

**Universidade de São Paulo
Instituto de Matemática e Estatística**

Centro de Estatística Aplicada

Relatório de Análise Estatística

RAE-CEA-20P21

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

**Interferência do Processo de Coagulação na Remoção de Compostos Orgânicos
Causadores de Gosto e Odor por Processos de Adsorção em Carvão Ativado em
Pó**

Guilherme Camargo Matos de Paula

Luís Gustavo Esteves

Pedro Paulo Sousa Argôlo da Silva

São Paulo, outubro de 2020

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “Interferência do Processo de Coagulação na Remoção de Compostos Orgânicos Causadores de Gosto e Odor por Processos de Adsorção em Carvão Ativado em Pó”.

PESQUISADOR(A): André Luiz Marguti

ORIENTADOR(A): Prof. Dr. Sidney Seckler Ferreira Filho

INSTITUIÇÃO: Escola Politécnica (POLI-USP)

FINALIDADE DO PROJETO: Doutorado

RESPONSÁVEIS PELA ANÁLISE: Guilherme Camargo Matos de Paula

Luís Gustavo Esteves

Pedro Paulo Sousa Argôlo da Silva

REFERÊNCIA DESTE TRABALHO: PAULA, G.C.M.; ESTEVES, L.G.; SILVA, P.P.S.A
Relatório de análise estatística sobre o projeto: “Interferência do Processo de Coagulação na Remoção de Compostos Orgânicos Causadores de Gosto e Odor por Processos de Adsorção em Carvão Ativado em Pó”. São Paulo, IME-USP, 2020. (RAE–CEA-20P21)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

MARGUTI, A.L. (2020). **Interferência do Processo de Coagulação na Remoção de Compostos Orgânicos Causadores de Gosto e Odor por Processos de Adsorção em Carvão Ativado em Pó**. São Paulo. 122p. Tese (Doutorado). Escola Politécnica - USP.

NASCIMENTO, G.J.C. e LUDOVICE, N.Z. (2020). **Interferência do Processo de Coagulação na Remoção de Compostos Orgânicos Causadores de Gosto e Odor por Processos de Adsorção em Carvão Ativado em Pó**. Relatório de Consulta do CEA. São Paulo, IME-USP, 10p.

FOGAÇA, J.R.V. “**Coagulação e Floculação**”; *Brasil Escola*. Disponível em: <<https://brasilescola.uol.com.br/quimica/coagulacao-floculacao.htm>>. Acesso em 19 de outubro de 2020.

MAGALHÃES, L. **Adsorção**; *Toda Matéria*. Disponível em: <<https://www.todamateria.com.br/adsorcao/>>. Acesso em 11 de dezembro de 2020.

UFRRJ. “**Turbidez**”. Disponível em: <<http://www.ufrrj.br/institutos/it/de/acidentes/turb.htm>>. Acesso em 19 de outubro de 2020.

GOMEZ, G.; CALLE, M.L.; OLLER, R.; LANGOHR, K. (2009) Tutorial on methods for interval-censored data and their implementation in R. **Statistical Modelling**; **9(4)**, 259-297.

COLOSIMO, E.A. e GIOLO, S.R. **Análise de sobrevivência aplicada**. 1a ed. Editora Blucher, ABE - Projeto Fisher, 392p.

SINGER, J.M.; NOBRE, J.S.; ROCHA, F.M.M. (2018). **Análise de dados longitudinais**. 306p.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Microsoft Word for Windows (versão 2019)

Microsoft Excel for Windows (versão 2016)

R (versão 4.0.3)

R Studio (versão 1.1.442)

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Unidimensional (03:010)

Análise Descritiva Multidimensional (03:020)

Análise de Sobrevida (13:070)

Outros (07:990)

ÁREA DE APLICAÇÃO

Ciências do Ambiente e Ecologia (14:060)

Resumo

O presente relatório trata da análise realizada sobre dados de um experimento que visa medir a influência do carvão ativado em pó (CAP), do coagulante utilizado e da turbidez prévia da água no processo de remoção de compostos orgânicos causadores de gosto e odor presentes nas águas poluídas. Os dados referentes à concentração ao longo do tempo do 2-metilisoborneol (MIB), partícula orgânica da água avaliada no estudo, depois da adição de diferentes níveis de CAP, de coagulante e da turbidez fixada *a priori* foram obtidos através de laboratórios. No total, temos 39 ensaios, cujos dados foram descritos e analisados através de duas modelagens: 1) uma regressão paramétrica com distribuição weibull para dados de sobrevivência com censura intervalar; 2) uma regressão para dados longitudinais. A primeira técnica evidenciou a influência do carvão e a presença de uma interação entre coagulante e turbidez na purificação total da água. Essa técnica nos permitiu estimar a probabilidade da água ficar totalmente limpa (dentro da sensibilidade dos aparelhos de medição) de acordo com os níveis fixados das três covariáveis e em função do tempo. A segunda técnica mostrou que o principal fator no processo de purificação da água é o CAP e o modelo por ela obtido serve como ferramenta de previsão para outros estudos com o carvão ativo.

Sumário

1. Introdução.....	8
2. Objetivo(s).....	9
3. Descrição do estudo.....	9
4. Descrição das variáveis	10
5. Análise descritiva	10
6. Analise Inferencial	12
7. Conclusão.....	20
APÊNDICE A	211
APÊNDICE B	25

1. Introdução

A questão da poluição das águas é de extrema importância no mundo atual. Com ela surgem cada vez mais pesquisas com o intuito de estudar como melhor tratar as águas nas estações de tratamento de água. Uma questão de extrema importância é a necessidade que a água tratada não apresente gosto nem odor.

Estudos mostram que “uma das principais causas de gosto e odor em águas de abastecimento tem como origem a presença de compostos orgânicos de origem biogênica, produzidos por cianobactérias e demais micro-organismos, sendo que os mais comumente encontrados em águas naturais são a geosmina e 2-metilisoborneol” (Marguti, 2020).

Alguns dos métodos mais comuns para tratamento de águas de abastecimento envolve o uso do Carvão ativado em pó (CAP). E é comum também o uso de coagulantes de modo a desestabilizar partículas coloidais de impureza, de modo a que formem flocos e possam ser mais facilmente retiradas da água (Fogaça, 2020); um desses coagulantes é o sulfato de alumínio.

Ao processo pelo qual o carvão é responsável dá-se o nome de adsorção pois, diferente da absorção, onde uma substância é embebida pela outra, ocasionando mudança de volume, na adsorção, por outro lado, uma substância somente fica retida na superfície da outra, sem fazer parte do seu volume (Magalhães, 2020).

A grande maioria dos estudos que avaliavam o efeito do coagulante no tratamento das águas não considerou a concentração de partículas de impureza na água bruta, mais especificamente, os seus valores de turbidez (Marguti, 2020). A turbidez pode ser definida como a medida da dificuldade de um feixe de luz atravessar uma certa quantidade de água, conferindo uma aparência turva à mesma. Essa medição é feita com o *turbidímetro* (UFRRJ, 2020).

2. Objetivo(s)

Avaliar a interferência da dosagem de coagulante e do nível de turbidez no comportamento do processo de adsorção em carvão ativado em pó (CAP), que é utilizado para a remoção de compostos causadores de gosto e odor.

Mais especificamente, temos os seguintes objetivos:

- Avaliar a influência do hidróxido metálico (CAP) na eficiência de remoção de compostos orgânicos causadores de gosto e odor em águas de abastecimento por processos de adsorção.
- Estudar a cinética do processo de adsorção, tendo em vista a influência das características da água bruta (turbidez) e da dosagem de coagulante na eficiência da remoção de compostos orgânicos causadores de gosto e odor.

3. Descrição do estudo

O estudo foi conduzido considerando águas com diferentes níveis de turbidez. Para cada nível, foram aplicadas combinações de níveis de coagulante e CAP. As únicas combinações que não foram consideradas são aquelas onde a dose de coagulante é zero e o nível da turbidez é maior que zero. O composto usado como coagulante foi o sulfato de alumínio. Avaliou-se, então, em diferentes instantes do tempo a quantidade de 2-metilisoborneol (MIB) restante na água, através da coleta de amostras.

Definiu-se uma medida que foi chamada eficiência, que consiste no percentual de MIB retirado da água no intervalo de tempo até a coleta daquela amostra. É de importância ímpar avaliar a eficiência aos 10 minutos.

Foi considerado que todo valor de MIB inferior a 4 ng/L é desprezível e, portanto, considerado como zero. Isso se deve a limitações do aparelho medidor.

4. Descrição das variáveis

Variáveis que caracterizam a amostra:

- MIB remanescente ao longo do tempo (ng/L);
- Turbidez na água: 0, 5, 20 ou 100 UNT;
- Coagulante: 0, 10, 20 ou 50 mg/L;
- Carvão ativado em pó (CAP): 10, 20 ou 40 mg/L;
- Tempo de coleta das amostras: 0, 5, 10, 20, 30, 45 ou 60 min.

5. Análise descritiva

Nesta seção, apresentamos a análise descritiva dos dados, que nos permite ter uma visão inicial dos resultados do estudo.

Na Tabela A.1, tem-se primeiramente as medidas descritivas das concentrações de MIB em função do tempo. Observa-se, através da mediana e média, um decaimento da concentração ao longo do tempo. Através da análise da Figura B.3, este comportamento de decaimento é novamente observado.

Em sequência, a Tabela A.2 mostra o comportamento das razões entre cada concentração de MIB ao longo do tempo e a respectiva concentração inicial. O comportamento do decaimento ao longo do tempo é novamente evidenciado. A Figura B.4 mostra o mesmo decaimento.

Finalizada a primeira análise, o próximo objetivo é avaliar essas mesmas variáveis segundo as variáveis explicativas na base de dados: turbidez, dosagem de coagulante e carvão ativado em pó (CAP). Nas Figuras B.5 à B.7, são apresentadas as curvas das concentrações ao longo do tempo, por cada nível de fator.

Na Figura B.5, observa-se que o nível de turbidez da água aparentemente não influencia as curvas de decaimento das concentrações. Na Figura B.6, nota-se uma pequena diferença entre as curvas nos instantes finais, o que sugere que os níveis de coagulante têm pouca ou nenhuma influência na concentração. Por fim, na Figura B.7, as curvas aparentam ser diferentes, do que se conclui que o nível de CAP influencia o decaimento das concentrações de MIB.

A próxima etapa da análise descritiva concentra-se na variável eficiência após 10 minutos.

Na Tabela A.3 e na Figura B.8 têm-se as medidas descritivas da eficiência após 10 minutos. A média é 84% e a mediana é 87%. Vemos que os dados apresentam assimetria à esquerda.

A próxima fase corresponde à análise da eficiência após 10 minutos considerando turbidez, dosagem de coagulante e CAP a fim de se determinar possíveis influências e interação entre esses fatores na eficiência.

Na Tabela A.4 e na Figura B.9, é possível notar que o nível de turbidez tem pouca influência na eficiência após 10 minutos. Isso pode ser visto através das medianas e médias, que mostram poucas diferenças.

No caso da Tabela A.5 e da Figura B.10, nota-se que em relação à eficiência, a dosagem alta de coagulante apresenta um pequeno acréscimo na mediana; desta forma, reforça-se a tese de que a dosagem de coagulante tem pouca influência no processo de decaimento do MIB.

Por fim, através da análise da Tabela A.6 e da Figura B.11, verifica-se que a eficiência após 10 minutos cresce com o aumento do nível do CAP. Ou seja, há indícios de influência do carvão ativo na eficiência após 10 minutos.

A Figura B.12 mostra a eficiência após 10 minutos em função da dosagem de coagulante e dosagem de CAP. Podemos afirmar que, descritivamente, não há indícios de interação, dado o aparente paralelismo das curvas. Sobre cada ponto está uma barra com um erro padrão para cima e para baixo. Nota-se que o tamanho dos erros padrões não parece mudar para os diferentes valores das variáveis. Os pontos sem barras são pontos onde só há uma unidade amostral e, portanto, não é possível calcular o erro padrão.

Na Figura B.13, o gráfico de perfis sugere que não há indícios de interação entre turbidez da água e CAP. Vê-se no gráfico que os erros padrões para a dosagem de

CAP igual a quarenta são menores que os erros padrões dos outros níveis de CAP quando a turbidez é 5, 20 e 100 UNT.

Por fim a Figura B.14 mostra que as linhas do gráfico de perfis se cruzam em diversos pontos, mostrando um indício de interação entre as variáveis turbidez e coagulante. Entretanto, devido aos erros padrões, não fica totalmente claro se há, de fato, interação entre tais variáveis. Esse gráfico dá algum embasamento à restrição explicada na Descrição do Estudo: que só há a combinação da turbidez com a dosagem de coagulante zero quando a turbidez não é zero. Porém, como citado acima, devido à sobreposição das barras de erro padrão, a explicação não é evidente. Também é possível perceber que o erro padrão é mais alto do que nos outros dois gráficos de interação anteriormente apresentados.

6. Análise inferencial

Nesta seção, apresenta-se as duas modelagens realizadas de maneira a relacionar a concentração de MIB ao longo do tempo com as covariáveis já apresentadas: dosagem de carvão ativado em pó (CAP), dosagem de coagulante e nível de turbidez da água.

O primeiro modelo trata-se de uma análise de sobrevivência (Colosimo e Giolo, 2006) e tem como interesse avaliar o tempo de limpeza total da água em função das variáveis explicativas. Tal modelo permite, por exemplo, obter probabilidades estimadas de limpeza total da água em cada instante de tempo.

O segundo modelo permite estimar a concentração absoluta de impureza (MIB) presente na água em cada instante de tempo, dados os valores das variáveis explicativas. O modelo utilizado nesse caso é uma regressão com dados longitudinais (Singer et al., 2018).

Note que os modelos não têm o mesmo objetivo (fornecem estimativas para quantidades distintas), portanto, o intuito é que juntos eles possam entregar um melhor

resultado geral para o pesquisador, respondendo a mais perguntas que possam ser-lhe de interesse. O restante dessa seção está dividido em duas subseções.

6.1 Modelo paramétrico de sobrevivência com censura intervalar

A análise de sobrevivência é uma área da regressão que costuma ter como variável resposta o tempo até a ocorrência de um evento de interesse, por exemplo, a morte de um paciente ou a cura de uma doença. A sobrevivência no instante t define-se como a probabilidade do evento de interesse ocorrer depois desse instante t . Uma característica de importância ímpar dos modelos de sobrevivência são os dados censurados. Diz-se que um indivíduo tem seu dado censurado quando se tem uma observação parcial, por exemplo, quando se sabe que um evento ocorreu depois ou antes de um determinado instante no tempo, mas não se tem o instante preciso da sua ocorrência.

Pode-se dividir as censuras em três tipos: à direita, à esquerda e intervalar. A primeira ocorre quando se sabe que o evento ocorrerá depois de um dado instante no tempo. A segunda ocorre quando se sabe que o evento ocorreu antes de um dado instante no tempo. A terceira censura ocorre quando se sabe que o evento ocorreu entre dois instantes, digamos I_i e I_s .

Um caso especial da censura intervalar são os dados grupados. Isso ocorre quando todos os indivíduos da amostra têm os mesmos instantes de avaliação.

Neste trabalho, o evento de interesse, como já dito, é a limpeza total da água. Define-se aqui limpeza total quando a água contém menos que 4 ng/L de MIB. Tal definição é necessária em virtude de limitações do equipamento utilizado para a medição da concentração de MIB na água.

Os nossos indivíduos são os 39 ensaios realizados. Nossos dados apresentam censura à direita, ou seja, há ensaios que, após todos os instantes de avaliação, ainda contém concentração de MIB maior ou igual a 4 ng/L. Nossos dados também apresentam censura intervalar e, em especial, tratam-se de dados grupados, uma vez

que todos os ensaios foram avaliados nos mesmos instantes: 0, 5, 10, 20, 30, 45 e 60 minutos após a aplicação do CAP e/ou do coagulante.

Uma vez que um modelo de sobrevivência foi ajustado, apresentam-se gráficos das sobrevivências estimadas pelo método de Kaplan Meier para cada nível de cada covariável, a fim de entender *a priori* como as sobrevivências podem se comportar de acordo com as covariáveis. Os gráficos estão presentes nas Figuras B.15, B.16 e B.17.

É possível perceber que a sobrevivência estimada aparenta ser diferente para os três níveis de CAP presentes no estudo, sendo menor conforme é maior a concentração de carvão ativado em pó. Como já definido acima, a sobrevivência no instante t é a probabilidade do evento de interesse não ocorrer até o instante t , o que no nosso estudo pode ser interpretado como a probabilidade da água não estar totalmente limpa até o instante t ou a probabilidade da água estar totalmente limpa somente após esse instante t . Logo, pela análise dos gráficos, quanto maior a concentração do carvão, menor a sobrevivência e, portanto, maior a probabilidade da água ficar limpa num dado instante t . Note que esta afirmação é válida para qualquer instante t , uma vez que no gráfico, quando uma curva de sobrevivência torna-se menor que a outra, mantém-se menor que ela até o fim do período de experimento.

O gráfico da Figura B.16 nos mostra que as curvas de sobrevivência estimadas para as diferentes dosagens de coagulante sobrepõem-se e cruzam-se umas às outras, de modo que é difícil afirmar *a priori* com exatidão alguma relação existente entre a limpeza da água e as dosagens de coagulante. A mesma conclusão é obtida ao examinar-se o gráfico presente na Figura B.17, que estima as curvas de sobrevivência para os diferentes níveis de turbidez.

Nesta análise, foi utilizado um modelo de regressão linear para dados de sobrevivência onde supõe-se que o tempo tenha distribuição Weibull (Colosimo e Giolo, 2006), que é uma distribuição comumente utilizada para ajuste de modelos de sobrevivência. Inicialmente, por razões experimentais, ajustou-se um modelo com as covariáveis dispostas como fatores, uma vez que no estudo foram pré-definidos níveis de cada uma das covariáveis para cada um dos ensaios. Deseja-se avaliar num

primeiro momento se é razoável supor que as covariáveis tenham comportamento linear e possam, portanto, serem ajustadas como variáveis contínuas. Uma questão é que o modelo não aceita nenhum limite inferior igual a zero, mas temos uma observação cuja limpeza total se deu entre 0 e 5 minutos (limite inferior igual a zero). Para contornar esse problema, consideramos o tempo inicial como 0,01 minutos.

Na Tabela A.7 estão presentes os resultados do ajuste desse modelo inicial. Vemos que há um problema com a estimação do erro padrão do parâmetro correspondente à covariável CAP (40 mg/L), o que pode tornar a sua estimativa pouco confiável. Entretanto, vê-se que as estimativas dos parâmetros diminuem conforme aumenta a concentração de CAP, o mesmo ocorre para os níveis de coagulante. Entretanto, isso não ocorre com os níveis da covariável turbidez. O nível de cada covariável que não figura com parâmetro estimado no modelo é aquele que foi considerado como categoria de referência. Assim, decide-se que é razoável supor que a covariável coagulante tenha comportamento linear. O mesmo é sugerido pelas estimativas dos parâmetros referentes ao CAP. Porém, devido ao mal ajuste de um deles, foi feito um gráfico para analisar se há indícios da relação entre tempo até limpeza total da água e CAP terem relação linear.

O gráfico está presente na Figura B.18. Trata-se de um gráfico de dispersão entre o l_i (limite inferior do intervalo de tempo onde cada evento ocorre, que corresponde ao instante da última avaliação para os indivíduos com censura à direita) e os níveis da covariável CAP. Através desse gráfico, vemos que, ao menos graficamente, a relação entre essas duas variáveis parece ser linear. Assim, suporemos que tempo até limpeza total da água e CAP têm relação linear.

A fim de não poluir o relatório com muitas informações, não serão apresentados os demais modelos intermediários, mas será somente comentado o procedimento utilizado em cada etapa do processo até chegarmos ao modelo final, o qual será apresentado.

Em seguida, ajustou-se o modelo com as covariáveis CAP e coagulante como contínuas, mas com a variável turbidez ainda como fator. Além disso, adicionou-se ao

modelo as três interações duas a duas entre as covariáveis principais. Deseja-se testar a hipótese de ausência de influência das covariáveis a 5% de nível de significância. Se algum dos níveis de uma covariável tiver parâmetro significativo, os demais níveis serão mantidos no estudo. Foram então excluídas do modelo as interações entre CAP e turbidez e entre CAP e coagulante. Apesar das covariáveis turbidez e coagulante não aparentarem ser significativas, serão mantidas no estudo devido à presença da sua interação.

O novo modelo ajustado, com todas as covariáveis principais (onde a turbidez ainda é fator) e a interação entre turbidez e coagulante, apresentou significância para CAP, dois dos três níveis de turbidez e para um dos três níveis da interação. Agora, a estimativa dos parâmetros referentes à turbidez é de 0,2, 1,99 e 2,03, para os níveis referentes a 5, 20 e 100 UNT de turbidez, respectivamente, evidenciando crescimento das estimativas em função da turbidez. Como o comportamento é crescente, avaliaremos as estimativas dos parâmetros referentes aos níveis da interação, para verificar se é possível afirmar que não é absurdo ajustar um modelo com a turbidez como variável contínua. As estimativas dos parâmetros relativos à interação são 0,01, -0,04 e -0,05, para coagulante contra 5, 20 e 100 UNT de turbidez, respectivamente. O comportamento não é totalmente linear, devido ao crescimento do primeiro parâmetro em relação à categoria de referência, porém como o primeiro parâmetro estimado é muito baixo, vamos considerar que tal quebra na linearidade não é grande e que, portanto, podemos supor que a variável turbidez tenha relação linear com a variável resposta.

Finalmente, ajustamos o modelo final, com a presença das três covariáveis principais e da interação entre turbidez e coagulante. O ajuste do modelo encontra-se na tabela A.8.

Podemos interpretar os parâmetros em função do risco relativo (RR). O risco não é uma probabilidade, logo não está restrito ao intervalo (0,1), mas é uma quantidade que é diretamente proporcional à probabilidade de ocorrência do evento de interesse, no caso, a limpeza total da água. Assim, a quantidade RR mede o quanto aumenta o risco da água ficar totalmente limpa, dados dois casos que estão sendo comparados.

Por exemplo, podemos interpretar o parâmetro β_1 , referente ao CAP, ao calcular o risco relativo entre uma dosagem x qualquer de CAP, entre 10 e 40 mg/L (limites do estudo), e $x + 1$. Tal valor é calculado como $\exp(-\beta_1/\alpha) = 1,20601$, onde α é a escala estimada. Ou seja, podemos dizer que o risco cresce 20% sempre que aumentamos em uma unidade a dosagem de CAP. E, portanto, quanto maior o CAP, maior a probabilidade de eliminarmos totalmente o MIB da água. Note que tal análise não depende do tempo, ou seja, o risco cresce igualmente quando o CAP é aumentado, independentemente do instante de tempo.

A interpretação dos outros dois parâmetros relativos às covariáveis principais não pode ser feita dessa forma devido à presença de interação. Assim, mostramos os gráficos presentes nas Figuras B.19 e B.20. Devido à interação, o risco relativo depende da outra variável. Assim, por exemplo, temos que em relação ao RR para duas dosagens de coagulante, ao aumentar uma unidade de coagulante, quando a turbidez é mínima, o risco cresce cerca de 3%, e quando a turbidez é máxima (100 UNT, dentro dos limites do estudo), cresce cerca de 19%.

Já em relação ao RR entre dois níveis de turbidez, temos que ao aumentar 1 UNT de turbidez, quando a dosagem de coagulante é mínima, o risco diminui cerca de 5%, ao passo que quando a dosagem de coagulante é máxima (50 mg/L, nos limites do modelo), o risco aumenta cerca de 2%.

As Figuras B.21 e B.22 apresentam gráficos de resíduos, com o objetivo de avaliar a qualidade do ajuste do modelo. Vemos que o modelo está bem ajustado. O gráfico de resíduos padronizados, que para o nosso modelo corresponde ao gráfico de resíduos de Cox-Snell, mostra-nos a qualidade do ajuste geral do modelo, nele os pontos devem estar no entorno da reta $y=x$. Isso parece ocorrer. Esse gráfico foi feito utilizando o início e o fim do intervalo como ocorrência do evento, calculando os seus resíduos, e dividindo por dois. Além disso, podemos observar com o gráfico na direita que a curva de sobrevivência estimada dos resíduos aproxima-se de uma *weibull* padrão, conforme é suposto neste modelo. Já no gráfico de resíduos *deviance*, avalia-se se existe algum ponto atípico. Espera-se que os resíduos estejam entre -2 e 2. Isso igualmente parece estar satisfeito nesse gráfico.

6.2 Modelo de regressão para dados longitudinais

A análise de dados longitudinais consiste em analisar medidas repetidas ao longo de uma condição de avaliação, por exemplo, quando o objetivo do estudo é analisar o comportamento do peso de bebês durante os 6 meses de vida, ou, a quantidade de absorção de água em um determinado tecido ao longo de 4 instantes de tempo.

Como a condição de avaliação é feita poucas vezes em mais de uma unidade amostral, a técnica de análise de séries temporais se mostra ineficiente neste tipo de contexto.

Desta forma, no presente projeto, o interesse em prever a quantidade de MIB suspensa na água ao longo de um determinado instante de tempo, caracteriza-se em um possível problema de análise de dados longitudinais.

As covariáveis utilizadas neste contexto serão: dosagem de CAP, nível de turbidez da água, dosagem de coagulante, concentração de MIB no instante inicial e o tempo. A variável resposta será a concentração de MIB.

Inicialmente, foi ajustado um modelo com interação dos fatores de dosagem de coagulante e nível de turbidez. Mas o algoritmo não conseguiu estimar os parâmetros e, portanto, a interação destes fatores foi descartada.

Após a retirada das interações, o algoritmo retornou os resultados mostrados na Tabela A.9. Analisando a tabela, tem-se que nível de turbidez e dosagem de coagulante não se mostram estatisticamente significantes ao nível de 5% para a maioria dos níveis das variáveis e, portanto, foram removidas para o ajuste final. A variável MIB inicial apresentou um valor-p um pouco maior que 5%, mas foi mantido a fim de se obter uma interpretação melhor do modelo.

O ajuste final do modelo, cujos resultados estão na Tabela A.10, mostrou que todos os coeficientes são estatisticamente significantes ao nível de 5%.

A quantidade de CAP 20 mg/L faz que o MIB decaia, em média, 10,133 ng/L em relação ao CAP 10 mg/L e o CAP 40 mg/L faz que o MIB decaia 17,797 ng/L em relação ao CAP 10 mg/L. A quantidade de MIB final aumenta 0,141 ng/L a cada aumento de uma unidade na quantidade de MIB inicial.

Por fim, o tempo 10 minutos tem um decaimento de -6,082 ng/L no MIB em relação ao instante de tempo 5 minutos; 9,554 ng/L de decaimento no instante 20 minutos em relação ao instante 5 minutos; no instante 30 minutos temos um decaimento de 12,459 ng/L em relação ao instante de 5 minutos; 14,915 ng/L no instante 45 minutos em relação ao instante 5 minutos; e 16,518 de decaimento no instante 60 minutos em relação ao instante de 5 minutos.

Em sequência, foi conduzida a análise de resíduos para verificar o ajuste do modelo.

A primeira análise a ser feita é acerca de valores discrepantes para a estrutura de covariância intraunidades amostrais, isto é mostrado na Figura B.23, e nota-se que só duas unidades ficam ligeiramente fora do limite de confiança. A Figura B.24 mostra um gráfico de dispersão dos resíduos marginais e seu histograma. A figura mostra que a maioria dos resíduos estão dentro do Limite de confiança e que a forma do histograma está próxima a uma distribuição normal, assim, indicando que a suposição de normalidade para a estrutura de covariância está adequada.

Em sequência, para avaliar a suposição de normalidade dos erros, temos as Figuras B.27 e B.28. A primeira figura mostra um gráfico de dispersão e um histograma com indícios de normalidade e homocedasticidade, ou seja, variância constante para as observações. O gráfico de dispersão mostra que a maioria dos resíduos estão dentro de uma faixa de confiança e o histograma tem forma próxima a uma distribuição normal. A segunda figura faz uma transformação dos resíduos condicionais para confirmar a suposição de normalidade. Nota-se no gráfico QQ que a maioria dos erros estão dentro da faixa de confiança e o histograma tem forma de uma distribuição normal com caudas mais pesadas.

As Figuras B.25 e B.26 analisam a existência de valores discrepantes e a suposição de normalidade para os efeitos aleatórios. Na primeira figura nota-se que não há valores discrepantes, e na segunda, o gráfico QQ mostra que há indícios de normalidade para os efeitos aleatórios.

Por fim, temos que a análise de resíduos mostra que o modelo está bem ajustado.

A do Figura B.29 mostra como ficaram as curvas de predições do processo de decaimento do MIB. É possível notar que curvas de decaimento apresentaram comportamento linear, isso mostra que o modelo pode auxiliar na predição do MIB em relação ao tempo, nível de carvão e o valor Inicial do MIB.

7. Conclusão

O modelo final de análise de sobrevivência indica que a variável CAP tem influência significativa no evento de purificação total de água ao longo do tempo, e que existe uma interação significativa entre turbidez e coagulante, o que pode explicar o porquê de algumas unidades amostrais apresentarem a purificação total rapidamente enquanto noutras não foi observado esse evento. A interação encontrada reforça a tese inicial do pesquisador de que o coagulante só tem efeito quando a turbidez não é baixa.

O modelo de dados longitudinais confirmou as suposições que surgiram após a análise descritiva de que o CAP é o principal fator no processo de remoção de impurezas na água e ainda mostrando um decaimento linear ao longo do tempo.

Com ambos os modelos, será possível ao pesquisador estimar a concentração de MIB em um dado instante do tempo, uma vez definido o nível de CAP aplicado na água, e será possível estimar a probabilidade da água não ficar limpa até um dado instante do tempo, definidos os níveis das covariáveis CAP, turbidez e coagulante aplicados naquela amostra de água.

APÊNDICE A

Tabelas

Tabela A.1 Medidas descritivas da concentração de MIB por instante de tempo.

Tempo (min)	Min.	1º Qu.	Mediana	Média	3º Qu.	Max.	NA's	Tabela A.2 Medidas
0	79	93	99	100	110	130	0	
5	7	11	22	23	32	44	1	
10	4	9	17	17	24	39	3	
20	4	8	16	16	21	34	8	
30	4	9	11	14	19	30	12	
45	4	7	9	11	13	26	14	
60	4	5	8	9	11	22	14	

as descritivas da razão C/C_0 de MIB por instante de tempo.

Tempo (min)	Min.	1º Qu.	Mediana	Média	3º Qu.	Max.	NA's
0	100%	100%	100%	100%	100%	100%	0%
5	0%	11%	22%	22%	32%	43%	0%
10	0%	8%	13%	16%	24%	34%	0%
20	0%	6%	11%	12%	19%	33%	0%
30	0%	0%	10%	9%	14%	29%	0%
45	0%	0%	6%	7%	10%	23%	0%
60	0%	0%	5%	5%	8%	20%	0%

Tabela A.3 Medidas descritivas para a eficiência após 10 minutos.

	Min.	1º Qu.	Mediana	Média	3º Qu.	Max.
Eficiência após 10 Min	66%	76%	87%	84%	92%	100%

Tabela A.4 Medidas descritivas para a eficiência após 10 minutos, por turbidez.

Tempo (min)	Min.	1º Qu.	Mediana	Média	3º Qu.	Max.
0	72%	77%	87%	86%	93%	100%
5	67%	75%	87%	84%	92%	96%
20	67%	78%	81%	84%	92%	96%
100	66%	76%	81%	83%	92%	100%

Tabela A.5 Medidas descritivas para a eficiência após 10 minutos, por dosagem de

Tempo (min)	Min.	1º Qu.	Mediana	Média	3º Qu.	Max.
0	1,40	1,58	1,75	2,02	2,33	2,90
10	1,45	1,50	1,78	1,90	2,14	3,20
20	1,35	1,44	1,60	1,78	2,06	2,80
50	1,65	1,78	2,35	2,33	2,65	3,50

coagulante.

Tabela A.6 Medidas descritivas para a eficiência após 10 minutos, por dosagem de CAP.

Tempo (min)	Min.	1º Qu.	Mediana	Média	3º Qu.	Max.
10	66%	69%	74%	73%	78%	79%
20	75%	81%	87%	85%	88%	93%
40	91%	92%	94%	95%	96%	100%

Tabela A.7 Ajuste do modelo de sobrevivência com as covariáveis como fatores.

	Valor	Erro padrão	Estatística do teste	Valor-p
Intercepto	20,237	0,534	37,89	<0,001
CAP (20 mg/L)	-14,921	0,415	-35,95	<0,001

CAP (40 mg/L)	-16,97	< 0,001	$-\infty$	<0,001
Coagulante (10 mg/L)	0,046	0,735	0,06	0,95
Coagulante (20 mg/L)	-1,056	0,709	-1,49	0,136
Coagulante (50 mg/L)	-1,526	0,643	-2,37	0,018
Turbidez (5 UNT)	0,862	0,469	1,84	0,066
Turbidez (20 UNT)	1,142	0,495	2,31	0,021
Turbidez (100 UNT)	0,798	0,504	1,58	0,113
Log (escala)	-0,651	0,233	-2,80	0,005

Tabela A.8 Ajuste do modelo final de sobrevivência.

	Valor	Erro padrão	Estatística do teste	Valor-p
Intercepto	7,239	0,797	9,08	< 0,001
CAP	-0,097	0,019	-5,09	< 0,001
Coagulante	-0,014	0,01	-1,44	0,149
Turbidez	0,029	0,015	1,83	0,067
Coagulante × Turbidez	-0,001	< 0,001	-2,26	0,023
Log (escala)	-0,66	0,23	-2,87	0,004

Tabela A.9 Ajuste do modelo Parcial de Análise de dados Longitudinais.

	Valor	Erro Padrão	estatística de Teste	Valor - p
Intercepto	19,04	7,754	2,455	0,015
CAP (20 mg/L)	-10,125	1,252	-8,084	< 0,001
CAP (40 mg/L)	-17,846	1,27	-14,051	< 0,001
Turbidez (5 UNT)	0,7	1,506	0,465	0,646
Turbidez (20 UNT)	1,761	1,586	1,11	0,276
Turbidez (100 UNT)	1,672	2,044	0,818	0,42
Coagulante (10 mg/L)	1,261	2,408	0,524	0,604
Coagulante (20 mg/L)	-0,192	2,621	-0,073	0,942

Coagulante (50 mg/L)	-5,136	2,412	-2,13	0,042
MIB Inicial (Ng/l)	0,126	0,069	1,827	0,078
tempo 10 (minutos)	-6,082	0,851	-7,148	< 0,001
tempo 20 (minutos)	-9,554	0,851	-11,229	< 0,001
tempo 30 (minutos)	-12,459	0,851	-14,643	< 0,001
tempo 45 (minutos)	-14,915	0,851	-17,53	< 0,001
tempo 60 (minutos)	-16,518	0,851	-19,414	< 0,001

Tabela A.10 Ajuste do modelo final de Análise de dados Longitudinais

	Valor	Erro Padrão	Valor - p
Intercepto	17,183	5,798	0,003
CAP (20 mg/L)	-10,113	1,602	< 0,001
CAP (40 mg/L)	-17,797	1,611	< 0,001
MIB Inicial (Ng/l)	0,141	0,056	0,016
tempo 10 (minutos)	-6,082	0,851	< 0,001
tempo 20 (minutos)	-9,554	0,851	< 0,001
tempo 30 (minutos)	-12,459	0,851	< 0,001
tempo 45 (minutos)	-14,915	0,851	< 0,001
tempo 60 (minutos)	-16,518	0,851	< 0,001

APÊNDICE B

Figuras



Figura B.1 Aplicação de CAP na ETA Guandu no início de 2020 (Reprodução / TV Globo).



Figura B.2 Turbidez da água para três níveis, Da esquerda para a direita: do mais alto ao mais baixo (Reprodução/ hidrogeron.com).

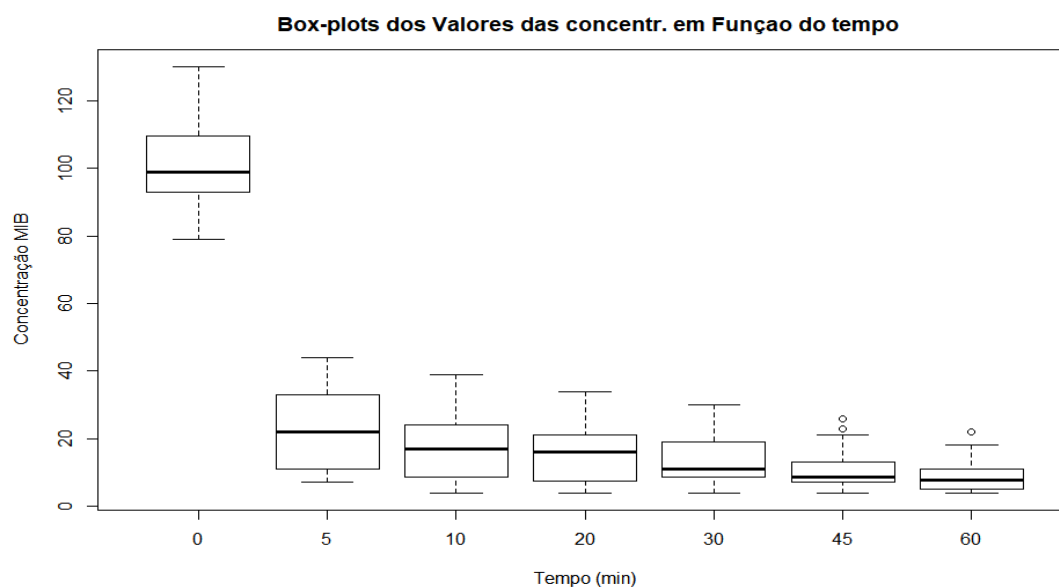


Figura B.3 Box plots das concentrações de MIB em função do tempo.

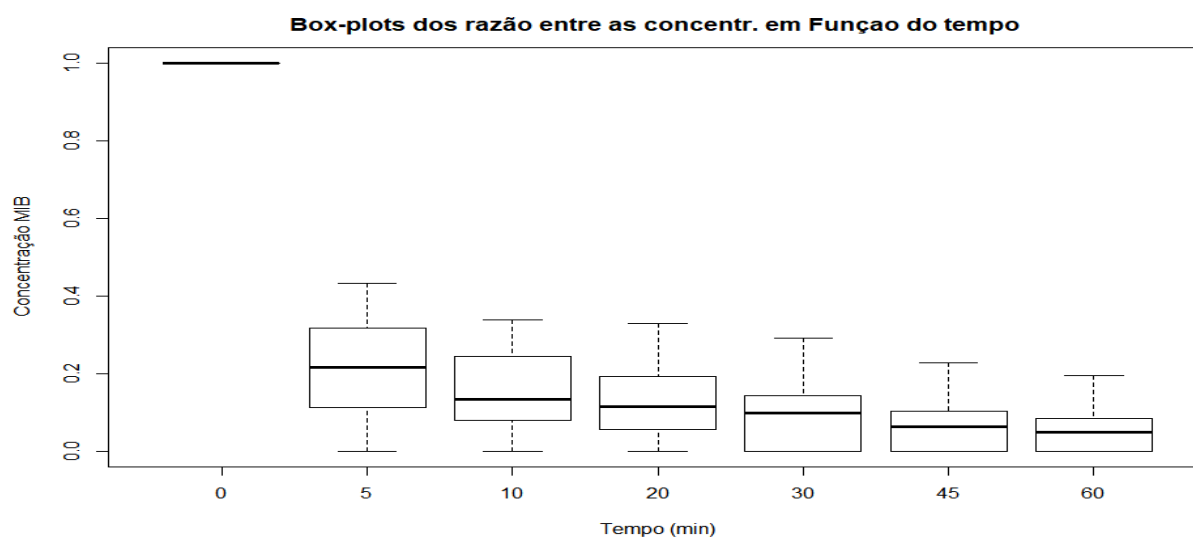


Figura B.4 Box plots das razões entre cada concentração ao longo do tempo de MIB em relação à primeira.

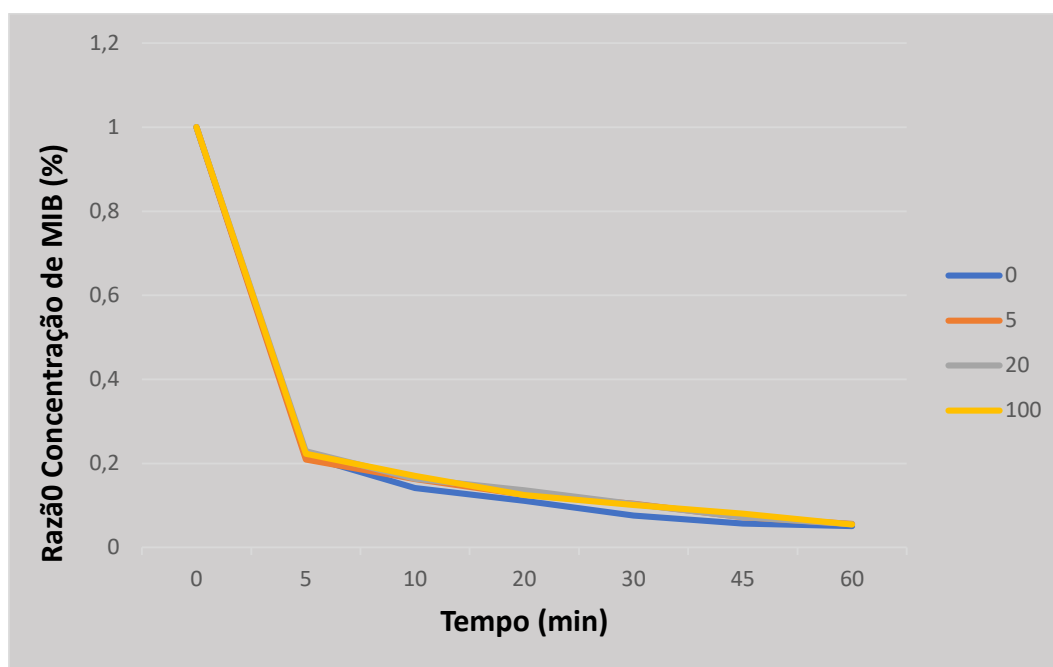


Figura B.5 Razões entre cada concentração ao longo do tempo de MIB em relação à primeira para cada nível de turbidez.

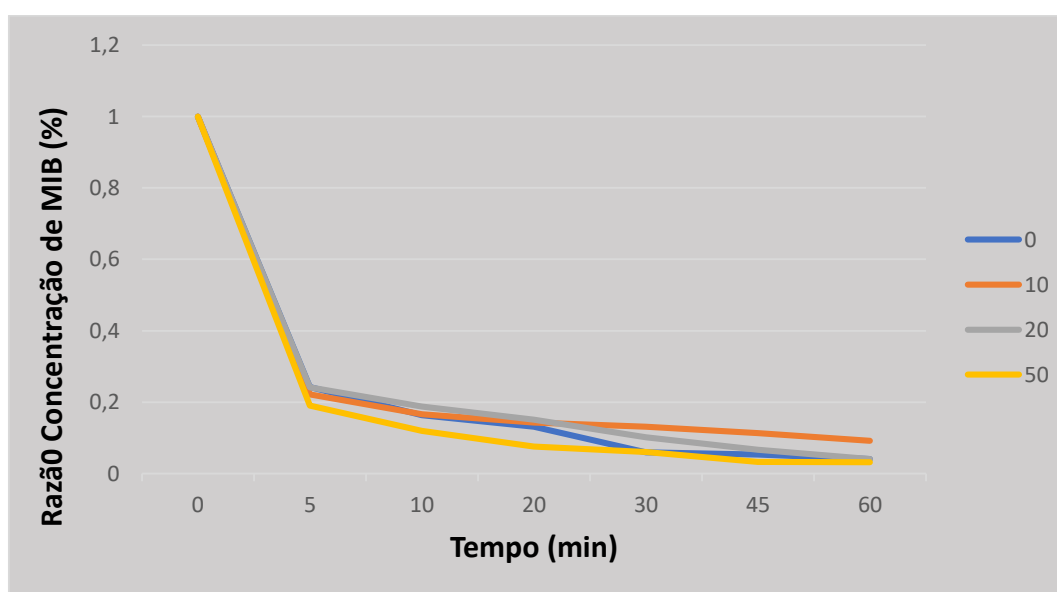


Figura B.6 Razões entre cada concentração ao longo do tempo de MIB em relação à primeira para cada nível de dosagem de coagulante.

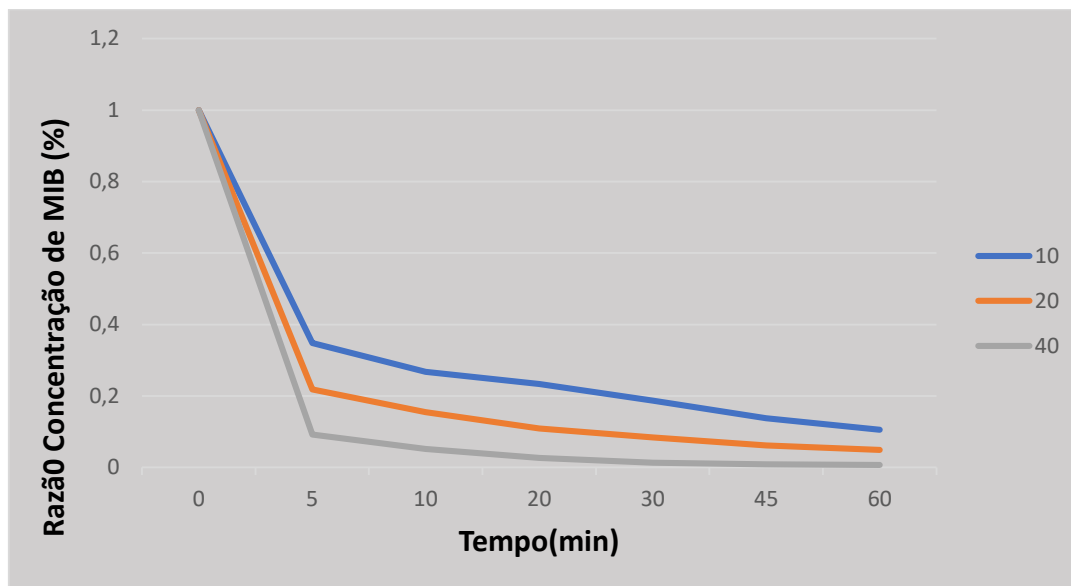


Figura B.7 Razões entre cada concentração ao longo do tempo de MIB em relação à primeira para cada nível de dosagem de CAP.

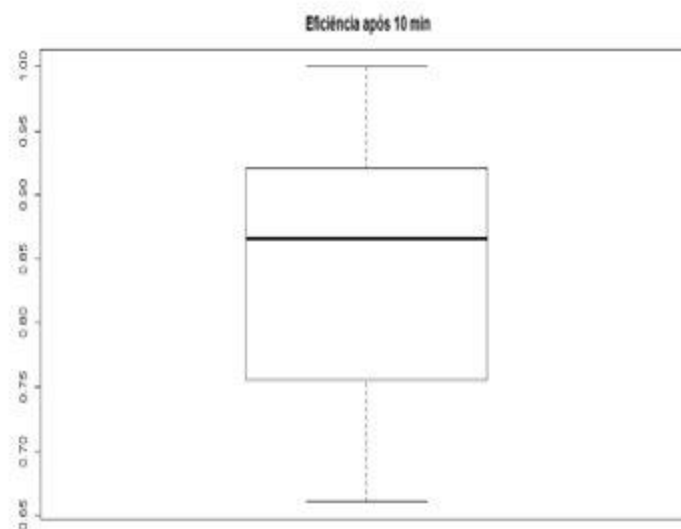


Figura B.8 Box plot dos valores de eficiência após 10 minutos.

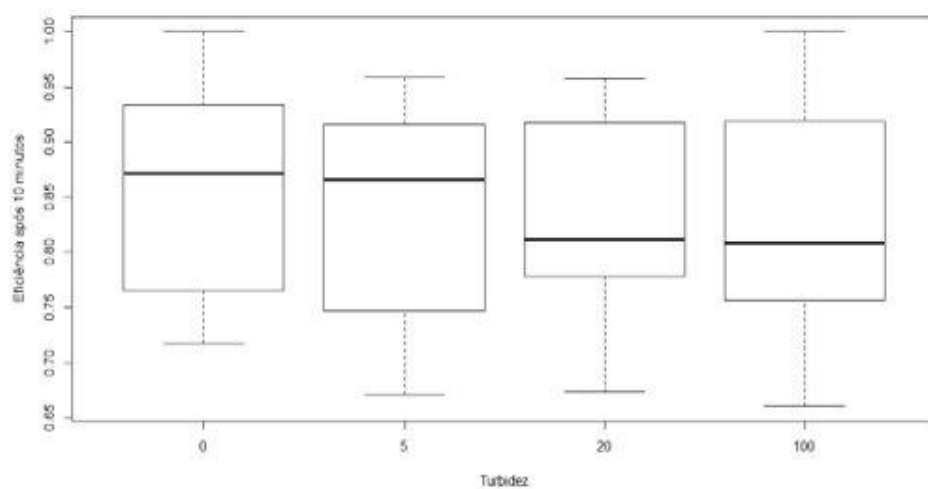


Figura B.9 *Box plots* dos valores de eficiência após 10 minutos para cada nível de turbidez.

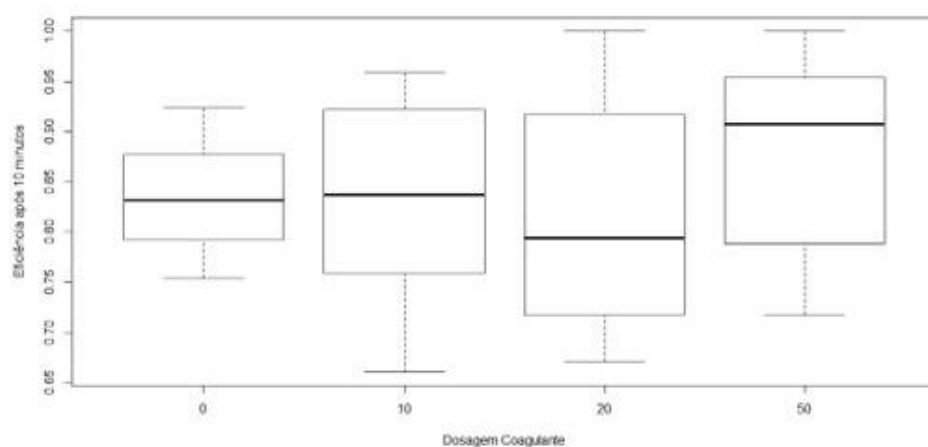


Figura B.10 *Box plots* dos valores de eficiência após 10 minutos para cada dosagem de Coagulante.

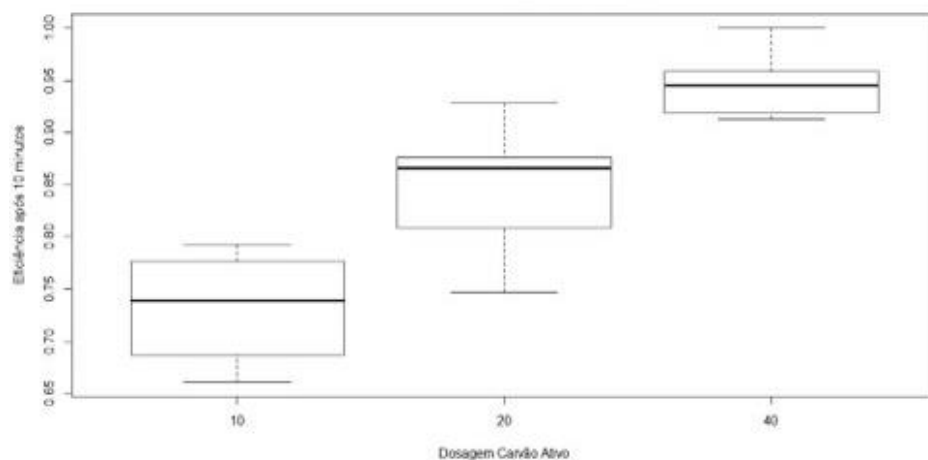


Figura B.11 Box plots dos valores de eficiência após 10 minutos para cada dosagem de CAP.

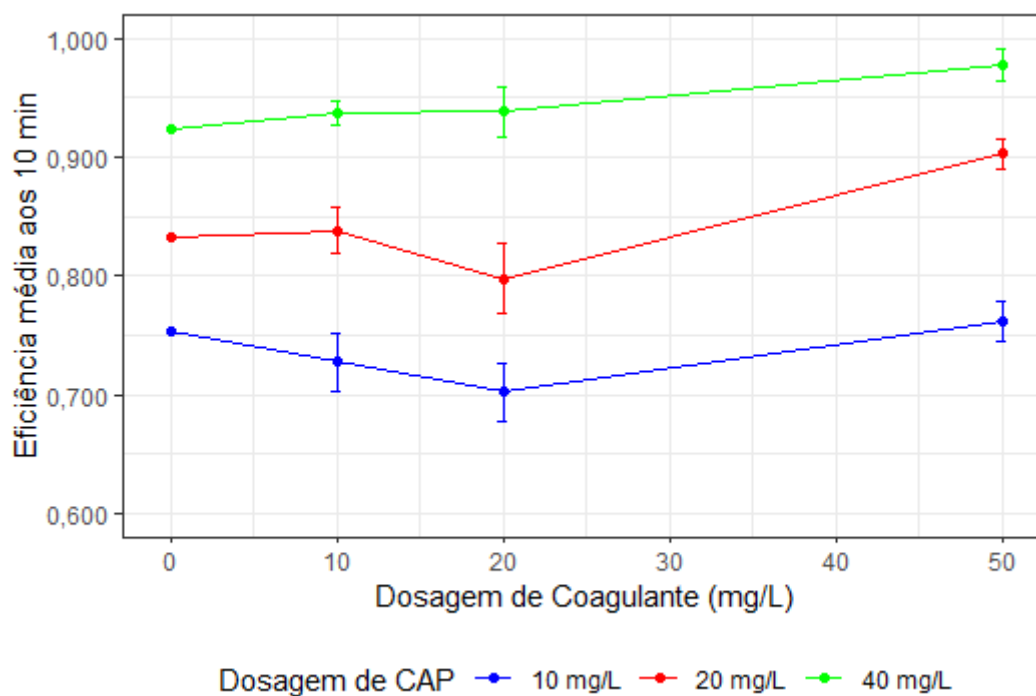


Figura B.12 Gráfico de eficiência média aos 10 minutos segundo dosagem de CAP e dosagem de coagulante.

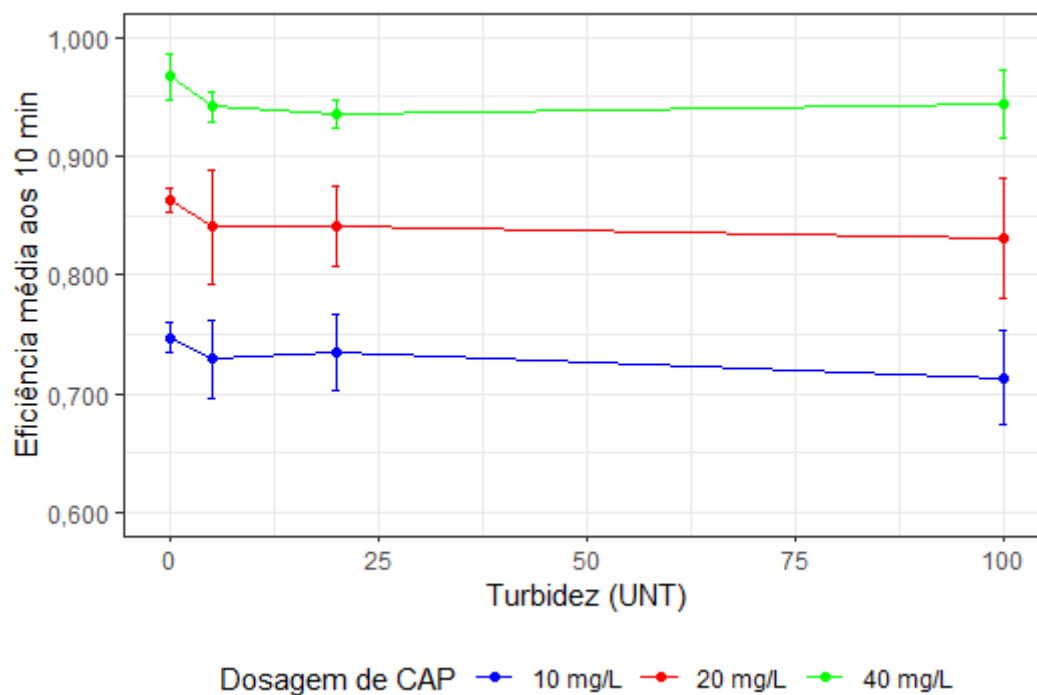


Figura B.13 Gráfico de eficiência média aos 10 minutos segundo dosagem de CAP e turbidez da água.

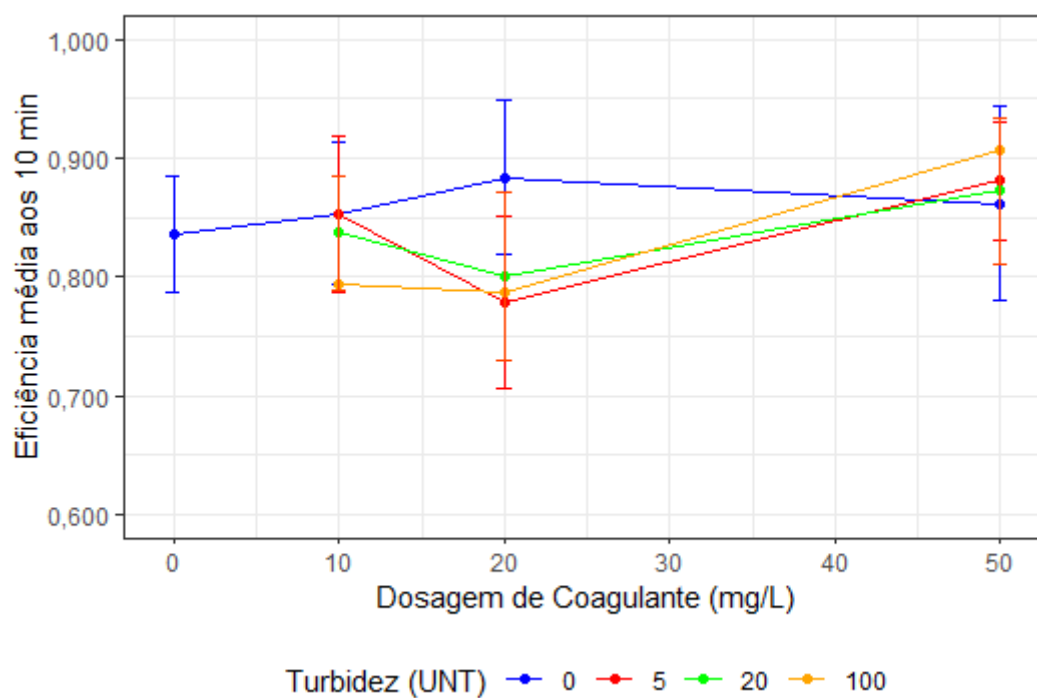


Figura B.14 Gráfico de eficiência média aos 10 minutos segundo turbidez da água e dosagem de coagulante.

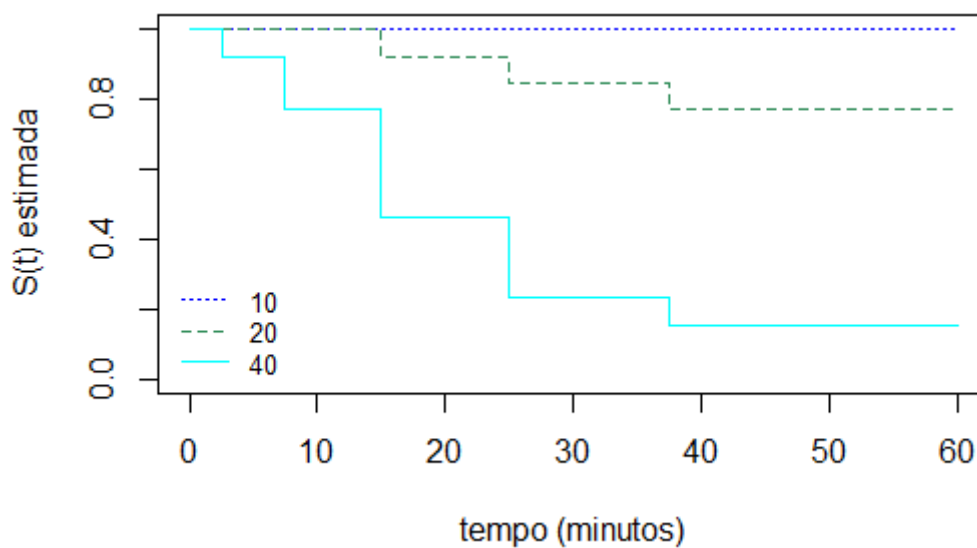


Figura B.15 Gráfico de sobrevivências estimadas pelo método de Kaplan-Meier ao longo do tempo para três dosagens de CAP.

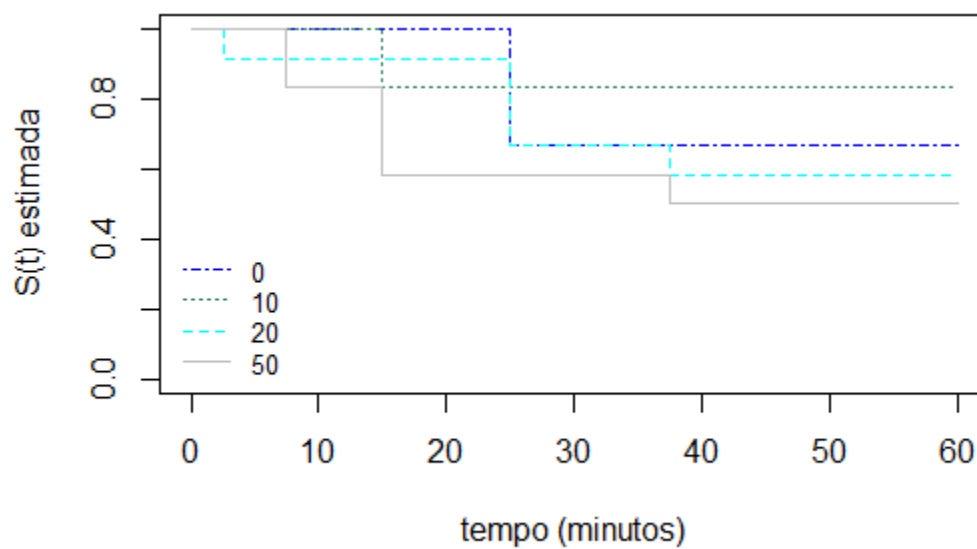


Figura B.16 Gráfico de sobrevivências estimadas pelo método de Kaplan-Meier ao longo do tempo para quatro dosagens de coagulante.

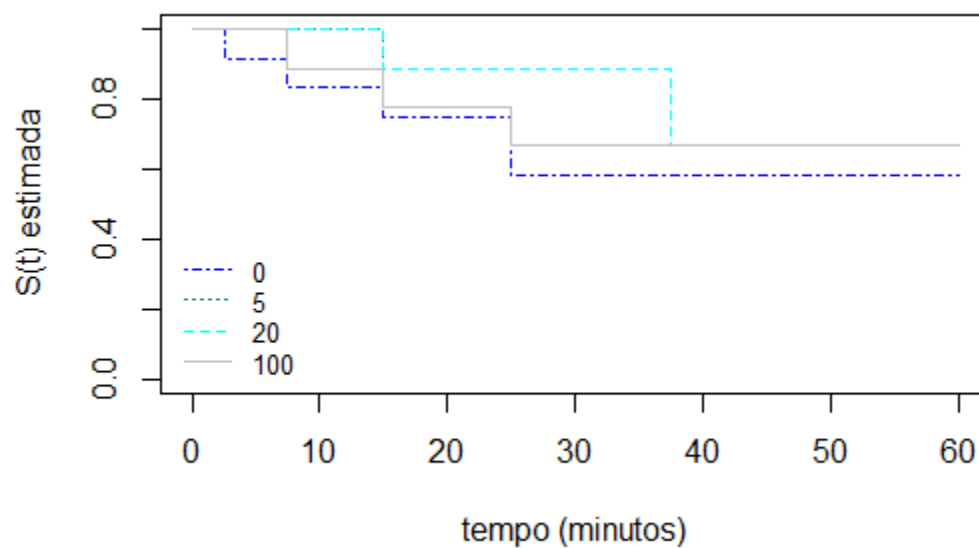


Figura B.17 Gráfico de sobrevivências estimadas pelo método de Kaplan-Meier ao longo do tempo para quatro níveis de turbidez.

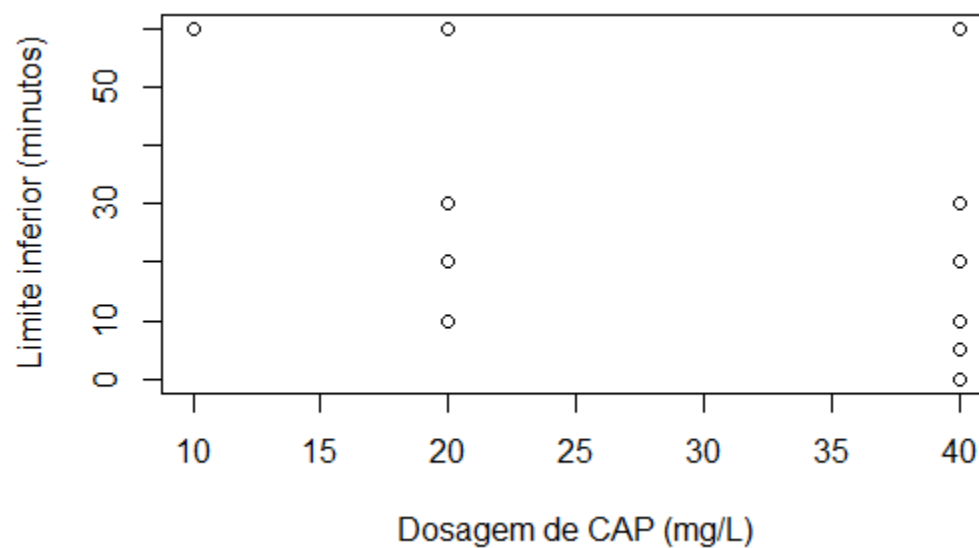


Figura B.18 Gráfico de pontos entre l_i e dosagem de CAP

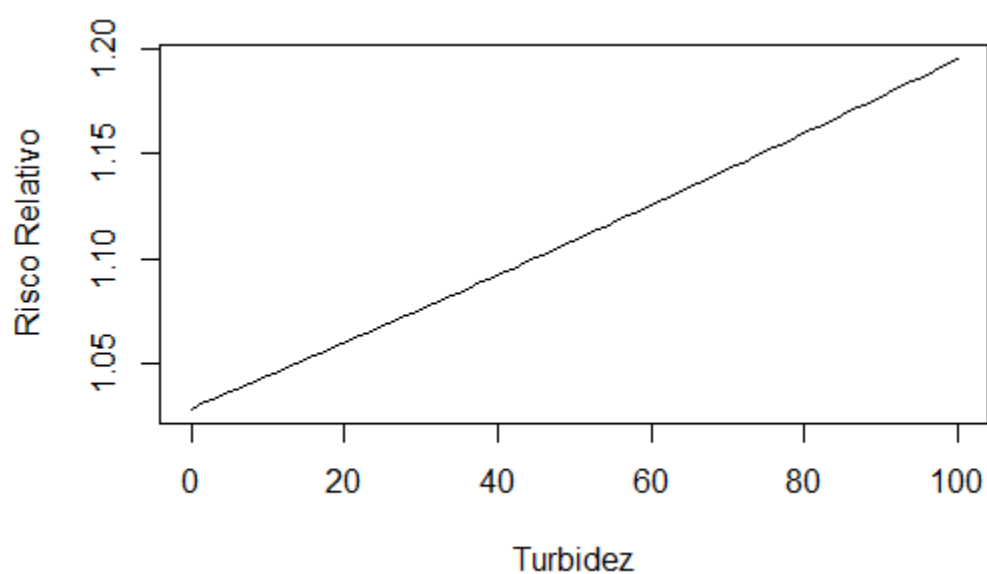


Figura B.19 Gráfico do risco relativo entre duas dosagens de coagulante com uma unidade de diferença. para diferentes concentrações de turbidez.

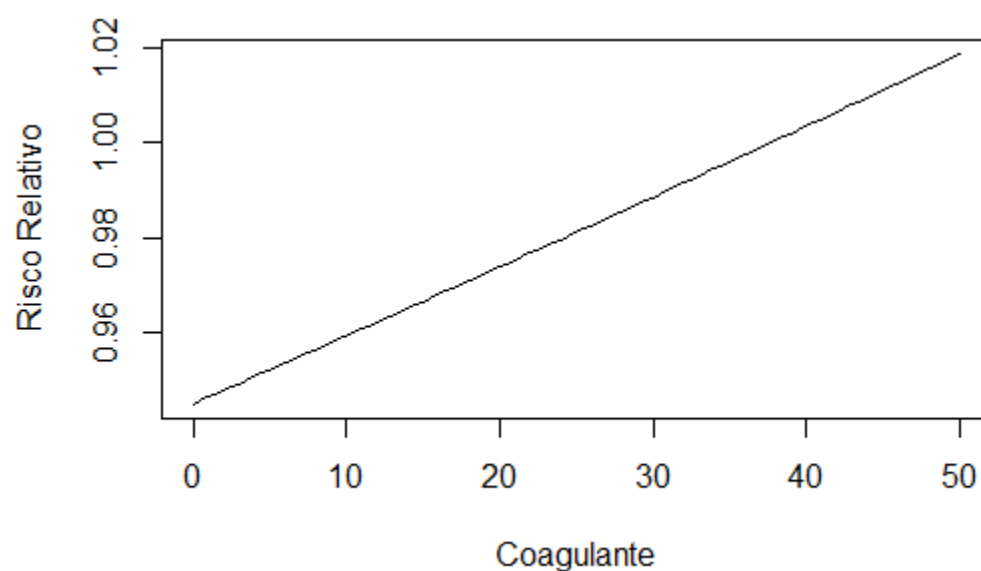


Figura B.20 Gráfico do risco relativo entre duas dosagens de turbidez com uma unidade de diferença. para diferentes concentrações de coagulante.

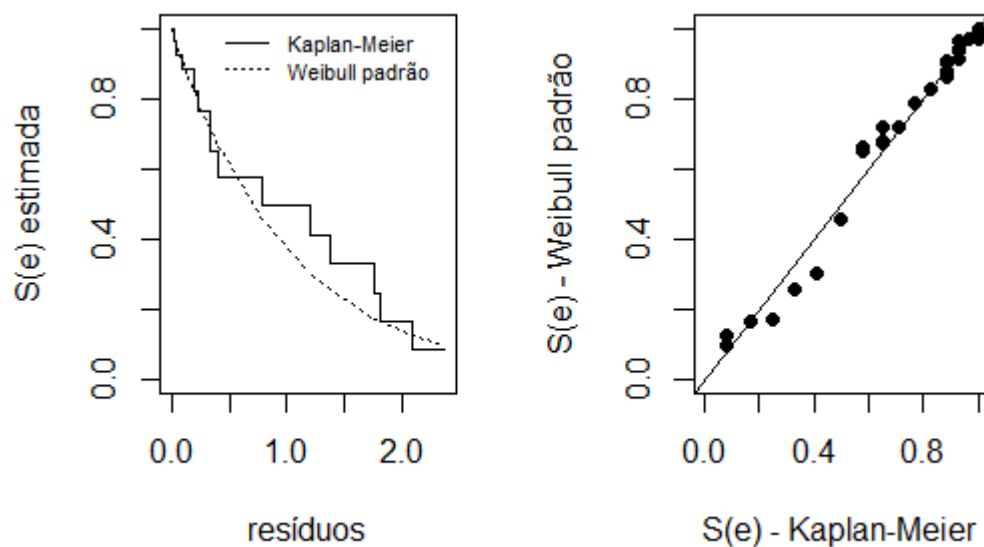


Figura B.21 Gráfico da sobrevivência estimada dos resíduos contra a Weibull padrão e gráfico de resíduos de Cox-Snell ou padronizados.

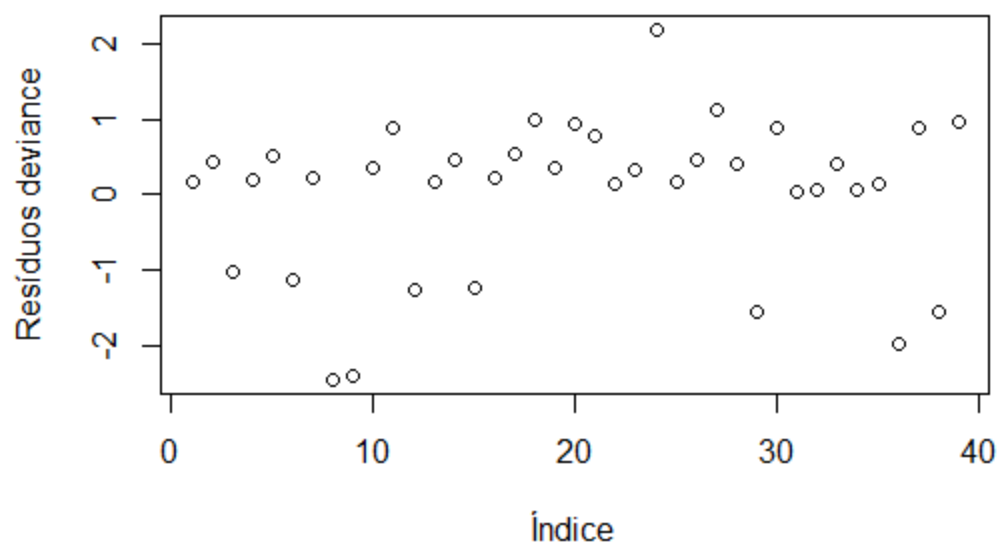


Figura B.22 Gráfico de resíduos *deviance*.

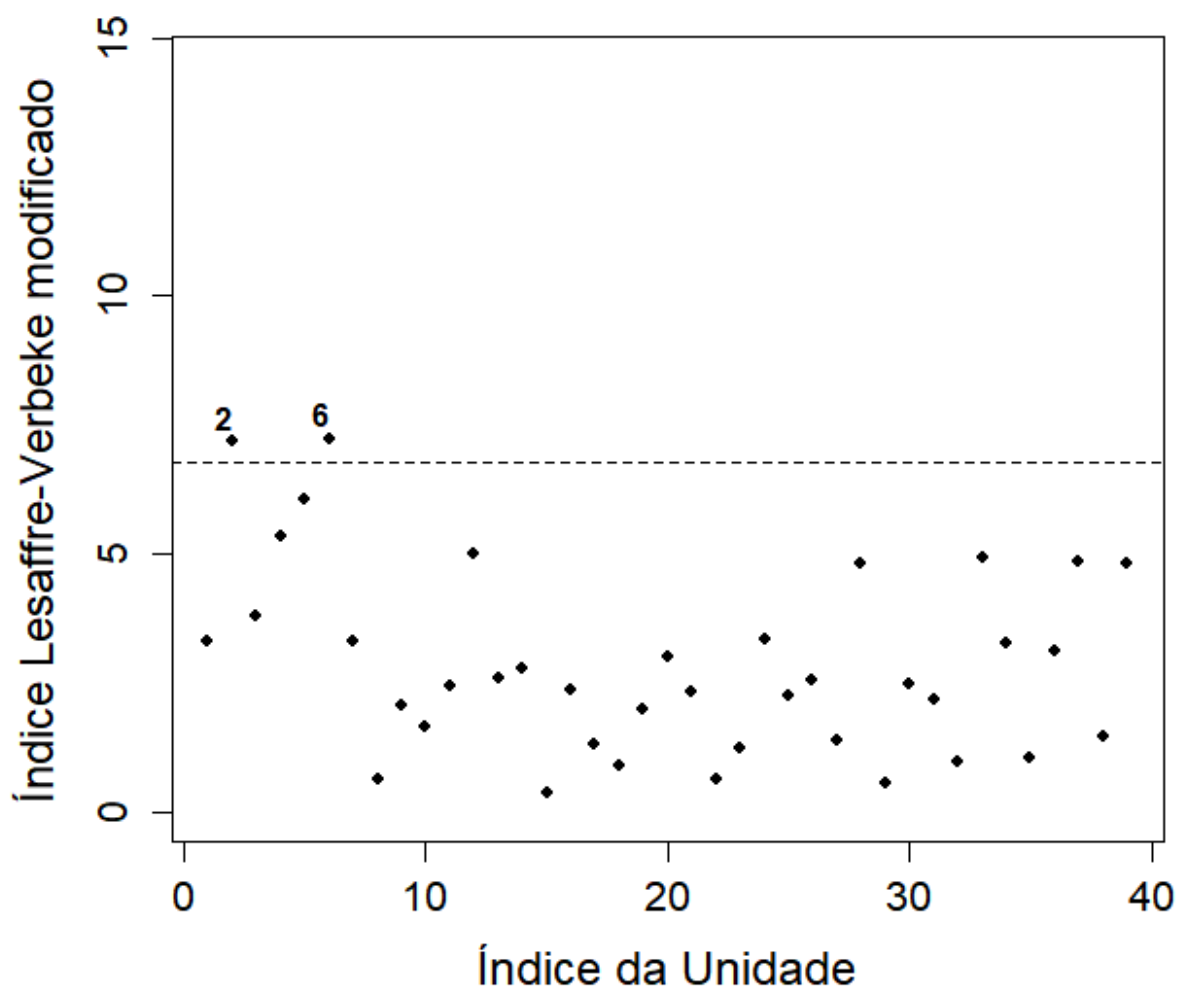


Figura B.23 Gráfico de resíduos Marginais para Observações discrepantes.

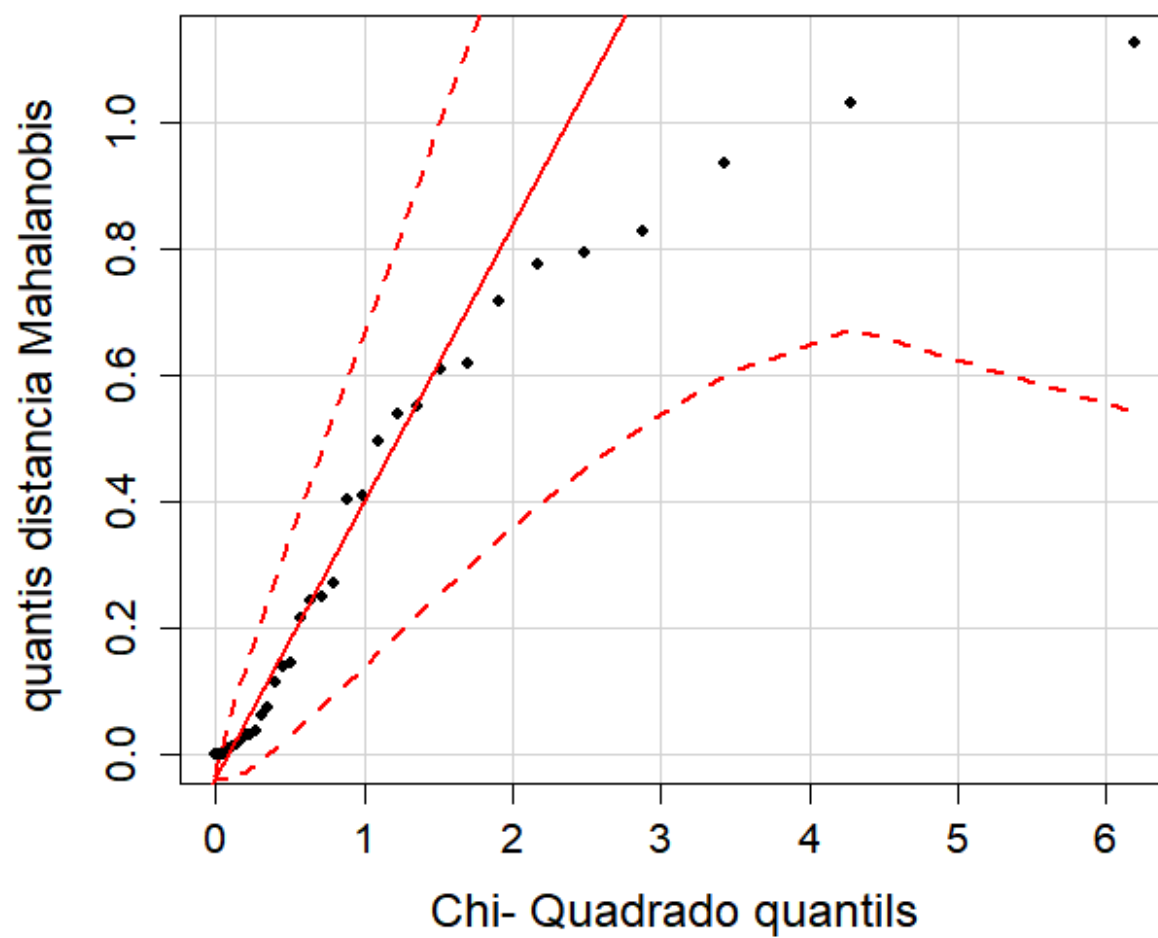


Figura B.25 Gráfico de distância de Mahalanobis *versus* quantis da Qui-quadrado.

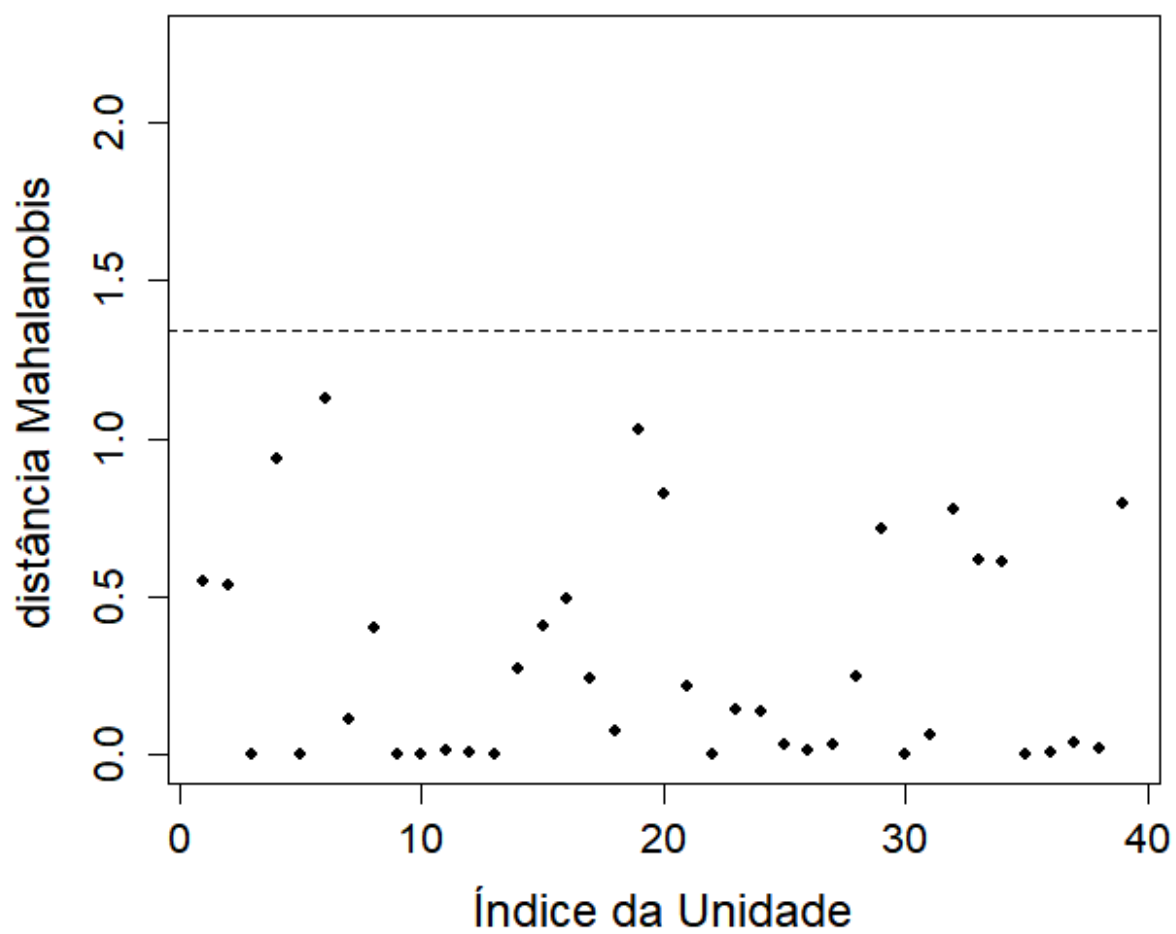


Figura B.26 Gráfico de resíduos da distância de Mahalanobis.

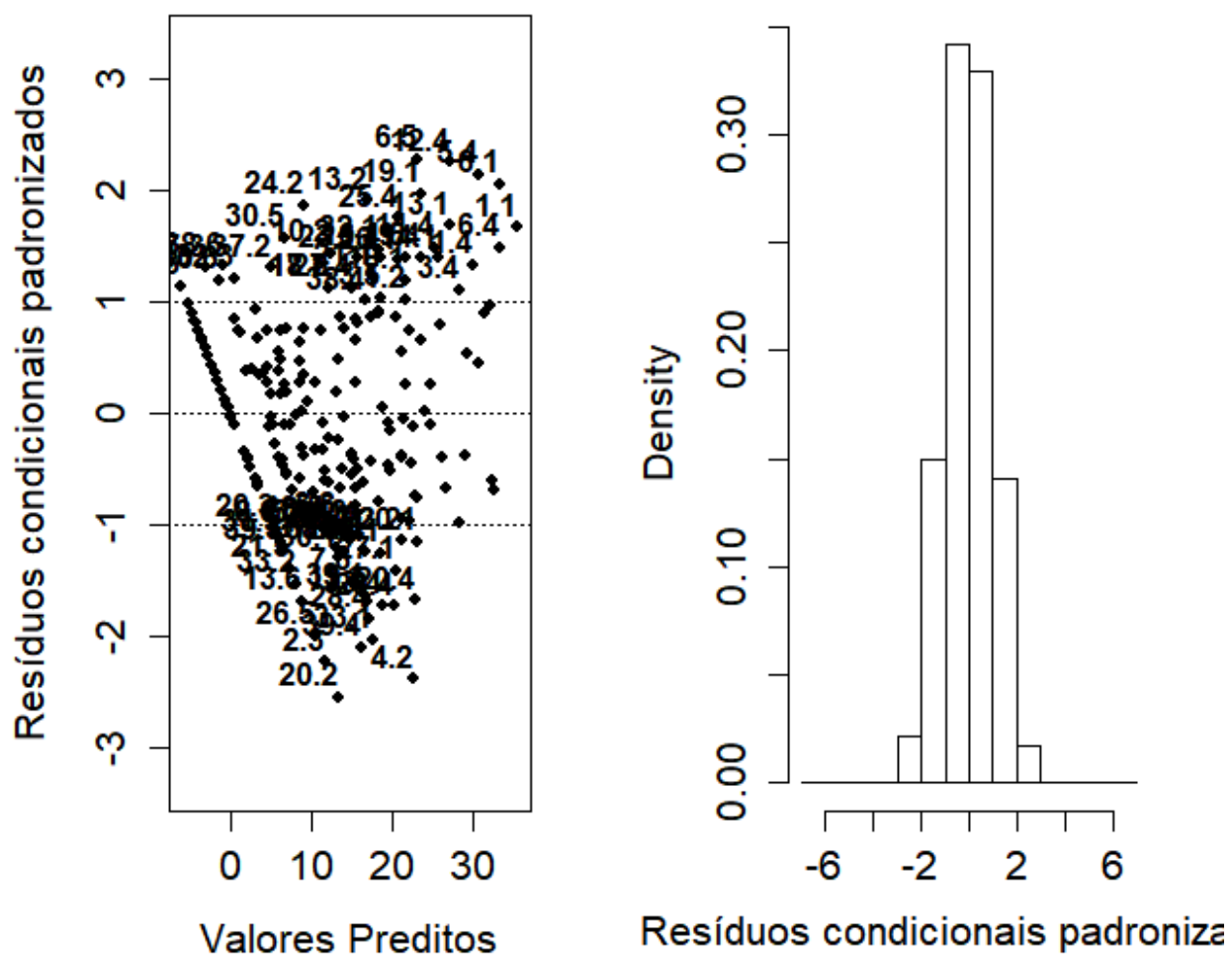


Figura B.27 Gráfico de resíduos condicionais (dispersão e densidade).

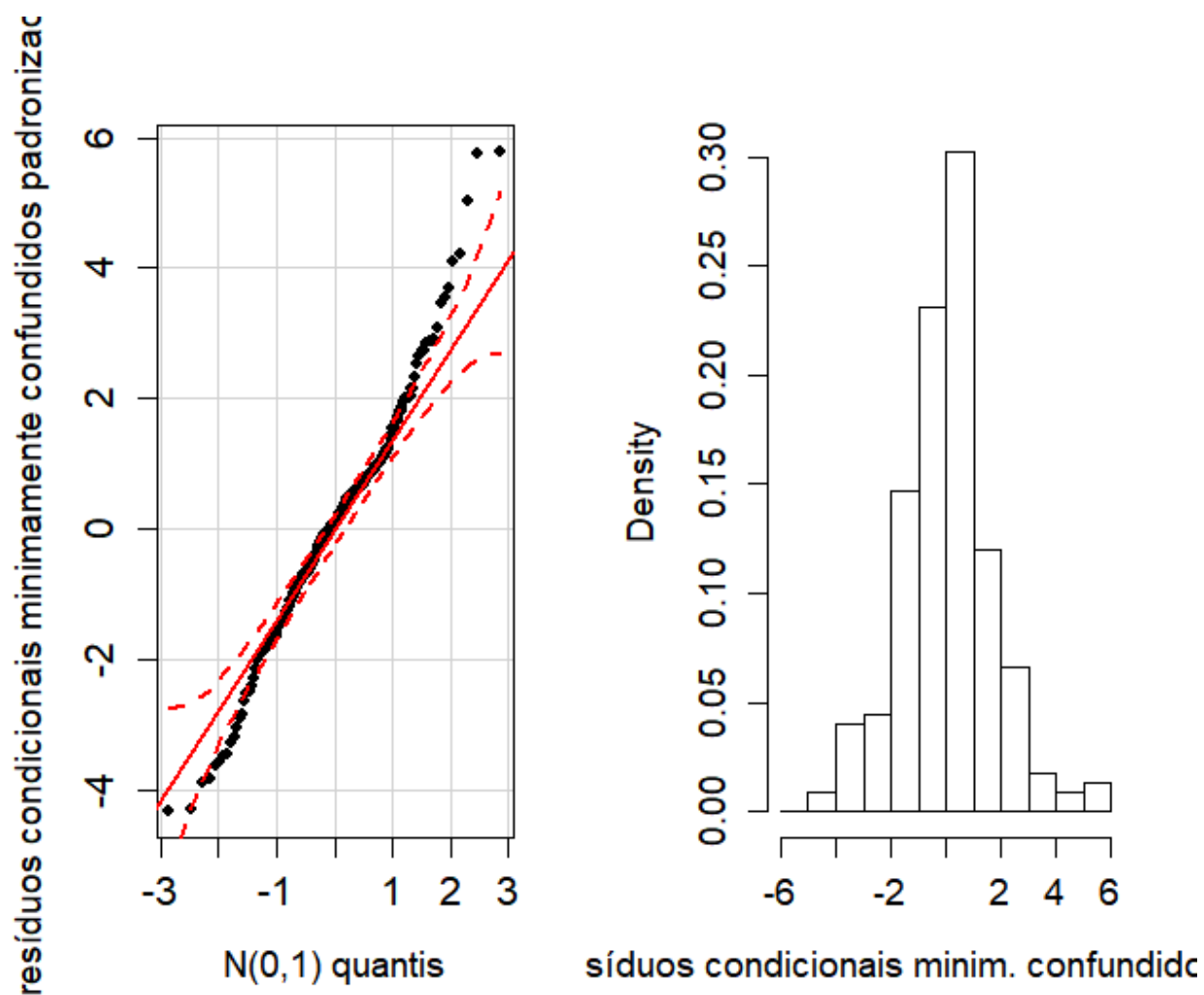


Figura B.28 Gráfico de resíduos condicionais minimamente confundidos (dispersão e densidade).

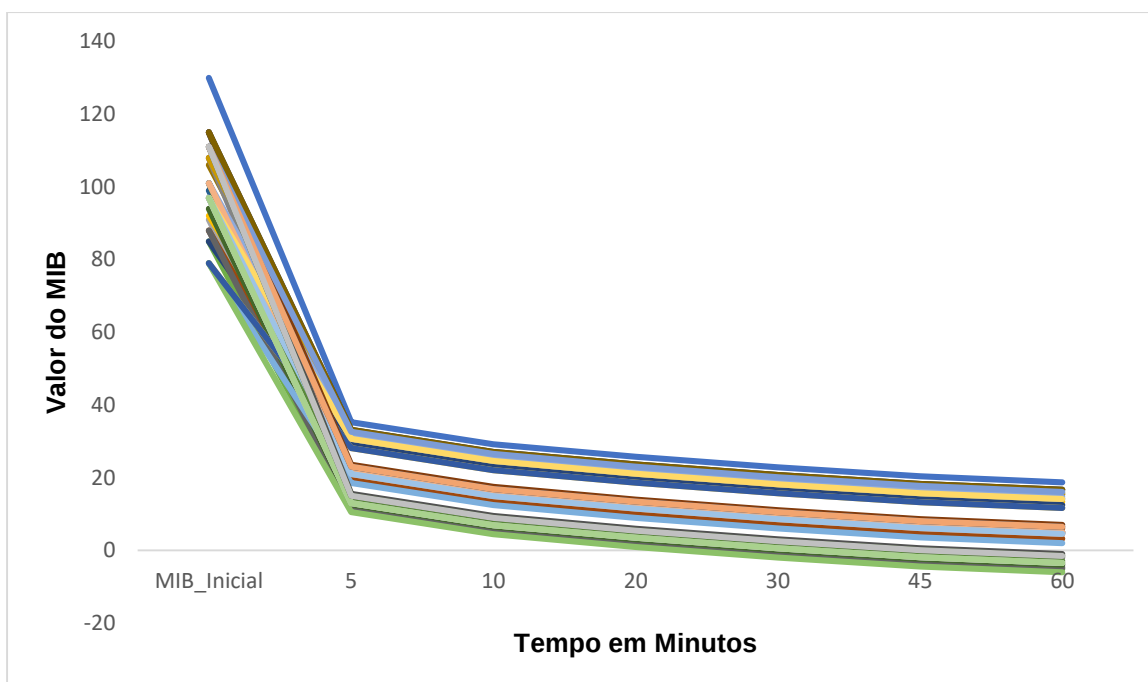


Figura B.29 Gráfico de Valores Preditos