

RAE-CEA-9710
RELATÓRIO DE ANÁLISE ESTATÍSTICA
SOBRE O PROJETO: RELAÇÃO
ESTRUTURA-ATIVIDADE DE ANESTÉSICOS
LOCAIS N,N [DIMETILAMINA] ETIL
BENZOATOS PARA-SUBSTITUÍDOS

**Carmen Diva Saldiva de André
Silvia Nagib Elian
Adriana Bruscato**

- São Paulo, Junho de 1997 -

CENTRO DE ESTATÍSTICA APLICADA

RELATÓRIO DE ANÁLISE ESTATÍSTICA - NÚMERO 9710 - CÓDIGO 97P10

TÍTULO : Relatório de análise estatística sobre o projeto: “Relação estrutura-atividade de anestésicos locais N,N [dimetilamina] etil benzoatos para-substituídos”.

PESQUISADORA: Walkyria Sigler

ORIENTADOR: Prof. Dr. Antonio Carlos Oliveira

INSTITUIÇÃO: Instituto de Ciências Biomédicas - USP

FINALIDADE: Tese de mestrado.

RESPONSÁVEIS PELA ANÁLISE :

Prof^a Dr^a Carmen Diva Saldiva de André
Prof^a Dr^a Silvia Nagib Elian
Adriana Bruscato.

REFERÊNCIA DESTE TRABALHO:

André, C. D. S., Elian, S. N., Bruscato, A. **Relatório de análise estatística sobre o projeto: Relação estrutura-atividade de anestésicos locais N,N [dimetilamina] etil benzoatos para-substituídos.** São Paulo, IME-USP, 1997.

FICHA TÉCNICA

BIBLIOGRAFIA:

BMDP Statistical Software Manual. Versão 7.0. Volume 1. (1992). Los Angeles: *University Press of California & John Wiley & Sons.* 678p.

BUSSAB, W. O. & MORETTIN, P. A. (1987). **Estatística Básica - Métodos Quantitativos.** 4. ed. São Paulo: Atual. 321p.

DRAPER, N. R. & SMITH, H. (1981). **Applied Regression Analysis.** 2.ed. New York: John Wiley. 709p.

HAHN, G. J. & SHAPIRO, S. S. (1967). **Statistical Models in Engineering.** 1.ed. New York: John Wiley. 355p.

JOHNSON, R. A.. & WICHERN, D. W. (1982). **Applied Multivariate Statistical Analysis.** 2.ed. Englewood Cliffs: Prentice Hall . 594p.

MINITAB for Windows ® : reference manual, Versão 10.0. (1996). USA: Minitab Corporation. 940p.

MONTGOMERY, D. C. & PECK, E. A. (1992). **Introduction to Linear Regression Analysis.** 2. ed. New York: John Wiley. 527p.

SPSS / PC + for Windows ® : Reference Manual, Versão 6.0. (1993). SPSS Inc.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

BMDP *for Windows* ® , versão 7.0.

Microsoft Word *for Windows* ® , versão 6.0.

MINITAB *for Windows* ® , versão 10.0.

SPSS *for Windows* ® , versão 6.0.

TÉCNICAS ESTATÍSTICAS UTILIZADAS: (entre parênteses a classificação ISI)

Análise Descritiva Unidimensional (03:010)

Análise de Conglomerados (06:120)

Análise de Regressão Clássica (07:020)

ÁREA DE APLICAÇÃO

Ciências Biomédicas (14:990)

ÍNDICE

Resumo	6
1. Introdução	7
2. Objetivo	8
3. Descrição do Processo de Obtenção dos Dados	8
4. Descrição das Variáveis	12
5. Análise Descritiva	13
5.1. Relações entre a Potência e as Variáveis Independentes	13
5.2. Relações entre as Variáveis Independentes	14
6. Análise Inferencial	15
6.1. Modelo e Interpretação dos Parâmetros	15
6.2. Seleção de Variáveis Independentes	16
6.3. Modelo com as Variáveis Independentes Sigma e LogPAPP	17
6.4. Modelo com as Variáveis Independentes Sigma, LogPAPP e MR4	21
7. Conclusões	24
Apêndice A - Gráficos	26
Apêndice B - Matriz	31
Apêndice C - Multicolinearidade	33
Apêndice D - Análise de Resíduos	37

RESUMO

O projeto é da área farmacológica e investiga o efeito de diversos tipos de anestésicos locais sobre o coração de ratos. O interesse deste estudo consiste em verificar quais características físico-químicas da molécula de determinada droga influenciam mais em sua potência tóxica. Para isto, foram utilizados setenta e dois ratos, homogêneos entre si, divididos em quatorze grupos, contendo de três a oito ratos. Cada grupo foi submetido a uma droga de interesse.

A análise descritiva indicou que todas as características físico-químicas das drogas consideradas no projeto apresentaram relação linear com a variável potência. Notou-se, também, que muitas destas características eram correlacionadas linearmente entre si, duas a duas, gerando o problema da multicolinearidade.

Na análise inferencial, algumas variáveis independentes foram eliminadas, a fim de barrar a multicolinearidade, o que resultou na seleção de dois modelos de regressão linear múltipla. Com base na análise dos resíduos, escolheu-se o modelo mais adequado para explicar a variável potência.

1. INTRODUÇÃO

O efeito de um fármaco é o resultado de sua interação com o organismo vivo. A nível molecular, esta interação, muitas vezes, se dá através da ligação da droga com sítios específicos no organismo alvo: receptores.

As drogas estudadas neste trabalho são anestésicos locais. Eles são usados na prática médica e odontológica para inibir a dor, permitindo, assim, a realização de procedimentos cirúrgicos. O alvo dos anestésicos locais são os tecidos excitáveis do organismo. Os anestésicos locais impedem a transmissão dos impulsos elétricos nestes tecidos, bloqueando (reversivelmente) os canais iônicos responsáveis pela transferência de carga positiva através da membrana celular.

A aplicação da droga em local eleito para ser anestesiado garante que o fármaco atinja apenas seu âmbito de ação. No entanto, se o anestésico local for inadvertidamente injetado em algum grande vaso e ganhar a circulação corpórea, ele atingirá outros tecidos excitáveis, entre eles, o coração. Resultará, então, uma ação indesejada, tóxica, do anestésico local sobre o tecido cardíaco. Este trabalho investiga a ação de vários anestésicos locais diretamente sobre o coração, e atribui a cada droga um valor de potência tóxica cardíaca. Procurou-se correlacionar a atividade tóxica com os valores conhecidos das variáveis físico-químicas de cada droga, que estão intrinsecamente relacionadas à sua estrutura molecular.

Para este estudo, foram sintetizados quatorze novos anestésicos locais que possuem, entre si, um núcleo comum. A única diferença estrutural entre estas moléculas é um grupamento substituinte no anel benzênico. As variáveis físico-químicas decorrentes desta única substituição foram avaliadas quanto à sua capacidade de prever a potência tóxica destas drogas sobre o coração.

A compreensão de quais características físico-químicas dos anestésicos locais estão implicadas no aumento da toxicidade sobre o coração pode ajudar no planejamento de drogas mais seguras.

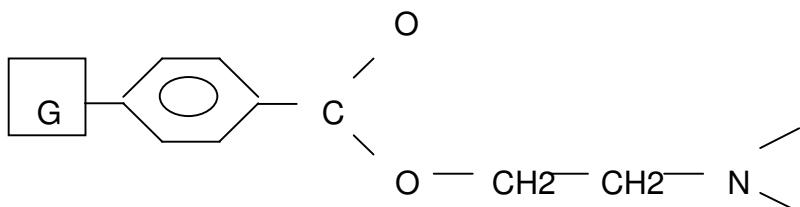
2. OBJETIVO

Investigar quais características físico-químicas de certos anestésicos locais são mais importantes para explicar a potência das diferentes drogas.

3. DESCRIÇÃO DO PROCESSO DE OBTENÇÃO DOS DADOS

Neste projeto, foram utilizadas 14 drogas com fórmulas moleculares parecidas, que diferem apenas em um grupamento molecular (ver Figura 3.1). As diferenças de efeitos biológicos observados para as drogas devem, por conseguinte, relacionar-se ao único grupamento que as diferencia.

Figura 3.1 - Molécula de uma droga. G : grupamento molecular substituinte que varia de uma droga para outra.

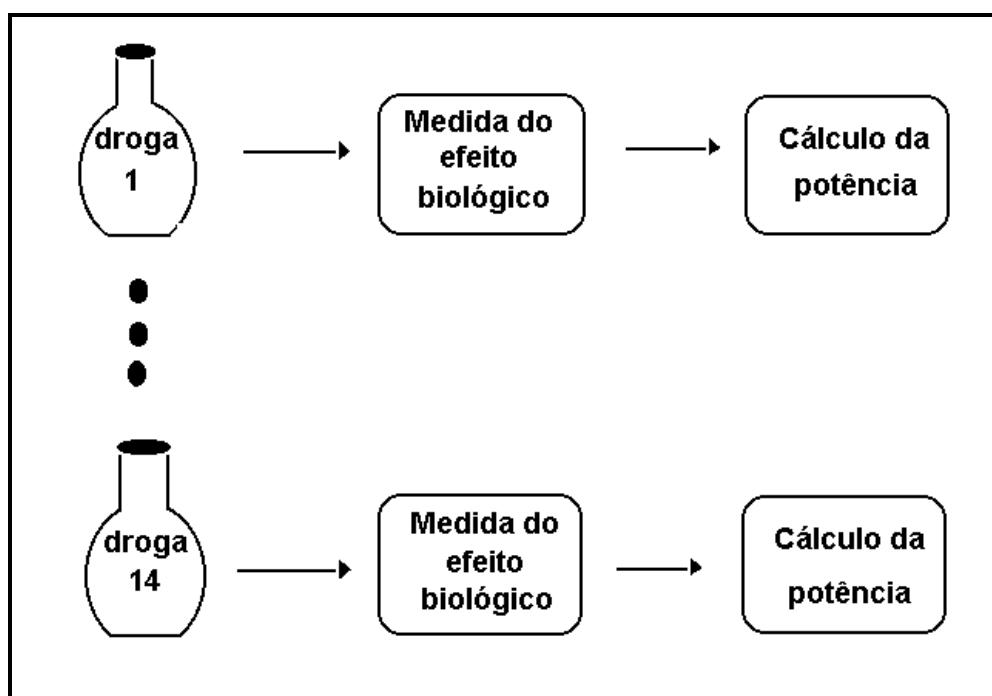


Os grupamentos utilizados foram:

- não substituído
- n-butil
- metil
- metóxi
- etil
- ciano
- n-hexil
- n-hexilóxi
- metilsulfonil
- CF₃
- n-butóxi
- n-propil
- acetil
- cloro

Tem-se, então, quatorze drogas com suas características físico-químicas próprias. Cada uma delas foi aplicada em grupos de três a oito ratos. Tomaram-se medidas dos seus efeitos biológicos e, posteriormente, determinou-se a potência da droga em cada rato. O processo de obtenção dos dados está esquematizado na Figura 3.2.. Cada uma de suas etapas é descrita a seguir.

Figura 3.2 - Processo de obtenção dos dados.



Medida do efeito biológico

O experimento a seguir, foi realizado, nas mesmas condições, para setenta e dois ratos.

O coração do rato foi retirado e o átrio esquerdo, isolado. Este átrio foi preso aos pólos de um eletrodo, para provocar choques, e colocado numa cuba com 15 ml de solução fisiológica. A cuba, por sua vez, foi colocada num recipiente maior com água aquecida para manter a temperatura do átrio em torno de 37°C, já que o mesmo é uma estrutura frágil.

O átrio também foi preso a um fio de algodão que está em contato com um transductor, ligado a um osciloscópio.

Quatro doses, em ml, da droga à qual o rato foi designado foram ministradas. Tanto para o controle (ausência da droga) quanto para cada uma das doses, provocaram-se dois choques no átrio, mantendo-se entre estes choques o tempo mínimo para que houvesse reação do átrio. A cada choque corresponde uma contração percebida pelo fio de algodão, transformada em energia elétrica pelo transductor e transmitida ao osciloscópio. O tempo, em segundos, entre as duas contrações foi medido. Este tempo foi transformado para freqüência (medida em Hertz).

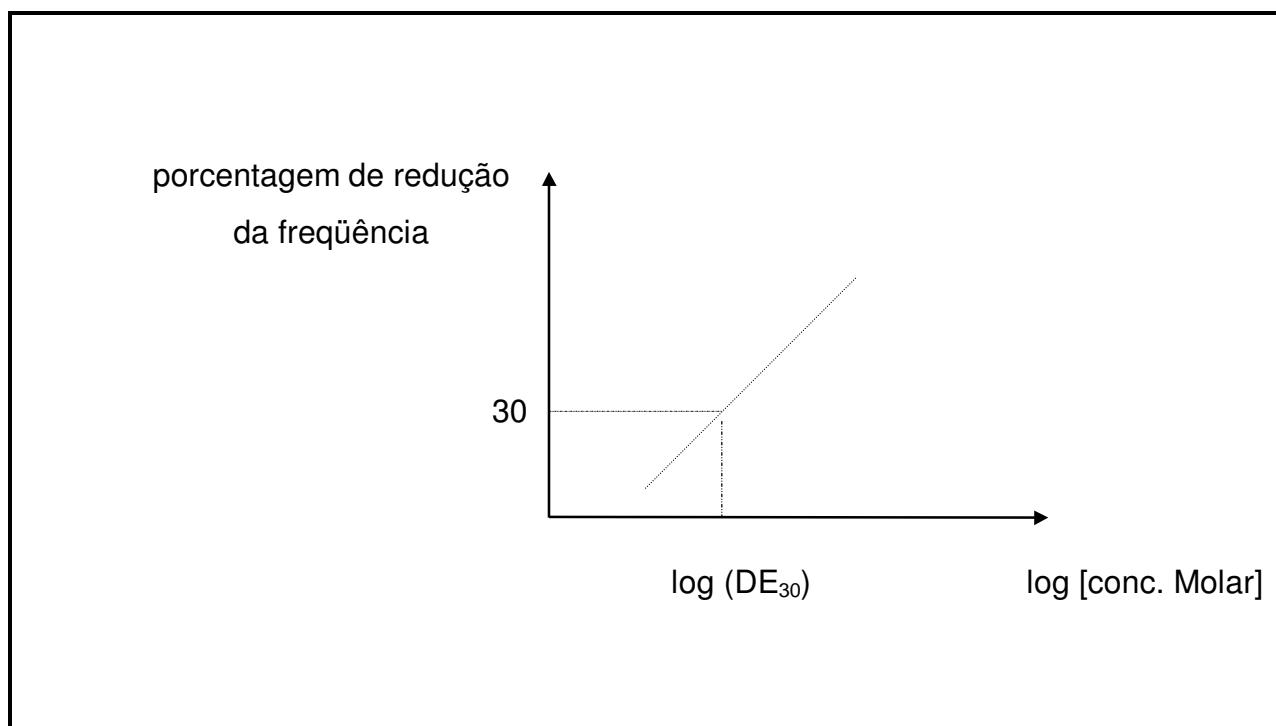
Ao fim do experimento, dispunha-se das medidas do tempo e freqüência para o controle e para as quatro doses ministradas a cada rato.

Cálculo da potência

Para a obtenção da potência de cada droga, para cada animal, calculou-se a porcentagem de redução da freqüência do átrio em relação ao controle, para cada uma das doses da droga.

Construiu-se o gráfico: porcentagem de redução da freqüência em função do logaritmo da concentração molar da droga, e ajustou-se uma reta aos dados. Desta reta, obteve-se o logaritmo da concentração molar ($\log DE_{30}$), como mostra a Figura 3.3, da droga relativo a 30% de redução da freqüência do átrio.

Figura 3.3 - Gráfico da porcentagem de redução da freqüência em função do logaritmo da concentração molar.



A potência relativa à droga, para cada rato, será o valor $-\log DE_{30}$.

Ao final de todos os experimentos, foram obtidos setenta e dois valores para a potência, divididos entre as quatorze drogas utilizadas, dependendo do número de ratos usados para cada droga.

4. DESCRIÇÃO DAS VARIÁVEIS

As dez variáveis independentes são características físico-químicas das próprias drogas, isto é, para uma dada droga, seus valores são determinados sem erro. Estão divididas em quatro grupos, de acordo com o efeito ao qual se referem.

As variáveis do grupo do *efeito estérico* medem um comprimento relacionado ao grupamento substituinte. São elas :

- **L** : comprimento do grupo substituinte ao longo do eixo da ligação com o esqueleto da molécula (medido em Ångstron);
- **B1 e B4** : larguras do grupo substituinte a partir do eixo da ligação, perpendiculares a ele (medidas em Ångstron).

As variáveis do grupo do *efeito eletrônico* medem o efeito de dispersão de carga em torno do anel benzênico. São elas :

- **F** : componente de campo (adimensional);
- **R** : componente de ressonância (adimensional);
- **SIGMA** : constante de Hammet - combinação linear das duas anteriores (adimensional);
- **C (CARBONILA)** : freqüência de estiramento da carbonila medida por espectroscopia no infra-vermelho (em cm^{-1}).

As variáveis do grupo do *efeito hidrofóbico* medem quanto as moléculas da droga misturam-se na água. São elas :

- **logPAPP** : log do coeficiente de partição óleo-água medido (adimensional);
- **PI** : coeficiente de partição óleo-água calculado (adimensional).

A variável do grupo de *outros efeitos* é :

- **MR4** : refratividade molar (adimensional).

A variável resposta é :

- **POTÊNCIA** : - log (DE_{30}), onde DE_{30} é a dose de droga necessária para ocorrer uma redução de 30% na freqüência do átrio em relação ao controle (adimensional).

5. ANÁLISE DESCRIPTIVA

A análise descritiva foi dividida em duas partes, uma para analisar se havia relação entre a variável potência e cada uma das dez variáveis independentes e a outra para verificar se existiam relações lineares entre as variáveis independentes, duas a duas.

Foram construídos gráficos de dispersão e calculados coeficientes de correlação de Pearson (Bussab & Morettin (1987)) para as variáveis envolvidas, utilizando-se o pacote computacional SPSS . Estas informações encontram-se nos Apêndices A e B.

5.1. RELAÇÕES ENTRE A POTÊNCIA E AS VARIÁVEIS INDEPENDENTES.

As variáveis B1, B4 e L parecem relacionar-se linearmente com a variável potência, sendo que B1 é a variável menos correlacionada com a potência, no grupo do efeito estérico. Os valores dos coeficientes de correlação linear de Pearson (Matriz B₁) de cada uma das variáveis deste grupo com a potência evidenciam os fatos citados acima.

Há indícios da existência de relação linear entre as variáveis potência e sigma, potência e R, potência e F, potência e C (Gráficos A₄, A₅, A₆, A₇,

respectivamente). Assim, no grupo do efeito eletrônico, percebe-se que esta relação linear é decrescente para todos os gráficos, ou seja, quanto maior o valor das variáveis independentes deste grupo, menor será a potência observada da droga.

Os fatos mencionados acima podem ser confirmados através dos valores dos coeficientes de correlação linear de Pearson contidos na matriz de correlação (Matriz B₁). Todas as correlações apresentam sinais negativos, sendo que a variável sigma é a mais correlacionada com a potência, no grupo do efeito eletrônico.

As variáveis logPAPP e PI apresentam uma forte relação linear positiva com a variável potência (Gráficos A₈ e A₉, respectivamente), ou seja, quando se aumenta o valor destas duas variáveis, a potência da droga aumenta.

As variáveis do grupo do efeito hidrofóbico possuem coeficientes de correlação linear de Pearson elevados e positivos (Matriz B₁) com a variável potência, indicando forte correlação das variáveis independentes deste grupo com a variável de interesse.

No Gráfico A₁₀, percebe-se a existência de relação linear entre as variáveis MR4 e potência. Esta correlação é positiva (Matriz B₁).

Conclui-se, pois, que todas as variáveis independentes consideradas no estudo parecem relacionar-se linearmente com a variável resposta (potência).

As variáveis do grupo do efeito hidrofóbico (logPAPP e PI) são as que possuem a mais forte relação com a variável potência, enquanto que as do grupo do efeito estérico são as menos correlacionadas com a variável resposta.

5.2. RELAÇÕES ENTRE AS VARIÁVEIS INDEPENDENTES.

A multicolinearidade, ou seja, a existência de relação linear entre as variáveis independentes é um grave problema para a análise inferencial,

Na própria descrição das variáveis independentes, percebe-se que algumas delas são correlacionadas naturalmente, como é o caso da variável sigma, que é combinação linear das variáveis F e R; e as variáveis logPAPP e PI, que são obtidas de formas diferentes, mas medem o mesmo fenômeno.

A matriz de correlação (Matriz B_1) contém os valores de todos os coeficientes de correlação linear de Pearson entre as variáveis independentes, duas a duas. As conclusões a seguir foram obtidas com base nesta matriz.

No grupo do efeito estérico, as variáveis B4 e L são as mais correlacionadas entre si (0,97); as demais, são pouco correlacionadas. No grupo do efeito eletrônico, a variável sigma é altamente correlacionada com F, R e C; e a correlação de R com C é elevada (0,94). No grupo do efeito hidrofóbico, logPAPP e PI têm correlação linear quase perfeita (0,99), como era esperado pela própria relação natural entre estas variáveis.

A variável MR4 é altamente correlacionada com B4 (0,98), L (0,99), logPAPP (0,78) e PI (0,76).

Há correlações altas, também, entre variáveis de grupos diferentes. Por exemplo, B1 e sigma (0,81), B4 e logPAPP (0,80), B4 e PI (0,77), L e logPAPP (0,79), L e PI (0,76).

Foi feito um dendrograma (Gráfico A₁₁) utilizando o método do centróide (Johnson & Wichern (1982)), de modo a agrupar as variáveis independentes altamente correlacionadas entre si segundo o coeficiente de correlação linear de Pearson. Realizando o corte, por exemplo, ao nível 7,5; foram obtidos 5 grupos: logPAPP e PI; L, MR4 e B4; R e C; B1 e Sigma; F. Esta última variável ficou sozinha, já que não é tão correlacionada com as demais variáveis.

Vale observar que os valores entre parênteses representam os coeficientes de correlação linear de Pearson.

6. ANÁLISE INFERENCIAL

6.1. MODELO E INTERPRETAÇÃO DOS PARÂMETROS

O objetivo é relacionar a variável Y, potência da droga, com as variáveis independentes descritas na Seção 4, e para isto, serão utilizados modelos de regressão linear múltipla (Montgomery & Peck (1992)).

De acordo com esta técnica, admitir-se-á que o valor da potência da droga à qual o i-ésimo rato foi submetido é dado por:

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j (x_{ij} - \bar{x}_j) + \varepsilon_i , \quad (6.1)$$

onde

x_{ij} é o valor da j-ésima variável independente associado à droga à qual o i-ésimo rato foi submetido;

\bar{x}_j é a média da j-ésima variável independente;

$\beta_0, \beta_1, \dots, \beta_k$ são os parâmetros do modelo;

ε_i são erros aleatórios, $j = 1, \dots, k$ e $i = 1, \dots, 72$.

Nestas condições, β_0 corresponde ao valor médio ou valor esperado da potência quando as k variáveis independentes assumem valores iguais às suas respectivas médias, simultaneamente, e β_j é a variação média ou variação esperada na potência quando aumenta-se $x_{ij} - \bar{x}_j$ em uma unidade, mantendo-se as demais variáveis independentes fixadas.

Admite-se, ainda, que as variáveis aleatórias ε_i são independentes com distribuição normal de média zero e variância σ^2 . A validade destas suposições será analisada posteriormente.

6.2 SELEÇÃO DE VARIÁVEIS INDEPENDENTES

Primeiramente, foram considerados três grupos de variáveis:

Grupo 1: todas as variáveis independentes consideradas no projeto;

Grupo 2: formado pelas variáveis B4, Sigma, logPAPP e MR4, que são as variáveis independentes mais correlacionadas linearmente com a variável potência em cada um dos grupos fornecidos pela pesquisadora (vide Seção 4);

Grupo 3: formado pelas variáveis B4, Sigma, F, R e logPAPP, que são as variáveis independentes mais correlacionadas linearmente com a variável potência dentro de cada um dos grupos obtidos a partir do dendrograma (vide Seção 5.2).

A seleção das variáveis independentes em cada grupo foi feita através do procedimento *stepwise* (Montgomery & Peck (1992)), com o auxílio do aplicativo BMDP-2R, fixando-se valores para a tolerância (vide apêndice C) de 0,3 e 0,4 para impedir a multicolinearidade nos modelos selecionados.

Nos grupos 1 e 2, foram selecionadas as variáveis:

Sigma e logPAPP para tolerância igual a 0,4;

Sigma, logPAPP e MR4 para tolerância igual a 0,3.

No grupo 3, selecionaram-se as variáveis Sigma e logPAPP para ambos os valores da tolerância.

Logo, os dois modelos selecionados foram:

1) modelo com as variáveis independentes Sigma e logPAPP

2) modelo com as variáveis independentes Sigma, logPAPP e MR4.

A seguir, apresenta-se uma análise completa de cada modelo selecionado, obtida com o auxílio do aplicativo MINITAB.

6.3. MODELO COM AS VARIÁVEIS INDEPENDENTES SIGMA E LOGPAPP

O modelo ajustado foi:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \text{Sigma}^* + \hat{\beta}_2 \log PAPP^* = \\ = 3,650 - 0,625 \text{Sigma}^* + 0,244 \log PAPP^* \quad (1)$$

A Tabela 6.1 apresenta a análise de variância associada ao modelo selecionado.

Tabela 6.1: Análise de Variância para o modelo com as variáveis Sigma e logPAPP

Fonte	g.l.	S.Q.	Q.M.	F	p
Rregressão	2	16,370	8,185	185,48	< 0,001
Resíduo	69	3,045	0,044		
Total	71	19,415			

A estatística F desta tabela é utilizada para testar:

$$H_0: \beta_1 = \beta_2 = 0; \quad (6.2)$$

contra

$$H_1: \text{pelo menos um } \beta_i \text{ é diferente de zero, } i = 1, 2.$$

O nível descritivo associado ao teste F, neste caso, foi menor que o nível de significância de 5%, de modo que rejeita-se a hipótese (6.2). Conclui-se, então, que pelo menos uma das variáveis presentes é importante na explicação da variável potência.

Na Tabela 6.2, apresentam-se os erros-padrão estimados dos estimadores dos parâmetros do modelo, bem como as estatísticas t para testar a hipótese:

¹ Obs.: * indica que as variáveis independentes foram centradas na média.

$$H_0: \beta_i = 0; \quad (6.3)$$

contra

$$H_1: \beta_i \neq 0, \quad i = 1, 2.$$

Estes testes verificam, isoladamente, a significância (importância) de cada uma das variáveis dado que a outra já está no modelo. Verifica-se que cada uma das variáveis é significante para o modelo, dado que as demais já foram incluídas ($p < 0,001$).

Tabela 6.2: Resultados inferenciais relativos ao modelo com as variáveis Sigma e logPAPP

Parâmetros	Coef.	e.p.	t	p	VIF
Constante	3,650	0,025	147,48	< 0,001	-
Sigma	-0,625	0,093	-6,73	< 0,001	1,8
logPAPP	0,244	0,027	8,93	< 0,001	1,8

Para avaliar o grau de multicolinearidade presente no modelo selecionado, além da tolerância fixada preliminarmente, foram analisados os fatores de inflação da variância (VIF) e os auto-valores da matriz contendo as correlações lineares de Pearson entre as variáveis Sigma e logPAPP (vide Apêndice C).

Na Tabela 6.2 são apresentados os fatores de inflação da variância relativos às variáveis Sigma e logPAPP. Ambos são menores que 5, donde conclui-se que existe baixa multicolinearidade no modelo de regressão linear múltipla com estas duas variáveis independentes. Conforme apresentado na Matriz B_1 , o coeficiente de correlação entre as variáveis Sigma e logPAPP é -0,668.

Finalizando, calculou-se L , o quociente entre o maior e o menor auto-valor da matriz de correlações, que resultou em 5,027. Como L é menor que 100, conclui-se

que não há série multicolinearidade neste modelo, de acordo com o critério apresentado no Apêndice C.

Portanto, com base nos três métodos para detecção de multicolinearidade, verificou-se que ocorre baixa multicolinearidade no modelo com as variáveis independentes Sigma e logPAPP.

Como a variável independente com maior valor absoluto da estatística t é logPAPP (Tabela 6.2), tem-se indicação de que esta seria a variável independente mais importante do modelo. Além disso, lembrando da equivalência entre esta estatística e a correspondente estatística F-parcial (Montgomery & Peck (1992)), conclui-se que logPAPP é a variável que produz a maior contribuição na explicação da Potência, dado que a outra variável independente está presente no modelo. É importante notar que, dentre todas as variáveis independentes envolvidas, logPAPP é a mais correlacionada com a resposta.

O coeficiente de explicação do modelo ajustado é $R^2 = 84,3\%$ e o coeficiente de explicação ajustado pelos graus de liberdade é $R^2(\text{aj.}) = 83,9\%$.

Foi empregada a distância de Cook (Montgomery & Peck (1992)) para verificar a eventual existência de pontos influentes. A maior distância encontrada foi 0,070 de sorte que nenhum ponto pode ser considerado influente.

As suposições de homocedasticidade (homogeneidade de variâncias), bem como de normalidade e independência dos erros foram verificadas com base em gráficos de resíduos e testes estatísticos.

No Gráfico D₁, observou-se que os resíduos padronizados distribuem-se numa estreita faixa em torno do zero, sugerindo que não há grandes desvios da suposição de homocedasticidade dos erros do modelo ajustado.

Quanto à normalidade, o Gráfico D₂ não sugere falha nesta suposição, já que os pontos estão dispostos ao redor da bissetriz do primeiro quadrante. Para reforçar esta conclusão, foi aplicado um teste equivalente ao de Shapiro-Wilk (Hahn & Shapiro (1967)), que não apontou acentuado desvio da normalidade ($p = 0,079$).

Para avaliar a independência dos resíduos, foi feito um teste de aleatoriedade (Draper & Smith (1981)), concluindo-se que ocorreu falha desta suposição ($p = 0,004$).

6.4. MODELO COM AS VARIÁVEIS INDEPENDENTES SIGMA, LOGPAPP E MR4

O modelo ajustado foi:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \text{Sigma}^* + \hat{\beta}_2 \log PAPP^* + \hat{\beta}_3 \text{MR4}^* = \\ = 3,650 - 0,614 \text{Sigma}^* + 0,318 \log PAPP^* - 0,124 \text{MR4}^* \quad (2)$$

A Tabela 6.3 apresenta a análise de variância associada ao modelo selecionado.

Tabela 6.3: Análise de Variância associada ao modelo com as variáveis Sigma, logPAPP e MR4

Fonte	g.l.	S.Q.	Q.M.	F	p
Rregressão	3	16,728	5,576	141,09	< 0,001
Resíduo	68	2,687	0,040		
Total	71	19,415			

A estatística F da Tabela 6.3 é utilizada para testar:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0; \quad (6.4)$$

contra

$$H_1: \text{pelo menos um } \beta_i \text{ é diferente de zero, } i = 1, 2, 3.$$

² Obs.: * indica que as variáveis independentes foram centradas na média.

O nível descritivo associado ao teste F, neste caso, foi menor que o nível de significância de 5%, de modo que rejeita-se a hipótese (6.4). Conclui-se, então, que pelo menos uma das variáveis presentes é importante na explicação da variável potência.

Na Tabela 6.4, apresentam-se os erros-padrão estimados dos estimadores dos parâmetros do modelo, bem como as estatísticas t para testar a hipótese:

$$H_0: \beta_i = 0; \quad (6.5)$$

contra

$$H_1: \beta_i \neq 0, \quad i = 1, 2, 3.$$

Estes testes verificam, isoladamente, a significância (importância) de cada uma das variáveis dado que as outras já estão no modelo. Verifica-se que cada uma das variáveis é significante para o modelo, dado que as demais já foram incluídas ($p \leq 0,004$).

Tabela 6.4: Resultados inferenciais relativos ao modelo com as variáveis Sigma, logPAPP e MR4

Parâmetros	Coef.	e.p.	t	p	VIF
Constante	3,650	0,023	155,84	< 0,001	-
Sigma	-0,614	0,088	-6,98	< 0,001	1,8
logPAPP	0,318	0,036	8,91	< 0,001	3,4
MR4	-0,124	0,041	-3,01	0,004	2,5

Para avaliar o grau de multicolinearidade presente no modelo selecionado, além da tolerância fixada preliminarmente, foram analisados os fatores de inflação da

variância (VIF) e os auto-valores da matriz contendo as correlações lineares de Pearson entre as variáveis Sigma, logPAPP e MR4 (vide Apêndice C).

Na Tabela 6.4 são apresentados os fatores de inflação da variância relativos às variáveis Sigma, logPAPP e MR4. Todos são menores que 5, donde conclui-se que existe baixa multicolinearidade no modelo de regressão linear com essas três variáveis independentes.

Finalizando, calculou-se L, o quociente entre o maior e o menor auto-valor da matriz de correlações, que resultou em 12,461. Como L é menor que 100, conclui-se que não há séria multicolinearidade neste modelo, de acordo com o critério apresentado no Apêndice C.

Portanto, com base nos três métodos para detecção da multicolinearidade, verificou-se que ocorre baixa multicolinearidade no modelo com as variáveis independentes Sigma, logPAPP e MR4.

Como a variável independente com maior valor absoluto da estatística t é logPAPP (Tabela 6.4), tem-se indicação de que esta seria a variável independente mais importante do modelo. Além disso, lembrando da equivalência entre esta estatística e a correspondente estatística F-parcial, conclui-se que logPAPP é a variável que produz a maior contribuição na explicação da Potência, dado que as outras variáveis independentes estão presentes no modelo. É importante notar que, dentre todas as variáveis independentes envolvidas, logPAPP é a mais correlacionada com a resposta.

O coeficiente de explicação do modelo ajustado é $R^2 = 86,2\%$ e o coeficiente de explicação ajustado pelos graus de liberdade é $R^2(\text{aj.}) = 85,5\%$.

Novamente, a distância de Cook foi empregada na tentativa de detecção de pontos influentes. A maior distância encontrada foi 0,064 e, assim, nenhum ponto pode ser considerado influente.

As suposições de homocedasticidade (homogeneidade de variâncias), bem como de normalidade e independência dos erros foram verificadas com base em gráficos de resíduos e testes estatísticos.

O Gráfico D₄, sugere a existência de igualdade das variâncias residuais, uma vez que os resíduos padronizados distribuem-se numa estreita faixa em torno do zero.

O Gráfico D₅ não sugere falha na suposição de normalidade, já que os pontos estão dispostos ao redor da bissetriz do primeiro quadrante. Para reforçar esta conclusão, foi aplicado o mesmo teste da Seção 6.3, que também não apontou acentuado desvio da normalidade ($p > 0,100$).

Para avaliar a independência dos resíduos, foi feito o teste de aleatoriedade. Concluiu-se que não ocorreu falha desta suposição ($p = 0,156$).

7. CONCLUSÕES

O modelo com as variáveis independentes Sigma e logPAPP apresenta baixa multicolinearidade e coeficiente de explicação igual a 83,4%. Quanto às suposições de homocedasticidade, normalidade e independência dos resíduos, apenas a última não foi verificada.

Por outro lado, no que tange ao modelo que tem como variáveis independentes Sigma, logPAPP e MR4, observou-se a existência de baixa multicolinearidade, embora um pouco maior que a do modelo citado no parágrafo anterior. Isto é natural, pois a introdução de uma nova variável no modelo tende a aumentar a multicolinearidade. Seu coeficiente de explicação foi 86,2%. Todas as suposições relativas ao modelo de regressão linear foram satisfeitas.

Estatisticamente, o modelo mais indicado é aquele com as variáveis Sigma, logPAPP e MR4, uma vez que a introdução da variável MR4 contribuiu significativamente para a explicação da potência das drogas em estudo.

A variável independente mais importante, em ambos os modelos considerados, é logPAPP, haja vista que esta apresentou maior contribuição na explicação da potência, quando as demais variáveis independentes consideradas

estão presentes no modelo. Além disto, esta variável é aquela que, isoladamente, é a mais correlacionada com a potência.

Finalizando, destaca-se que a conclusão sobre a baixa multicolinearidade nos modelos selecionados foi tirada com base na amostra observada. Desta forma, para um outro tipo de planejamento, ou outro conjunto de dados, uma nova análise deverá ser realizada.