, 2, 3) e; 2) utilizar os nomes das variáveis que estamos reestruturando. A segunda opção é indicada quando trabalhamos com a reestruturação de mais de uma variável. O nome da variável auxiliar pode ser editado no campo “Editar o nome e o rótulo da variável de índice”. Indique o nome da variável índice como “Tempo”, por exemplo. Clique em “Próximo”.

Na etapa 6 de 7, o padrão do SPSS é reestruturar o banco de dados em um novo arquivo. Em relação a valores omissos, a opção padrão é considerar todos os sujeitos, mesmo que o desfecho tenha a casela vazia. Uma opção é descartar os valores omissos do banco de dados. Clique em "Próximo". Por fim, na última etapa, é possível salvar a sintaxe desse comando ("Colar a sintaxe gerada pelo assistente em uma janela de sintaxe"). Clique em "Concluir". Selecione o comando no arquivo de sintaxe e execute-o.

No banco de dados, o resultado é reestruturado em duas variáveis, uma para o desfecho ("trans1") e outra indicando o tempo (1, 2, 3), além das variáveis fixadas ("Identificação do paciente" e "Sexo"). Agora, cada sujeito apresenta três linhas no banco de dados. É importante salvar esse banco de dados reestruturado com outra denominação, para não se perder os dados originais do banco de dados.

Com o banco de dados reestruturado, podemos realizar a análise de GEE. Para isso, vamos novamente dividir o arquivo por sexo, no menu "Dados", "Dividir arquivo" e "Organizar a saída por grupos" baseados pela variável "Sexo". Em seguida, no menu "Analisar", "Modelos Lineares Generalizados", selecione "Equações de estimativas generalizadas". Na janela "Equações de estimativa generalizadas", na aba "Repetido", defina a variável que identifica o sujeito ("Identificação do paciente") em "Variáveis de assunto", e insira a variável "Tempo" em "Variáveis dentro do assunto".

Na aba "Tipo de Modelo", defina a distribuição da variável dependente como "Linear" em "Resposta de escala", pois estamos trabalhando com variável de distribuição normal (teste de Shapiro-Wilk). Na aba "Resposta", selecione como "Variável dependente" a "Massa 1 [trans1]". Na aba "Preditores", selecione a variável categórica ("Tempo") em "Fatores" e, se necessário, pode-se inserir as variáveis contínuas em "Covariáveis". Na aba "Modelo", selecione a variável "Tempo" em "Fatores e covariáveis", selecione o "Tipo: fatorial", e a insira em "Modelo". Caso haja outras variáveis no modelo, pode-se criar uma variável de interação entre as variáveis ("Por*") na opção "Criar termo aninhado" e incluí-la no modelo. Na aba "Médias EM", selecione os "Fatores e interações" e insira em "Exibir médias para",

selecionando a opção de comparação “Por par” em “Contraste”. Em “Ajuste para comparações múltiplas”, selecione a opção “Bonferroni”. Clique em “Ok” ou “Colar”.

No arquivo de saída, encontramos os resultados para ambos os sexos. No entanto, considerando a distribuição não normal para o sexo feminino, iremos apenas interpretar o valor do teste para o sexo masculino. No primeiro quadro (*Case Processing Summary*), são indicados os valores absoluto (N) e relativo (Percent) de casos incluídos, excluídos e total. No segundo quadro (*Categorical Variable Information*), temos os valores absoluto (N) e relativo (Percent) para cada medida no tempo da massa corporal. Também são apresentados os valores mínimo e máximo, a média e o desvio padrão para a variável dependente (*Continuous Variable Information*).

No quadro seguinte (*Tests of Model Effects*), o teste de Qui-quadrado de Wald fornece o valor de p para efeitos principais e interação. Neste exemplo, somente o efeito principal foi avaliado, pois não houve interação com outra variável. Esse efeito principal foi significativo ($p < 0,001$), ou seja, ao longo do tempo houve uma mudança significativa da massa corporal no sexo masculino.

A partir desse resultado, precisamos verificar a análise de *post hoc* para definir onde ocorre essa mudança e como ocorre. No quadro “*Estimates*”, são apresentados a média (*Mean*), erro padrão (*Std. Error*) e intervalo de confiança de 95% para cada medida de massa corporal no tempo (1, 2 e 3). Em seguida (*Pairwise Comparisons*), temos os resultados do teste de Bonferroni. Fixando-se os tempos, há diferença significativa entre as três medidas de massa corporal. Com p-valor menor que 0,001, há diferença significativa entre cada medida de massa corporal nos três tempos para o sexo masculino. Ou seja, cada medida de massa corporal no tempo é diferente uma da outra. Lembre-se: quando a interação for significativa, os efeitos principais não devem ser analisados. Podemos apresentar os resultados da análise em uma tabela (Tabela 3), utilizando o sistema de letras sobrescritas.

Tabela 3. Comparação da massa corporal antes e após a cirurgia bariátrica no sexo masculino. Os dados são representados por média \pm desvio padrão. As unidades de medida e número de pacientes são apresentadas com as variáveis correspondentes.

	Tempo 1 (n = 31)	Tempo 2 (n = 31)	Tempo 3 (n = 31)	P-valor
Massa corporal (em kg)	133,3 \pm 16,7 ^a	96,62 \pm 14,3 ^b	85,3 \pm 12,7 ^c	<0,001*

Letras sobrescritas distintas representam média de massa corporal significativamente diferentes.

n, número absoluto.

* GEE. Significância estatística considerada de $p \leq 0,050$.

Teste de Friedman

O teste de Friedman substitui a análise GEE quando a variável estudada não apresentar os pré-requisitos para tal. É utilizado para a comparação de variáveis contínuas entre três ou mais tempos (amostras pareadas). Assim como no teste para amostras independentes, uma diferença significativa é a indicação de que pelo menos dois tempos diferem na variável de interesse. Como nos outros testes não paramétricos, o teste de Friedman também é baseado no ranqueamento dos dados.

Para exemplificar essa análise, no “banco de dados 4.sav”, vamos comparar as medianas de massa corporal (“Massa corporal 1”, “Massa corporal 2” e “Massa corporal 3”) para o sexo feminino, uma vez que o teste de Shapiro-Wilk indicou uma distribuição não normal. Para isso, vamos novamente dividir o arquivo por sexo, no menu “Dados”, “Dividir arquivo” e “Organizar a saída por grupos” baseados pela variável “Sexo”. Em seguida, no menu “Analisar”, “Testes não-paramétricos”, clique em “Amostras relacionadas”.

Na janela “Testes Não Paramétricos: Duas ou Mais Amostras Relacionadas”, na aba “Objetivo”, a opção padrão do SPSS é “Compara automaticamente dados observados e hipotetizados”. Clique na aba “Campos”, selecione e insira as variáveis a serem testadas (“Massa corporal 1”, “Massa corporal 2” e “Massa corporal 3”) em “Testar Campos”, com o auxílio do botão da seta. É importante destacar que somente variáveis classificadas como escalares poderão ser adicionadas ao teste.

Na aba “Configurações”, deve-se marcar a opção “Escolhe automaticamente os testes com base nos dados” e o SPSS irá escolher automaticamente se realizará o teste para dois grupos (teste de Wilcoxon) ou para três ou mais grupos (teste de Friedman) baseado no número de variáveis no campo “Testar campos”. Você também pode definir o teste, selecionando a opção “Friedman (amostras k)” em “Customizar testes”. Por este caminho, é possível alterar o nível de significância (α) em “Opções de teste”. Por fim, clique em “Executar” ou “Colar” para obter o resultado do teste estatístico.

O teste utilizado é informado na saída, bem como a hipótese de nulidade, o p-valor e a decisão baseada em uma significância de 5%. Novamente, teremos os resultados de teste para ambos os sexos, mas iremos considerar apenas o resultado para o sexo feminino. 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 das medianas de massa corporal seja a mesma. Ou seja, com p-valor inferior a 0,001, há diferença significativa nas medianas de massa corporal entre os tempos pré- e pós-operatórios no sexo feminino. A mediana de massa corporal basal foi de 106,5 (98,0-115,7) kg e, nos períodos pós-operatórios, as medianas foram de 78,4 (71,2-87,2) kg e de 69,3 (63,0-77,8) kg.

Como o resultado do teste de Friedman foi menor que 0,001, ou seja, pelo menos duas medidas diferem entre si, precisamos verificar na análise de *post-hoc* quais medianas de massa corporal são diferentes. Para isso, clique duas vezes sobre o teste de hipótese no arquivo de saída. Na janela de “Visualizador de modelo”, no campo inferior direito, em “Visualização”, deve-se selecionar a opção “Comparações por pares”. Neste caso, a massa corporal foi diferente em todas as comparações por pares. Podemos apresentar os dados em uma tabela (Tabela 4), utilizando o sistema de letras sobrescritas.

Tabela 4. Comparação da massa corporal antes e após a cirurgia bariátrica no sexo 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.

	Tempo 1 (n = 164)	Tempo 2 (n = 164)	Tempo 3 (n = 164)	P-valor
Massa corporal (em kg)	106,5 (98,0-115,7) ^a	78,4 (71,2-77,8) ^b	69,3 (63,0-77,8) ^c	<0,001*

Letras sobrescritas distintas representam medianas de massa corporal significativamente diferentes.

n, número absoluto.

*Teste de Friedman. Significância estatística considerada de $p \leq 0,050$.

Vamos supor que, considerando a perda de massa corporal significativa em ambos os sexos, o pesquisador resolva comparar as medidas entre os sexos masculino e feminino. Podemos comparar os dados da forma que estão apresentados: média (distribuição normal) para o sexo masculino e mediana (distribuição não-normal) para o sexo feminino? A resposta é não, pois a distribuição dos dados é diferente, logo, o teste de hipótese e as medidas são diferentes. Neste caso, deve-se optar pela análise não paramétrica e expressar os dados em ambos os sexos pela mediana e percentis.

Tamanho de efeito e sua implicação no cálculo amostral

Determinar o tamanho da amostra, que é um subconjunto da população, é crucial para o desenvolvimento de um estudo por aspectos de viabilidade, custos, estatística, além de possibilitar maior precisão nos resultados. Para o cálculo do tamanho amostral, considere: qual a pergunta da pesquisa? Qual a variável primária? Como será a análise estatística? Qual é o resultado de estimativa no grupo controle? Qual a diferença a ser detectada e o grau de confiança?

Adicionalmente, deve-se entender que o cálculo propiciará a quantidade de sujeitos necessários para responder à pergunta de pesquisa, sendo essencial ainda identificar quatro fatores: 1) tamanho de efeito que pode ser obtido utilizando a média e a variância; 2) distribuição da amostra; 3) poder do teste e; 4) nível de significância.

Para se calcular o tamanho da amostra, pode-se: a) realizar um estudo-piloto de tamanho pequeno definido arbitrariamente e viável de ser investigado em curto espaço de tempo; b) comparar duas médias quando o desfecho é medido através de escores de escalas; c) comparar duas proporções (por exemplo, estudos transversais, coorte); d) risco relativo (RR), que estima a magnitude da associação entre a exposição ao fator de risco e o desfecho, indicando quantas vezes a ocorrência do desfecho nos expostos é maior do que aquela entre os não-expostos; e) razão de chances (*Odds Ratio*), empregado em estudos de caso-controle, é uma estimativa do risco relativo entre o percentual de exposição nos controles (pode-se usar dados da população geral sadia) e o número de controles por cada caso.

Além disso, outros três pontos devem ser destacados no cálculo do tamanho da amostra: a) estudos longitudinais (coortes prospectivas e ensaios clínicos), devido às características de seguimento dos voluntários por determinado tempo, podem apresentar perdas (por desistência, morte, exclusão), recomendando por consequência, a correção desse cálculo da amostra inicial com aumento de até 30%; b) mediante mais de um desfecho de interesse, o tamanho da amostra deve ser calculado para cada um deles, adotando-se o maior valor e; c) se houver modificação no delineamento de pesquisa (por exemplo, de transversal para estudo de coorte) será necessário realizar um novo cálculo amostral.

Similarmente, caso a coleta de dados já tenha ocorrido, é possível calcular o tamanho do efeito, a fim de se verificar se a amostra obtida foi válida. Apenas o valor de p ser significativo não garante que o efeito é significativo ou importante. É essencial sinalizar que o tamanho do efeito deve ser entendido como uma estatística descritiva que serve como complemento ao teste de significância estatística da hipótese de nulidade. Ele propicia realizar a medição de uma potencial significância real de um efeito em uma intervenção por meio da descrição do tamanho dos efeitos observados, independentemente de um possível efeito enganoso em função do tamanho da amostra. O tamanho do efeito também é útil para se comparar resultados de estudos com diferentes escalas de medidas. O tamanho do efeito é medido em

unidades de desvio padrão, descrevendo a distância entre dois valores de média.

O cálculo do tamanho de efeito baseado em diferenças padronizadas entre as médias engloba as opções d de Cohen, g de Hedges e Δ de Glass. Por sua vez, se as variáveis independentes possuem mais de duas categorias ou são contínuas, pode-se empregar Epsilon quadrado (ϵ^2), Eta quadrado (η^2), Eta quadrado parcial ($\eta^2 p$), Eta quadrado generalizado ($\eta^2 G$) e f de Cohen. Já quando as variáveis independente e dependente são categóricas, sugere-se ϕ (f), V de Cramer (ou f_c), λ de Goodman-Kruskal e w de Cohen.

Dessa forma, com os dados já coletados, poder-se-ia aplicar a avaliação do tamanho de efeito d de Cohen, obtido pela divisão da diferença das médias pelo desvio-padrão, cuja interpretação seria: tamanho de efeito grande (valores superiores ou iguais a 0,8), tamanho de efeito médio (entre 0,8 a 0,2) e tamanho de efeito pequeno (inferiores a 0,2). Quando a variância dos grupos é similar, pode-se utilizar o desvio padrão de um dos grupos. Caso as variâncias sejam diferentes, utiliza-se o desvio padrão agrupado, calculado como a raiz-quadrada da média dos desvios padrão dos dois grupos. O teste d de Cohen é apropriado para dados de distribuição normal e não indicado para amostras pequenas por superestimar os resultados.

O g de Hedge é utilizado nas amostras pequenas em substituição ao d de Cohen. Este teste também utiliza médias e desvios padrão, porém faz um ajuste para tamanho de amostra. Para estudos experimentais, o tamanho de efeito é calculado pelo Δ de Glass, que é obtido da mesma maneira do teste d de Cohen, porém utilizando o desvio padrão do grupo controle.

O Eta quadrado é um índice de força de associação entre um fator e a variável dependente, sendo a proporção da variação total atribuída ao fator, calculado como a razão da variância do fator sobre a variância total. Os valores obtidos variam de 0 a 1. No SPSS, pode-se obter no menu "Analisar", "Comparar médias", e em "Médias, nas "Opções", selecione "Tabelas ANOVA" e "Eta". A partir do Eta quadrado pode se obter o f de Cohen, dividindo a

raiz quadrada do quadrado de Eta sobre um menos Eta quadrado. Essa conversão é apropriada quando os tamanhos dos grupos são semelhantes. Para o efeito f de Cohen, um valor de 0,1 é considerado um tamanho de efeito pequeno, 0,25 é considerado médio e 0,4 é considerado um efeito grande.

Referências

Barton, B, Peat, J. Medical Statistics: A guide to SPSS. Data analysis and Critical Appraisal. Second edition. John Wiley & Sons, Ltda. 2014. Disponível em: www.wiley.com/go/barton/medicalstatistics2.e

Coutinho ESF, Cunha GM. Conceitos básicos de epidemiologia e estatística para a leitura de ensaios clínicos controlados. Revista Brasileira de Psiquiatria, v.27, n.2, p.146-51, 2005.

Espírito-Santo H, Daniel F. Calcular e apresentar tamanhos do efeito em trabalhos científicos (1): As limitações do $p < 0,05$ na análise de diferenças de médias de dois grupos. Revista Portuguesa de Investigação Comportamental e Social, v.1, n. 1, p. 3-16, 2015.

Field, A. Discovering statistics using SPSS. Third Edition. SAGE, 2009.

Guimarães LSP, Hirakata VN. Uso do Modelo de Equações de Estimções Generalizadas na análise de dados longitudinais. Revista HCPA, v. 32, n.4, p. 503-511, 2012.

Lindenau J Dal-Ri, Guimarães LSP. Calculando o tamanho de efeito no SPSS. Revista HCPA, v. 32, n.3, p. 363-381, 2012.

Miot HA. Tamanho da amostra em estudos clínicos e experimentais. Jornal Vascular Brasileiro, v.10, n.4, p 275-278, 2011.

Pallant, J. SPSS Survival Manual. Mc Graw Hill Open University Peress. 2007.

Pereira DA de A et al. Aspectos metodológicos na construção de projetos de pesquisa em Nutrição Clínica. Revista de Nutrição, v. 27, n. 5, p. 597-604, 2014.

Exercícios propostos

1. No "banco de dados 4.sav", há diferença nas medidas de circunferência da cintura e pressórica (PAS e PAD) entre os três tempos avaliados? Construa a tabela apresentando os dados da análise descritiva, o teste estatístico utilizado e o p-valor encontrado e, interprete os resultados considerando um nível de significância de 5%.