



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Luis Gabriel Marques da Silva

APLICAÇÃO DO MODELO EMAX EM ENSAIOS CLÍNICOS DE BIOEQUIVALÊNCIA

Luis Gabriel Marques da Silva

APLICAÇÃO DO MODELO EMAX EM ENSAIOS CLÍNICOS DE BIOEQUIVALÊNCIA

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Estatística da Universidade Federal de Pernambuco como requisito parcial para obtenção do título de Bacharel em Estatística.

Orientador (a): Cristiano Ferraz

Coorientador (a): Leila Leal

Recife

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Silva, Luis Gabriel Marques da.

Aplicação do modelo Emax em ensaios clínicos de bioequivalência / Luis Gabriel Marques da Silva. - Recife, 2025.

30p : il., tab.

Orientador(a): Cristiano Ferraz

Coorientador(a): Leila Leal

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Ciências Exatas e da Natureza, Estatística - Bacharelado, 2025.

Inclui apêndices, anexos.

1. Regressão não linear. 2. Blocos completos. 3. Modelo Emax. I. Ferraz, Cristiano. (Orientação). II. Leal, Leila. (Coorientação). IV. Título.

310 CDD (22.ed.)

LUIS GABRIEL MARQUES DA SILVA

APLICAÇÃO DO MODELO EMAX EM ENSAIOS CLÍNICOS DE BIOEQUIVALÊNCIA

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Estatística da Universidade Federal de Pernambuco como requisito parcial para obtenção do título de Bacharel em Estatística.

Aprovado em: 20/07/2025

BANCA EXAMINADORA

Prof. Dr. Cristiano Ferraz (Orientador)

Universidade Federal de Pernambuco

Prof. Dr. Roberto Ferreira Manghi

Universidade Federal de Pernambuco

Prof. Dr. Getulio Jose Amorim do Amaral

Universidade Federal de Pernambuco

DEDICATÓRIA

Dedico este trabalho a Deus, por cuidar de mim todos os dias. A minha família, por caminhar no sol para que eu pudesse chegar na sombra. A meus pais, por não desistir de mim mesmo antes de eu nascer.

RESUMO

Este trabalho abrange as várias etapas de um estudo clínico que tem como objetivo aferir a bioequivalência entre dois corticoides, um de referência e um genérico, tendo como foco a fase piloto do estudo. As etapas que serão estudadas incluem, a realização de um experimento utilizando a aleatorização por blocos completos, para poder isolar variabilidade causada pelos diferentes braços onde os corticóides serão aplicados, em seguida a aplicação do método dos trapézios para transformar as medidas obtidas a partir do experimento na variável resposta, em seguida esses dados são usados para a estimação do modelo de regressão E_{max} , e em seguida as estimativas do modelo E_{max} são usadas para o cálculo do D_1 e D_2 , medidas que serão posteriormente usadas para definir os parâmetros do experimento realizado fase pivotal do estudo.

Palavras-chaves: Regressão não Linear, Blocos completos, Modelo Emax

ABSTRACT

This work covers the various stages of a clinical study aimed at assessing the bioequivalence between two corticosteroids, one reference and one generic, with a focus on the pilot phase of the study. The stages to be examined include conducting an experiment using complete block randomization to isolate variability caused by the different arms in which the corticosteroids will be applied. Following that, the trapezoidal rule is applied to transform the measurements obtained from the experiment into the response variable. These data are then used to estimate the E_{max} regression model, and subsequently, the E_{max} model estimates are used to calculate D_1 and D_2 , measures that will later be used to define the parameters of the experiment conducted in the pivotal phase of the study.

Keywords: Nonlinear Regression, Complete Block Design, Emax Model

Lista de figuras

Figura 1 - Exemplo de aleatorização do estudo piloto	15
Figura 2 - Perfil de branqueamento	17
Figura 3 - Histograma dos resíduos.....	23
Figura 4 - Dispersão entre resíduos e valores preditos	24

Lista de Quadros

Quadro 1 - ANOVA	12
Quadro 2 - Áreas dos trapézios.....	18
Quadro 3 - Estimativas	24

Sumário

1. Introdução	8
2. PLANO EXPERIMENTAL ALEATORIZADO EM BLOCOS COMPLETOS	10
2.1 ESTRUTURA DO MODELO	10
2.2 ANOVA	12
2.3 PLANO EXPERIMENTAL PARA ESTUDO PILOTO DE BIOEQUIVALÊNCIA.....	14
3. ÁREA SOB A CURVA PELA SOMA DOS TRAPÉZIOS.....	16
3.1 EXEMPLO DE CÁLCULO DE AUC.....	17
4. modelos DE REGRESSÃO NÃO-LINEAR	19
4.1 MÍNIMOS QUADRADOS NÃO LINEARES	19
4.2 MODELO <i>E_{max}</i>	20
4.2.1 ESCOLHA DE VALORES INICIAIS DO MODELO	21
4.3 ANÁLISE DE RESÍDUOS.....	22
4.4 RESULTADOS DO MODELO <i>E_{max}</i>	24
5. CONCLUSÕES	26
Referências	27
Apêndice A – códigos dos valores iniciais do modelo	28
Apêndice B – códigos da estimação do modelo <i>E_{max}</i>	28
Apêndice c – códigos das figuras 2, 3 e 4, e testes de hipótese	29

1. INTRODUÇÃO

A área da bioequivalência é central na avaliação de medicamentos genéricos, pois visa demonstrar que duas formulações possuem desempenhos farmacocinéticos semelhantes, visando a segurança e eficácia de produtos utilizados por pacientes. Dois produtos farmacêuticos são considerados bioequivalentes se, após administração na mesma dose e tempo, apresentarem perfil de biodisponibilidade, que é a fração da dose que chega no organismo de forma ativa, semelhante, especialmente em termos da concentração máxima do medicamento num organismo e do tempo que leva para essa concentração máxima ser atingida.

O estudo de bioequivalência em corticoides é composto por duas fases, a fase piloto, em que é feito um experimento usando um o medicamento genérico, e os resultados obtidos a partir desse experimento são usados na fase pivotal, que por sua vez é composta de um experimento que inclui tanto o medicamento genérico quanto o medicamento de referência, para que se possa comparar ambos os fármacos. Este trabalho cobre a fase piloto do estudo de bioequivalência.

As etapas da fase piloto incluem a realização de um experimento aleatorizado em blocos completos, em que os braços dos voluntários assumem o papel de blocos, diferentes locais nos braços representam as unidades experimentais e os tratamentos são definidos pela quantidade de tempo que o corticoide fica no braço do indivíduo (tempo-dose), durando de 0 até 6 horas. Após esse tempo, o corticoide é retirado e periodicamente os locais de aplicação são medidos 4 vezes consecutivas, usando um cromômetro, este equipamento mede o clareamento da pele no local indicado. O período entre a retirada do corticoide e o início das medidas dura de 0 até 48 horas.

Após todas as medidas serem recolhidas, é calculada a média das 4 medidas coletas em cada ponto para cada unidade experimental e cada voluntário, em seguida, pelo método dos trapézios, é calculada a área sob a curva (*AUC*) para cada uma das unidades experimentais de cada um dos voluntários, essas áreas são os pontos que serão usados na estimação do modelo E_{max} , um modelo de regressão não linear usado para dados de dose-resposta, onde a dose é o tempo que o corticoide passou aplicado e a resposta é a *AUC* do clareamento observado ao longo do tempo. O objetivo de achar as estimativas para o modelo E_{max} , em especial o parâmetro ED_{50} e a obtenção dos valores $D_1 = ED_{50}/2$ e $D_2 = 2 * ED_{50}$. O ED_{50} representa a dosagem necessária para que um indivíduo tenha metade da maior resposta possível a um medicamento. Definimos, a partir do D_1 e D_2 , a janela sensível da curva tempo-dose, que começa em D_1 e termina em D_2 , e é o intervalo de duração da aplicação do corticoide no qual pequenas variações no tempo de exposição provocam

mudanças detectáveis e significativas. Os tratamentos do experimento da fase pivotal, ou seja, os tempos de aplicação do corticoide, serão definidos dentro desta janela. Outro uso de D_1 e D_2 no experimento pivotal é a detecção de indivíduos não responsivos, esses indivíduos devem ser excluídos do estudo. Um indivíduo é considerado não responsivo se no experimento realizado no estudo pivotal, for constatado que:

$$\frac{AUC_{D_2}}{AUC_{D_1}} \geq 1.25$$

Onde:

AUC_{D_1} = média das AUCs no tempo-dose D_1 em ambos os braços

AUC_{D_2} = média das AUCs no tempo-dose D_2 em ambos os braços

Todas as etapas realizadas seguem as diretrizes criada pela FDA para a validação de bioequivalência de corticoides, essas mesmas etapas têm sido recomendadas pela ANVISA pois elas garantem a segurança e eficácia dos experimentos realizados e resultados obtidos, além de produzirem um padrão a ser seguido, já que as metodologias indicadas pelo FDA são validadas e reprodutíveis, permitindo que os resultados sejam confiáveis e comparáveis entre diferentes estudos e laboratórios.

Este trabalho está estruturado em 5 capítulos, cada um se aprofundando na teoria por trás dos métodos utilizados na fase piloto. A ordem das seções segue mesma ordem em que cada uma das etapas ocorre na fase piloto do estudo, sendo elas o planejamento do experimento aleatorizado por blocos completos, seguido do cálculo da área sob a curva pela soma dos trapézios, e a estimação do modelo E_{max} , que estão descritos respectivamente nas seções 2, 3 e 4.

2. PLANO EXPERIMENTAL ALEATORIZADO EM BLOCOS COMPLETOS

Nesta seção está descrito o processo usado para que as medidas de clareamento possam ser coletadas. o experimento que será realizado para essa coleta deve ser planejado de forma a evitar possíveis vieses e aumentar a validade dos resultados obtidos, para isso será empregado o desenho experimental de blocos completos.

O desenho experimental de blocos completos é um tipo de plano experimental usado quando existe uma fonte de variação que interfere na variável resposta, mas que não faz parte do tratamento. No experimento que desejamos realizar, essa fonte de variação é conhecida e controlável, sendo determinada pelo braço dos voluntários. Neste plano experimental, considera-se disponíveis 20 unidades experimentais e 2 níveis de blocos, sendo os níveis de blocos os diferentes braços de um voluntário. Como os braços de um mesmo indivíduo podem estar sujeitos a diferentes condições ao longo do tempo, eles podem reagir de forma distinta ao branqueamento, caracterizando uma situação em que uma variável que não é parte do tratamento influencia a variável resposta.

Um plano aleatorizado em blocos (ou Randomized Complete Block Design – RCBD) é um procedimento que combina o controle da variação conhecida com a aleatorização dos tratamentos dentro de cada bloco. A ideia central é que, ao dividir as unidades experimentais em blocos homogêneos (ou relativamente homogêneos) segundo uma fonte de variação conhecida, podemos reduzir a variabilidade experimental e aumentar a precisão das comparações entre os tratamentos. Cada bloco deve conter todas as condições ou tratamentos do experimento, permitindo que a comparação entre tratamentos seja feita de forma mais justa, minimizando o efeito das diferenças entre blocos.

O procedimento para criar um plano experimental em blocos consiste em:

1. Isolar as unidades experimentais dentro de blocos, de forma que cada unidade pertença a um único bloco;
2. Realizar aleatorizações dentro de cada bloco, sendo cada aleatorização independente entre os diferentes blocos.

O plano é chamado de completo porque cada bloco contém todos os tratamentos, garantindo que as comparações sejam balanceadas. Os blocos representam, portanto, uma fonte de variação conhecida e controlada, enquanto a aleatorização dentro deles garante que os efeitos residuais (não explicados pelo bloco ou pelo tratamento) não introduzam viés.

A seguir será abordada a estrutura do modelo, que representa tanto os efeitos dos tratamentos quanto os efeitos dos blocos na variável resposta, permitindo separar e quantificar as diferentes fontes de variação presentes no experimento.

2.1 ESTRUTURA DO MODELO

Assumindo que tenhamos t tratamentos e b blocos, com apenas uma unidade experimental por tratamento dentro de cada bloco. O modelo de efeitos para o Desenho de blocos completos pode ser escrito como:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, i = 1, \dots, t, j = 1, \dots, b$$

Onde:

i representa índice dos tratamentos

j representa índice dos blocos

Y_{ij} é a resposta do i -ésimo tratamento no j -ésimo bloco

μ é a média geral

τ_i é o efeito do i -ésimo tratamento

β_j é o efeito do j -ésimo bloco

ε_{ij} é o erro aleatório do i -ésimo tratamento no j -ésimo bloco, assumimos que os diferentes ε_{ij} 's são independentes entre si e $\varepsilon_{ij} \sim N(0; \sigma^2)$

Outra suposição que fazemos num modelo quando não temos réplicas em um experimento aleatorizado com blocos completos é a aditividade, significa que não há interação entre tratamentos e blocos, ou seja, o efeito de um tratamento é o mesmo em todos os blocos. Assim, o efeito de um tratamento se soma ao efeito de um bloco, e não muda dependendo do bloco. Essa suposição é necessária porque quando não possuímos réplicas não conseguimos estimar a interação entre blocos e tratamentos, denotada por $(\tau\beta)_{ij}$, logo não é possível realizar as análises referentes as diferenças de tratamentos em um modelo não aditivo.

Para que possamos estimar os parâmetros, adicionamos algumas restrições, sendo elas:

$$\sum_{i=1}^t \tau_i = 0 \text{ e } \sum_{j=1}^b \beta_j = 0$$

Os estimadores para μ, τ_i e $\beta_j, i = 1, \dots, t, j = 1, \dots, b$ são dados por:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^t \sum_{j=1}^b Y_{ij} = \bar{Y}_{..}$$

$$\hat{\tau}_i = \frac{1}{n} \sum_{j=1}^b Y_{ij} - \hat{\mu} = \bar{Y}_{i.} - \bar{Y}_{..}$$

$$\hat{\beta}_j = \frac{1}{n} \sum_{i=1}^t Y_{ij} - \hat{\mu} = \bar{Y}_{.j} - \bar{Y}_{..}$$

Para testar se os efeitos dos tratamentos são significativos, é necessário comparar a variabilidade entre grupos com a variabilidade dentro dos grupos, para isso é usada a ANOVA, descrita a seguir.

2.2 ANOVA

A ANOVA é usada para separar a variabilidade total observada nos dados em três componentes principais: a variabilidade entre tratamentos, que representa o efeito dos diferentes níveis do fator de interesse, a variabilidade entre blocos, que reflete diferenças sistemáticas entre os blocos, como lotes de material ou períodos de tempo, e o erro residual, que representa a variação não explicada. A ANOVA permite testar estatisticamente se há diferenças significativas entre os tratamentos, levando em conta a variabilidade entre os blocos e aumentando assim a precisão da comparação entre os tratamentos ao controlar fontes conhecidas de variabilidade.

A ANOVA é dada da seguinte forma:

Quadro 1 - ANOVA

FV	g.l	SQ	QM	E(QM)
Tratamento	$I - 1$	$\sum_i^I \sum_j^J (\bar{Y}_{i.} - \bar{Y}_{..})^2$	$\frac{SQT}{I - 1}$	$\sigma^2 + \frac{J \sum_{i=1}^I (\tau_i^2)}{I - 1}$
Bloco	$J - 1$	$\sum_i^I \sum_j^J (\bar{Y}_{.j} - \bar{Y}_{..})^2$	$\frac{SQB}{J - 1}$	$\sigma^2 + \frac{I \sum_{j=1}^J (\beta_j^2)}{J - 1}$
Erro	$(I - 1)(J - 1)$	$\sum_i^I \sum_j^J (Y_{ij} - \bar{Y}_{.j} - \bar{Y}_{i.} + \bar{Y}_{..})^2$	$\frac{SQE}{(I - 1)(J - 1)}$	σ^2
Total corrigido	$N - 1$	$\sum_i^I \sum_j^J (Y_{ij} - \bar{Y}_{..})^2$		

Fonte: Douglas Montgomery (2012)

a ANOVA separa:

- SQ Tratamento: mede a variabilidade devida às diferenças entre os tratamentos aplicados. Reflete o quanto as médias dos tratamentos se afastam da média geral.
- SQ Bloco: mede a variabilidade devida às diferenças entre os blocos. Como os blocos representam uma fonte de variação conhecida (ex.: braços dos voluntários), esse termo quantifica o quanto eles influenciam a variável resposta.
- SQ Erro (ou resíduo): representa a variabilidade não explicada pelos tratamentos nem pelos blocos. É a variação aleatória associada a fatores não controlados.
- SQ Total: soma de todas as fontes de variação, refletindo a variabilidade total dos dados em torno da média geral.

No Quadro da ANOVA temos:

- Fonte de Variação (FV): identifica cada componente de variação (Tratamentos, Blocos, Erro, Total).
- g.l. (graus de liberdade): número de informações independentes usadas para estimar cada componente de variação.
- SQ (Soma de Quadrados): quantifica a variação associada a cada fonte.
- QM (Quadrado Médio): é a variância estimada para cada fonte de variação, obtida dividindo-se SQ pelos respectivos graus de liberdade.
- E(QM): valor esperado do quadrado médio sob as suposições do modelo. Ele mostra a decomposição teórica, usado para definir o teste F.

O teste F de Fisher é usado para verificar se há diferenças significativas entre os tratamentos.

Hipóteses:

$H_0: \tau_1 = \dots = \tau_i = 0$ (não há efeito de tratamento, todos são iguais).

$H_1: \tau_i \neq 0$ para pelo menos um τ_i (pelo menos um tratamento difere dos demais).

Estatística de teste:

$$F = \frac{QMT}{QME}$$

onde:

- QMT estima a variabilidade entre tratamentos,
- QME estima a variabilidade residual (erro experimental).

Distribuição de referência:

Sob H_0 , a estatística F segue aproximadamente uma distribuição F de Snedecor com $t - 1$ e $(t - 1)(b - 1)$ graus de liberdade (dependendo do número de tratamentos t e blocos b).

Regra de decisão:

Se $F_{calculado} > F_{crítico}(\alpha; gl_1, gl_2)$ rejeitamos H_0 e concluímos que existe diferença significativa entre os tratamentos.

Caso contrário, não rejeitamos H_0 .

Essa regra é equivalente a comparar o p-valor do teste com o nível de significância α .

Apesar da tabela ANOVA poder ser usada para comparar diferentes tratamentos, ela não está inclusa no estudo de bioequivalência.

A seguir vemos como as unidades experimentais são definidas e como os tratamentos são aleatorizados na fase piloto do estudo.

2.3 PLANO EXPERIMENTAL PARA ESTUDO PILOTO DE BIOEQUIVALÊNCIA

Neste experimento os blocos designados são os diferentes braços dos voluntários, já que a pele de cada braço pode estar exposta a diferentes situações, como o sol, que faz com que os diferentes braços de um mesmo indivíduo reajam de maneira diferente ao corticoide.

Essa aleatorização se refere ao estudo piloto, onde o corticoide de referência será aplicado por diferentes quantidades de tempos em ambos os braços. Ao todo são 10 locais em cada braço, em 8 desses locais o corticoide será aplicado, os outros 2 serão locais onde o corticoide não será aplicado, os locais onde o corticoide não é aplicado são chamados de locais brancos, os locais brancos serão medidos em relação ao seu branqueamento e as médias das medidas obtidas nesses dois locais serão subtraídas das medidas dos demais locais dos braços, esse passo tem como objetivo corrigir o efeito basal das demais medidas. Os tempos de aplicação são os seguintes:

T1 (0,25h): 1 POR BRAÇO

T2 (0,5h): 1 POR BRAÇO

T3 (0,75h): 1 POR BRAÇO

T4 (1,0h): 1 POR BRAÇO

T5 (1,5h): 1 POR BRAÇO

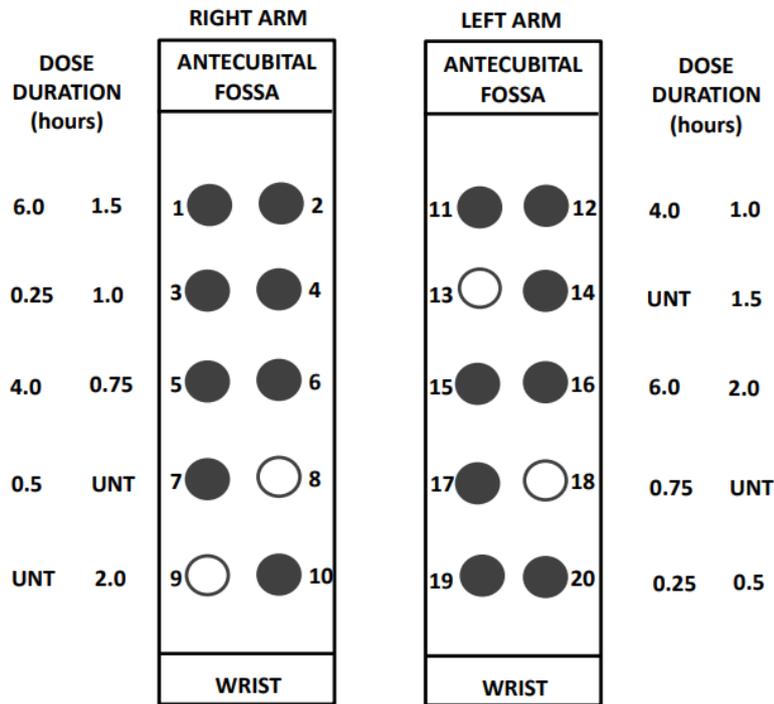
T6 (2h): 1 POR BRAÇO

T7 (4h): 1 POR BRAÇO

T8 (6h): 1 POR BRAÇO

BRANCO: 2 POR BRAÇO

Figura 1 - Exemplo de aleatorização do estudo piloto



Fonte: U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (2022)

Na figura acima vemos um exemplo de aleatorização por blocos, onde cada braço age como um bloco, e as escolhas das doses nos diferentes locais de cada braço acontecem de forma independente um do outro.

Uma vez obtidas as medidas de branqueamento corrigidas, é necessário convertê-las numa medida de área sob a curva, pois esse tipo de medida é usual para representar todo o efeito de um medicamento ou droga num organismo. Para o cálculo de área sob a curva é usada o método da soma dos trapézios, descrito na seção seguinte.

3. ÁREA SOB A CURVA PELA SOMA DOS TRAPÉZIOS

Esse método serve para aproximar integrais definidas de funções que não possuam uma função exata ou em situações em que possuímos apenas um conjunto de pontos, por exemplo dados experimentais, este método é uma alternativa ao método de integração tradicional, que usa a soma dos retângulos, usual para o cálculo de *AUC* quando conhecemos e conseguimos integrar a função que define a curva dos dados. A ideia do método da soma dos trapézios é aproximar a área sob uma curva dividindo o intervalo de integração em vários subintervalos e, em cada um, aproximar a área pela de um trapézio.

Para um conjunto de n pontos (x_i, y_i) , a *AUC* desses pontos é calculada pelo método dos trapézios usando a fórmula:

$$\sum_{i=1}^{n-1} (y_i + y_{i+1})(x_{i+1} - x_i)/2 = \sum_{i=1}^{n-1} A_i$$

Onde:

$(y_i + y_{i+1})$ representam as somas das alturas do trapézio i

$(x_{i+1} - x_i)$ representa a largura do trapézio i

No contexto do experimento temos que:

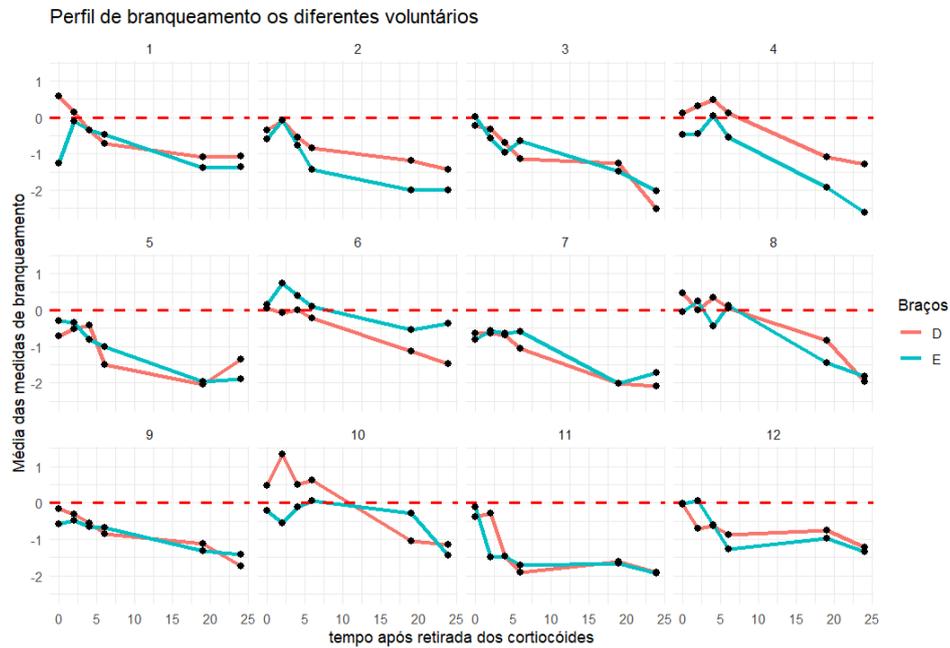
x_i = i -ésimo tempo de leitura após a retirada do corticoide

y_i = o i -ésima média do valor de branqueamento ajustado pela linha de base

A_i = área do i -ésimo trapézio.

A figura 2 mostra a curva da qual as *AUCs* são calculadas.

Figura 2 - Perfil de branqueamento



Fonte: Luis Gabriel (2025)

As médias das medidas de branqueamento da figura 2 são calculadas levando em conta as médias de todos os tratamentos de cada voluntário.

A seguir temos um exemplo de aplicação do método dos trapézios, usando as medidas de um dos voluntários no estudo piloto.

3.1 EXEMPLO DE CÁLCULO DE AUC

Aqui um exemplo, considerando o braço esquerdo de um dos voluntários, considerando que o tempo-dose, ou seja, o tempo que corticoide passou em contato com a pele, é 0,25h horas.

Quadro 2 - Áreas dos trapézios

Branqueamento (Y_I)	tempo de medida (X_I)	$Y_I + Y_{I+1}$	$X_{I+1} - X_I$	AREA (A_I)
-2,02125	0	-2,7825	2	-2,7825
-0,76125	2	-1,125	2	-1,125
-0,36375	4	-1,18	2	-1,18
-0,81625	6	-2,675	13	-17,38
-1,85875	19	-3,54125	5	-8,85
-1,6825	24	-	-	-

Fonte: Luis Gabriel (2025)

$$AUC = \sum_{i=1}^{n-1} A_i = -31,33$$

Logo, o AUC é igual a $-31,33$. Esse mesmo processo é repetido para todos os diferentes tempos-dose em ambos os braços.

Uma vez que temos todas as $AUCs$ de todos os tratamentos para todos os voluntários, usamos essas medidas para estimar o modelo de regressão não linear E_{max} , muito usado para modelar dados de dose-resposta, onde a dose é representada pelo tempo que o corticóide passou em contato com a pele, e a resposta é o AUC encontrado para essa dose. A teoria sobre modelos de regressão não lineares e sobre o modelo E_{max} estão descritas na próxima seção.

4. MODELOS DE REGRESSÃO NÃO-LINEAR

Esta seção explica a teoria envolta dos modelos de regressão não lineares, com foco no modelo E_{max} , que é a classe de modelos recomendada pelo FDA por ser uma boa representação do fenômeno estudado, além de mostrar as estimativas do modelo E_{max} obtidas a partir das $AUCs$ dos voluntários do estudo.

Ao contrário dos modelos de regressão padrões que são definidos da forma:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Podemos reescrever essa equação como:

$$y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$$

Onde $\mathbf{x}' = [1, x_1, \dots, x_k]$

Ou, de forma mais geral:

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_r z_r + \varepsilon$$

Onde z_i pode representar qualquer função das variáveis independentes originais x_1, \dots, x_k e ε é um vetor de erros aleatórios não correlacionados com $E(\varepsilon) = 0$ e $Var(\varepsilon) = \sigma^2$. Também assumimos que os erros seguem uma distribuição normal.

Modelos como esse capturam a dependência entre a variável y e as demais variáveis x_i , mas estão limitados a forma como conseguem representar essa relação. Modelos de regressão não linear são mais flexíveis, podem assumir um conjunto maior formas e representar uma gama mais ampla de relacionamentos. Eles são definidos da seguinte forma:

$$y = f(\mathbf{x}, \boldsymbol{\theta}) + \varepsilon$$

Onde $\boldsymbol{\theta}$ é um vetor de parâmetros desconhecidos com dimensões $p \times 1$ e f é uma função conhecida e derivável em relação aos parâmetros do vetor $\boldsymbol{\theta}$ podendo ser linear ou não linear em relação aos parâmetros, sendo modelos de regressão linear são um caso particular da classe de modelos de regressão não lineares.

4.1 MÍNIMOS QUADRADOS NÃO LINEARES

Aqui vemos como achar estimativas para $\boldsymbol{\theta}$ pelo método dos mínimos quadrados.

A função dos mínimos quadrados é definida por:

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2$$

Para encontrar as estimativas dos mínimos quadrados, é necessário derivar em relação a cada um dos parâmetros em $\boldsymbol{\theta}$, gerando as seguintes equações normais:

$$\sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})] \left[\frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \theta_j} \right] = 0 \text{ para } j = 1, 2, \dots, p$$

Como a solução muitas vezes não tem uma forma fechada, as implementações desse método usam o método Gauss-Newton, como é o caso da implementação no R, ou algum outro método iterativo.

Dando continuidade a questão de modelos de regressão não lineares, veremos agora a formulação do modelo E_{max} .

4.2 MODELO E_{max}

O modelo E_{max} foi formalizado com base na observação de que muitos efeitos biológicos seguem uma curva em forma de hipérbole. Este modelo passou a ser muito utilizado quando ensaios clínicos começaram a demandar modelos matemáticos mais precisos para descrever e prever os efeitos de medicamentos.

O E_{max} é um modelo não linear usado em análises dose-resposta. O modelo é dado por:

$$R_i = E_0 + \frac{D_i^N E_{max}}{D_i^N + ED_{50}^N} + \varepsilon_i \quad (1)$$

Onde:

i = o indicador do indivíduo

R_i = o valor da resposta do indivíduo i

D_i = o nível, dose ou concentração da droga no indivíduo i

E_0 = o efeito correspondente a resposta quando a dose da droga é zero

E_{max} = O efeito máximo atribuível a droga

ED_{50} = A dose que produz metade do E_{max}

N é o fator de inclinação (fator Hill), mede a sensibilidade da resposta ao intervalo de dose da droga, determinando a inclinação da curva dose-resposta ($N > 0$). No estudo da fase piloto, é assumido que $N = 1$.

ε_i = O termo de erro aleatório para o paciente i . Uma suposição padrão, adotada aqui, é que os ε_i termos são independentes e identicamente distribuídos com média 0 e variância σ^2 . Uma suposição padrão adicional é a de que os erros são normalmente distribuídos.

A curva do modelo E_{max} dose-resposta mostrada na Eq. (2) é o valor esperado do modelo E_{max} .

$$R = E_0 + \frac{D^N E_{max}}{D^N + ED_{50}^N} \quad (2)$$

A curva do modelo E_{max} pode crescer ou diminuir em relação a um aumento na dosagem. Se a resposta está diminuindo, o valor do parâmetro do E_{max} será negativo. Aqui D_i Representa o tempo que o corticoide passou em contato com a pele.

4.2.1 ESCOLHA DE VALORES INICIAIS DO MODELO

O método Gauss-Newton, usado por para estimar o modelo E_{max} , necessita que valores iniciais sejam fornecidos para as estimativas. Os valores iniciais devem ser suficientemente próximos dos valores verdadeiros dos parâmetros, cujo valor real é desconhecido, para garantir a convergência do modelo. A escolha feita do método usado para se chegar a um valor inicial para as estimativas dos parâmetros teve como base a natureza do parâmetro que estamos tentando estimar.

Como o E_0 representa o valor da resposta quando a dose é 0, o valor inicial do E_0 é escolhido como sendo $\hat{\beta}_0$, o estimador do intercepto (β_0) do modelo de regressão linear dado por:

$$\hat{Área} = \beta_0 + Tempo * \beta_1 + \varepsilon_i$$

Onde $\hat{Área}$ representa o AUC e $Tempo$ representa a dose-tempo.

Como o E_{max} é a resposta máxima atribuída a dose, o valor inicial escolhido para esse parâmetro é a maior resposta (positiva ou negativa) encontrada nos dados.

O ED_{50} corresponde a dose que produz metade do efeito do E_{max} , logo o valor inicial escolhido para ele é a metade do valor do tempo-dose que está na mesma linha que o valor que foi usado como valor inicial do E_{max} .

Tendo os dados e os valores iniciais das estimativas, podemos calcular as estimativas finais do modelo. Os códigos necessários para realizar esse cálculo são mostrados na próxima seção.

4.3 ANÁLISE DE RESÍDUOS

Nesta seção vamos checar, por meio da análise de resíduos, se as suposições feitas nos erros do modelo E_{max} são violadas. Os resíduos usados foram os resíduos ordinários, definidos como:

$$\hat{\epsilon}_i = y_i - \hat{y}_i$$

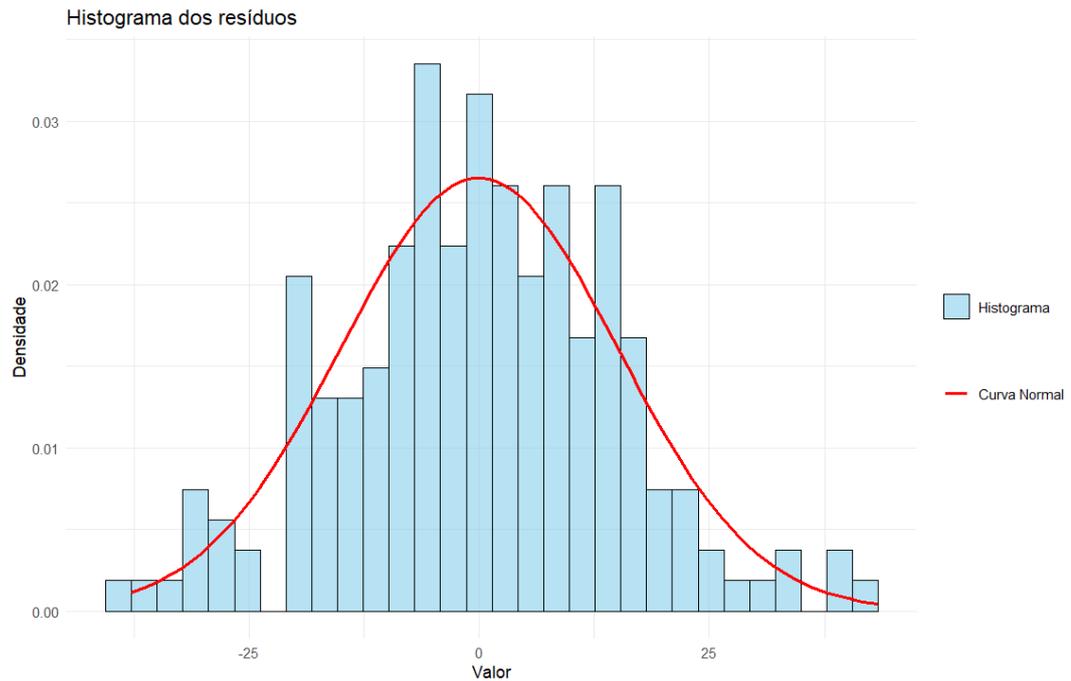
Onde:

y_i representa a i -ésima resposta;

\hat{y}_i a previsão a de y_i feita pelo modelo.

Primeiro vamos checar a suposição de Normalidade. Faremos isso tanto graficamente, comparando um histograma dos erros com uma curva normal, quanto por meio de um teste de hipóteses.

Figura 3 - Histograma dos resíduos



Fonte: Luis Gabriel (2025)

Usando o teste Shapiro-Wilk para testar a normalidade. As hipóteses do teste são as seguintes:

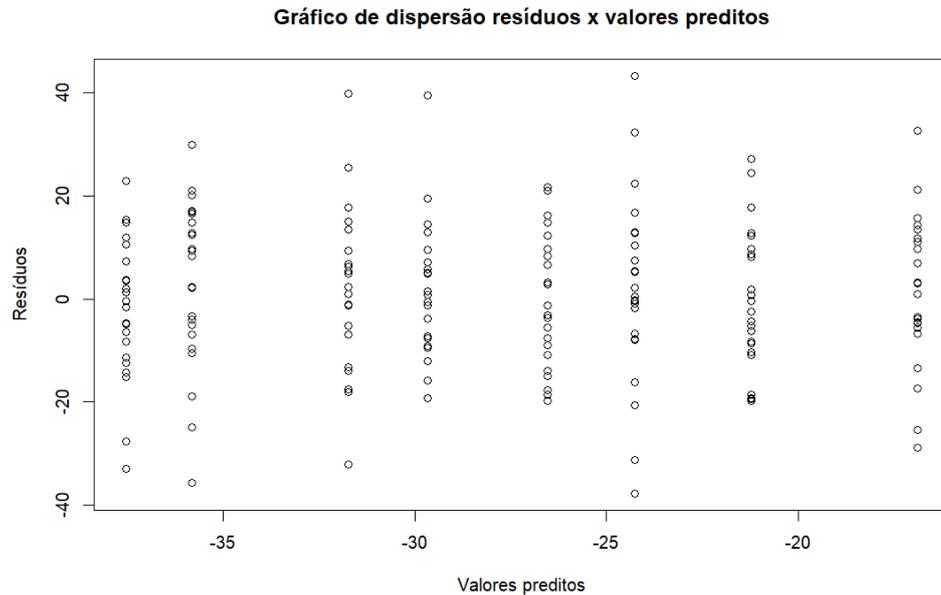
H_0 : Os resíduos seguem uma distribuição normal

H_1 : Os resíduos não seguem uma distribuição normal

Valor de $p = 0.621$. Logo nós não rejeitamos a hipótese nula. Tanto o teste de hipóteses quanto o histograma indicam que a suposição de normalidade não foi violada. É verificada a seguir a suposição de homoscedasticidade.

A análise da homoscedasticidade é feita por meio de um gráfico de dispersão considerando os resíduos com os valores preditos. Nessa análise é verificado se a variância dos resíduos é constante ao longo dos valores ajustados, que se revela a partir de uma forma de funil no gráfico de dispersão. A partir do gráfico abaixo vemos que não existem sinais de heteroscedasticidade e nenhuma estrutura sistemática nos resíduos, portanto a suposição de homoscedasticidade parece estar confirmada.

Figura 4 - Dispersão entre resíduos e valores preditos



Fonte: Luis Gabriel (2025)

Uma vez que não existem sinais de que as suposições tenham sido violadas, logo podemos olhar os resultados das estimativas e ver o quão bem o modelo representa os dados.

4.4 RESULTADOS DO MODELO E_{max}

Seguem as estimativas do modelo, juntamente com os valores estimados de D_1 e D_2 :

Quadro 3 - Estimativas

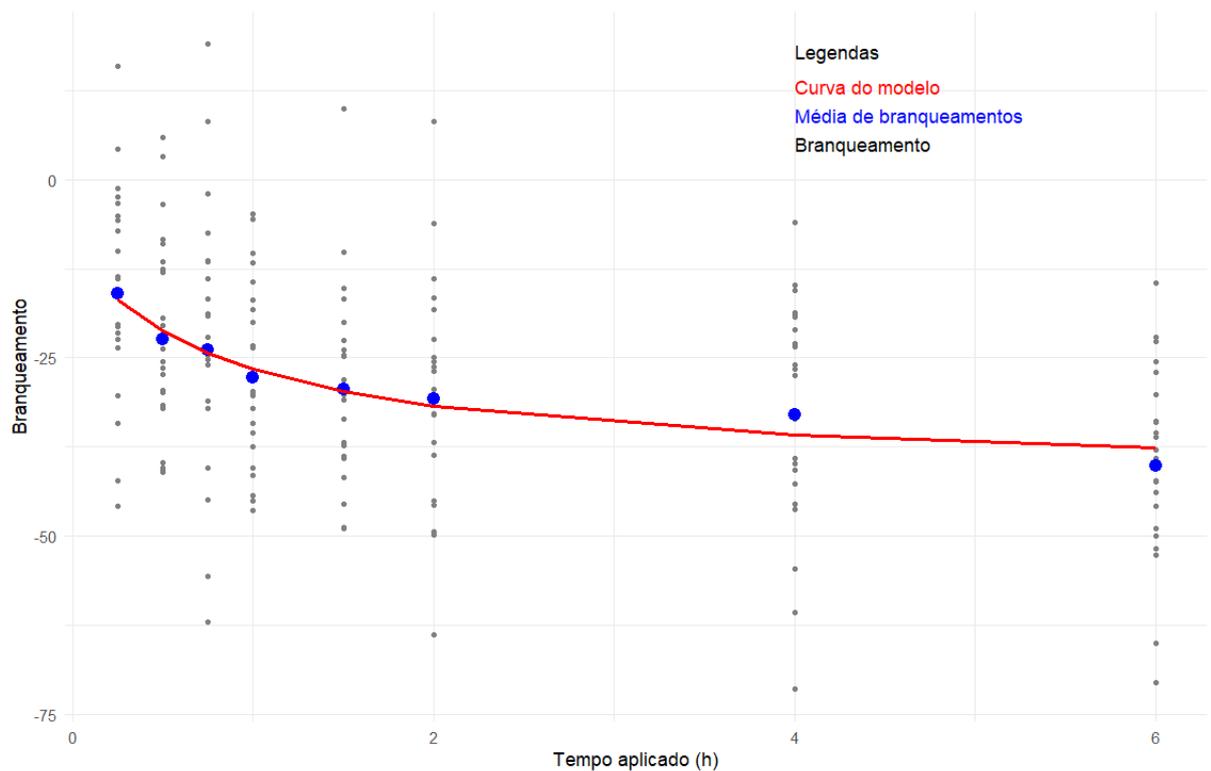
Parâmetros	Estimativa	Erro padrão	CV	Valor inicial
E_0	-10.203	7.401	72.54%	-21.379
E_{max}	-31.603	5.771	18.27%	-71.501
ED_{50}	0.935	0.775	82.89%	2
D_1	0.4675	-	-	-
D_2	1.87	-	-	-

Fonte: Luis Gabriel (2025)

Daqui podemos observar que o efeito basal é -10.203, indo até um máximo de -31.603 e a metade do efeito máximo é obtido em 0.935 horas. Os coeficientes de variação oscilam entre 18%-83%.

A seguir um gráfico comparando os valores observados com a curva do modelo E_{max} com as estimativas dos parâmetros.

Figura 5 - branqueamento ao longo do tempo



Fonte: Luis Gabriel (2025)

Pelo gráfico vemos que o modelo se ajusta bem aos dados, isso porque as médias dos branqueamentos nos diferentes tempos-dose estão próximas da curva ajustada.

5. CONCLUSÕES

Este trabalho abordou a fase piloto de um estudo de bioequivalência in vivo de corticoides, com base nas diretrizes da Food and Drug Administration (FDA). A fase piloto tem papel fundamental na estruturação metodológica do estudo pivotal, sendo responsável por estimar os parâmetros essenciais para a comparação entre o produto teste e o produto de referência.

Por meio da aplicação do modelo E_{max} à resposta de branqueamento induzida pelo corticoide de referência, foi possível estimar os parâmetros ED_{50} , D_1 e D_2 , os quais definem a janela sensível da curva do tempo-dose. Essa janela é crucial para a seleção de sujeitos responsivos e para garantir a sensibilidade do estudo que ocorrerá posteriormente na fase pivotal.

Referências

ATKINSON, K. *Theoretical numerical analysis: a functional analysis framework*. 2. ed. New York: Springer, 2009.

FOOD AND DRUG ADMINISTRATION (FDA). *Topical dermatologic corticosteroids: in vivo bioequivalence*. Silver Spring, MD: U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, 2022.

GAIL, M.; KRIEGER, K. *Dose finding in drug development*. New York: Springer, 2006.

MONTGOMERY, D. C. *Design and analysis of experiments*. 8. ed. Hoboken, NJ: John Wiley & Sons, 2012.

MONTGOMERY, D. C. *Introduction to linear regression analysis*. 5. ed. Hoboken, NJ: John Wiley & Sons, 2012.

APÊNDICE A – CÓDIGOS DOS VALORES INICIAIS DO MODELO

```
e0_inicial1 = function(dados){
  e0.inicial = lm(dados$area ~ dados$tempo)$coefficients[1]
  return(e0.inicial)
}
```

```
emax_inicial1 = function(dados){
  emax.inicial = ifelse(max(dados$area) == max(abs(dados$area)),
                        max(dados$area),
                        -max(abs(dados$area)))

  return(emax.inicial)
}
```

```
ed50.inicial1 = function(dados){
  emax.iniciais = emax_inicial1(dados)
  return((dados$tempo[dados$area == emax.iniciais])/2)
}
```

APÊNDICE B – CÓDIGOS DA ESTIMAÇÃO DO MODELO E_{max}

#Aqui os dados são importados e as são feitas manipulações para que eles possam ser

#usados na estimação do modelo.

```
compilado_area <- read.csv("D:/compilado_area.csv")
```

```
head(compilado_area)
```

```
compilado_area$tempo <- as.numeric(gsub(",", ".", compilado_area$tempo))
```

```
compilado_area$area <- as.numeric(gsub(",", ".", compilado_area$area))
```

```
#Aqui são calculados os valores iniciais para as estimativas. Os códigos relacionados
#as funções usadas para achar os valores iniciais estão no apêndice A.
```

```
e0_ini = e0_inicial1(compilado_area)
```

```
emax_ini = emax_inicial1(compilado_area)
```

```
ed50_ini = ed50.inicial1(compilado_area)
```

```
#As linhas seguintes mostram como ajustar o modelo e como acessar as estimativas
#as quais o processo de convergência resultou.
```

```
emax_model <- nls(area ~ E0 + (Emax * tempo) / (ED50 + tempo), data =
compilado_area,
```

```
start = list(E0 = e0_ini, Emax = emax_ini, ED50 = ed50_ini))
```

```
summary(emax_model)
```

APÊNDICE C – CÓDIGOS DAS FIGURAS 2, 3 E 4, E TESTES DE HIPÓTESE

```
plot(x=compilado_area$tempo, y=compilado_area$area, col = "black",
```

```
  xlab = "Tempo aplicado (h)", ylab = "Branqueamento")
```

```
coefs = coef(emax_model)
```

```
curve(coefs[1] + (coefs[2] * x) / (coefs[3] + x), from = 0, to = 50,
```

```
  col = "black", add = TRUE)
```

```
legend("topright",          # posição
      legend = c("Dados observados", "Curva Emax ajustada"),
      col = c("black", "black"),      # cores
      pch = c(1, NA),                # símbolo dos pontos (NA = nada)
      lty = c(NA, 1),                # tipo da linha (1 = contínua, NA = nada)
      bty = "n")                     # sem borda na legenda
```

```
residuos = resid(emax_model)
```

```
plot(residuos, main = "Gráfico de dispersão dos resíduos")
```

```
hist(residuos, main = "Histograma dos resíduos")
```

```
shapiro.test(residuos)
```

```
boxplot(area~tempo, data = compilado_area)
```

```
plot(residuos~fitted(emax_model), data = compilado_area,
      main = "Gráfico de dispersão resíduos x valores preditos",
      xlab = "Valores preditos",
      ylab = "Resíduos")
```