Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Aplicação de Modelos de Regressão Logística Ordinal na Aferição do Tratamento com Plasma Convalescente para a COVID-19

por

Pedro Henrique Monteiro Moreira

Brasília, 12 de junho de 2024

Aplicação de Modelos de Regressão Logística Ordinal na Aferição do Tratamento com Plasma Convalescente para a COVID-19

por

Pedro Henrique Monteiro Moreira

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientadora: Profa. Dra. Joanlise M. de Leon Andrade

Brasília, 12 de junho de 2024

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Estatística.

Texto aprovado por:

Profa.Dra. Joanlise M. de Leon Andrade

Orientadora, EST/UnB

Prof.Dr. André Luiz Fernandes Cançado

EST/UnB

Profa.Dra. Carla Almeida Vivacqua

EST/UFRN

You can't go back and make a new start, but you can start right now and make a brand new ending.

(James R Sherman)

Para minha família, minha fortaleza.

Agradecimentos

Meus sinceros agradecimentos aos professores do PPGEST/UnB, em especial, Profa. Joanlise M. de Leon Andrade; ao prezado professor André Nicola (FM - UnB); à NYU Grossman School of Medicine, pela disponibilização e autorização para o uso do banco de dados, aos meus colegas de trabalho, minha família e amigos que sempre me incentivaram e me motivaram a não desistir.

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

Resumo

Objetivos: Avaliar a resposta ao tratamento da COVID-19 por meio do plasma convalescente monitorado utilizando modelos de regressão logística ordinal. Identificar se o desempenho do tratamento com plasma convalescente, no combate à COVID-19, depende mais de características do doador ou do receptor e identificar quais variáveis são as mais importantes. **Métodos:** Para avaliar o desempenho do tratamento com plasma convalescente foram utilizados modelos de regressão logística ordinal, incluindo modelos de chances proporcionais (MCP), modelo de chances proporcinais parciais (MCPP), modelo de razão contínua (MRC) e modelo *estereótipo* (ME) para verificar qual deles melhor descreve o banco de dados estudado. **Resultados:** O conjunto de dados envolveu 2.369 pacientes e foi dividido em 6 grupos menores de acordo com resultados preliminares de outros estudos. Em todos os grupos analisados houve significância, mas o ajuste não foi satisfatório. **Conclusão:** Não foi possível encontrar evidência estatística que comprovasse a eficácia do tratamento utilizando os modelos MCP, MCPP, MRC e ME. Todos os modelos classificaram, no mínimo, 98,7% dos casos na categoria de menor severidade, por estar em maior proporção na base, evidenciando um desequilíbrio considerável na distribuição das categorias. Considerando o conjunto de dados, as variáveis características do doador não se mostraram tão relevantes para os modelos quanto as variáveis do receptor do plasma. Neste cenário, mesmo sem um resultado preditivo satisfatório, algumas variáveis como "Grau de severidade ao ser hospitalizado" e "Score OMS ao ser hospitalizado" foram incluídas em praticamente todos os modelos. Com isso, futuras investigações podem considerar abordagens alternativas, explorando melhor tais variáveis ou a inclusão de variáveis adicionais para com-

preender melhor os fatores que influenciam os desfechos dos pacientes submetidos a este tipo de tratamento.

Abstract

Objectives: To evaluate the response to COVID-19 treatment through monitored convalescent plasma using ordinal logistic regression models. To determine whether the effectiveness of convalescent plasma treatment in combating COVID-19 depends more on donor or recipient characteristics, and to identify the most important variables. **Methods:** To assess the performance of convalescent plasma treatment, ordinal logistic regression models were used, including proportional odds models (POM), partial proportional odds models (PPOM), continuation ratio models (CRM), and stereotype models (SM) to determine which best describes the studied dataset. **Results:** The dataset included 2,369 patients and was divided into 6 smaller groups according to preliminary results from other studies. In all the groups analyzed, there was significance, but the fit was not satisfactory. **Conclusion:** Statistical evidence to prove the effectiveness of the treatment using POM, PPOM, CRM, and SM was not found. All models classified at least 98.7% of cases in the lowest severity category, due to the higher proportion of this category in the database, highlighting a considerable imbalance in category distribution. Considering the dataset, donor characteristics were not as relevant to the models as recipient variables. In this scenario, even without satisfactory predictive results, some variables such as "Severity level upon hospitalization"and "WHO score upon hospitalization"were included in almost all models. Therefore, future investigations may consider alternative approaches to better explore these variables or include additional variables to better understand the factors influencing patient outcomes subjected to this type of treatment.

Abreviações e Siglas

COMPILE	Continuous Monitoring of Pooled International Trials of Convalescent Plasma for COVID-19 Hospitalized Patients
RCT	Randomized Controlled Trial
SAP	Statistical Analysis Plan
DSMB	Data and Safety Monitoring Board
DSA	Data Sharing Agreement
COVID-19	Coronavirus Disease 2019
SARS	Severe Acute Respiratory Syndrome
NYU	New York University
ECMO	Oxigenação por Membrana Extracorpórea

Sumário

Capítulo 1

Introdução

A COVID-19 *(coronavirus disease 2019)*, doença que mudou a história recente do mundo, originou-se na cidade de Wuhan, na China. Até o presente momento, maio de 2024, são quase 700 milhões de casos confirmados e 7 milhões de mortos[1](Dong et al., 2022).

No dia 11 de março de 2020 a OMS (Organização Mundial da Saúde) decretou o início oficial da pandemia[2], desde então países de todos os cantos do mundo tentaram encontrar maneiras de combater o vírus. Como se tratava de uma doença nova e que se espalhava rapidamente, foi uma corrida contra o tempo. Foram testados remédios, tratamentos alternativos e vacinas, grande parte sem sucesso até encontrarem melhores formas de se combater a doença, apesar de ainda estarem longe do ideal.

Uma das formas de tratamento avaliadas foi o tratamento com plasma convalescente, que utiliza o plasma sanguíneo de pacientes que se recuperaram de uma infecção, como a COVID-19, e desenvolveram anticorpos contra o vírus. Esse plasma é coletado através de doação de sangue, processado para separar os componentes sanguíneos e o plasma rico em anticorpos é então transfundido em pacientes que estão atualmente lutando contra a mesma infecção.

[1]Dados retirados do COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)

[2]Fonte:https://g1.globo.com/bemestar/coronavirus/noticia/2020/03/11/oms-declara-pandemia-de-coronavirus.ghtml

O tratamento se mostrava promissor, uma vez que em outras duas epidemias com milhares de mortos, a SARS- 1 *(Severe Acute Respiratory Syndrome)* em 2003 e a síndrome respiratória do Oriente Médio (MERS) em 2012, ele foi utilizado com sucesso (Casadevall e Pirofski, 2020). A epidemia de SARS-1 foi controlada, mas a de MERS acabou migrando e desencadeou um grande surto secundário na Coreia do Sul. Em ambas as epidemias, a alta mortalidade e a ausência de terapias eficazes levaram ao uso do tratamento em questão, que se mostrou eficiente na diminuição da carga viral, auxiliando no combate nas formas mais graves da doença (Petkova, Antman e Troxel, 2020).

A *New York University* (NYU) - *Grossman School of Medicine*, por meio do COMPILE *(The Continuous Monitoring of Pooled International Trials of Convalescent Plasma for COVID-19 Hospitalized Patients)* compilou os dados de estudos realizados com 2.369 pacientes hospitalizados em 6 países: Bélgica, Brasil, Espanha, EUA, Índia e Holanda. Os dados foram coletados de abril de 2020 a março de 2021. Os pacientes foram divididos em dois grupos: um grupo que recebeu o tratamento com o plasma convalescente e o outro que foi o grupo controle.

Por meio da técnica de regressão logistíca ordinal espera-se mensurar o efeito do tratamento e quais as variáveis que mais contribuíram para o seu resultado. Tal técnica se mostra promissora, pois a resposta do tratamento também é apresentada em escala ordinal. Estudos a respeito do peso das características dos doadores versus as características dos receptores no sucesso do tratamento ainda não foram publicados, algo que se espera medir com esse trabalho.

Capítulo 2

Objetivos

Geral

O objetivo principal deste trabalho é avaliar a resposta ao tratamento da COVID-19 por meio do plasma convalescente monitorado utilizando modelos de regressão logística ordinal com base no banco de dados COMPILE - *The Continuous Monitoring of Pooled International Trials of Convalescent Plasma for COVID-19 Hospitalized Patients*.

Específicos

• Descrever o banco de dados utilizado;

• Avaliar o desempenho do tratamento de plasma convalescente na COVID-19;

• Identificar se o desempenho do tratamento depende mais de características do doador ou do receptor com base nos modelos de regressão logística ordinal;

• Identificar quais variáveis são as mais importantes no momento de escolha desse tratamento com base nos modelos de regressão logística ordinal.

Capítulo 3

Metodologia

Neste capítulo, é apresentada uma análise detalhada do banco de dados utilizado, começando pela motivação por trás de sua criação e seguindo pelo processo de consolidação e compartilhamento dos dados. Além disso, é descrita a técnica de regressão logística ordinal aplicada neste estudo, com uma descrição aprofundada dos modelos: MCP, MCPP, MRC e ME, executados na tentativa de responder as questões levantadas no Capítulo 2.

3.1 Banco de Dados

3.1.1 Motivação para criação do Banco de Dados

O ritmo rápido da propagação do coronavírus 2019 (COVID-19) fez com que muitos esforços de investigação fossem iniciados rapidamente. Porém, variações nos picos de casos por tempo e local dificultaram a execução dos Ensaios Clínicos Randomizados, do inglês, RCTs *(randomized controlled trials)* de qualidade logo no início da pandemia (Petkova, Antman e Troxel, 2020).

Para que tratamentos eficazes fossem avaliados o mais rápido possível, dados de diferentes estudos foram integrados com critérios muito rigorosos. O modelo de integração de dados po-

deria ser aplicado para diversos tipos de tratamento e instituições. Entretanto, o foco do estudo foi a estimação dos efeitos com o tratamento de plasma convalescente nos pacientes hospitalizados por conta da COVID-19. Foram considerados vários estudos simultâneos, formando uma base única. Isoladamente, muitos desses estudos não teriam poder para detectar associações por falta de participantes.

Foi necessária a criação de regras específicas para o compartilhamento e publicação dos resultados, o financiamento, propriedade dos dados e outras partes sensíveis de propriedade intelectual. Tudo isso teve que ser definido em um tempo muito curto para o estudo ser iniciado o quanto antes. Para tanto foi feito um acordo de comparpartilhamento de dados entre os participantes, levando em conta seu monitoramento e segurança.

Os desafios bioestatísticos foram grandes, pois deveriam considerar: variações em protocolos de tratamento, condições de controle, diferentes documentos de consentimento a depender da instituição, populações alvo e até mesmo os resultados propriamente ditos. Mesmo com esse cenário, a combinação de dados de diferentes estudos possibilitaria inferências mais críveis do que estimativas em estudos individuais (Tierney et al., 2020).

3.1.2 Integração dos dados

A Figura 1 ilustra resumidamente como os estudos foram integrados. Os acordos de partilha de dados foram executados entre os investigadores participantes por meio de um documento de governança estabelecido previamente. Uma quantidade mínima de variáveis dos pacientes seria enviada para um repositório único e seguro, no qual os estudos, os pacientes e os centros de estudos teriam identificações únicas. A cada duas semanas os dados acumulados seriam enviados a esse repositório. Em paralelo, o SAP *(statistical analysis plan)* era finalizado de forma colaborativa entre todas as equipes participantes do estudo. As premissas do SAP estabelecendo regras de segurança e eficácia para os dados, seriam encabeçadas por bioestatísticos independentes e não cegos.

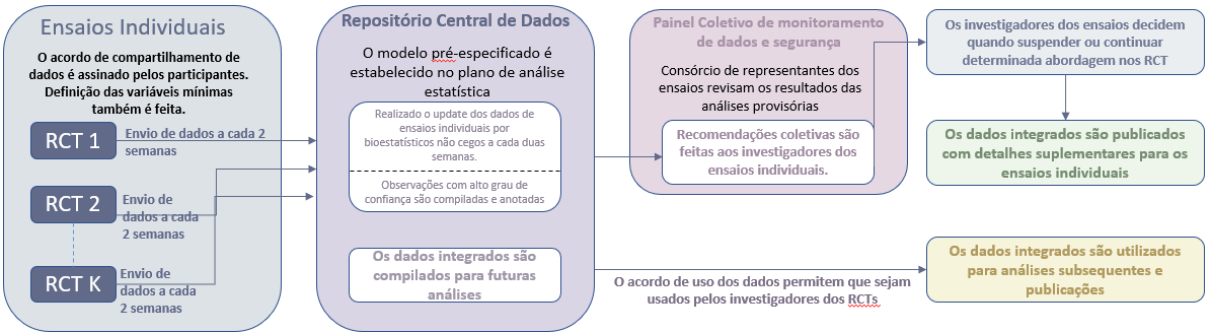

Figura 1 - Planejamento do *Pool* de Dados para Estudos de Plasma Convalescente em Pacientes Hospitalizados com COVID-19[1]

Foi estabelecido o repositório central seguro para os dados integrados, com atualização contínua de novos dados em intervalos de 2 semanas. Bioestatísticos realizaram as análises preliminares e apresentaram relatórios a um DSMB *(Data and Safety Monitoring Board)* coletivo. Caso surgissem evidências estatísticas com um elevado grau de confiança, o DSMB faria uma recomendação conjunta à liderança de todos os ensaios contra ou a favor de determinado protocolo.

O comitê independente de monitoramento de dados e segurança (DSMB), representava todos os estudos participantes. Eles se reuniam quinzenalmente para rever os resultados e fazer recomendações coletivas. Caso provas com um elevado grau de confiança sobre a eficácia, ou falta dela, e segurança do tratamento tivessem sido suficientemente acumuladas, era recomendado o encerramento ou ampliação de determinado estudo. Em uma pandemia esse é um fator de muita importância, pois assim, tratamentos eficazes poderiam ser rapidamente identificados e tratamentos ineficazes ou prejudiciais abandonados com a velocidade adequada.

O fornecimento de detalhes acerca dos dados das equipes dos RCTs era feito por meio de relatórios. Os resultados dos mesmos poderiam ser incorporados como anexos no artigo principal que relatava os resultados globais dos experimentos. Os pesquisadores responsáveis por cada RCT avaliaram a necessidade de suspenção, ou não, do seu estudo, de acordo com as

[1] Adaptado de Petkova, Antman e Troxel, 2020

recomendações do DSMB.

Os investigadores do estudo analisavam a recomendação do DSMB e determinavam a suspenção da inscrição no seu ensaio específico. O repositório de dados seria colocado à disposição dos investigadores participantes para análises adicionais que foram aprovadas por um comitê de publicações.

3.1.3 Estratégia estatística para o compartilhamento de dados

Seria necessária uma abordagem bioestatística para analisar e monitorar os dados individuais dos RCTs que eram imputados de maneira acumulada no banco principal. O monitoramento foi contínuo, com o uso, inclusive, de regras de parada Bayesianas (Lewis e Angus, 2018). A cada análise interina, a distribuição a *posteriori* do parâmetro que descrevia o efeito do tratamento poderia ser relatada juntamente com os critérios de parada. Direcionando assim, as recomendaçoes do DSMB. O processo envolveu a estimativa da probabilidade a *posteriori* de um resultado (sucesso ou fracasso) expresso em razão de chances ou risco relativo. As regras de parada de determinado estudo se baseavam na probabilidade a *posteriori* da razão de chance exceder um limite pré-estabelecido (Petkova, Antman e Troxel, 2020).

Como diversos fatores influenciaram os resultados, o modelo estatístico deveria, além de conter tais fatores, ser generalizável a múltiplos tratamentos para a COVID-19. O desfecho primário foi a escala da OMS (Figura 2). Nos ensaios de plasma convalescente podem ser aplicados diversos tipos de controle (plasma não convalescente, cuidado padrão, soro fisiológico). Para termos de comparação, foi utilizado o tratamento com plasma convalescente versus qualquer dos controles mencionados acima. Todos os controles foram tabulados apenas como "controle", não havendo diferenciação entre eles no banco de dados.

Estudos apontaram diversos fatores relevantes, tais como estado clínico no início do experimento, idade, sexo, dentre outros fatores que estavam no conjunto de dados original. Um problema encontrado foi a ausência de alguns desses dados. Como vieram de diversas fontes,

algumas variáveis apresentaram falhas e não foram recolhidas em todos os estudos. Para mitigar esse problema foi estabelecido um conjunto de dados mínimos.

3.1.4 Conjunto mínimo de dados

Para fins de padronização foi definido um conjunto de informações mínimas a serem recolhidas em todos os ensaios. Foi necessário informar o tipo de estudo (Cego, Duplo-cego, etc...); tipos de tratamento; características básicas do paciente como: características físicas, histórico médico; uso de medicamentos durante o experimento; eventos adversos devido aos testes; tipo de plasma do doador e a escala ordinal de 11 pontos da OMS a 2 e a 4 semanas após a randomização. Tal escala é utilizada para medir de forma padronizada o grau de severidade da Covid-19 em determinado paciente.

Status do Paciente	Descrição	Score
Não-infectado	Não-infectado; sem carga viral detectada	0
Ambulatorial - doença leve	Assintomático; carga viral detectada	1
	Sintomático; independente	2
	Sintomático; necessita de auxílio	3
Hospitalizado - doença moderada	Hospitalizado, sem tratamento com oxigênio	4
	Hospitalizado, com tratamento de oxigênio	5
Hospitalizado - doença severa	Hospitalizado, com tratamento de oxigênio em grande fluxo	6
	Intubação e ventilação mecânica	7
	Ventilação mecânica e vasopressores	8
	Ventilação mecânica, vasopressores, diálise ou ECMO	9
Óbito	Óbito	10

Figura 2 - Escala de progressão clínica de Covid da OMS[2]

A principal variável resposta é a escala da OMS, porém, como ela conta com 11 categorias, dificultaria a análise e interpretação dos modelos. Com isso, uma nova categorização foi proposta em conjunto com o Dr. André Nicola, com menos categorias no intuito de facilitar

[2] Adaptado de Marshall et al., 2020

análises e interpretações de resultados. Os limites de referência para a mudança de categoria foram respectivamente o tratamento com oxigênio e o uso de vasopressores.

Status do Paciente	Descrição	Score	Novo Score
Não-infectado	Não-infectado; sem carga viral detectada	0	1
Ambulatorial - doença leve	Assintomático; carga viral detectada	1	
	Sintomático; independente	2	
	Sintomático; necessita de auxílio	3	
Hospitalizado - doença moderada	Hospitalizado, sem tratamento com oxigênio	4	
	Hospitalizado, com tratamento de oxigênio	5	2
Hospitalizado - doença severa	Hospitalizado, com tratamento de oxigênio em grande fluxo	6	
	Intubação e ventilação mecânica	7	
	Ventilação mecânica e vasopressores	8	3
	Ventilação mecânica, vasopressores, diálise ou ECMO	9	
Óbito	Óbito	10	

Figura 3 - Nova categorização da escala OMS.

Com base na nova classificação da escala da OMS, foram conduzidas análises tendo essa nova categorização como referência.

3.1.5 Plataforma Ortho

A plataforma Ortho é um sistema de gerenciamento de dados eletrônicos (EDMS) utilizado pela *Food and Drug Administration* (FDA) dos Estados Unidos para ajudar a avaliar a segurança e eficácia dos dispositivos médicos.

A plataforma Ortho permite que a FDA colete e gerencie informações sobre dispositivos médicos, desde a fase de desenvolvimento até a comercialização. Ela é uma ferramenta importante para ajudar a FDA a monitorar a segurança e eficácia dos dispositivos médicos comercializados nos Estados Unidos (FDA, 2020).

A partir de janeiro de 2021, os laboratórios passaram a utilizar o score de 12 na plataforma Ortho como corte, sendo que acima de 12 o plasma possui alto valor (Ortigoza et al., 2022).

3.2 Regressão Logística Clássica

Os modelos de regressão são importantes instrumentos para se descrever a relação entre uma variável resposta e uma ou mais variáveis explicativas. Essa ferramenta é comumente utilizada em diversos campos do conhecimento como medicina, *marketing* , informática, engenharia, farmácia, psicologia, entre outros.

Quando a variável resposta é binária, ou seja, só possui como resposta "sucesso" ou "fracasso, a regressão logística é a abordagem mais utilizada. Supondo que a variável resposta assuma valores 0 ou 1, transforma-se a resposta em uma probabilidade que varia conforme as variáveis explicativas do modelo. Essas últimas podem ser de natureza nominal, ordinal, contínua ou discreta.

Tomando o modelo mais simples para esse caso de uma resposta binária temos a probabilidade de sucesso, $\pi(x)$ dada por

$$\pi(x) = \alpha + \beta x \quad . \tag{3.1}$$

O modelo de probabilidade é linear, pois a probabilidade de sucesso muda linearmente em x. O parâmetro β representa a alteração da probabilidade por unidade de mudança em x, e α é o intercepto do modelo. Apesar de ter fácil interpretação o modelo possui um problema, a variável $\pi(x)$ só assume valores entre 0 e 1, enquanto regressões lineares convencionais podem assumir valores em toda a reta real.

As relações entre $\pi(x)$ e as variáveis explicativas são geralmente, em casos práticos, não-lineares. Uma mudança fixa na variável explicativa pode ter menos impacto quando $\pi(x)$ está mais perto de 0 ou 1 do que quando $\pi(x)$ está distante das extremidades. A transformação mais importante para sair da dependência linear assumida em (3.1) é dada por

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \quad . \tag{3.2}$$

Partindo da Equação (3.2) pode-se obter a transformação logito que é assim definida (Agresti, 2018)

$$logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x \quad . \tag{3.3}$$

Considerando múltiplas variáveis explicativas em (3.3) temos

$$logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k = x'\beta \quad . \tag{3.4}$$

A função $logit$, também chamada de logito, é baseada na razão de chances. O termo $\left(\frac{\pi}{1-\pi}\right)$ é conhecido como chance de sucesso. A cada incremento unitário de x_k a chance de sucesso será multiplicada por $e^{(\beta k)}$, caso as demais variáveis explicativas do modelo permaneçam constantes (Agresti, 2018).

Os parâmetros do modelo são estimados por Máxima Verossimilhança. assumindo que cada uma das n respostas avaliadas é uma variável independente com distribuição $Bernouli$ onde

$$P(Y_i = 1|x) = \pi_i \quad \text{e}$$
$$P(Y_i = 0|x) = 1 - \pi_i \quad .$$

Logo, uma resposta da variável tem a distribuição dada por

$$P_{Y_i}(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} \quad y_i = 0, 1; \quad i = 1, 2, .., n. \tag{3.5}$$

Como as respostas são independentes, a distribuição conjunta é dada por

$$P_{(y_1,...,y_n)} = \prod_{i=1}^{n} = P_{Y_i}(y_i) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \quad . \tag{3.6}$$

Ao substituir π_i na Equação (3.6) por sua forma descrita na Equação (3.2) obtém-se a função de Máxima Verossimilhança dada por

$$P_{(y_1,...,y_n)|\beta} = \prod_{i=1}^{n} \left(\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} \right)^{y_i} \left(\frac{1}{1+e^{x_i'\beta}} \right)^{1-y_i} \quad . \tag{3.7}$$

Os valores dos parâmetros que maximizam a função acima serão as estimativas de máxima verossimilhança e não existe um cálculo fechado para sua resolução, sendo necessário o uso computacional de algoritmos numéricos (Kutner et al., 2004).

3.3 Regressão Logística Multi-Categórica

A regressão logística tem seu foco na predição de variáveis dicotômicas. Ela acaba sendo limitada, uma vez que muitas variáveis respostas podem assumir mais de uma categoria.

A regressão logística multi-categórica representa uma generalização da regressão logística usual, em que a variável resposta assume mais de duas categorias (Kutner et al., 2004).

Partindo da regressão logística clássica pode-se generalizar a regressão logística nominal e ordinal, ambas permitem a análise de variáveis não-dicotômicas. A primeira acarreta perda de informação quando as variáveis são ordinais, sendo indicada quando a variável resposta não possui relação de grandeza direta (Kutner et al., 2004). Por exemplo, um estudo de *marketing* em que o cliente pode optar por uma de 3 ou mais marcas distintas seria um caso ideal para o uso dessa técnica. Para um melhor entendimento, será descrita primeiramente a regressão logistíca nominal, seguida da regressão logística ordinal.

3.4 Regressão Logística Nominal (Multinomial)

O modelo de regressão logística nominal, também chamado de multinomial, pressupõe uma variável resposta com J categorias, sendo que $[\pi_1, \pi_2, ..., \pi_J]$ são as probabilidades de todas as

categorias, com a condição de que $\sum_{i=1}^{i=J} \pi_J = 1$.

Tomando n observações independentes, a distribuição de probabilades é a multinominal, que evidencia todas as comparações entre as J categorias (Agresti, 2007).

O valor da variável resposta para cada uma das n observações pode ser escrito em forma binária, onde

$$Y_{ij} = \begin{cases} 1, & \text{se a } i\text{-ésima resposta está na categoria j} \\ 0, & \text{c.c.} \end{cases}$$

Seja, agora, π_{ij} a probabilidade da categoria j ser selecionada para a i-ésima resposta, dada por

$$\pi_{ij} = P(Y_{ij} = 1) \quad .$$

Levando em conta a regressão logística clássica, pode-se considerar a regressão nominal com $J = 2$ de tal maneira que $Y_i = 1$ se a i-ésima resposta for da categoria 1, e $Y_i = 0$ se a i-ésima resposta for da categoria 2. Como visto acima, nesse caso $\pi_i = \pi_{i1}$ e $1 - \pi_i = \pi_{i2}$.

Dessa forma, existe apenas a comparação entre π_{i1} e π_{i2} na função logito, como descrito a seguir:

$$\pi_i' = log\left[\frac{\pi_i}{1-\pi_i}\right] = log\left[\frac{\pi_{i1}}{\pi_{i2}}\right] = \pi_{i12}' = x_i'\beta_{12} \quad .$$

Realizando a análise de 3 categorias, são realizadas 3 comparações. Para 4 categorias seriam 6 comparações. Generalizando para J categorias são $\frac{J(J-1)}{2}$ comparações. Exemplificando o

caso com 4 categorias, as comparações são:

$$\begin{aligned}
\pi'_{i12} &= log\left[\frac{\pi_{i1}}{\pi_{i2}}\right] = x'_i\beta_{12} \quad , \\
\pi'_{i13} &= log\left[\frac{\pi_{i1}}{\pi_{i3}}\right] = x'_i\beta_{13} \quad , \\
\pi'_{i14} &= log\left[\frac{\pi_{i1}}{\pi_{i4}}\right] = x'_i\beta_{14} \quad , \\
\pi'_{i23} &= log\left[\frac{\pi_{i2}}{\pi_{i3}}\right] = x'_i\beta_{23} \quad , \\
\pi'_{i24} &= log\left[\frac{\pi_{i2}}{\pi_{i4}}\right] = x'_i\beta_{24} \quad , \\
\pi'_{i34} &= log\left[\frac{\pi_{i3}}{\pi_{i4}}\right] = x'_i\beta_{34} \quad .
\end{aligned} \tag{3.8}$$

Nesse modelo de regressão todos os possíveis pares de categorias são comparados simultaneamente. Assim se identifica qual categoria é mais provável para um determinado conjunto de variáveis explicativas.

A priori, não é necessário que se compare todos os pares de categorias. Pode-se escolher, arbitrariamente, uma categoria de referência (Agresti, 2007). Assim, o total de comparações diminui de $\frac{J(J-1)}{2}$ para $J-1$. Com isso, a função $logito$ para a categoria J pode ser escrita como

$$\pi'_{ijJ} = log\left[\frac{\pi_{ij}}{\pi_{iJ}}\right] = x'_i\beta_{jJ} \quad J = 1, 2, ..., J-1 \quad . \tag{3.9}$$

Como somente a categoria J está sendo usada como base, a Equação (3.9) pode ser reescrita como

$$\pi'_{ij} = log\left[\frac{\pi_{ij}}{\pi_{iJ}}\right] = x'_i\beta_j \quad j = 1, 2, ..., J-1 \quad . \tag{3.10}$$

Nessa situação, apesar de se fixar apenas uma categoria como base, qualquer outra comparação pode ser obtida por meio de uma combinação dos logitos encontrados. Por exemplo, a comparação entre as categorias 1 e 3 mantendo a categoria 4 como categoria de referência pode ser realizada por:

$$log\left[\frac{\pi_{i1}}{\pi_{i3}}\right] = log\left[\frac{\pi_{i1}}{\pi_{i4}} X \frac{\pi_{i4}}{\pi_{i3}}\right] = log\left[\frac{\pi_{i1}}{\pi_{i4}}\right] - log\left[\frac{\pi_{i3}}{\pi_{i4}}\right] = x_i'\beta_1 - x_i'\beta_3 \quad . \tag{3.11}$$

Generalizando para a comparação de duas categorias quaisquer, o logito é dado por

$$log\left[\frac{\pi_{ia}}{\pi_{ib}}\right] = x_i'(\beta_a - \beta_b) \quad . \tag{3.12}$$

Uma vez que se realize a comparação par a par, a interpretação segue a mesma regra da regressão logística clássica, baseada na razão de chances.

Considerando todas as $J-1$ comparações que são realizadas na Equação (3.10), pode-se obter uma expressão para o cálculo das probabilidade em cada uma das categorias, expressão essa dada por

$$\pi_{ij} = \frac{e^{x_i'\beta_j}}{1 + \sum_{k=1}^{J-1} e^{x_i'\beta_k}} \quad . \tag{3.13}$$

Assim como em (3.7), a estimação dos parâmetros é realizada pelo método da Máxima Versossimilhança. Supondo as 4 categorias descritas anteriormente e uma resposta de interesse na categoria de número 2 tem-se que

$$Y_{i1} = 0 \quad , \quad Y_{i2} = 1,$$
$$Y_{i3} = 0 \quad \text{e} \quad Y_{i4} = 0.$$

A probabilidade do evento descrito é dada por:

$$P(Y_i = 2) = \pi_{i2} = [\pi_{i1}]^0 \times [\pi_{i2}]^1 \times [\pi_{i3}]^0 \times [\pi_{i4}]^0 = \prod_{j=1}^{4} [\pi_{ij}]^{y_{ij}} \quad . \tag{3.14}$$

Com n respostas independentes, a função de distribuição de probabilidade conjunta é dada por

$$P(y_1, y_2, ..., y_n) = \prod_{i=1}^{n} P_{Yi}(y_i) = \prod_{i=1}^{n} \left[\prod_{j=1}^{J} [\pi_{ij}]^{y_{ij}} \right] \quad . \tag{3.15}$$

Substituindo $\pi_{iJ} = 1 - \sum_{j=1}^{J-1} \pi_{ij}$ e $y_{iJ} = 1 - \sum_{j=1}^{J-1} y_{ij}$, a Equação (3.15) pode ser reescrita da seguinte maneira

$$P(y_1, y_2, ..., y_n) = \prod_{i=1}^{n} P_{Yi}(y_i) = \prod_{i=1}^{n} \left[\left(\prod_{j=1}^{J-1} [\pi_{ij}]^{y_{ij}} \right) \left(1 - \sum_{j=1}^{J-1} \pi_{ij} \right)^{1-\sum_{j=1}^{J-1} y_{ij}} \right] \quad . \tag{3.16}$$

Por fim, ao se substituir π_{ij} pela Equação (3.13) têm-se a função de verossimilhança. Tomando uma categoria J como categoria de referência segue:

$$P(y_1, y_2, ..., y_n | \beta_1, \beta_2, ..., \beta_{J-1}) = \prod_{i=1}^{n} \left[\prod_{j=1}^{J-1} \left(\frac{e^{x_i' \beta_j}}{1 + \sum_{k=1}^{J-1} e^{x_i' \beta_k}} \right)^{y_i j} \right] \left[\left(\frac{1}{1 + \sum_{k=1}^{J-1} e^{x_i' \beta_k}} \right)^{1-\sum_{j=1}^{J-1} y_{ij}} \right] \quad . \tag{3.17}$$

Como visto na Equação (3.7), não existem formas analíticas fechadas para a resolução da equação da verossimilhança maximizando todos os parâmetros $\beta_J s$. Faz-se necessário o auxílio de técnicas numéricas como a de Newton-Raphson (Kutner et al., 2004).

3.5 Regressão Logística Ordinal

Quando a variável resposta é ordinal, ela pode ser originada de duas maneiras distintas. Pode ser fruto de uma variável numérica, contínua ou discreta, que foi categorizada. Por exemplo, a variável peso em quilogramas pode ser transformada na variável ordinal [0kg,20kg), [20kg,40kg), [40kg,60kg), [60kg,80kg), [80kg ou mais).

A segunda maneira é a avaliação de uma informação não quantíficavel, porém associada a níveis. Por exemplo, nível de interesse dos alunos a uma determinada matéria, nível esse que poderia ser medido em "baixo", "médio" ou "elevado". Tais níveis podem ser divididos em n novos subníveis, aproximando-se de uma variável contínua.

Nas duas maneiras descritas acima a variável resposta pode ser considerada como a categorização de uma variável contínua (Anderson, 1984).

Por exemplo, considerando as eleições de 2022, pode-se realizar uma pesquisa sobre a ideologia política dos eleitores. Para tanto a variável pode ser dividida em cinco categorias: esquerda, centro-esquerda, centro, centro-direita e direita. De maneira prática, pessoas dentro da mesma categoria podem ter ideologias ainda diferentes. A variável resposta é a representação de uma variável contínua implícita, também chamada de variável latente.

É descrito a seguir o tipo de modelo mais comum para o desenvolvimento da regressão lógística multi-categórica ordinal, ou apenas, regressão logística ordinal.

Variáveis respostas ordinais podem ser analisadas com a técnica de regressão logística multi-categórica nominal. Porém, levar em conta a escala ordinal das categorias resulta em um modelo mais fácil de se interpretrar e mais parcimonioso (Kutner et al., 2004).

Diferentes modelos são utilizados para se analisar a regressão logística ordinal, tais modelos são descritos a seguir.

3.5.1 Modelo de chances proporcionais (MCP)

O MCP, também conhecido por modelo do logito cumulativo, produz estimativas de simples compreensão. Seu uso é recomendado quando a variável resposta originalmente era uma variável contínua que foi agrupada.

A ordenação das categorias da variável resposta faz com que a modelagem da probabilidade de ocorrência de uma determinada classe possa ser feita considerando as probabilidades acumuladas das categorias (Agresti, 2007).

A forma do modelo MCP é

$$log\left(\frac{\sum_1^j P(Y=j|x)}{\sum_{j+1}^k P(Y=j|x)}\right) = \alpha_j + (\beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k). \tag{3.18}$$

A probabilidade acumulada de uma variável Y é a probabilidade de Y ser menor ou igual a um determinado ponto.

Considerando J classes pertencentes a esta variável Y, a probabilidade de se observar a classe j ou uma classe inferior em um vetor de observações independentes é dada por

$$P(Y \leq j|x) = \pi_1 + ... + \pi_j \quad j = 1, ..., J \quad . \tag{3.19}$$

Com $\pi_1 = P(Y = 1|x), \pi_2 = P(Y = 2|x), \pi_j = P(Y = J|x)$, conclui-se que $P(Y \leq 1|x) \leq P(Y \leq 2|x) \leq P(Y \leq J-1|x)$, uma vez que a probabilidade é acumulada à medida em que se aumentam os valores da categoria J. A última categoria J não é considerada, uma vez que a probabilidade acumulada sempre será igual a 1, não necessitando de modelagem. A variável ordinal pode ser descrita como a representação de alguma outra variável contínua não mensurada (latente). Ou seja, a variável resposta Y resulta da divisão da variável latente Y^* em J classes ordinais e distintas entre si (Okura, 2008).

Toma-se $-\infty = \alpha_0 < \alpha_1 < ... < \alpha_J = \infty$ como os pontos de divisão da variável Y^* vistos

sob uma perspectiva contínua. Com isso, a variável resposta satisfaz a equação

$$\begin{aligned} Y = j \qquad & \text{se} \\ \alpha_{j-1} < Y^* < \alpha_j \quad & j = 1, 2, ..., J \quad . \end{aligned} \tag{3.20}$$

Ou seja, Y se encontra na categoria j quando a variável latente estiver contida no j-ésimo intervalo de pontos.

Relembrando a Equação (3.1), parte-se da suposição que Y^* pode ser descrita de uma forma linear, dada por

$$Y^* = x'\beta + \epsilon \quad , \tag{3.21}$$

em que $\beta = (\beta_1, ..., \beta_k)$ é o vetor de parâmetros e ϵ é uma variável aleatória com distribuição F. Partindo destes pressupostos a variável resposta Y é dada por

$$P(Y \leq j|x) = F(\alpha_j - x'\beta) \quad . \tag{3.22}$$

Para compreender a Equação (3.22), segue que $P(Y \leq j|x) = P(Y = 1|x) + P(Y = 2|x) + ... + P(Y = J|x) = P(\alpha_0 \leq Y^* = \alpha_1|x) + P(\alpha_1 \leq Y^* = \alpha_2|x) + ... + P(\alpha_{j-1} \leq Y^* = \alpha_j|x) = F_{Y^*|x}(\alpha_j) - F_{Y^*|x}(\alpha_0)$. Como $\alpha_0 = -\infty$ têm-se que $F_{Y^*|x}(\alpha_0) = 0$. Dessa forma $F_{Y^*|x}(\alpha_j) = P(x'\beta + \epsilon \leq \alpha_j) = P(\epsilon \leq \alpha_j - x'\beta) = F(\alpha_j - x'\beta)$, onde F é a distribuição da variável aleatória ϵ.

A distribuição F tem como sua inversa a distribuição F^{-1}, chamada de função de ligação, ou função *Link*. Tal nome se deve ao fato da função associar linearmente a parte aleatória do modelo $P(Y \leq k)$ e a parte sistemática $(x'\beta)$. A associação é dada por

$$Link(P(Y \leq j)) = \alpha_j - x'\beta \quad . \tag{3.23}$$

Existem diversos tipos de funções de ligação. O uso de cada uma vai depender do tipo de distribuição de probabilidade que a variável dependente apresenta. A escolha da função é muito importante, uma vez que uma escolha não adequada pode acarretar em perda da capacidade preditiva do modelo. As cinco principais funções de ligações em modelos cumulativos são apresentadas no Quadro 3.1 (Agresti, 2007).

Quadro 3.1 - Principais funções de ligação utilizadas em modelos de regressão logística ordinal

Nome	**Função de Ligação *(Link)*** F^{-1}
Logit	$log\left[\frac{P(Y<j)}{P(Y>j)}\right]$
Complemento Log-log	$log(-log(1-P(Y\leq j))$
Log-log negativo	$-log(-log(P(Y\leq j))$
Cauchit	$Tan(\pi(P(Y\leq j)-0,5))$
Probit	$\phi^{-1}(P(Y\leq j))$, em que ϕ tem distribuição $N(0,1)$

3.5.2 Função de Ligação (*Link) - Logit*

A função de ligação mais utilizada é a função de ligação $Logit$. O modelo proposto é baseado em uma analogia com a regressão logística clássica, de forma que o Logit das probabilidades para modelos cumulativos é dado por

$$Logit[P(Y_i \leq j|x)] =$$
$$log\left[\frac{P(Y_i \leq j|x)}{1-P(Y_i \leq j|x)}\right] = \alpha_j - \beta_1 x_{i1} - ... - \beta_k x_{ik} \quad j = 1, 2, ..., J-1 \quad . \tag{3.24}$$

Logo,

$$P(Y_i \leq j|x) = \frac{e^{(\alpha_j - x'\beta)}}{1+e^{(\alpha_j - x'\beta)}} \quad j = 1, 2, ..., J-1 \quad . \tag{3.25}$$

O modelo ordinal exemplificado em (3.25) permite estimar o logaritmo da probabilidade da

variável resposta assumir determinados valores. Estes são inferiores ou iguais a j, em comparação com a probabilidade de assumir valores de classes superiores a j.

Dado $J = 3$, o modelo para se estimar as probabilidades acumuladas usaria $Logit[P(Yi \leq 1)|x] = log[\pi_1/(\pi_2 + \pi_3)]$ e $Logit[P(Yi \leq 2)|x] = log[(\pi_1 + \pi_2)/\pi_3)]$. Cada um dos logitos cumulativos usa todas as categorias possíveis da variável resposta (Kutner et al., 2004).

No modelo descrito em 3.24 os coeficientes de regressão $(\beta = \beta_1, ..., \beta_k)$ não variam conforme a classe j. O modelo com chances proporcionais pressupõe que o efeito das variáveis independentes é o mesmo para todas as classes (Kutner et al., 2004).

Dessa maneira, a resposta em cada classe j é deslocada apenas em função de α_j. Para $\beta_k > 0$, um aumento em alguma variável independente X_k diminui a probabilidade da variável resposta assumir valores iguais ou menores que a classe j, considerando as demais variáveis explicativas constantes. Em resumo, quando x_k aumenta, Y aumenta, no caso em que $\beta_k < 0$ quando x_k aumenta, Y diminui.

O modelo pode ser interpretado pelo uso das razões de chance para as probabilidade cumulativas. Para dois valores distintos x_1 e x_2 de uma das variáveis explicativas do modelo, a razão de chances compara as probabilidades cumulativas para todas as classes da variável resposta. Mantendo as demais variáveis explicativas contantes têm-se que

$$\frac{P(Y \leq j|X_k = x_2)/P(Y > j|X_k = x_2)}{P(Y \leq j|X_k = x_1)/P(Y > j|X_k = x_1)} \quad . \tag{3.26}$$

O logarítmo da razão de chances em questão será a diferença entre os logitos cumulativos para os dois valores de x_k, ou seja, $-\beta_k(x_2 - x_1)$. Caso $x_2 - x_1 = 1$ a chance da variável dependente assumir valores menores para qualquer classe é multiplicado por $e^{-\beta_k}$ para cada unidade adicionada a X_k.

Todos os parâmetros $(\alpha_1, \alpha_2, ...\alpha_{J-1}$ e $\beta s)$ são estimados pelo método de Máxima Verossimilhança, sendo necessário encontrar a sua função. Assumindo o pressuposto anterior do modelo considerar que as curvas de probabilidade das $J - 1$ classes são iguais, independente

da classe faz com que sejam calculadas de forma cumulativa.

Tomando a Equação (3.15) como partida para J classes e n observações independentes, a função de verossimilhança é dada por (Agresti, 2018)

$$P(y_1, y_2, ..., y_n) = \prod_{i=1}^{n} P_{Yi}(y_i) = \prod_{i=1}^{n} \left[\prod_{j=1}^{J} [\pi_{ij}]^{y_{ij}} \right] = \prod_{i=1}^{n} \left[\prod_{j=1}^{J} [P(Y_i \leq j|x) - P(Y_i \leq j-1|x)]^{y_{ij}} \right] \quad . \quad (3.27)$$

Considerando $P(Y_i \leq J|x) = 1, P(Y_i \leq 0|x) = 0$ e $P(Y_i \leq j|x), j = 1, , J-1$ e inserindo na Equação (3.25), encontra-se a função de verossimilhança para $(\alpha_1, \alpha_2, ...\alpha_{J-1}$ e $\beta)$

$$P(y_1, y_2, ..., y_n|\alpha_1, \alpha_2, ...\alpha_{J-1}, \beta) = \prod_{i=1}^{n} \left[\left(\frac{e^{(\alpha_j - x'\beta)}}{1 + e^{(\alpha_j - x'\beta)}} \right)^{y_{i1}} \left(\prod_{j=2}^{J-1} \left(\frac{e^{(\alpha_j - x'\beta)}}{1 + e^{(\alpha_j - x'\beta)}} - \frac{e^{(\alpha_j - x'\beta)}}{1 + e^{(\alpha_j - x'\beta)}} \right)^{y_{i1}} \right) \left(\frac{1}{1 + e^{(\alpha_j - x'\beta)}} \right) \right] \quad (3.28)$$

As estimativas de máxima verossimilhança maximizam (3.28). Assim como visto em (3.7) e (3.21) não existem formas fechadas de cálculo, sendo necessário o auxílio de métodos numéricos para obtenção dos parâmetros (Kutner et al., 2004).

3.6 Modelo de chances proporcionais parciais (MCPP)

De forma prática, dificilmente os parâmetros $(\beta = \beta_1, ..., \beta_k)$ possuem a mesma característica para todos os níveis de resposta da variável dependente (Peterson e Harrell Jr, 1990).

O modelo permite que algumas covariáveis sejam modeladas com a suposição de chance proporcional, e outras não. Existem dois tipos de MCPP, sem e com restrição. Tais modelos serão descritos a seguir.

3.6.1 Modelo de chances proporcionais parciais não restrito (MCPP-NR)

O modelo considera que para uma variável resposta Y com J categorias, dentre os k preditores ($\beta = \beta_1, ..., \beta_k$) apenas alguns tenham chances proporcionais (Peterson e Harrell Jr, 1990).

Assumindo uma variável explicativa x_i em que não vale a propriedade de chances proporcionais, a equação $\alpha+\beta x_i$ é incrementada pelo coeficiente γ_{ji}. Tal incremento é o efeito associado a cada logito cumulativo. Assim, a equação é reescrita como $\alpha + \beta x_i + \gamma_{ji}$.

Para o modelo são estimados $k-1$ parâmetros, dos quais p são independetes. São estimados $((k-1)-p)(k-1)$ parâmetros gama (γ). Caso todos os parâmetros gama (γ) sejam nulos o modelo retorna ao MCP.

No MCPP-NR, as q primeiras covariáveis possuem seu coeficiente angular dependente de j. Isso implica que a relação entre x e Y é dependente na categoria j. Com isso, são estimadas razões de chances para todas as comparações das categorias da variável resposta. Para as demais variáveis independentes, apenas uma razão de chances é estimada. O modelo é dado por

$$log\left(\frac{\sum_1^j P(Y=j|x)}{\sum_{j+1}^k P(Y=j|x)}\right) = \alpha_j + [(\beta_1+\gamma_{j1})x_1 + ... + (\beta_q+\gamma_{jq})x_q + (\beta_{q+1}x_{q+1}) + ... + (\beta_k x_k)]. \quad (3.29)$$

3.6.2 Modelo de chances proporcionais parciais restrito (MCPP-R)

É esperado algum tipo de tendência entre a variável resposta e a variável explicativa quando a relação não é proporcional. Foi proposto um modelo para quando existe relação linear entre o logito da covariável e a variável resposta (Peterson e Harrell Jr, 1990).

Para uma dada covariável, o coeficiente γ_j não depende dos pontos de corte, porém é multiplicado por um coeficiente τ que é específico para cada logito (Lall et al., 2002).

A restrição pode ser escolhida de várias maneiras. Comumente, elas são determinadas usando um banco de dados de um estudo piloto ou um valor predefinido a *priori*.

3.7 Modelo de razão contínua (MRC)

O modelo MCR compara a probabilidade de uma resposta igual a uma determinada categoria, por exemplo, $Y = j$ *versus* $Y > J$.

Esse modelo possui constantes e coeficientes específicos para cada comparação. O MRC pode ser ajustado através de k modelos de regressão logística clássica. Para cada categoria $(j = 1, ...k)$, o intercepto do modelo é α_j e os coeficientes das variáveis explicativas são os coeficientes β_j.

O modelo de razão contínua é afetado pelo sentido determinado para modelar a variável. Por exemplo, a razão de chances obtida quando se modela o crescimento na gravidade não é equivalente ao recíproco que é obtido quando se modela o decrescimento na gravidade (Abreu, 2007). A forma do modelo é

$$log\left(\frac{P(Y=j|x)}{\sum_{j+1}^{k} P(Y=j|x)}\right) = \alpha_j + (\beta_{j1}x_1 + \beta_{j2}x_2 + ... + \beta_{jk}x_k). \tag{3.30}$$

3.8 Modelo *estereótipo* (ME)

O ME deve ser utilizado quando a variável resposta é intrínseca como no exemplo abordado a respeito da pesquisa eleitoral na Seção 3.5 (Anderson, 1984).

O modelo pode ser considerado uma extensão do modelo de regressão multinomial. O mais flexível para análise de respostas ordinais.

Por conta do caráter ordinal dos dados originais, é imposta uma estrutura linear ao logito desse modelo. São atribuídos pesos aos coeficientes dados por $\beta_{jl} = \omega_j\beta_l$.

Para cada variável explicativa existe um parâmetro β e um peso ω. Os pesos são diretamente relacionados com o efeito das variáveis explicativas. Logo, a razão de chances encontrada terá uma tendência de crescimento, uma vez que os pesos geralmente possuem ordenação $(\omega_1 < \omega_2 < ... < \omega_j)$. Sendo assim, o efeito das covariáveis é menor na primeira razão de chance em

comparação com a segunda e assim sucessivamente. O modelo é escrito da seguinte forma

$$log\left(\frac{P(Y=j|x)}{P(Y=0|x)}\right) = \alpha_j + \omega_j(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k). \tag{3.31}$$

3.9 Revisão da Literatura

Por se tratar de um tema recente, ainda existe pouca literatura sobre os benefícios do tratamento com plasma convalescente para a COVID-19. Alguns artigos estudaram e revisaram seus efeitos no combate à Covid-19. Os principais resultados de alguns deles estão descritos a seguir por terem muitas semelhanças com o estudo conduzido neste trabalho.

O artigo "*Association of Convalescent Plasma Treatment With Clinical Status in Patients Hospitalized With COVID-19: A Meta-analysis*" (Troxel et al., 2022) utilizou para sua avaliação o mesmo banco de dados deste trabalho, o COMPILE. O estudo foi realizado durante o monitoramento contínuo dos dados descritos em 3.1.2 e 3.1.3. O artigo avaliou, por meio da estatística Bayesiana, o desempenho geral do tratamento no 14º e 28º dias, além de estudar o tempo até a morte ou até a alta do paciente. Também foi investigada a ocorrência, ou não, de eventos adversos ocasionados pelo procedimento de transfusão.

De acordo com as técnicas utilizadas e variáveis analisadas, não foram encontradas evidências que apontem especificamente os benefícios do tratamento com plasma convalescente. Porém, o banco de dados disponível é de grande utilidade, pois permite análises posteriores, como as realizadas neste trabalho.

O artigo "*COVID-19 convalescent plasma for the treatment of immunocompromised patients: a systematic review*" (Senefeld et al., 2022) realizou uma revisão sistemática em estudos que utilizaram o tratamento de plasma convalescentes em pacientes imunossuprimidos. A hipótese de um efeito benéfico significativo do tratamento com plasma convalescente na redução da mortalidade em doentes imunocomprometidos não pôde ser definitivamente demonstrada com os dados atuais, mas elementos muito fortes sugerem a sua eficácia. Existe a possibilidade de

que o tratamento seja adequado somente para alguns agrupamentos, levando em consideração, pioritariamente, o tempo entre o surgimento de sintomas e o início do tratamento.

O artigo *"The effect of convalescent plasma therapy on mortality among patients with COVID-19: systematic review and meta-analysis"* (Klassen et al., 2021) apresentou revisão sistemática de estudos cujo desfecho de interesse foi a mortalidade dos pacientes tratados com plasma convalescente. Tal artigo conclui que a taxa de mortalidade de pacientes transfundidos com COVID-19 foi inferior à de pacientes não transfundidos com COVID-19. Sugere, também, que a transfusão precoce de plasma de alto teor representa o cenário de utilização ideal para reduzir o risco de mortalidade entre pacientes com COVID-19. Tal conclusão está em conformidade com a obtida em Senefeld et al., 2022. Os resultados favoreceram a eficácia do plasma convalescente como agente terapêutico da COVID-19.

O artigo *"Association of Convalescent Plasma Treatment With Clinical Outcomes in Patients With COVID-19: A Systematic Review and Meta-analysis"* (Janiaud et al., 2021) realizou a revisão sistemática do desempenho geral do tratamento com plasma convalescente, sem o foco em efeitos específicos. Sem considerar agrupamentos específicos, o tratamento com plasma convalescente em comparação com placebo não foi significativamente associado a benefícios em resultados clínicos. A evidência foi baixa ou moderada para a mortalidade por todas as causas e baixa para outros desfechos.

O artigo *"A randomized controlled study of convalescent plasma for individuals hospitalized with COVID-19 pneumonia"* (Bar et al., 2021) consistiu em um estudo randomizado com a aplicação de plasma convalescente em indíviduos hospitalizados com pneumonia associada à COVID-19. Dois locais distintos aplicaram o tratamento, precocemente, nos pacientes. Tal aplicação mostrou um benefício significativo na pontuação de gravidade clínica e na mortalidade aos 28 dias. Os resultados sugeriram que o tratamento com plasma convalescente pode beneficiar grupos específicos, especialmente com comorbidades e que são tratadas no início da doença.

O último artigo "Fatores associados a maior risco de ocorrência de óbito por COVID-19:

análise de sobrevivência com base em casos confirmados" (Galvão e Roncalli, 2021) foi uma investigação realizada no Rio Grande do Norte, utilizando dados até 24 de agosto de 2020, mapeou os fatores que aumentam o risco de óbito por COVID-19. Através da análise de sobrevivência, o estudo identificou que indivíduos com 80 anos ou mais $(HR = 8,06; p < 0,001)$, do sexo masculino $(HR = 1,45; p < 0,001)$, com cor de pele não branca $(HR = 1,13; p < 0,033)$ ou sem informação $(HR = 1,29; p < 0,001)$, portadores de comorbidades $(HR = 10,44; p < 0,001)$ ou com informação incompleta sobre comorbidades $(HR = 10,87; p < 0,001)$ apresentaram maior probabilidade de falecimento pela doença. Esses resultados ressaltam a necessidade de estratégias direcionadas a esse perfil de risco, visando prevenir o óbito.

Capítulo 4

Resultados

Neste capítulo, são apresentados os resultados das análises realizadas no banco de dados COMPILE. São discutidas tanto as análises descritivas quanto os resultados dos modelos de regressão logística ordinal aplicados aos seis agrupamentos distintos.

4.1 Análise Descritiva

Ao todo, foram analisados dados de 2.369 pacientes de oito diferentes RCTs. Os pacientes foram divididos em dois grupos. O controle (1.138; 48%) e o grupo tratado (1.231; 52%). Antes de seguir para a modelagem, faz-se necessário um melhor entedimento da base de dados. A análise descritiva se baseou em quatro aspectos principais: características demográficas, condições pré-existentes dos pacientes, gravidade do paciente ao ser internado e desfecho do paciente. Todas as análises foram realizadas separando os grupos de tratamento para evidenciar o efeito da randomização da amostra.

De acordo com a revisão da literatura, realizada na seção 3.9, o tratamento com plasma tende a ser mais efetivo em períodos iniciais, além disso, também é importante observar a qualidade do plasma utilizado. Por esse motivo, os modelos foram avaliados considerando 6 diferentes agrupamentos conforme apresentado no Quadro 4.1.

Quadro 4.1 - Subdivisões da amostra utilizadas.

Agrupamento	Descrição	Referência	Tamanho (n)
Agrupamento 1	Geral, sem considerar nenhum tipo de separação	14º dia	2.008
Agrupamento 2	Geral, sem considerar nenhum tipo de separação	28º dia	1.881
Agrupamento 3	Pacientes com até 6 dias desde o início dos sintomas	14º dia	1.009
Agrupamento 4	Pacientes com até 6 dias desde o início dos sintomas	28º dia	935
Agrupamento 5	Pacientes que receberam plasma de alto valor	14º dia	224
Agrupamento 6	Pacientes que receberam plasma de alto valor	28º dia	211

Para a construção dos agrupamentos foram seguidos os seguintes passos:

Passo 1 - Identificação de todos os casos em que se usou o plasma com o filtro da descrição e a referência de dia aplicados.

Passo 2 - Para comparação, foram adicionados pacientes que passaram pelo tratamento de controle, seguindo a proporção de pacientes por RCT assumida no **Passo 1**. Exemplo: Para o agrupamento 6, foram considerados pacientes que receberam plasma de alto valor, porém era necessário um grupo para comparação. Foi extraído um grupo com as mesmas características demográficas e hospitalares do agrupamento 6, exceto pelo tratamento com plasma de alto valor.

4.1.1 Características Demográficas

Os EUA foram o país de origem mais comum entre os pacientes, com (44,5%), o que era esperado já que 37,5% RCTs eram americanos. Em relação à idade, foi observada uma entrada nos hospitais de pacientes mais velhos, como visto na Tabela 4.1. A média de idade foi de 60,4 anos, com um desvio padrão de 15 anos. Era de se esperar uma idade mais avançada nos pacientes, como visto na Seção 3.9.

Nos demais agrupamentos a idade se manteve em proporção semelhante à da base completa, não aparentando diferenças consideráveis. Em relação aos RCTs, a divisão entre o tratamento

Tabela 4.1: Distribuição percentual dos grupos Controle e Tratamento, por Idade, RCT de origem e Agrupamento

Variáveis	Agrup. 1		Agrup. 2		Agrup. 3		Agrup. 4		Agrup. 5		Agrup. 6	
	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.
Idade	**n = 998**	**n = 1.010**	**n = 939**	**n = 942**	**n = 501**	**n = 508**	**n = 464**	**n = 471**	**n = 112**	**n = 112**	**n = 105**	**n = 106**
Até 30	2,3	1,8	2,2	1,9	2,3	1,5	2,4	1,4	1,0	2,7	3,8	2,9
30 a 39	7,3	7,4	7,4	7,4	5,6	7,4	5,8	7,6	2,7	1,8	1,0	1,9
40 a 49	13,7	15,1	14,1	14,0	11,1	13,3	11,3	12,4	12,6	15,2	17,1	13,3
50 a 59	24,5	22,0	24,2	22,3	21,9	17,9	22,1	18,1	29,7	19,6	19,0	21,9
60 a 69	22,8	23,7	22,9	24,4	23,6	23,5	23,8	24,2	24,3	26,8	27,6	25,7
70 a 79	17,2	19,1	17,2	18,8	18,9	22,3	18,4	22,0	13,5	19,6	17,1	20,0
80+	12,3	10,9	12,0	11,2	16,6	14,1	16,2	14,4	16,2	14,3	14,3	14,3
RCT	**n = 998**	**n = 1.010**	**n = 939**	**n = 942**	**n = 501**	**n = 508**	**n = 464**	**n = 471**	**n = 112**	**n = 112**	**n = 105**	**n = 106**
Bélgica	9,3	15,5	5,0	9,4	7,7	14,4	4,8	8,8	23,9	23,7	19,5	19,5
Brasil	1,5	1,8	1,5	2,0	0,4	1,1	0,4	1,1	-	-	-	-
Espanha	17,2	17,6	17,7	19,0	25,9	26,4	26,4	28,1	37,9	37,7	39,5	39,5
EUA[1]	45,0	45,0	47,0	49,0	46,5	43,0	47,5	45,0	8,5	8,5	9,0	9,0
EUA[2]	3,2	3,0	3,0	2,9	3,0	2,8	2,5	2,2	4,2	4,1	3,5	3,6
EUA[3]	1,8	0,8	1,1	0,2	2,1	1,2	1,3	0,4	-	-	-	-
Holanda	0,5	0,6	1,0	0,7	0,6	0,6	1,3	0,7	3,6	4,5	4,8	4,8
Índia	21,5	15,7	23,7	16,8	14,8	10,5	15,8	13,7	21,5	21,9	23,7	23,7

controle e o tratamento com plasma é semelhante, uma vez que os RCTs foram utilizados como variável de estratificação na criação dos agrupamentos assim como evidenciado na Tabela 4.1.

4.1.2 Condições Pré-existentes

Frequências de fatores associados a outras doenças pré-existentes que agravam o quadro da COVID-19, como diabetes, doenças pulmonares ou cardiovasculares são apresentadas na Tabela 4.2.

Nota-se frequências similares entre os grupos controle e o grupo tratado no que diz respeito às condições pré-existentes avaliadas. Tal divisão se fez necessária, uma vez que fatores de risco com alta concentração em determinado grupo poderiam enviesar os resultados. Os percentuais de pacientes que apresentaram diabetes, doenças pulmonares e doenças cardiovasculares foram, respectivamente, 33,6%, 11,8% e 42,5% na base geral.

Como observado na Tabela 4.2, as 3 enfermidades avaliadas possuem uma distribuição similar em todos os agrupamentos, exceto pelos agrupamentos 5 e 6 que possuem uma distribuição um pouco distinta dos demais.

Tabela 4.2: Distribuição percentual dos grupos Controle e Tratamento, por Condições Pré-existentes e Agrupamento

Variáveis	Agrup. 1		Agrup. 2		Agrup. 3		Agrup. 4		Agrup. 5		Agrup. 6	
	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.
Diabetes	**n = 998**	**n = 1.010**	**n = 939**	**n = 942**	**n = 501**	**n = 508**	**n = 464**	**n = 471**	**n = 112**	**n = 112**	**n = 105**	**n = 106**
Sim	33,7	36,1	33,6	36,3	34,5	35,4	34,3	35,7	26,1	33,0	19,0	35,2
Pulmonar	**n = 994**	**n = 995**	**n = 935**	**n = 937**	**n = 497**	**n = 503**	**n = 460**	**n = 466**	**n = 108**	**n = 107**	**n = 101**	**n = 101**
Sim	11,3	11,3	11,4	11,2	13,0	13,5	12,8	13,3	12,6	20,5	21,4	20,0
Cardio	**n = 994**	**n = 997**	**n = 935**	**n = 939**	**n = 497**	**n = 505**	**n = 460**	**n = 468**	**n = 108**	**n = 109**	**n = 101**	**n = 103**
Sim	40,9	42,0	39,8	40,6	48,7	47,1	47,3	44,9	45,9	57,1	44,7	56,2

4.1.3 Sintomas pré-hospitalização

A velocidade do início do tratamento é um fator importante para o seu sucesso, como visto na seção 3.9. Para mensuração da velocidade foi observado o número de dias em que os sintomas apareceram antes do paciente ser internado no hospital.

As proporções foram semelhantes entre os agrupamentos, onde cerca de 70% dos pacientes sentiram os primeiros sintomas de 4 a 10 dias antes da randomização. Para os agrupamentos 3 e 4 só existem duas categorias, pois a variável foi utilizada como filtro para o grupo, como mostra a Tabela 4.3.

Tabela 4.3: Distribuição percentual dos grupos de Controle e Tratamento, por Dias de sintomas até a randomização e Agrupamento

Variáveis	Agrup. 1		Agrup. 2		Agrup. 3		Agrup. 4		Agrup. 5		Agrup. 6	
	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.	Cont.	Trat.
Dias de Sintomas	**n = 992**	**n = 998**	**n = 933**	**n = 930**	**n = 495**	**n = 496**	**n = 458**	**n = 459**	**n = 106**	**n = 100**	**n = 99**	**n = 94**
0 a 3	13,0	12,6	13,5	12,8	26,6	24,7	27,4	25,1	9,2	12,8	6,7	11,7
4 a 6	35,8	38,4	35,7	38,3	73,4	75,3	72,6	74,9	41,3	41,3	34,3	40,8
7 a 10	35,7	34,9	34,8	34,5	-	-	-	-	34,9	32,1	36,2	33,0
11 a 14	10,7	8,5	10,8	8,7	-	-	-	-	7,3	7,3	11,4	6,8
14+	4,9	5,5	5,2	5,8	-	-	-	-	7,3	6,4	11,4	7,8

4.1.4 Desfechos de interesse

As seguintes variáveis resposta foram consideradas prioritárias para se mensurar o desfecho dos pacientes: A nova escala da OMS, no 14º e 28º dias.

Observando a Tabela 4.4, tem-se um leve indicativo de que o tratamento cause algum bene-

fício, pois a nova escala da OMS está muito concentrada na 1ª categoria, com o score de 4 ou menos. A última categoria (score 8 ou mais), é a segunda em proporção de pacientes, exceto nos grupos 5 e 6. Os grupos com plasma de alta qualidade possuem uma maior concentração na categoria intermediária (score de 5 a 7) do que na última categoria.

A Tabela 4.4 também apresenta semelhança na proporção de óbitos e altas para todos os agrupamentos observados, próximo de 20% de óbitos considerando todo o período do estudo.

Tabela 4.4: Distribuição percentual dos grupos de Controle e Tratamento, por Nova Categorização da OMS, Desfecho e Agrupamento

Variáveis	**Agrup. 1**		**Agrup. 2**		**Agrup. 3**		**Agrup. 4**		**Agrup. 5**		**Agrup. 6**	
	Cont.	**Trat.**	**Cont.**	**Trat.**	**Cont.**	**Trat.**	**Cont.**	**Trat.**	**Cont.**	**Trat.**	**Cont.**	**Trat.**
Nova categorização OMS	**n = 998**	**n = 1.010**	**n = 939**	**n = 942**	**n = 501**	**n = 508**	**n = 464**	**n = 471**	**n = 112**	**n = 112**	**n = 105**	**n = 106**
4 ou menos	70,3	69,4	75,4	75,7	67,7	67,4	72,8	74,5	58,6	53,6	75,2	64,8
5 a 7	11,9	14,0	4,5	5,2	9,1	9,5	2,9	4,6	21,6	26,8	13,3	13,3
8 ou mais	17,8	16,6	20,1	19,1	23,2	23,1	24,3	20,9	19,8	19,6	11,4	21,9
Desfecho	**n = 998**	**n = 1.010**	**n = 939**	**n = 942**	**n = 501**	**n = 508**	**n = 464**	**n = 471**	**n = 112**	**n = 112**	**n = 105**	**n = 106**
Alta	81,4	83,0	80,1	81,4	78,5	81,9	77,2	80,6	83,5	81,3	86,7	79,0

4.1.5 Características do Plasma

Um total de 475 amostras de plasma foram avaliadas na plataforma Ortho. Considerando a Tabela 4.5, o estudo contou com 22% de plasma de alto valor $(Score >= 12)$ tendo em vista a média das amostras.

A proporção foi semelhante entre os agrupamentos, exceto nos agrupamentos 5 e 6, em que só foi utilizado plasma de alto valor na análise.

Um dos fatores que dificultou a análise da qualidade do Plasma foi a ausência de medição em todas as amostras coletadas. Tal dificuldade, se deu, entre outros fatores, pelo aumento exponencial do preço dos insumos para realizar a medição durante o período do experimento.

Tabela 4.5: Distribuição percentual do Score Médio do Plasma, por Agrupamento

Variáveis	Agrup. 1	Agrup. 2	Agrup. 3	Agrup. 4	Agrup. 5	Agrup. 6
Score Médio Plasma	**n = 432**	**n = 401**	**n = 248**	**n = 231**	**n = 112**	**n = 106**
0 a 2	12,5	13,2	12,5	13,0	-	-
2 a 4	13,9	14,0	16,9	16,5	-	-
4 a 6	10,6	11,5	11,3	12,1	-	-
6 a 8	16,4	15,7	16,9	16,9	-	-
8 a 10	13,4	12,5	14,1	13,9	-	-
10 a 12	8,6	8,2	7,7	7,4	-	-
12 a 14	13,0	13,7	12,9	13,4	54,7	57,6
14 a 16	4,2	4,2	4,4	4,8	19,5	18,2
16+	7,4	7,0	3,2	2,2	25,9	24,2

Obs: valores destacados em verde indicam as faixas em que os pacientes receberam plasma de alto valor.

4.2 Construção dos Modelos

Esta seção tem por objetivo apresentar os resultados da modelagem dos dados de acordo com as variáveis fornecidas pelo Dr. André Nicola, médico participante do estudo. As variáveis explicativas foram: local de origem dos dados; quantidade de dias entre o início dos sintomas e a randomização; quantidade de dias entre o diagnóstico da COVID-19 e a randomização; presença de doenças pulmonares, cardíacas ou diabetes pré-existentes; sexo; idade; tipo de tratamento recebido; gravidade do paciente ao ser randomizado (escala OMS); necessidade de UTI no momento da randomização e Características do Plasma.

Os modelos foram submetidos a uma análise comparativa em relação à nova categorização da OMS nos seis agrupamentos distintos. O estudo foi iniciado com o emprego exclusivo da variável de tratamento e, posteriormente, foram introduzidas outras variáveis explicativas, com o objetivo de identificar o modelo que melhor executa a predição dos grupos de interesse.

Para todos os modelos analisados foi utilizada uma abordagem de validação cruzada para avaliar a estabilidade dos coeficientes. Para isso, o conjunto de dados foi dividido aleatoriamente em uma proporção de 80% para treinamento e 20% para teste, seguindo uma prática comum na literatura (Hastie et al., 2009). Esse processo foi repetido 10 vezes em cada modelo a fim de propiciar uma adequada estimativa dos parâmetros.

Cada iteração do processo envolveu a seleção aleatória de conjuntos de treinamento e teste, seguida pelo ajuste de um modelo de regressão logística aos dados de treinamento e a obtenção dos coeficientes associados. Esses coeficientes foram então registrados para cada iteração.

Após a conclusão das 10 iterações, foram calculadas a média e o desvio padrão dos coeficientes em todas as repetições. Essas estatísticas fornecem uma síntese da tendência central e da dispersão dos parâmetros do modelo em todas as diferentes divisões dos dados (Refaeilzadeh, Tang e Liu, 2009).

Para determinar quais variáveis fariam parte do modelo final em cada método e agrupamento, utilizou-se o método *backward stepwise*. Todas as variáveis foram inicialmente incluídas nos modelos e, em seguida, retiradas com base em seu nível de significância, até que apenas as variáveis com um p-valor $< 0,05$ permanecessem.

4.2.1 Análise do tratamento - MCP

Para cada um dos 6 agrupamentos descritos no Quadro 4.1, foi construído um modelo levando em conta apenas o fato do paciente ter recebido, ou não, o plasma. Ou seja, utilizando-se como variável explicativa o grupo do paciente (tratamento ou controle).

Tabela 4.6: Razões de chances, IC95% e p-valores de modelos de regressão logística ordinal com logito cumulativo e chances proporcionais utilizando a variável "Tratamento" como preditora para a Nova Categorização da OMS

Agrupamentos	**RC**	**IC (95%)**	**P-valor**
Agrupamento 1	1,01	(0,82 - 1,25)	0,46
Agrupamento 2	0,86	(0,68 - 1,09)	0,25
Agrupamento 3	1,04	(0,84 - 1,29)	0,39
Agrupamento 4	0,97	(0,76 - 1,23)	0,61
Agrupamento 5	1,16	(0,66 - 2,06)	0,49
Agrupamento 6	1,43	(0,73 - 2,84)	0,61

Em todos os agrupamentos a variável "Tratamento" atendeu o pressuposto de chances proporcionais, porém quando analisada a razão de chances e seu intervalo de confiança percebe-se

que todos os intervalos contém o valor 1, o que significa que não existe evidência estatística para afirmar que a variável "tratamento" está associada a um aumento ou diminuição significativa nas chances do desfecho de interesse.

Dentro do modelo inicial foram testadas as funções de ligação: *logistc, probit, loglog* e *cauchit*. A eficácia das funções foi quantificada através da avaliação do critério AIC, conforme ilustrado na Tabela a seguir:

Tabela 4.7: Mensuração do AIC para as diferentes funções de ligação

Agrupamentos	***logistc***	***probit***	***loglog***	***cauchit***
Agrupamento 1	2.574,52	2.574,57	2.574,46	2.594,14
Agrupamento 2	2.057,98	2.085,55	2.085,52	2.104,36
Agrupamento 3	2.575,22	2.575,26	2.575,22	2.595,33
Agrupamento 4	2.078,46	2.044,50	2.044,44	2.063,14
Agrupamento 5	375,06	375,22	374,81	376,78
Agrupamento 6	271,10	270,81	271,38	274,53

Como mostra a Tabela 4.7, as funções de ligação praticamente não diferem entre si dentro de cada agrupamento. A diferença no AIC só é visível entre os grupos, porém tal diferença é explicada pela diferença do tamanho amostral dos mesmos.

4.2.2 Análise das demais variáveis - MCP e MCPP

Multicolinearidade

Uma vez que o tratamento com plasma, observado isoladamente, não evidenciou significância estatatística, fez-se necessária a construção de modelos com mais variáveis preditoras. Para dar início à analise do modelo, é necessário verificar a multicolinearidade entre as variáveis independentes listadas na seção anterior. Para verificar a presença de multicolinearidade foi utilizado o VIF (Fator de Inflação da Variância). Um valor de VIF = 1 indica que as variáveis do modelo não possuem correlação entre si. Se o valor do VIF estiver entre 1 e 5, indica uma

correlação de fraca a moderada. Se o VIF estiver entre 5 e 10, indica a presença de multicolinearidade entre os preditores na regressão do modelo e um VIF > 10 indica que os coeficientes de regressão serão fracamente estimados com a grande presença de multicolinearidade (Shrestha, 2020).

Tabela 4.8: Cálculo do VIF das variáveis preditoras para os diferentes grupos

Variáveis	Agrupamento					
	1	2	3	4	5	6
RCT de Origem	2,83	2,99	2,53	2,62	1,39	1,43
Período de Coleta	1,74	1,79	1,63	1,70	1,30	1,31
Dias de Sintoma	1,28	1,32	1,07	1,08	1,26	1,30
Doença Cardíaca	1,45	1,46	1,48	1,47	1,36	1,36
Diabetes	1,11	1,12	1,15	1,15	1,15	1,15
Doença Pulmonar	1,06	1,05	1,06	1,06	1,09	1,09
Sexo	1,04	1,03	1,04	1,04	1,04	1,04
Tratamento	1,02	1,02	1,03	1,02	1,03	1,03
Grau de Hospitalização	1,26	1,08	1,36	1,05	1,04	1,02
Faixa Etária	1,44	1,46	1,52	1,53	1,45	1,45
Escala OMS Inicial	1,33	1,30	1,30	1,34	1,07	1,07

Como mostra a Tabela 4.8, nenhuma variável preditora em nenhum agrupamento apresentou um valor de VIF superior a 5, sendo assim, a primeira parte de análise do modelo é satisfeita, já que as variáveis independentes não apresentaram multicolinearidade.

Análise dos modelos

Uma vez que somente o tratamento com plasma convalescente não foi estatisticamente significante para determinar a melhora dos pacientes, outras variáveis serão analisadas para se descobrir quais teriam influência na condição final dos pacientes.

Primeiramente, foi realizado o teste de Brant para verificar a hipótese de chances proporcionais (Brant, 1990) para todas as variáveis nos 6 agrupamentos. Os resultados são exibidos na Tabela a seguir:

Tabela 4.9: Suposição de Chances Proporcionais das variáveis independentes para os diferentes agrupamentos

Variáveis	Agrupamento					
	1	2	3	4	5	6
RCT de origem	Não	Não	Não	Não	Não	Não
Período de coleta	Sim	Sim	Sim	Sim	Sim	Sim
Dias de sintomas	Sim	Não	N.A	N.A	Não	Sim
Doença cardíaca	Sim	Sim	Sim	Sim	Sim	Sim
Diabetes	Sim	Sim	Sim	Sim	Sim	Sim
Doença pulmonar	Sim	Sim	Sim	Sim	Sim	Sim
Tratamento	Sim	Sim	Sim	Sim	Sim	Sim
Grau de hospitalização	Sim	Sim	Sim	Sim	Sim	Sim
Faixa etária	Não	Sim	Não	Sim	Não	Sim
Escala OMS inicial	Sim	Sim	Sim	Sim	Sim	Sim

As variáveis preditoras assumiram chances proporcionais, exceto “RCT de Origem”, “Faixa Etária” e “Dias de Sintoma”. Como nem todas as variáveis assumiram chances proporcionais, o MCPP foi aplicado e as razões de chances finais são descritas na Tabela 4.10.

As variáveis “Grau de Hospitalização” e “Score OMS no dia 0” forma incluídas em 4 e 5 modelos respectivamente. Tal fato pode ser um indicativo de que a gravidade do paciente ao chegar no hospital exerça efeito na evolução do quadro.

Outro ponto que talvez mereça um estudo mais aprofundado é o da variável “Tratamento” ter aparecido como variável preditora no Agrupamento 6, de acordo com a Tabela 4.10. Esse agrupamento mensurou a escala OMS no 28º dia considerando pacientes que receberam plasma de alto valor em comparação aos pacientes de controle. Nos demais agrupamentos, em que todos os tipos de plasma foram considerados, o tratamento não foi significativo para o modelo.

Tabela 4.10: Razões de chances estimadas das variáveis significantes para os 6 agrupamentos - MCPP

Agrup.	Covariável	RC	IC (95%)	p-valor
1	Diabetes	1,61	(1,30 - 2,01)	0,04
	Grau de Hospitalização	1,58	(1,31 - 1,90)	0,02
	Score OMS dia 0	2,80	(2,20 - 3,61)	0,01
2	Grau de Hospitalização	1,67	(1,36 - 2,05)	0,03
	Score OMS dia 0	2,81	(1,82 - 3,78)	0,00
3	Grau de Hospitalização	1,58	(1,31 - 1,91)	0,04
	Score OMS dia 0	2,86	(2,24 - 3,70)	0,00
4	Grau de Hospitalização	1,70	(1,39 - 2,09)	0,04
	Score OMS dia 0	2,56	(1,96 - 3,41)	0,05
5	Score OMS dia 0	5,07	(2,03 - 18,82)	0,00
	Faixa Etária			
	30 a 39 vs. <30	0,92	(0,85 - 1,22)	0,13
	40 a 49 vs. <30	0,49	(0,32 - 1,97)	0,27
	50 a 59 vs. <30	1,08	(0,96 - 1,52)	0,11
	60 a 69 vs. <30	0,70	(0,55 - 1,15)	0,14
	70 a 79 vs. <30	1,35	(1,11 - 1,85)	0,02
	80+ vs. <30	0,50	(0,43 - 1,32)	0,22
6	Cardio	1,84	(1,03 - 3,75)	0,02
	Tratamento	1,59	(1,06 - 2,38)	0,05

Método de seleção - *Backward Stepwise*

4.2.3 Análise do tratamento - *estereótipo*

O modelo *estereótipo* também foi ajustado para mensurar a eficácia do tratamento com plasma convalescente. Para construção, a categoria "< 5" foi utilizada como referência e seus resultados podem ser observados na Tabela 4.11:

Os resultados não apresentaram significância. O número 1 estava contido em todos os intervalos de confiança, o que significa a ausência de efeito ou nenhuma diferença para ambas categorias em todos os grupos levando em conta a análise do efeito da variável "Tratamento".

Tabela 4.11: Razões de chances estimadas para os 6 agrupamentos pelo método *estereótipo* utilizando a variável "tratamento" como preditora (ref.: "< 5")

Agrup.	Categoria - Novo Score	RC	IC (95%)	p-valor
1	5 a 7	1,99	(0,91 - 4,36)	0,13
	8 ou mais	0,79	(0,59 - 1,06)	0,18
2	5 a 7	1,82	(0,83 - 4,00)	0,15
	8 ou mais	0,89	(0,66 - 1,19)	0,21
3	5 a 7	2,01	(0,92 - 4,41)	0,14
	8 ou mais	0,79	(0,59 - 1,06)	0,16
4	5 a 7	1,81	(0,91 - 3,60)	0,12
	8 ou mais	0,90	(0,67 - 1,20)	0,15
5	5 a 7	1,59	(0,88 - 2,86)	0,08
	8 ou mais	0,73	(0,49 - 1,08)	0,09
6	5 a 7	2,01	(0,92 - 4,41)	0,08
	8 ou mais	0,83	(0,61 - 1,11)	0,07

4.2.4 Análise das demais variáveis - *estereótipo*

Dado que, de acordo com o modelo *estereótipo*, apenas o tratamento com plasma convalescente não apresentou significância estatística na melhora dos pacientes, foi necessário expandir a investigação para incluir outras variáveis que poderiam potencialmente influenciar o desfecho clínico dos pacientes.

Assim, além das variáveis já examinadas na seção 4.2.2, foram analisadas outras variáveis relevantes. Aquelas que demonstraram significância estatística foram selecionadas para compor os modelos descritos na Tabela 4.12, contribuindo para uma compreensão mais abrangente dos fatores que podem afetar o resultado clínico dos pacientes

Tabela 4.12: Razões de chances estimadas das variáveis significantes para os 6 agrupamentos - *Estereótipo*

Agrup.	Covariável	"<5" vs. "5 a 7"			"<5" vs. "8 ou mais"		
		RC	**IC (95%)**	**p-valor**	**RC**	**IC (95%)**	**p-valor**
1	Cardio	1,17	(1,03-1,29)	0,02	1,01	(1,00-1,19)	0,04
	Sexo	2,00	(1,86-3,61)	0,00	1,99	(1,66-2,20)	0,04
	Grau de Hosp.	1,34	(1,10-1,58)	0,04	1,01	(1,00-1,90)	0,05
	Score OMS dia 0	1,18	(1,13-1,20)	0,02	1,01	(1,00-1,78)	0,05
2	Cardio	1,28	(1,17-1,37)	0,01	2,19	(1,14-4,02)	0,01
	Pulmonar	1,48	(1,33-1,64)	0,03	1,58	(1,10-2,75)	0,02
	Grau de Hosp.	1,35	(1,06-1,56)	0,02	1,93	(1,58-2,05)	0,03
	Score OMS dia 0	1,19	(1,17-1,26)	0,01	2,78	(1,52-3,20)	0,00
3	Cardio	1,19	(1,11-1,31)	0,04	1,96	(1,86-1,99)	0,04
	Grau de Hosp.	1,29	(1,08-1,30)	0,01	1,97	(1,33-2,99)	0,05
	Score OMS dia 0	1,17	(1,04-1,27)	0,00	1,95	(1,77-2,87)	0,03
4	Cardio	1,23	(1,18-1,45)	0,05	2,55	(1,23-3,68)	0,04
	Grau de Hosp.	1,34	(1,19-1,63)	0,04	1,90	(1,15-2,25)	0,01
	Score OMS dia 0	1,21	(1,02-1,32)	0,02	2,65	(1,30-4,95)	0,03
5	Score OMS dia 0	1,20	(1,15-1,51)	0,03	1,79	(1,63-2,94)	0,01
6	Cardio	1,34	(1,02-1,76)	0,04	1,85	(1,44-2,97)	0,05
	Score OMS dia 0	1,08	(1,00-1,16)	0,02	1,49	(1,20-1,87)	0,00

Método de seleção - *Backward Stepwise*

4.2.5 Análise do tratamento - Razão Contínua

Dado que a variável de resposta possui categorias ordenadas derivadas de uma variável latente contínua, ou seja, uma variável que não é diretamente observada ou medida, mas é inferida a partir de outras variáveis observadas, os modelos de chances proporcionais e razão contínua seriam os mais apropriados, a princípio (Agresti, 2010).

Embora os dois modelos estejam de acordo em relação ao critério de homogeneidade das razões de chances entre as categorias comparadas, eles diferem na avaliação das magnitudes dos riscos, uma vez que estão voltados para comparações distintas.

Assim como os demais modelos, o MRC não obteve um bom desempenho em classificar os pacientes do estudo. Os resultados obtidos podem ser vistos na Tabela 4.13.

Tabela 4.13: Razões de chances estimadas das variáveis significantes para os 6 agrupamentos - Razão Contínua

Agrup.	Covariável	"<5" versus "5 a 7"			"<5" versus "8 ou mais"		
		RC	IC (95%)	p-valor	RC	IC (95%)	p-valor
1	Grau de Hosp.	1,30	(1,17-1,69)	0,03	1,15	(1,06-1,79)	0,00
	Score OMS dia 0	1,92	(1,45-2,69)	0,01	1,88	(1,12-3,11)	0,03
2	Cardio	1,67	(1,46-2,48)	0,01	1,47	(1,30-2,67)	0,04
	Grau de Hosp.	1,11	(1,05-1,21)	0,05	1,73	(1,28-3,35)	0,04
	Score OMS dia 0	1,46	(1,20-1,65)	0,03	1,79	(1,41-3,49)	0,02
3	Grau de Hosp.	1,75	(1,65-2,49)	0,05	1,75	(1,55-2,08)	0,04
	Score OMS dia 0	1,35	(1,19-1,67)	0,00	1,62	(1,00-2,45)	0,01
4	Grau de Hosp.	1,62	(1,38-1,86)	0,05	1,08	(1,00-1,29)	0,05
	Score OMS dia 0	1,01	(1,00-1,97)	0,03	1,55	(1,20-2,77)	0,02
5	Score OMS dia 0	1,71	(1,35-2,61)	0,04	1,46	(1,08-2,63)	0,05
6	Cardio	1,74	(1,19-2,86)	0,01	1,37	(1,19-2,01)	0,04
	Score OMS dia 0	1,58	(1,30-2,92)	0,04	1,79	(1,46-2,89)	0,04

Método de seleção - *Backward Stepwise*

4.2.6 Comparação entre os modelos

Apesar de variáveis como "Escala OMS ao ser randomizado" e "Necessidade de UTI ao ser randomizado" estarem incluídas em quase todos os modelos, tais modelos não foram eficazes na predição. A variável "Doenças cardíacas" também foi selecionada para os modelos. Das 12 variáveis analisadas, 5 delas não foram significantes para nenhum dos modelos e outras 3 para somente 1 modelo cada.

Conforme mostra a Tabela 4.14, o modelo de chances proporcionais parciais classificou 100% das amostras como "< 5". Os modelos de razão contínua e *estereótipo* classificaram um pouco menos de 100% nessa categoria, porém foram muito aquém do esperado, se afastando muito da distribuição original das categorias em cada agrupamento.

Tabela 4.14: Distribuição percentual da classificação final da variável resposta, por agrupamento e modelo

Agrup.	Fonte	Novo Score OMS		
		<5	5 a 7	8+
1	**Real**	**69,8**	**12,9**	**17,2**
	Modelo - Chances Proporcionais Parciais	100,0	0,0	0,0
	Modelo estereótipo	100,0	0,0	0,0
	Modelo Razão Contínua	100,0	0,0	0,0
2	**Real**	**75,5**	**4,8**	**19,6**
	Modelo - Chances Proporcionais Parciais	100,0	0,0	0,0
	Modelo estereótipo	100,0	0,0	0,0
	Modelo Razão Contínua	100,0	0,0	0,0
3	**Real**	**67,5**	**9,3**	**23,1**
	Modelo - Chances Proporcionais Parciais	100,0	0,0	0,0
	Modelo estereótipo	98,7	0,0	1,2
	Modelo Razão Contínua	97,6	0,0	2,3
4	**Real**	**73,6**	**3,7**	**22,6**
	Modelo - Chances Proporcionais Parciais	100,0	0,0	0,0
	Modelo estereótipo	98,3	0,0	1,6
	Modelo Razão Contínua	97,2	0,0	2,7
5	**Real**	**66,7**	**9,1**	**24,1**
	Modelo - Chances Proporcionais Parciais	100,0	0,0	0,0
	Modelo estereótipo	97,8	0,0	2,1
	Modelo Razão Contínua	96,5	0,0	3,4
6	**Real**	**66,2**	**8,0**	**25,6**
	Modelo - Chances Proporcionais Parciais	100,0	0,0	0,0
	Modelo estereótipo	97,4	0,0	2,5
	Modelo Razão Contínua	96,7	0,0	3,2

Capítulo 5

Considerações Finais

Conforme detalhado nas seções anteriores, este estudo visava investigar, primariamente, a relação entre as variáveis de desfecho: “Escala da OMS no 14º dia” e “Escala da OMS no 28º dia”, por meio de análises de regressão logística ordinal. Ambas as variáveis foram recategorizadas para as análises feitas neste trabalho, conforme a Figura 3, na Seção 3.1.4.

Foram realizadas análises exploratórias para compreender melhor as características básicas do conjunto de dados antes da construção dos modelos. A maior parte dos dados originou-se dos Estados Unidos, e os pacientes apresentaram uma idade média de 60,4 anos com desvio padrão de 15,25 anos. Os pacientes foram randomizados com o intuito de tornar os dois grupos mais homogêneos, diferenciando-se apenas na aplicação, ou não, do tratamento com plasma convalescente. De acordo com as estatísticas descritivas não estava muito clara a influência do tratamento. Os dados mostraram que pacientes submetidos ao tratamento faleceram proporcionalmente menos do que pacientes do grupo de controle, por outro lado, a mensuração da escala OMS não diferiu muito em relação aos dois grupos comparados.

O estudo abordou a diferentes modelos na aplicação da regressão logística ordinal. Foram analisados em separado, além da base completa, pacientes que chegaram ao hospital com até 6 dias de sintomas e uma amostra com pacientes que receberam plasma de alto valor.

Dentre os modelos analisados, tanto no MCPP, quanto no MCR e *estereótipo*, duas variáveis

foram selecionadas em mais de 80% dos modelos: “Grau de hospitalização” e “Novo Score Dia 0”. As variáveis foram associadas com a gravidade da situação do paciente ao dar entrada no hospital, evidenciando uma importância desse quadro no resultado final do tratamento. A variável de tratamento só foi incluída em um modelo do grupo que mensurou a escala da OMS no 28º dia, considerando pacientes que receberam plasma de alto valor.

No entanto, ao finalizar o estudo, não foi possível encontrar resultados satisfatórios que comprovassem a eficácia do tratamento utilizando os modelos MCP, MCPP, razão contínua e o modelo *estereótipo*. Todos os modelos classificaram, no mínimo, 98,7% dos casos na categoria de menor severidade, por estar em maior proporção na base, evidenciando um desajuste considerável.

Considerando o conjunto de dados, as variáveis características do doador não se mostraram tão relevantes para os modelos quanto variáveis do receptor do plasma. Neste cenário, futuras investigações podem considerar abordagens alternativas ou a inclusão de variáveis adicionais para compreender melhor os fatores que influenciam os desfechos dos pacientes submetidos a este tipo de tratamento.

Bibliografia

Abreu, Mery Natali Silva (2007). "Uso de modelos de regressão logística ordinal em epidemiologia: um exemplo usando a qualidade de vida". Em.

Agresti, A (2007). *An Introduction to Categorical Data Analysis second edition, New Jersey, Jon Wiley and Sons*.

Agresti, Alan (2010). *Analysis of ordinal categorical data*. Vol. 656. John Wiley & Sons.

— (2018). *An introduction to categorical data analysis*. John Wiley & Sons.

Anderson, TW (1984). *An introduction to multivariate statistical analysis John Wiley & Sons New York*.

Bar, Katharine J. et al. (dez. de 2021). "A randomized controlled study of convalescent plasma for individuals hospitalized with COVID-19 pneumonia". Em: *The Journal of Clinical Investigation* 131.24. DOI: 10.1172/JCI155114. URL: https://www.jci.org/articles/view/155114.

Brant, Rollin (1990). "Assessing proportionality in the proportional odds model for ordinal logistic regression". Em: *Biometrics*, pp. 1171–1178.

Casadevall, Arturo e Liise anne Pirofski (abr. de 2020). "The convalescent sera option for containing COVID-19". Em: *The Journal of Clinical Investigation* 130.4, pp. 1545–1548. DOI: 10.1172/JCI138003. URL: https://www.jci.org/articles/view/138003.

Dong, Ensheng et al. (2022). “The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned”. Em: *The Lancet Infectious Diseases*.

FDA (2020). *US Food and Drug Administration COVID-19 Convalescent Plasma*. Accessed September 22, 2021. https://www.fda.govmedia/141480/download.

Galvão, Maria Helena Rodrigues e Angelo Giuseppe Roncalli (2021). “Fatores associados a maior risco de ocorrência de óbito por COVID-19: análise de sobrevivência com base em casos confirmados”. Em: *Revista brasileira de epidemiologia* 23.

Hastie, Trevor et al. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.

Janiaud, Perrine et al. (mar. de 2021). “Association of Convalescent Plasma Treatment With Clinical Outcomes in Patients With COVID-19: A Systematic Review and Meta-analysis”. Em: *JAMA* 325.12, pp. 1185–1195. ISSN: 0098-7484. DOI: 10.1001/jama.2021.2747. eprint: https://jamanetwork.com/journals/jama/articlepdf/2777060/jama_janiaud_2021_oi_210019_1616177281.46916.pdf. URL: https://doi.org/10.1001/jama.2021.2747.

Klassen, Stephen A et al. (2021). “The effect of convalescent plasma therapy on mortality among patients with COVID-19: systematic review and meta-analysis”. Em: *Mayo Clinic Proceedings*. Vol. 96. 5. Elsevier, pp. 1262–1275.

Kutner, M et al. (2004). “Applied linear statistical models. McGraw-Hill”. Em: *Irwin Series*.

Lall, R et al. (2002). “A review of ordinal regression models applied on health-related quality of life assessments”. Em: *Statistical methods in medical research* 11.1, pp. 49–67.

Lewis, Roger J e Derek C Angus (2018). “Time for clinicians to embrace their inner bayesian?: reanalysis of results of a clinical trial of extracorporeal membrane oxygenation”. Em: *Jama* 320.21, pp. 2208–2210.

Marshall, John C et al. (2020). “A minimal common outcome measure set for COVID-19 clinical research”. Em: *The Lancet Infectious Diseases* 20.8, e192–e197.

Okura, Roberta Irie Sumi (2008). “Modelos de regressao para variáveis categóricas ordinais com aplicaçoes ao problema de classificaçao”. Tese de dout. Universidade de Sao Paulo.

Ortigoza, Mila B et al. (2022). “Efficacy and safety of COVID-19 convalescent plasma in hospitalized patients: a randomized clinical trial”. Em: *JAMA internal medicine* 182.2, pp. 115–126.

Peterson, Bercedis e Frank E Harrell Jr (1990). “Partial proportional odds models for ordinal response variables”. Em: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 39.2, pp. 205–217.

Petkova, Eva, Elliott M. Antman e Andrea B. Troxel (ago. de 2020). “Pooling Data From Individual Clinical Trials in the COVID-19 Era”. Em: *JAMA* 324.6, pp. 543–545. ISSN: 0098-7484. DOI: 10.1001/jama.2020.13042. eprint: https://jamanetwork.com/journals/jama/articlepdf/2768851/jama_petkova_2020_vp_200152_1596830790.69304.pdf. URL: https://doi.org/10.1001/jama.2020.13042.

Refaeilzadeh, Payam, Lei Tang e Huan Liu (2009). “Cross-validation”. Em: *Encyclopedia of database systems*, pp. 532–538.

Senefeld, Jonathon W. et al. (2022). “COVID-19 convalescent plasma for the treatment of immunocompromised patients: a systematic review”. Em: *medRxiv*. DOI: 10.1101/2022.08.03.22278359. eprint: https://www.medrxiv.org/content/early/2022/08/16/2022.08.03.22278359.full.pdf. URL: https://www.medrxiv.org/content/early/2022/08/16/2022.08.03.22278359.

Shrestha, Noora (2020). “Detecting multicollinearity in regression analysis”. Em: *American Journal of Applied Mathematics and Statistics* 8.2, pp. 39–42.

Tierney, Jayne F et al. (2020). “Comparison of aggregate and individual participant data approaches to meta-analysis of randomised trials: An observational study”. Em: *PLoS medicine* 17.1, e1003019.

Troxel, Andrea B. et al. (jan. de 2022). “Association of Convalescent Plasma Treatment With Clinical Status in Patients Hospitalized With COVID-19: A Meta-analysis”. Em: *JAMA Network Open* 5.1, e2147331–e2147331. ISSN: 2574-3805. DOI: 10.1001/jamanetworkopen.2021.47331. eprint: https://jamanetwork.com/journals/jamanetworkopen/articlepdf/2788377/troxel_2022_oi_211300_1646239960.74013.pdf. URL: https://doi.org/10.1001/jamanetworkopen.2021.47331.

Apêndice A

Figuras

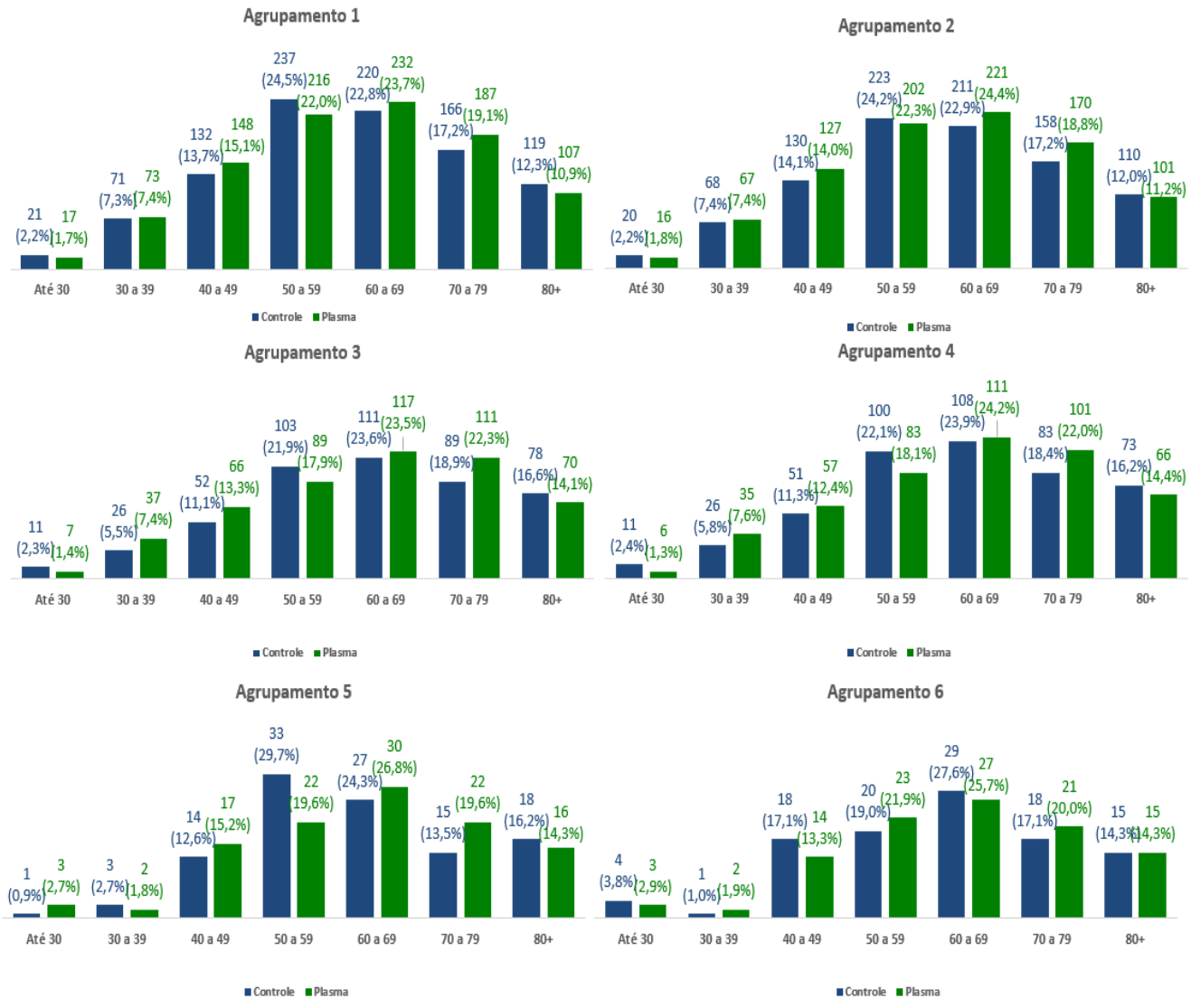

Figura A.1 - Distribuição etária percentual dos pacientes por tipo de tratamento e agrupamento

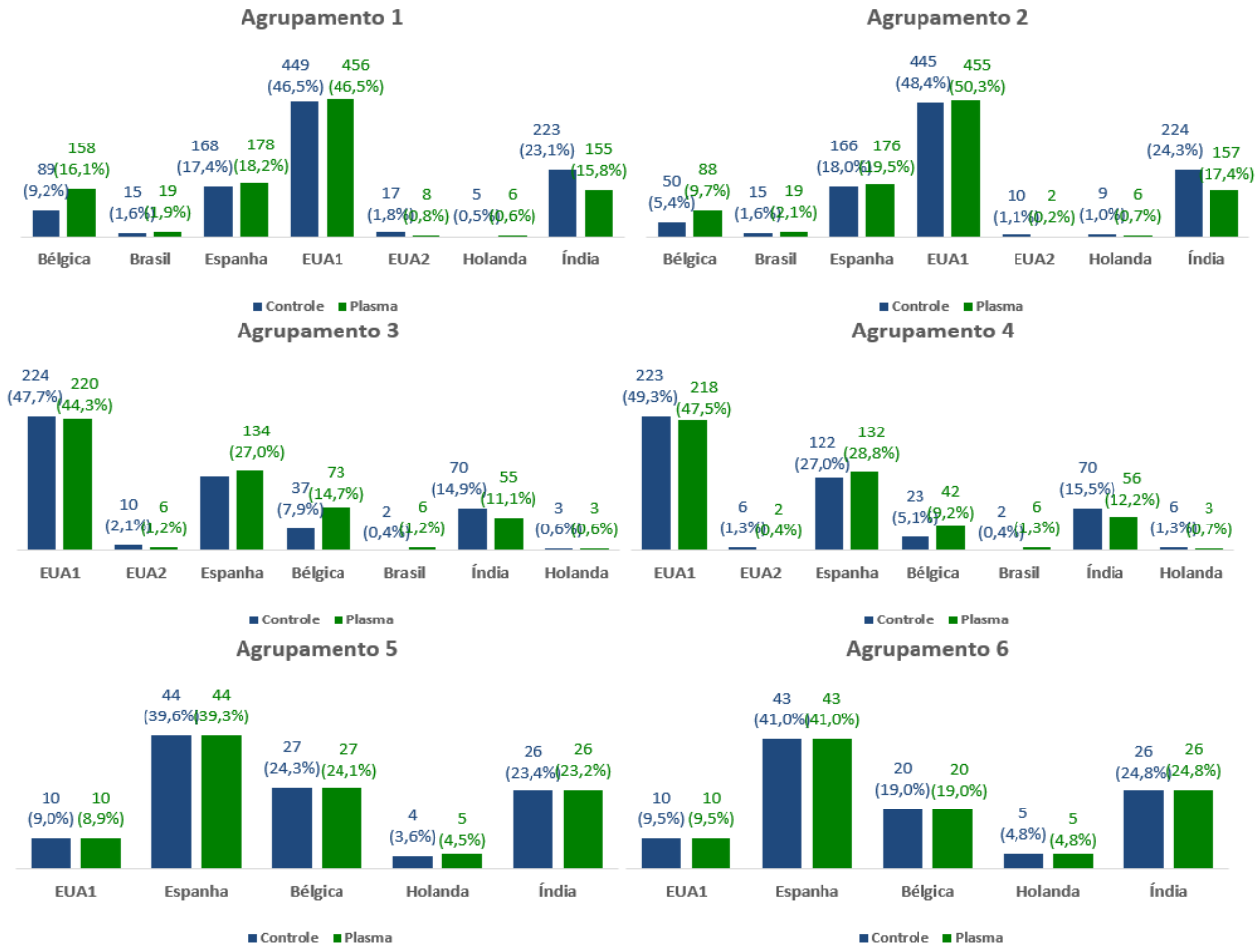

Figura A.2 - Quantidade de pacientes por RCT de origem, tipo de tratamento e agrupamento

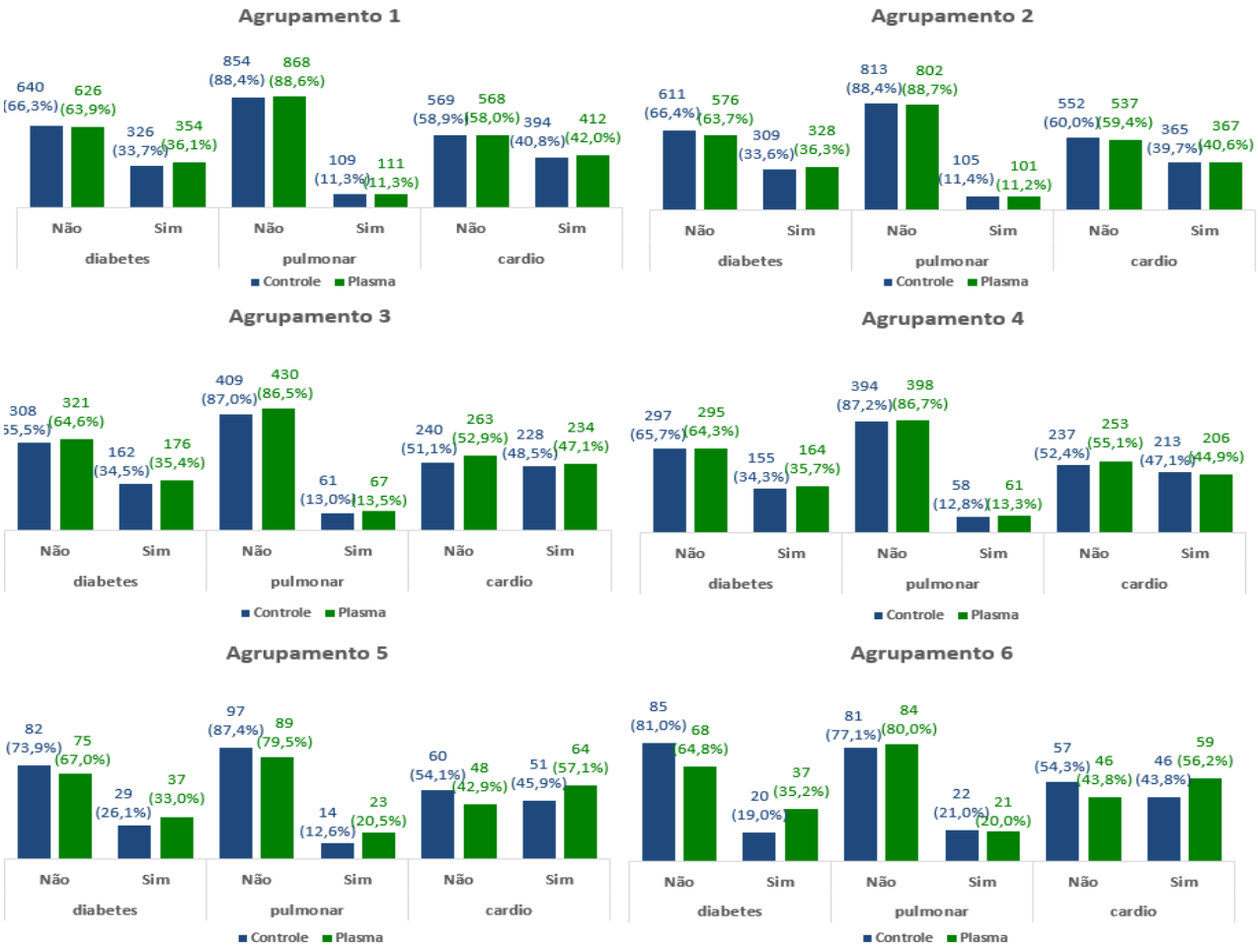

Figura A.3 - Ocorrência de doenças associadas a fatores de risco nos pacientes por tipo de tratamento e agrupamento

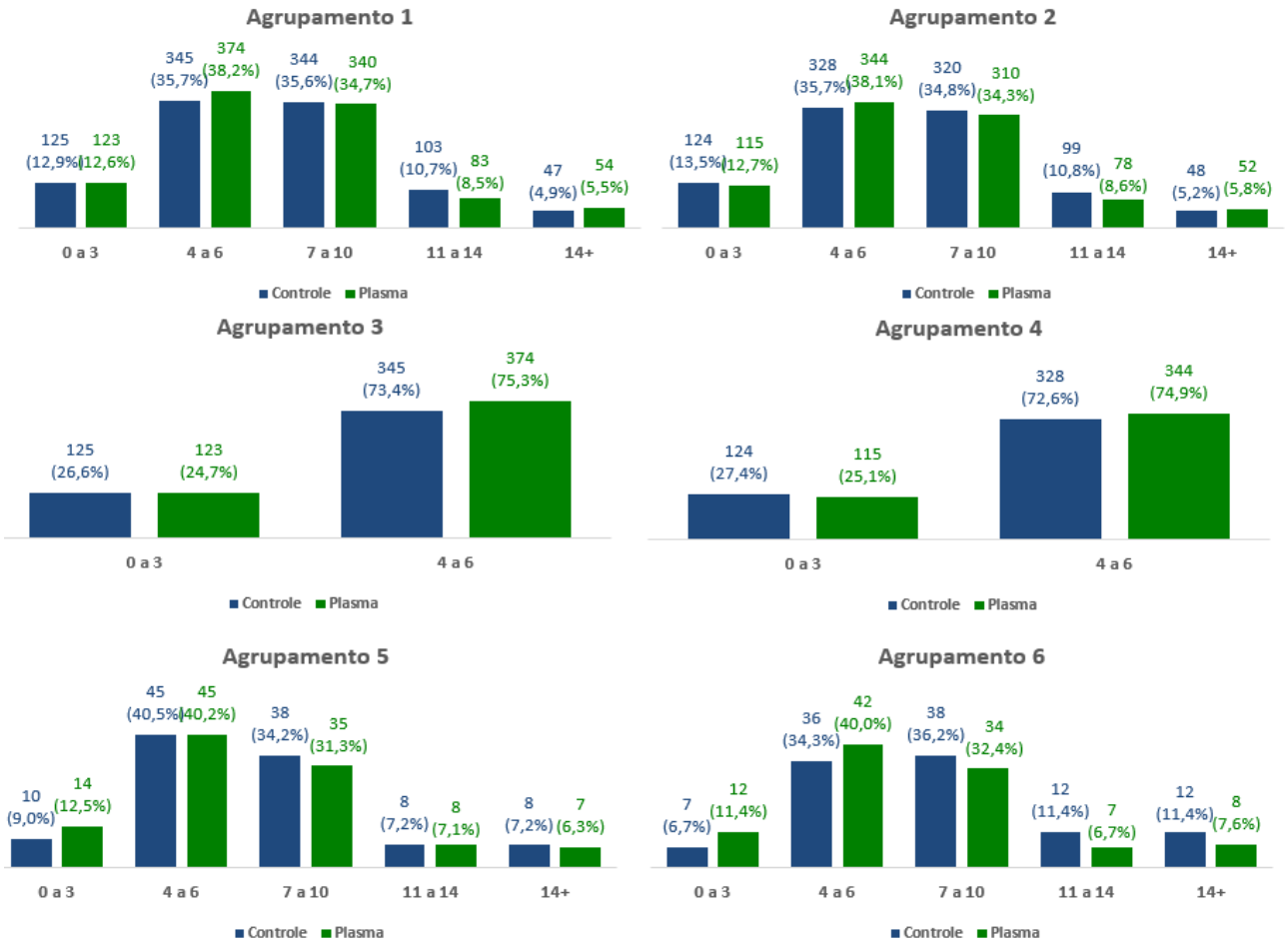

Figura A.4 - Quantidade de dias entre o início dos sintomas e a randomização por tipode de tratamento e agrupamento

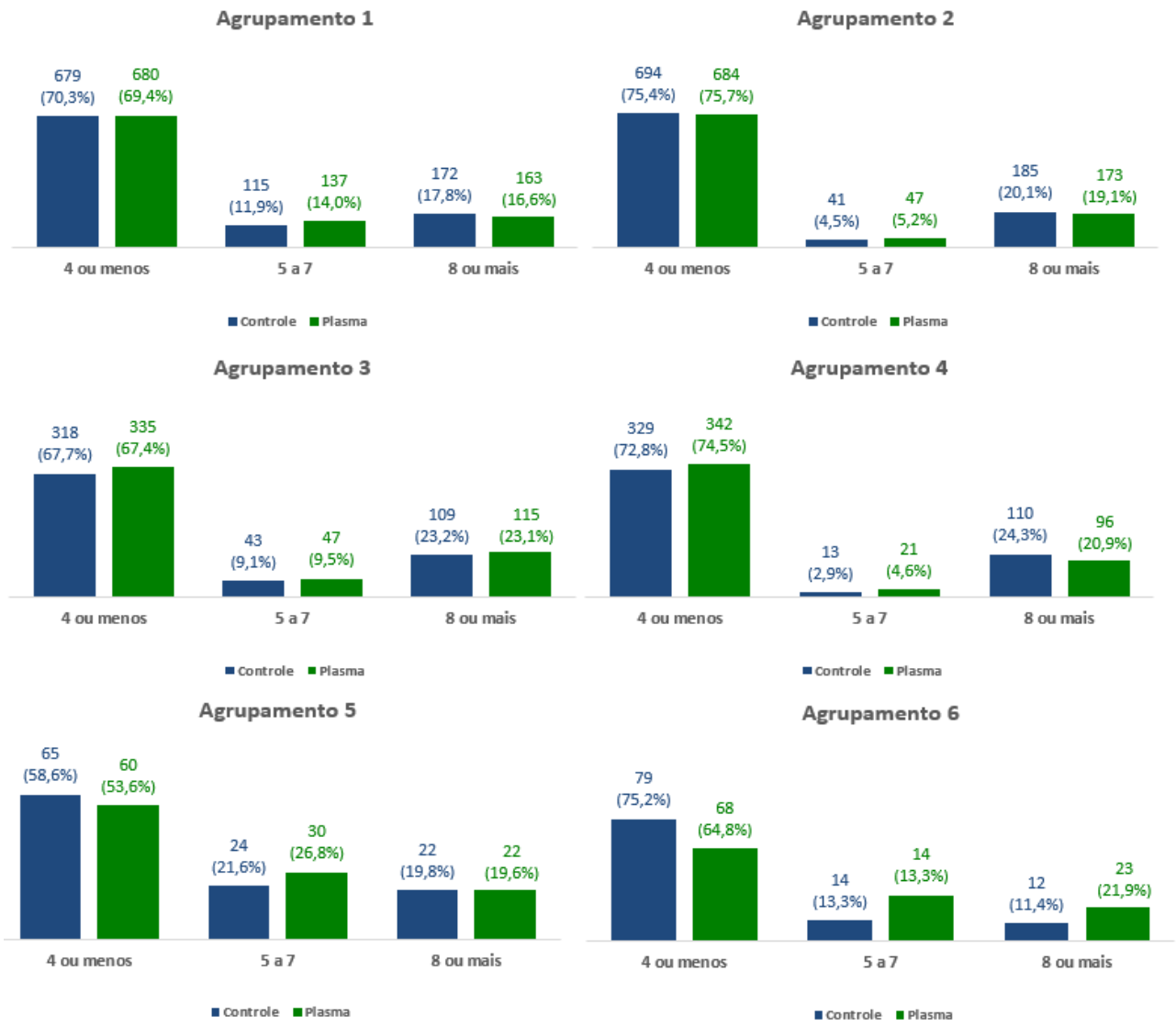

Figura A.5 - Escala da OMS na data de referência por tipo de tratamento e agrupamento

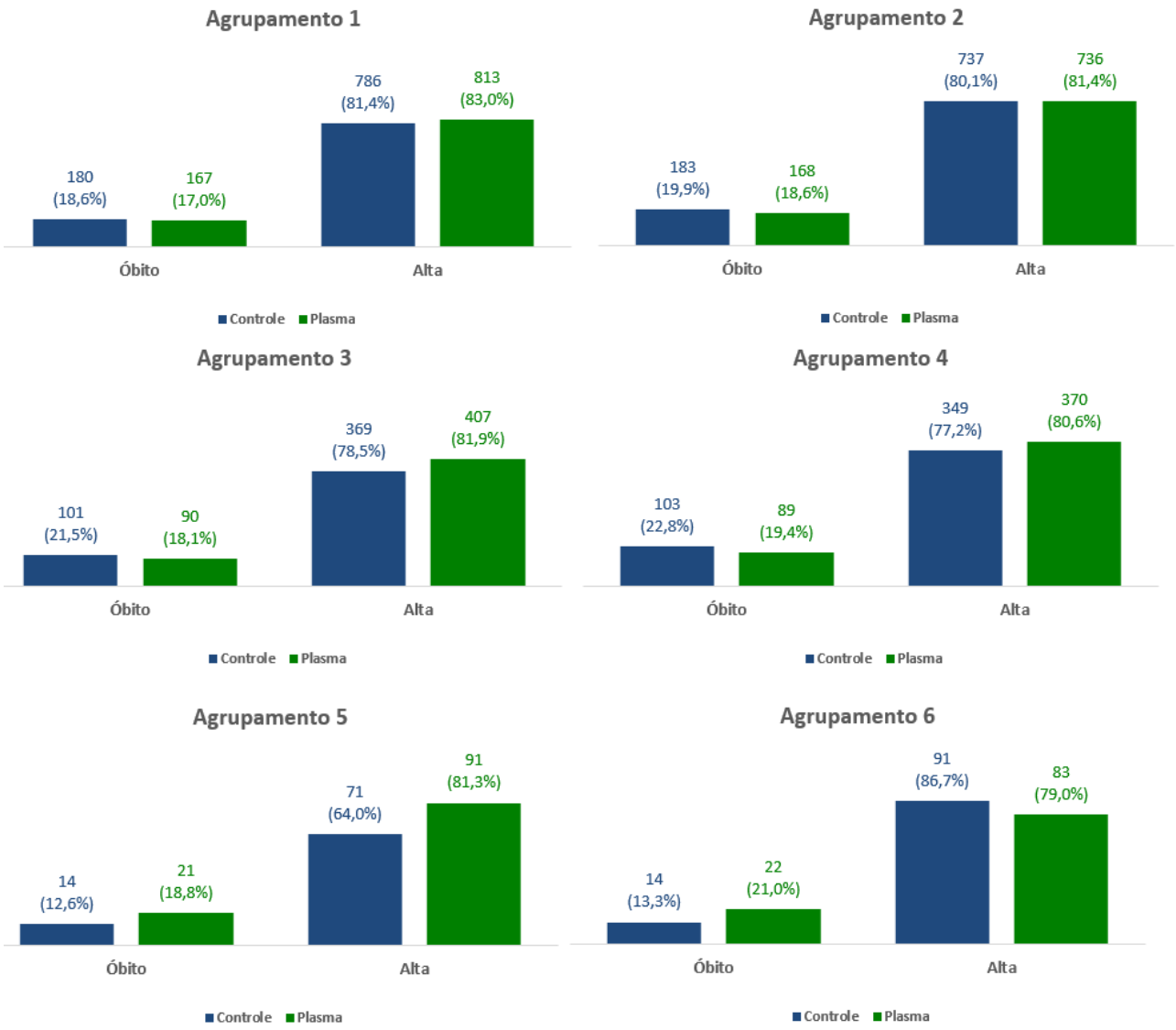

Figura A.6 - Proporção de óbitos e altas por tipo de tratamento e agrupamento

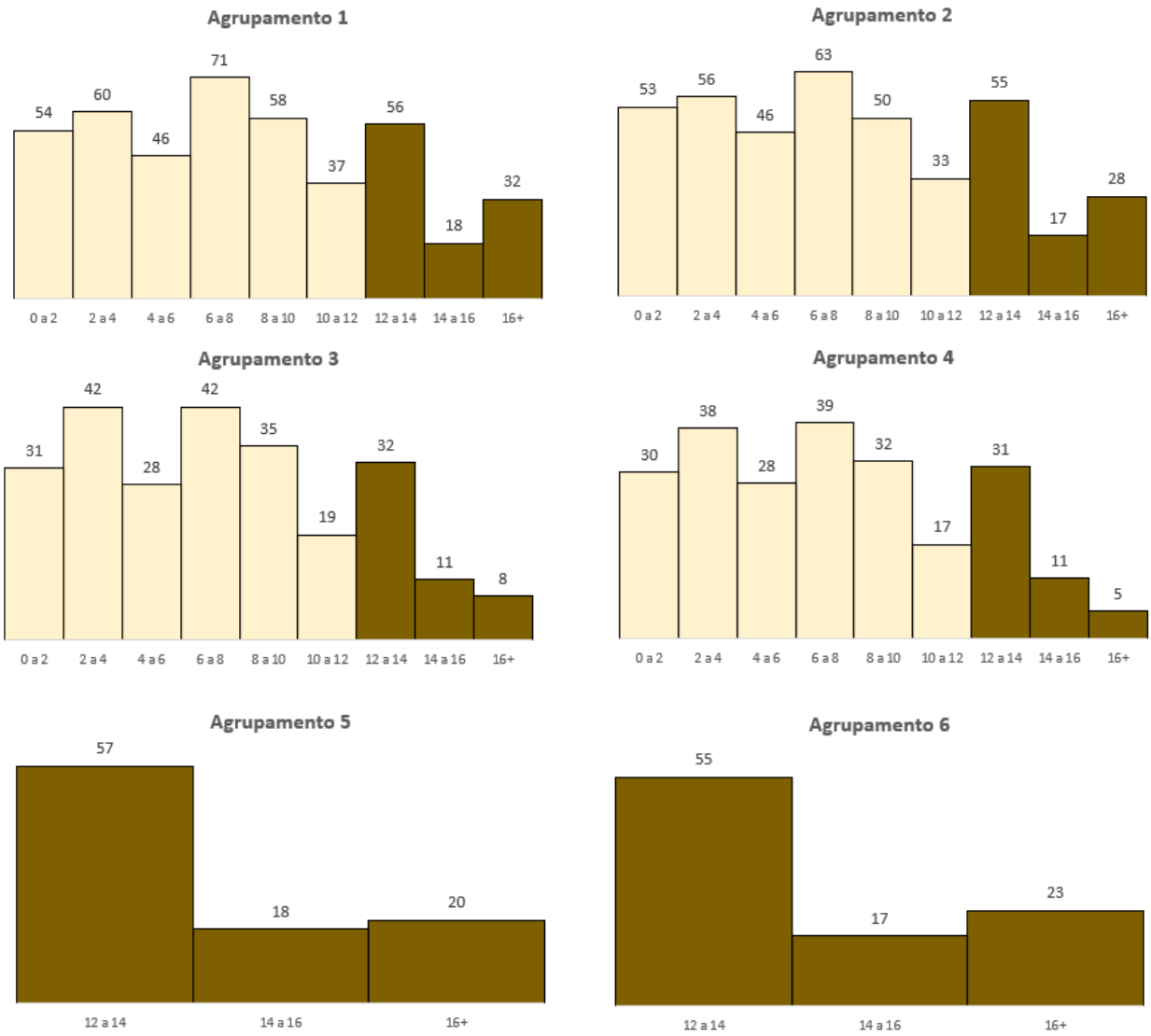

Figura A.7 - Score médio mensurado na plataforma Ortho por agrupamento

Apêndice B

Características Gerais dos Grupos

Tabela B.1: Características gerais dos grupos de Controle e Tratamento

Variável	Controle	%	Plasma	%	Variável	Controle	%	Plasma	%
Tratamento	1.138	48,0	1.231	52,0	**Escala OMS 28º dia**				
Sexo					<5	910	80,0	1.008	81,9
Masculino	730	64,1	794	64,5	5 a 7	55	4,8	66	5,4
Feminino	408	35,9	437	35,5	8 ou +	173	15,2	157	12,8
Faixa Etária					**Diabetes**				
<30	27	2,4	20	1,6	Não	770	67,7	804	65,3
30 a 39	82	7,2	91	7,4	Sim	368	32,3	427	34,7
40 a 49	151	13,3	188	15,3	**Doença Pulmonar**				
50 a 59	281	24,7	279	22,7	Não	998	87,7	1.082	87,9
60 a 69	265	23,3	293	23,8	Sim	136	12,0	144	11,7
70 a 79	197	17,3	233	18,9	(Vazio)	4	0,4	5	0,4
80 ou +	135	11,9	127	10,3	**Doença Cardíaca**				
RCT de Origem					Não	660	58,0	694	56,4
Bélgica	163	14,3	314	25,5	Sim	474	41,7	534	43,4
Brasil	15	1,3	19	1,5	(Vazio)	4	0,4	3	0,2
Espanha	171	15,0	179	14,5	**Início Sintomas até Randomização**				
EUA[1]	473	41,6	468	38,0	0 a 3	142	12,5	148	12,0
EUA[2]	39	3,4	41	3,3	4 a 6	394	34,6	441	35,8
EUA[3]	18	1,6	16	1,3	7 a 10	402	35,3	431	35,0
Holanda	35	3,1	37	3,0	11 a 14	136	12,0	125	10,2
Índia	224	19,7	157	12,8	14+	58	5,1	74	6,0
Status na Randomização					(Vazio)	6	0,6	12	1,0
Hospitalizado sem UTI	892	78,4	1.010	82,0	**Óbito**				
Hospitalizado com UTI	242	21,3	218	17,7	Não	953	83,7	1.058	85,9
(Vazio)	4	0,4	3	0,2	Sim	185	16,3	173	14,1
Escala OMS 14º dia					**Alta**				
<5	838	73,6	923	75,0	Não	185	16,3	173	14,1
5 a 7	149	13,1	169	13,7	Sim	953	83,7	1.058	85,9
8 ou +	151	13,3	139	11,3	**Score Médio**	**N.A**	**N.A**	**7,60**	**100,0**

Apêndice C

Script R

```
 #####Bibliotecas Usadas####
 #install.packages("RMySQL")
 library(RMySQL)
library(DBI)
 #install.packages("Rcpp")
library(Rcpp)
 #install.packages("MASS")
library(MASS)
#install.packages("ordinal")
library(ordinal)
#install.packages("readxl")
library(readxl)
#install.packages("coin")
library(coin)
#install.packages("pacman")
library(pacman)
#install.packages("VGAM")
```

```
library(VGAM)
#install.packages("dplyr")
library(dplyr)
library(readxl)
#install.packages("Amelia")
library(Amelia)
#if(!require(pacman)) install.packages("pacman")
library(pacman)
#install.packages("designr")
library(designr)
#install.packages("rms")
library(rms)
#install.packages("brant")
library(brant)
library(car)

###############Transformação de Variáveis####################
dados <- read_excel("C:/Users/pedro/Desktop/Base_COMPLETA_R.xlsx", sheet
= 'Plasma 28°dia')
dados$novo_score_14<-factor(dados$novo_score_14,levels=c(1,2,3),labels=c
("<5","5 a 7","8 ou mais"),ordered = T)
table(dados$novo_score_14)
dados$novo_score_28<-factor(dados$novo_score_28,levels=c(1,2,3),labels=c
("<5","5 a 7","8 ou mais"), ordered = T)
table(dados$novo_score_28)
dados$novo_score_0<-factor(dados$novo_score_0,levels=c(1,2),labels=c
("<5","5 a 7"), ordered = T)
```

```
table(dados$novo_score_0)
dados$origem<-as.factor(dados$origem)
dados$morte<-as.factor(dados$morte)
dados$alta<-as.factor(dados$alta)
dados$trimestre<-as.factor(dados$trimestre)
dados$dias_sintomas<-factor(dados$dias_sintomas,levels=c(1,2,3,4,5),
labels=c("0 a 3","4 a 6","7 a 10","11 a 14","14+"), ordered = T)
dados$cardio<-factor(dados$cardio,levels=c(0,1),labels=c("Não","Sim"))
dados$diabetes<-factor(dados$diabetes,levels=c(0,1),labels=c("Não","Sim")
dados$pulmonar<-factor(dados$pulmonar,levels=c(0,1),labels=c("Não","Sim")
dados$sexo<-factor(dados$sexo,levels=c(0,1),
labels=c("Masculino","Feminino"))
dados$tratamento<-factor(dados$tratamento,levels=c(0,1),
labels=c("Controle","Plasma"))
dados$grau_hospitalização<-factor(dados$grau_hospitalização,levels=c(1,2)
labels=c("Hospitalizado/sem UTI","Hospitalizado/com UTI"), ordered = T)
dados$idadec<-factor(dados$idadec,levels=c(1,2,3,4,5,6,7),
labels=c("<30","30 a 39","40 a 49","50 a 59","60 a 69","70 a 79","80+"),
ordered = T)
dados$ortho_max_Q<-factor(dados$ortho_max_Q, levels=c(0,1),
labels=c("Controle","Plasma"))
######## Verificando Multicolinearidade ##############################
m <- lm(as.numeric(novo_score_14) ~ origem + trimestre + dias_sintomas +
cardio + diabetes + pulmonar + sexo + tratamento + grau_hospitalização +
idadec + novo_score_0, dados)
car::vif(m)
######## Verificando as categorias de referência#######################
```

```
levels(dados$novo_score_0)
levels(dados$origem)
levels(dados$trimestre)
levels(dados$dias_sintomas)
levels(dados$cardio)
levels(dados$novo_score_14)
levels(dados$novo_score_28)
levels(dados$diabetes)
levels(dados$pulmonar)
levels(dados$sexo)
levels(dados$tratamento)
levels(dados$grau_hospitalização)
levels(dados$idadec)
levels(dados$morte)
levels(dados$alta)
levels(dados$ortho_max_Q)
# Criar uma matriz para armazenar os coeficientes de todos os modelos
coeficientes <- matrix(NA, nrow = 10, ncol = length(coef(mod)))

################### Divisão da base, 80% treino/ 20% teste###############
ind <- sample(sample(2,nrow(dados), replace = T, prob=c(0.8,0.2)))
treino<- dados[ind==1,]
teste <- dados[ind==2,]

######## Verificação gráfica da variável resposta nas diferentes bases###
barplot(table(dados$novo_score_14))
barplot(table(treino$novo_score_14))
```

```
barplot(table(teste$novo_score_14))


#Modelo MCP c/ função de lig. modificáve e teste de proporcionalidade####
mod <- MASS::polr(novo_score_14 ~ grau_hospitalização + novo_score_0 ,
data=treino, Hess = "T", method="logistic")

# Calcule o pseudo-R² de McFadden
null_deviance <- mod$null.deviance
residual_deviance <- mod$deviance
pseudo_R2_mcfadden <- 1 - (residual_deviance / null_deviance)

# Exiba o pseudo-R² de McFadden
pseudo_R2_mcfadden
mod <- MASS::polr(novo_score_14 ~ origem + trimestre + dias_sintomas +
cardio + diabetes + pulmonar + sexo + tratamento + grau_hospitalização +
idadec + novo_score_0 ,data=treino, Hess = "T", method="logistic")

AIC(mod)
car::poTest(mod)
################### Tabela de coeficientes com p-valor###########
ctable<- coefficients(summary(mod))
p<-pnorm(abs(ctable[,"t value"]),lower.tail = F)*2
ctable<-cbind(ctable,"p value" =p)
ctable
summary(mod)
```

```
############ Poder do Teste ###########
pred<-predict(mod,treino)
tab<-table(pred,treino$novo_score_14)
tab/colSums(tab)
sum(diag(tab)/sum(tab))
exp(cbind(OR = coef(mod), confint(mod)))
car::Anova(mod, type = "II", test = "Wald")
lmtest::coeftest(mod1)
 exp(coef(mod))
exp(confint(mod))
residuals (mod, type='score.binary', pl=TRUE)
fitted.values(mod1)
###### Modelo Estereótipo############
mod1 <- rrvglm(novo_score_28 ~ cardio +
             novo_score_0,data=treino, multinomial)
summary(mod1)
exp(coef(mod1))
fitted_probs <- fitted.values(mod1)
max_category_indices <- max.col(fitted_probs)
table(max_category_indices)
count_of_lt5_as_max <- sum(max_category_indices == 1)
count_of_lt5_as_max
red<-predict(mod1,treino)
tab<-table(red)
tab
tab/colSums(tab)
sum(diag(tab)/sum(tab))
```

```
exp(cbind(OR = coef(mod), confint(mod)))
########### MRC #################
  rc=rc.setup(treino$novo_score_14.ord)
  treino$novo_score_14.mrc=cr$novo_score_14
  treino$grau_hospitalização=rc$grau_hospitalização
  treino$novo_score_0.rc=rc$novo_score_0
  treino$cardio.rc=rc$cardio

  mrc=lrm(novo_score_14.mrc ~ grau_hospitalização.rc +
  novo_score_0.rc ,data=treino)
  mrc
```