

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

MANOVA Aplicada a Dados de Doenças Pulmonares

Douglas de Paula Nestlehner
Gabriel Cezar Lopes dos Santos
Giovanni Vicente Batista
Nícolas Córdoba Gusman

São Carlos
Dezembro de 2022

Sumário

1	Introdução	1
1.1	Objetivos	2
2	Metodologia	3
2.1	Análise de Variância Multivariada (MANOVA)	3
3	Resultados	5
3.1	Análise Exploratória	5
3.2	MANOVA	8
3.2.1	Verificação dos pressupostos	8
3.2.2	Aplicação do modelo	9
3.2.3	O2Hb_cerebral	9
3.2.4	O2Hb_respiratorio	10
3.2.5	O2Hb_periferico	11
3.2.6	HHb_cerebral	11
3.2.7	HHb_respiratorio	11
3.2.8	HHb_periferico	12
4	Conclusão	13
	Referências Bibliográficas	15
A	Figuras	16

Capítulo 1

Introdução

Segundo o Ministério da Saúde, a Doença Pulmonar Obstrutiva Crônica (DPOC) obstrui as vias aéreas, tornando a respiração difícil. A obstrução é causada por um estreitamento dos brônquios, tanto pelo engrossamento das paredes quanto pela presença de secreção dentro dos brônquios.

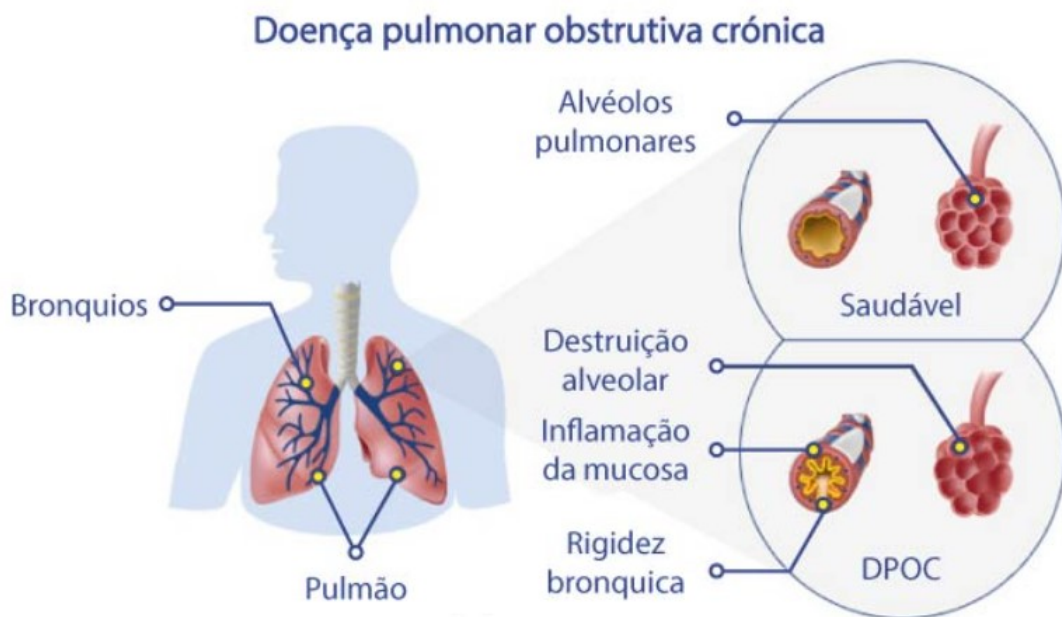


Figura 1.1: Ilustração dos efeitos da DPOC

A doença constitui um grupo de doenças respiratórias que inclui a bronquite crônica e o enfisema, e tem forte relação com o tabagismo, dado que sua origem está totalmente ligada ao efeito da fumaça nos pulmões. De acordo com o Ministério da Saúde e dados da Fundação Proar, de 80% a 90% das vezes que a DPOC se manifesta, o tabagismo é o responsável, principalmente quando falamos em pessoas com mais de 40 anos e fumantes

de 1 ou mais maços de cigarro por dia, por 10 anos ou mais.

Já a Insuficiência Cardíaca (IC), ocorre quando o coração não bombeia a quantidade de sangue suficiente para atender às necessidades do corpo. Pode ocorrer quando o músculo cardíaco está mais fraco ou quando existe algum defeito no coração que impede o sangue de sair para circular.

Os principais sintomas da Insuficiência Cardíaca são a falta de ar, cansaço, tosse, confusão mental e tontura. Em fases iniciais, algumas pessoas podem não apresentar sintomas ou ignorar sintomas como cansaço e falta de ar por serem considerados sintomas de envelhecimento.

A Doença Pulmonar Obstrutiva Crônica (DPOC) e a Insuficiência Cardíaca (IC) são bastante próximas, principalmente quando estamos falando dos sintomas. Ambas são condições de saúde que, se não forem tratadas, podem se tornar perigosas.

Neste contexto, no Capítulo 2 é apresentado com maiores detalhes a metodologia conhecida como Análise de Variância Multivariada (MANOVA) a qual será utilizada para realizar a análise dos dados de doença pulmonares. No capítulo 3 apresentaremos os resultados obtidos e no capítulo 4 serão apresentadas as conclusões sobre o estudo realizado.

1.1 Objetivos

O objetivo deste trabalho é verificar se existe diferença significativa do tratamento entre os grupos de pacientes com Doença Pulmonar Obstrutiva Crônica juntamente a Insuficiência Cardíaca (DPOC-IC) e Insuficiência Cardíaca (IC) em relação as variáveis respostas. Além disso, se possível, verificar dentro de cada um dos dois grupos de pacientes, se há diferença entre os momentos “Repouso” e “Tlim”.

Capítulo 2

Metodologia

2.1 Análise de Variância Multivariada (MANOVA)

A análise de variância multivariada (MANOVA) representa uma extensão da análise de variância para experimentos univariados (ANOVA). Para o nosso problema, estamos interessados em verificar se os tratamentos (“DPOC-IC” e “IC”) se diferem simultaneamente em relação a algumas variáveis resposta.

Dessa forma, a MANOVA permite obter conclusões a respeito da diferenciação entre tratamentos, considerando, simultaneamente, todo o conjunto de dados observados sobre as unidades experimentais. Nesta análise os testes de hipóteses consideram, além de variâncias, as covariâncias (ou correlações) entre as variáveis, como possíveis determinantes das diferenças entre os tratamentos. De acordo com [O’Brien e Kaiser \(1985\)](#), somente no caso de haver independência entre as variáveis respostas, as análises univariadas individuais resolveriam o problema.

Matricialmente, a análise baseia-se num modelo linear:

$$Y = X\beta + \epsilon \quad (2.1)$$

em que:

- Y é a matriz de observações
- X é a matriz do delineamento
- β é a matriz de parâmetros
- ϵ é a matriz dos erros aleatórios

A MANOVA possui algumas premissas que devem ser respeitadas para que o resultado obtido seja confiável, são essas: normalidade multivariada, ausência de *outliers* multivariados e homogeneidade das matrizes de covariâncias e variâncias.

Capítulo 3

Resultados

Neste capítulo apresentamos os resultados obtidos nas aplicações em estudo.

3.1 Análise Exploratória

A base de dados utilizada tem informações de 22 pacientes, sendo 11 pacientes com insuficiência cardíaca e os outros 11 com DPOC-IC. Os dados foram coletados em dois instantes de tempo (“Repouso” e “Tlim”) totalizando 44 observações.

A seguir apresentamos a descrição de cada variável:

- ***Paciente***: Variável responsável por identificar o paciente.
- ***Doença***: Variável indicativa do tipo de doença que paciente apresenta. (“Insuficiência cardíaca”(IC) ou “DPOC-IC”)
- ***Instante***: Variável indicativa do instante em que foi coletada as medidas dos pacientes. (“Repouso” ou “Tlim”)
- ***O2Hb_cerebral***: Medida de saúde. (Variável quantitativa)
- ***O2Hb_respiratorio***: Medida de saúde. (Variável quantitativa)
- ***O2Hb_periferico***: Medida de saúde. (Variável quantitativa)
- ***HHb_cerebral***: Medida de saúde. (Variável quantitativa)
- ***HHb_respiratorio***: Medida de saúde. (Variável quantitativa)
- ***HHb_periferico***: Medida de saúde. (Variável quantitativa)

Para analisar os dados, realizamos a reestruturação da base bruta, na Tabela 3.1 re-presentamos a base reestruturada, com informações de 2 pacientes.

Tabela 3.1: Informações de dois dos pacientes presente na base dados.

<i>Paciente</i>	<i>Doença</i>	<i>Instante</i>	<i>O2Hb_cerebral</i>	<i>...</i>	<i>HHb_perferico</i>
AA	Insuficiência cardíaca	Repouso	1.4688	...	1.2457
AA	Insuficiência cardíaca	Tlim	-3.9405	...	7.1117
DC	DPOC-IC	Repouso	0.9985	...	1.0126
DC	DPOC-IC	Tlim	-4.4301	...	3.7346

As Tabelas 3.2 e 3.3 apresentam algumas medidas descritivas de cada variável quantitativa responsável por indicar o status de saúde do paciente, para os dados coletados em “Repouso” e “Tlim”.

Tabela 3.2: Medidas descritivas das variáveis indicativas do status de saúde, coletadas quando os pacientes estavam em “Repouso”.

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
O2Hb_cerebral	-1.3896	-0.2124	0.3116	0.3020	1.0232	1.4688
O2Hb_respiratorio	-1.4681	-0.3157	0.2989	0.3564	1.0028	3.1457
O2Hb_periferico	-1.3799	-0.0671	0.4348	0.3612	0.8911	1.9890
HHb_cerebral	-1.5559	-0.4389	-0.0088	0.1300	0.7561	2.1424
HHb_respiratorio	-0.8838	-0.4108	0.1056	0.1735	0.5227	1.9900
HHb_periferico	-1.0518	-0.3073	0.0587	0.0263	0.4015	1.2878

Tabela 3.3: Medidas descritivas das variáveis indicativas do status de saúde, coletadas quando os pacientes estavam em “Tlim”.

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
O2Hb_cerebral	-10.1250	-4.5550	3.7300	-2.9670	-1.0480	7.6350
O2Hb_respiratorio	-11.1020	-9.1750	-5.0470	-5.0330	-2.180	1.4620
O2Hb_periferico	-15.2350	-10.430	-7.8550	-6.9090	-3.6130	4.1980
HHb_cerebral	-3.8280	1.6830	3.7730	2.8350	4.5970	5.6270
HHb_respiratorio	-1.5090	4.5680	5.3530	5.8390	5.930	16.1720
HHb_periferico	2.9560	4.8310	6.0950	6.8820	8.452	14.6540

Observando as medidas descritivas podemos notar que:

- Quando os pacientes estavam em “Tlim”, os indicadores apresentaram alta variabilidade em praticamente todas as variáveis;

- Ainda em Tlim, observamos indícios de *outliers* (max. e min. distantes da média/mediana), o que não notamos quando os pacientes estão em “Repouso”;
- Para os dados em “Repouso” notamos características bem diferentes ao “Tlim”, dentre as principais: diferença de grandeza dos indicadores (“Repouso” bem abaixo) e variabilidade baixa.

De modo geral, apenas com as medidas descritivas já pudemos notar diferença entre os dois instantes (“Repouso” e “Tlim”), o que já era esperado, uma vez que os pacientes foram expostos a um tratamento e, portanto, deseja-se obter mudanças nos indicadores de saúde.

Entretanto, também tivemos como interesse verificar se existe diferença nos indicadores considerando o tipo de doença (“Insuficiência cardíaca” e “DPOC-IC”).

Na Figura 3.1 representamos os box-plots construídos considerando a variável “O2Hb_cerebral” para cada tipo de doença nos dois instantes.

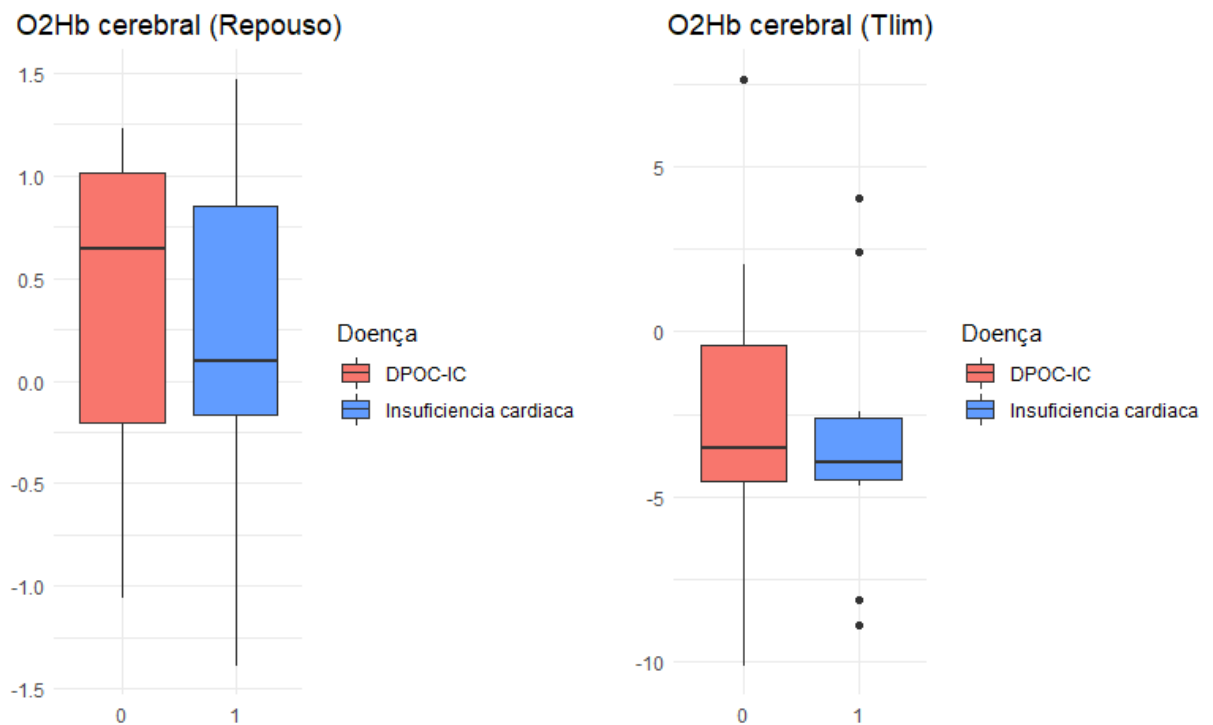


Figura 3.1: Box-plots da variável “O2hb_cerebral” por tipo de doença, considerando os diferentes instantes.

No instante “Tlim”, notamos que a variável “O2Hb_cerebral” tem maior variabilidade para a doença “DPOC-IC” e menor numero de outliers em relação a doença “Insuficiência Cardíaca”, entretanto a média apresentada para as duas doenças são bem próximas.

Já no instante “Repouso”, notamos comportamento semelhante em relação a variabilidade e outliers, porém com médias diferentes. Levantando o questionamento se o efeito do tratamento “Repouso”, na variável “O2Hb_cerebral”, é diferente para os diferentes tipos de doenças.

Realizando a mesma análise de box-plot para as demais variáveis, foi passível observar comportamentos distintos em cada variável, porém sempre apresentando resultados que indicam uma certa diferença na variável resposta dependendo do tipo de doença. Os gráficos de box-plot das demais variáveis estão representados no apêndice [A](#).

Contudo, poderíamos verificar se realmente existe diferença nas variáveis respostas dependendo da doença, avaliando cada variável por meio de uma ANOVA. Entretanto, todas as variáveis respostas (“O2Hb_cerebral”, “O2Hb_repiratorio”, “O2Hb_perferico”, “HHb_cerebral’’, “HHb_repiratorio” e “HHb_perferico”) são dependentes, consequentemente não teremos como interesse analisar variável por variável, e sim uma análise multivariada que considere todas.

3.2 MANOVA

Estamos interessados em avaliar se existe diferença significativa entre as duas doenças em relação às variáveis resposta nos dois instantes de tempo em que foram medidas. Nesse caso, como temos seis variáveis resposta e dois fatores, cada um com dois níveis, decidimos conduzir uma MANOVA bidirecional.

3.2.1 Verificação dos pressupostos

Primeiramente vamos verificar se os dados atendem à condição de normalidade multivariada por grupo de pacientes. Realizando o teste de Henze-Zirkler, obtemos os resultados apresentados na Tabela [3.4](#).

Tabela 3.4: Teste hz para normalidade multivariada.

Grupo	Normalidade
Normalidade Multivariada para DPOC-IC	Sim
Normalidade Multivariada para IC	Sim

Com isso, temos normalidade multivariada para os dois grupos. No entanto, ao verificar a normalidade univariada, pelo teste de Anderson-Darling, obtivemos que a variável

“HHb_respiratório” não possui normalidade para os grupos “DPOC-IC” e “IC”, além disso, a variável “HHb_cerebral” não possui normalidade para o grupo “IC”.

Ao testar a condição de *outliers* multivariados através da distância de mahalanobis, obtivemos que nenhuma das observações configura.

Por fim, para verificar a condição de homogeneidade das matrizes de covariâncias e variâncias, realizamos o teste de box que possui as seguintes hipóteses:

H_0 : As matrizes de variâncias-covariâncias são homogêneas.

H_1 : As matrizes de variâncias-covariâncias não são homogêneas.

Ao realizar o teste, obtemos o $p\text{-valor} = 0.418 > 0,001$, ou seja, as matrizes são homogêneas.

Portanto, os pressupostos necessários para a realização da MANOVA foram atendidos, apesar de não atingir a normalidade univariada para duas das variáveis dependentes. Sendo assim, podemos confiar nos resultados obtidos pela análise de variância multivariada.

3.2.2 Aplicação do modelo

Primeiramente, desejamos testar as seguintes hipóteses:

H_0 : Não existe interação entre as duas doenças e os instantes de tempo de interesse.

H_1 : Existe interação entre as duas doenças e os instantes de tempo de interesse.

Conduzindo a MANOVA a partir do teste de Wilks, obtemos um valor-p de 0.0007. Portanto, nós rejeitamos H_0 ao nível de significância de 5%, isto é, concluímos que de fato existe interação entre as doenças e os instantes de tempo de interesse.

Dessa forma, conduziremos também comparações múltiplas pelo teste de Tukey para cada uma das variáveis resposta, a fim de determinar qual combinação de doença e instante é maior ou menor do que outra, em média.

3.2.3 O2Hb_cerebral

Analisando os intervalos de confiança que contém o zero, vemos que a variável O2Hb_cerebral é maior, em média, para o grupo de pacientes com insuficiência cardíaca no instante de repouso do que para o grupo com DPOC-IC no instante de repouso, e também maior para o grupo com insuficiência cardíaca no instante tlim do que para o grupo com DPOC-IC no instante tlim. Além disso, a variável é, em média, menor para o grupo com DPOC-IC

Combinação de fatores	Diferença	Limite inferior	Limite superior
IC:repouso - DPOC-IC:repouso	0.1349215	-3.443836	3.71367881
DPOC-IC:tlim - DPOC-IC:repouso	-3.5047693	-7.083527	0.07398795
IC:tlim - DPOC-IC:repouso	-2.8974384	-6.476196	0.68131886
DPOC-IC:tlim - IC:repouso	-3.6396909	-7.218448	-0.06093360
IC:tlim - IC:repouso	-3.0323600	-6.611117	0.54639731
IC:tlim - DPOC-IC:tlim	0.6073309	-2.971426	4.18608817

Tabela 3.5: Comparações múltiplas de Tukey com 5% de confiança para a variável O2Hb_cerebral

no instante tlim do que para o grupo com DPOC-IC no instante repouso, menor para o grupo com IC no instante tlim do que para o grupo com DPOC-IC no repouso e também menor para o grupo com IC no instante tlim do que para o grupo com IC no instante repouso.

Portanto, vemos que a variável O2Hb_cerebral é, de forma geral, maior para pacientes com insuficiência cardíaca do que para pacientes com DPOC-IC, e também maior no instante de repouso do que no instante tlim.

3.2.4 O2Hb_respiratorio

Combinação de fatores	Diferença	Limite inferior	Limite superior
IC:repouso - DPOC-IC:repouso	-0.4252792	-3.442543	2.5919843
DPOC-IC:tlim - DPOC-IC:repouso	-7.8761965	-10.893460	-4.8589330
IC:tlim-DPOC - IC:repouso	-3.3285689	-6.345832	-0.3113054
DPOC-IC:tlim - IC:repouso	-7.4509174	-10.468181	-4.4336539
IC:tlim - IC:repouso	-2.9032897	-5.920553	0.1139738
IC:tlim - DPOC-IC:tlim	4.5476277	1.530364	7.5648912

Tabela 3.6: Comparações múltiplas de Tukey com 5% de confiança para a variável O2Hb_respiratorio

Nesse caso, temos apenas dois intervalos de confiança que não englobam o zero. Vemos que a variável O2Hb_respiratorio é, em média, maior para os pacientes que possuem DPOC-IC do que para os pacientes com IC, mas apenas no instante repouso. No tlim, não podemos afirmar isso. Da mesma forma, observamos que a média da variável é maior no repouso do que em tlim, mas apenas nos pacientes que possuem IC. Entre o grupo de pacientes que possuem DPOC-IC, não podemos afirmar isso.

Combinação de fatores	Diferença	Limite inferior	Limite superior
IC:repouso - DPOC-IC:repouso	0.2996808	-2.490445	3.0898062
DPOC-IC:tlim - DPOC-IC:repouso	-10.5228057	-13.312931	7.73268030
IC:tlim - DPOC-IC:repouso	-3.7184356	-6.508561	0.9283102
DPOC-IC:tlim - IC:repouso	-10.8224865	-13.612612	8.0323612
IC:tlim - IC:repouso	-4.0181164	-6.808242	1.2279911
IC:tlim - DPOC-IC:tlim	6.8043701	4.014245	9.5944954

Tabela 3.7: Comparações múltiplas de Tukey com 5% de confiança para a variável O2Hb_periferico

3.2.5 O2Hb_periferico

Vemos que, no instante repouso, a variável O2Hb_periferico é maior, em média, no grupo com IC do que no grupo com DPOC-IC. Além disso, o grupo com DPOC-IC no repouso possui valores maiores do que o grupo com DPOC-IC em tlim e do que o grupo com IC em tlim. Por fim, a variável é maior para o grupo IC:repouso do que os grupos DPOC-IC:tlim e IC:tlim.

3.2.6 HHb_cerebral

Combinação de fatores	Diferença	Limite inferior	Limite superior
IC:repouso - DPOC-IC:repouso	-0.5660475	-2.7035121	1.5714172
DPOC-IC:tlim - DPOC-IC:repouso	3.5988335	1.4613688	5.7362982
IC:tlim - DPOC-IC:repouso	1.2443897	-0.8930750	3.3818543
DPOC-IC:tlim - IC:repouso	4.1648810	2.0274163	6.3023456
IC:tlim - IC:repouso	1.8104371	-0.3270276	3.9479018
IC:tlim - DPOC-IC:tlim	-2.3544438	-4.4919085	-0.2169792

Tabela 3.8: Comparações múltiplas de Tukey com 5% de confiança para a variável HHb_cerebral

A combinação de fatores IC:repouso possui valores médios de HHb_cerebral maiores do que os grupos IC:tlim, DPOC-IC:tlim e DPOC-IC:repouso, que por sua vez possui valores maiores do que os grupos DPOC-IC:tlim e IC:tlim.

3.2.7 HHb_respiratorio

Nesse caso, em ambos os instantes o grupo com DPOC-IC apresenta valores médios de HHb_respiratorio maiores do que o grupo com IC.

Tabela 3.9: Comparações múltiplas de Tukey com 5% de confiança para a variável HHb_respiratorio

Combinação de fatores	Diferença	Limite inferior	Limite superior
IC:repouso - DPOC-IC:repouso	-0.8646716	-3.639598	1.910254
DPOC-IC:tlim - DPOC-IC:repouso	5.6217576	2.846832	8.396684
IC:tlim - DPOC-IC:repouso	4.8439652	2.069039	7.618891
DPOC-IC:tlim - IC:repouso	6.4864291	3.711503	9.261355
IC:tlim - IC:repouso	5.7086367	2.933711	8.483563
IC:tlim - DPOC-IC:tlim	-0.7777924	-3.552718	1.997134

3.2.8 HHb_periferico

Por fim, o grupo com DPOC-IC apresenta maiores valores de HHb_periferico do que o grupo com IC, tanto no instante repouso quanto tlim.

Tabela 3.10: Comparações múltiplas de Tukey com 5% de confiança para a variável HHb_periferico

Combinação de fatores	Diferença	Limite inferior	Limite superior
IC:repouso - DPOC-IC:repouso	0.331481	-2.080518	2.7434797
DPOC-IC:tlim - DPOC-IC:repouso	8.563690	6.151691	10.9756883
IC:tlim - DPOC-IC:repouso	5.478900	3.066901	7.8908986
DPOC-IC:tlim - IC:repouso	8.232209	5.820210	10.6442073
IC:tlim - IC:repouso	5.147419	2.735420	7.5594175
IC:tlim - DPOC-IC:tlim	-3.084790	-5.496788	-0.6727911

Capítulo 4

Conclusão

A partir do que foi exposto na análise descritiva, pudemos ter uma ideia de como seriam os resultados posteriormente. Analisando as medidas descritivas para os diferentes grupos “Repouso” e “Tlim”, vimos que os indicadores pertencentes ao grupo “Tlim” apresentaram uma maior variabilidade em praticamente todas as variáveis resposta, uma diferença que já era esperada dado que alguns pacientes foram expostos à um tratamento.

Ainda na análise descritiva, verificamos também indícios de diferença entre os tipos de doença “Insuficiência Cardíaca” e “DPOC-IC” para cada variável resposta. Os resultados obtidos foram distintos em cada uma delas, porém, apresentando também uma certa diferença entre as doenças.

Após realizado a validação dos pressupostos, foi conduzido a MANOVA a partir do teste de Wilks com o objetivo de descobrir se existe ou não interação entre as duas doenças e os instantes de tempo de interesse. O p-valor obtido foi de 0.0007 e, considerando um nível de significância de 5%, concluímos que de fato existe interação entre as doenças e os instantes de tempo de interesse. Sabendo disso, com o intuito de determinar qual combinação de doença é menor ou maior que a outra, em média, foi conduzido comparações múltiplas pelo teste de Tukey.

Em geral, concluímos que a variável O2Hb_cerebral é, de forma geral, maior para pacientes com insuficiência cardíaca do que para pacientes com DPOC-IC, e também maior no instante de “Repouso” do que no instante “Tlim”. Para a variável O2Hb_respiratorio, em média, vemos que é maior para os pacientes que possuem DPOC-IC do que para os pacientes com IC, mas apenas no instante ‘Repouso’. Já para a variável O2Hb_periferico, o resultado obtido foi o contrário pra variável O2Hb_respiratorio. Em relação à variável HHb_cerebral, a combinação de fatores IC:repouso possui valores médios maiores do

que os grupos IC:tlim, DPOC-IC:tlim e DPOC-IC:repouso. Por fim, para as variáveis HHb_respiratorio e HHb_periferico, em ambos os instantes, o grupo com DPOC-IC apresenta valores médios maiores do que o grupo com IC.

Também seria interessante verificar o impacto do tratamento dentro de cada um dos grupos nos dois instantes do estudo, no entanto, não tivemos tempo hábil para realizar a análise.

Referências Bibliográficas

O'Brien, R. G. e Kaiser, M. K. (1985). Manova method for analyzing repeated measures designs: an extensive primer. *Psychological bulletin*, **97**(2), 316.

Apêndice A

Figuras

A seguir representamos os box-plots das variáveis “O2Hb_repiratorio”, “O2Hb_perferico”, “HHb_cerebral’’, “HHb_repiratorio” e “HHb_perferico”, por tipo de doença.

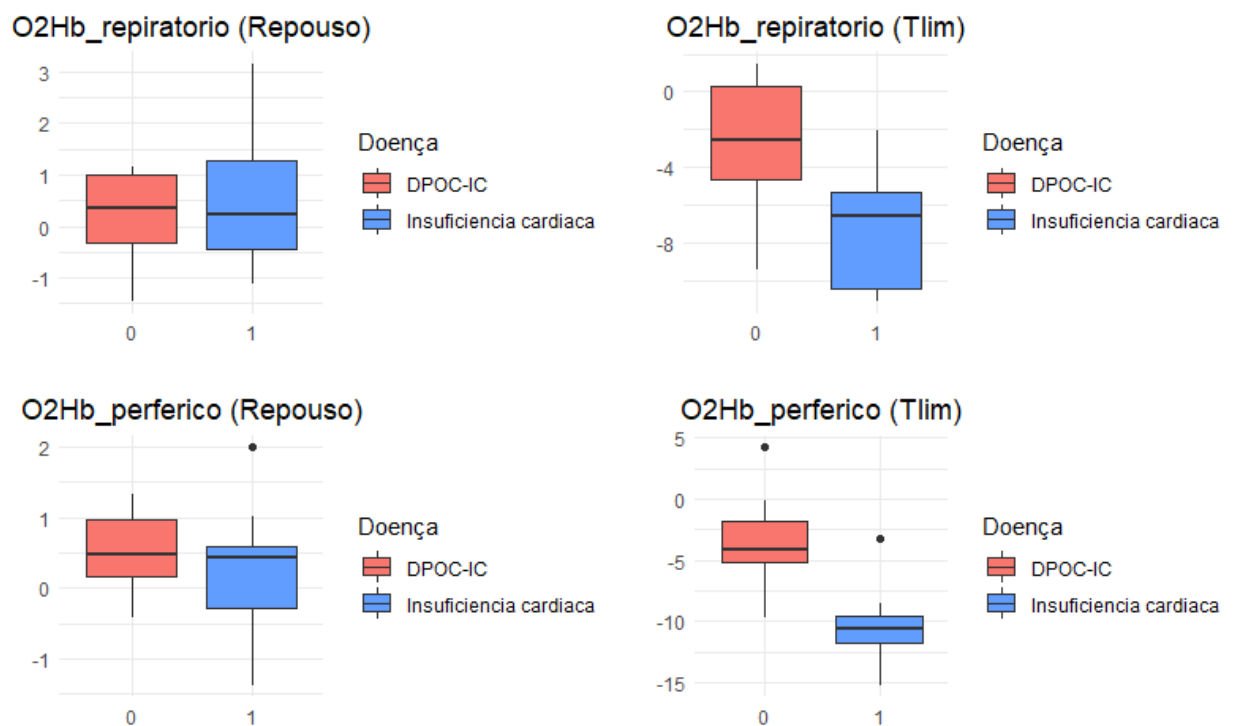


Figura A.1: Box-plots da variáveis “O2Hb_repiratorio” e “O2Hb_perferico” por tipo de doença, considerando os diferentes instantes.

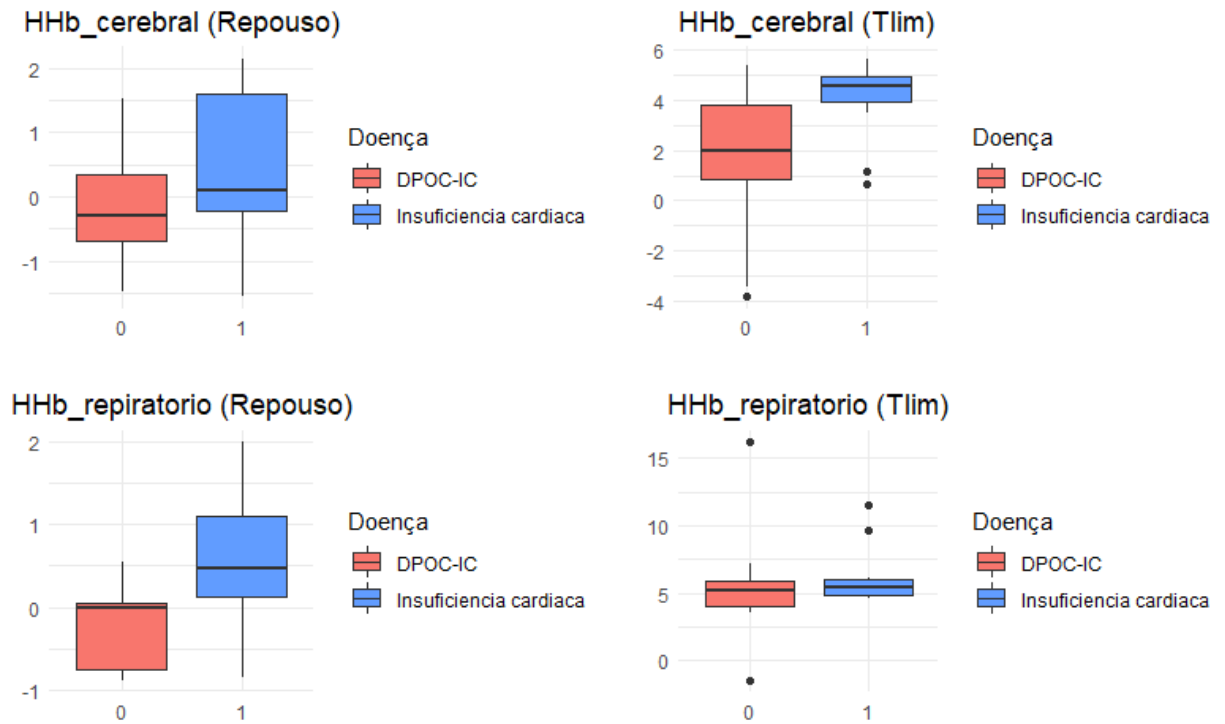


Figura A.2: Box-plots da variáveis “HHb_cerebral” e “HHb_repiratorio” por tipo de doença, considerando os diferentes instantes.

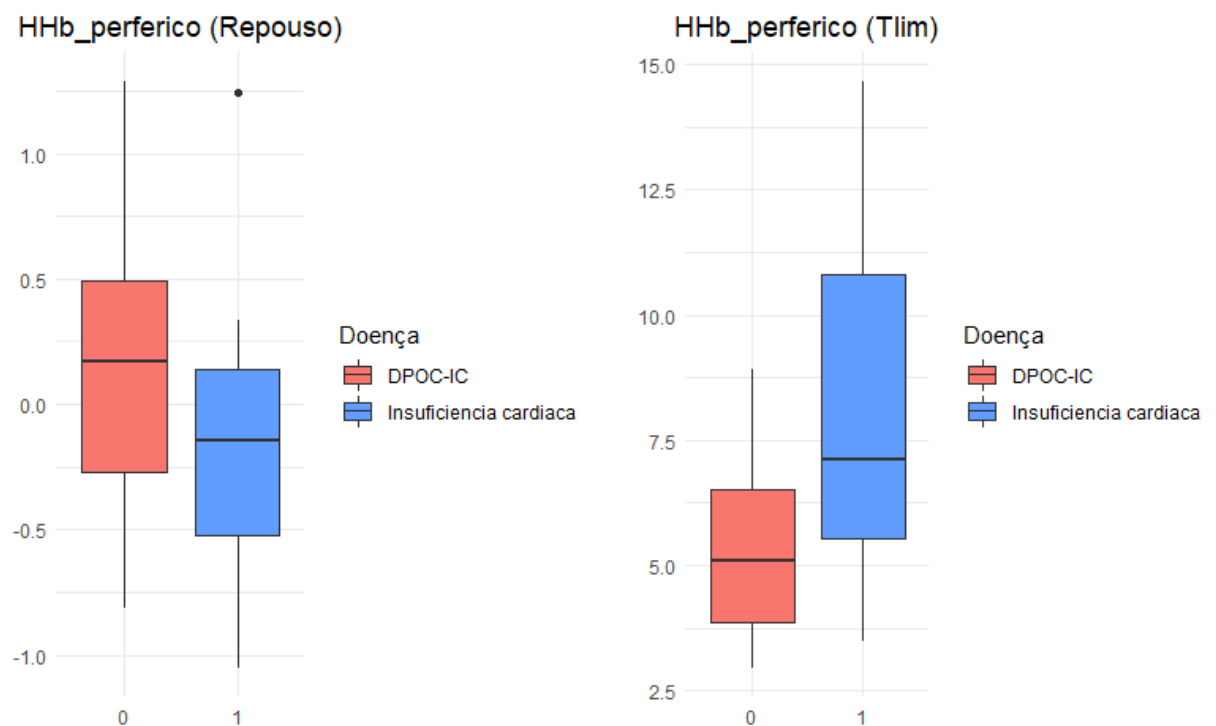


Figura A.3: Box-plots da variável “HHb_perferico” por tipo de doença, considerando os diferentes instantes.