UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE CIÊNCIAS EXATAS
CURSO DE ESTATÍSTICA

AVALIAÇÃO DE MÉTODOS NÃO PARAMÉTRICOS PARA PREDIÇÃO EM MODELOS ADITIVOS

Marco Aurelio Valles Leal

Maringá
2021

MARCO AURELIO VALLES LEAL

AVALIAÇÃO DE MÉTODOS NÃO PARAMÉTRICOS PARA PREDIÇÃO EM MODELOS ADITIVOS

Trabalho de conclusão de curso apresentado como requisito parcial para a obtenção do título de bacharel em Estatística pela Universidade Estadual de Maringá.

Orientador: Profº Drº George Lucas Moraes Pezzot
Coorientador: Profº Drº Willian Luís de Oliveira

Maringá
2022

AVALIAÇÃO DE MÉTODOS NÃO PARAMÉTRICOS PARA PREDIÇÃO EM MODELOS ADITIVOS

MARCO AURELIO VALLES LEAL

Trabalho de conclusão de curso apresentado como requisito parcial para a obtenção do título de bacharel em Estatística pela Universidade Estadual de Maringá.

Aprovado em: _____/_____/_____.

BANCA EXAMINADORA

_____

Orientador
Profº Drº George Lucas Moraes Pezzot
Universidade Estadual de Maringá

_____

Membro da banca
Nome do professor membro da banca
Intituição do professor membro da banca

_____

Membro da banca
Nome do professor membro da banca
Intituição do professor membro da banca

RESUMO

É comum, nas mais diversas áreas, investigar e modelar a relação entre variáveis. O modelo mais simples é denominado modelo de regressão linear simples e assume que a média da variável resposta é modelada como uma função linear das variáveis explicativas, supondo erros aleatórios com média zero, variância constante e não correlacionados. Entretanto, nem sempre a relação existente é perfeitamente linear. Neste contexto, é possível flexibilizar o modelo de regressão linear modelando a dependência da variável resposta com cada uma das variáveis explicativas em um contexto não paramétrico. Esta nova classe de modelos é dita modelos aditivos e mantêm a característica dos modelos de regressão lineares de serem aditivos nos efeitos preditivos. Portanto, este projeto visa apresentar os modelos aditivos, além de técnicas de suavização utilizadas para ajustar modelos no contexto não paramétrico. Por fim, a metodologia é aplicada em dados artificiais (simulados) e em dados reais, dando enfoque à qualidade das predições.

Palavras-chave : regressão, modelo aditivo, suavizadores

Sumário

1 Introdução

Análise de regressão é uma técnica amplamente utilizada na estatística para investigar e modelar a relação entre variáveis. Usualmente, é de interesse apenas uma variável, chamada de variável resposta ou dependente, e desejamos estudar como esta variável depende de um conjunto de variáveis observáveis, chamadas de variáveis explicativas ou independentes. Nesse contexto, os modelos de regressão (linear simples ou múltipla) podem ser utilizados.

Nota-se, porém, que em muitos casos a relação existente entre a variável resposta (média) e cada uma das variáveis explicativas não é perfeitamente linear e determinar uma função que representa a correta relação existente nem sempre é fácil. Uma alternativa é flexibilizar o modelo de regressão linear, modelando a dependência da variável resposta com cada uma das variáveis explicativas em um contexto não paramétrico. Esta nova classe de modelos é dita modelos aditivos e mantêm a característica dos modelos de regressão lineares de serem aditivos nos efeitos preditivos. **REFERÊNCIA**

Portanto, o objetivo do presente trabalho é apresentar os modelos aditivos e estudar as principais técnicas de suavização existentes utilizadas para ajustar modelos no contexto não paramétrico, em particular os modelos aditivos, apresentando suas principais características e aplicações. Além disso, pretende-se mostrar o ganho na predição, ao se utilizar modelos mais flexíveis, em determinadas situações.

1.1 Objetivo Geral

O objetivo deste projeto é apresentar os modelos aditivos, uma generalização dos modelos de regressão linear, descrevendo suas principais características e estudar algumas técnicas de estimação do modelo, no contexto não paramétrico.

1.2 Objetivos Específicos

- Introduzir os modelos aditivos, especificando sua principais características e apresentar os principais métodos de estimação dos parâmetros do modelo assim como as técnicas de diagnóstico;
- Apresentar técnicas de suavização utilizadas para estimar funções não paramétricas presentes nos modelos aditivos, identificando suas principais características;
- Apresentar métricas para verificar a qualidade de predição;
- Realizar um estudo de simulação para verificar a qualidade do ajuste dos modelos em alguns cenários, considerando os modelos aditivos;

- Aplicar a metodologia estudada a um conjunto de dados reais, comparando modelos e técnicas de estimação e predição.

2 Referencial Teórico

3 Metodologia

3.1 Regressão Linear

Análise de regressão é uma técnica estatística utilizada para investigar e modelar a relação entre variáveis. Aplicações de regressão são numerosas e ocorrem em quase todas as áreas do conhecimento, como engenharia, ciências físicas e químicas, economia, ciência biológicas, etc. Resumidamente a regressão tem como objetivo descrever uma relação entre uma variável de interesse, chamada de variável resposta ou dependente (Y) e um conjunto de variáveis preditoras ou independentes (X), as co-variáveis. Através do modelo é possível estimar parâmetros e fazer inferências sobre os mesmos, como testes de hipóteses e intervalos de confiança.

Além disso, o modelo de regressão pode ser usado para predição, onde se é esperado que grande parte da variação de Y seja explicado pelas variáveis X. Dessa forma, obtem-se valores esperados de Y correspondentes a valores de X que não estavam entre os dados.

O modelo de regressão linear simples, constitui uma tentativa de estabelecer uma equação matemática linear que descreve o relacionamento entre duas variáveis, X (preditora) e Y (resposta). O modelo de regressão linear populacional é definido por:

$$Y = \beta_0 + \beta_1 x + \varepsilon, \tag{1}$$

onde o intercepto β_0 e a inclinação da reta β_1 são parâmetros desconhecidos e ε é um erro aleatório. Pressupõe-se que os erros têm média zero, variância σ^2 desconhecida e são não correlacionados, assim, as respostas também não têm relação. A média da distribuição é dada por uma função linear de x:

$$E(Y|x) = \beta_0 + \beta_1 x \tag{2}$$

Os parâmetros β_0 e β_1, também chamados de coeficientes da regressão, tÊm uma interpretação simples e muitas vezes útil. A inclinação β_1 é a alteração média da distribuição de Y produzida por uma mudança unitária da variável X, ou seja, o quanto varia a média de Y para o aumento de uma unidade de X. Se os dados de X incluem

$x = 0$, então o intercepto β_0 é a média da distribuição da resposta Y quando $x = 0$. Porém, se a observação no zero não estiver incluída, β_0 não tem interpretação prática e é chamado de intercepto ou coeficiente linear, pois é o ponto onde a reta regressora corta o eixo y.

3.1.1 Estimação dos parâmetros pelo Métodos dos Mínimos Quadrados

Os parâmetros β_0 e β_1 são desconhecidos e devem ser estimados usando dados de uma amostra. Suponha que tem-se n pares de dados, $(x_1, y_1), ..., (x_n, y_n)$. Esses dados podem ter sido resultado de um experimento controlado feito especificamente para coletá-los, de um estudo observacional, etc. Uma maneira de estimar esses parâmetros, é utilizando o Método dos Mínimos Quadrados, onde não é necessário conhecer a distribuição dos erros. Esse método tem como objetivo encontrar os valores de β_0 e β_1 que minimizam a soma dos quadrados dos erros (ou desvios do modelo). A equação de regressão linear amostral é definida como:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ para i=1,2,...,n} \tag{3}$$

A partir de (3) tem-se:

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i \Rightarrow \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \Rightarrow S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Denomina-se os estimadores de mínimos quadrados de β_0 e β_1, $\hat{\beta}_0$ e $\hat{\beta}_1$, respectivamente. Para obtê-los deve-se encontrar os valores para $\hat{\beta}_0$ e $\hat{\beta}_1$ que minimizam a equação $S(\beta_0, \beta_1)$. Em geral, deriva-se a equação, iguala a zero e isola os parâmetros e,(**determinante?** se a segunda derivada em relação a cada parâmetro for positiva), os valores encontrados são os que minimizam a equação. Sendo assim é fácil verificar que:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

E portanto podemos definir, o modelo de regressão linear simples ajustado como sendo

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{4}$$

A diferença entre o valor observado y_i e o correspondente valor estimado é o resíduo, que é importante para verificar a adequação do modelo. Matematicamente, o i-ésimo resíduo é

$$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x), \quad i = 1, ..., n$$

FALAR DA PREDIÇÂO!

3.1.2 Regressão Linear Múltipla

FALAR BREVEMENTE

Como pode ser visto anteriormente, o modelo de regressão linear simples, com uma variável explicativa, aplica-se a várias situações. Entretanto, diversos problemas envolvem dois ou mais regressores influenciando o comportamento da variável resposta (dependente) y.

Suponha que o rendimento, em libras, de conversão em um processo químico dependa da temperatura e da concentração do catalisador. Um modelo de regressão múltipla que talvez descreva esse relacionamento é dado por

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

onde y é o rendimento, x_1 a temperatura e x_2 a concentração do catalisador. Esse é um modelo de regressão linear múltipla com duas variáveis regressoras. O termo linear é usado porque a equação é uma função linear dos parâmetros desconhecidos β_1, β_0 e β_2. Outro ponto a ser destacado é que esse modelo forma um plano no espaço tridimensional de y, x_1 e x_2. O parâmetro β_0 é o intercepto do plano; se os dados incluem $x_1 = x_2 = 0$, então β_0 é a média de y quando $x_1 = x_2 = 0$. Caso contrário, β_0 não tem interpretação prática. Já β_1 indica a mudança na resposta média y a cada mudança unitária de x_1 quando x_2 é constante. O parâmetro β_2 indica a mudança na resposta média a cada unidade de mudança em x_2, quando x_1 é constante. Em geral, a resposta y pode estar relacionada a k regressores ou variáveis preditoras. O modelo

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

é chamado de modelo de regressão linear múltipla com k regressores. Os parâmetros $\beta_j, j = 0, 1, ..., k$ são os coeficientes de regressão. Esse modelo descreve um hiperplano no espaço k-dimensional das variáveis regressoras x_j. O parâmetro β_j representa a mudança esperada na resposta y a cada mudança unitária de x_j quando quando todas as outras variáveis regressoras x_i, com $(i \neq j)$, são constantes. Por isso, os parâmetros β_j são frequentemente chamados de coeficientes parciais de regressão.

Os modelos de regressão linear múltipla são frequentemente usados como modelos empíricos ou funções aproximadas. Isso é, a verdadeira função que descreve o relacionamento entre y e $x_1, x_2, ..., x_k$ é desconhecida, mas em certos intervalos das variáveis regressoras, o modelo de regressão linear é uma aproximação adequada para a verdadeira função desconhecida.

FALAR BREVEMENTE DOS EMQ'S

3.2 Modelo Aditivo

Uma das mais populares e úteis ferramentas em análise de dados é o modelo de regressão linear. Se a dependência de de Y em X é linear ou quase linear, então o modelo de regressão linear é útil. Caso esta dependência não seja linear, não iremos querer resumi-la em uma linha reta. Poderíamos adicionar um termo quadrático, mas geralmente é dificultoso encontrar a forma mais apropriada. Nesse contexto, tem-se os modelos aditivos, que podem ser vistos como uma flexibilização do modelo de regressão linear, considerando a dependência da variável resposta com cada uma das variáveis explicativas em um contexto não paramétrico.

Para isto, considera-se que cada uma das variáveis explicativas está relacionada à média da variável resposta Y através de uma função univariada desconhecida (função suave) não especificada de uma forma paramétrica, ou seja, o componente sistemático é formado por uma soma de funções suaves não especificadas das variáveis explicativas. Esta nova classe de modelos é dita modelos aditivos e mantêm a característica dos modelos de regressão lineares de serem aditivos nos efeitos preditivos. Os modelos aditivos são um caso particular de uma classe mais geral denominada modelos aditivos generalizados (HASTIE & TIBSHIRANI, 1990). Um modelo aditivo é definido por

$$y = \alpha + \sum_{j=1}^{p} f_j(X_j) + \epsilon,$$

onde os erros ϵ são independentes, com $E(\epsilon) = 0$ e $var(\epsilon) = \sigma^2$. As $f_j s$ são funções univariadas arbitrárias, uma para cada preditor.

Os modelos aditivos mantêm muitas das boas propriedades dos modelos lineares, porém são mais flexíveis. Como visto, Uma das vantagens de modelos lineares é sua simplicidade na interpretação: caso o interesse seja em saber como a previsão muda conforme mudanças em x_j, só é necessário saber o valor de β_j, embora a função de resposta parcial f_j desempenha esse mesmo papel em um modelo aditivo.

3.3 Suavizadores

Essas funções do componente sistemático podem ser estimadas através de um suavizador (*smoother*), uma ferramenta que representa a tendência da variável resposta como função das covariáveis disponíveis. No caso em que apenas uma covariável está disponível para predizer a variável resposta, um suavizador do gráfico de dispersão é frequentemente utilizado.

Um suavizador (*smoother*) pode ser definido como uma ferramenta para resumo da tendência das medidas Y como função de uma ou mais medidas X. É importante destacar que as estimativas das tendências terão menos variabilidade que as variáveis respostas observadas, o que explica o nome de suavizador para a técnica aplicada (HASTIE & TIBSHIRANI, 1990). Chamamos a estimativa produzida por um suavizador (*smoother*) de "*smooth*". O caso de uma variável preditora é chamado de suavizador em diagrama de dispersão.

Os suavizadores possuem dois usos principais, sendo o primeiro uso a descrição. Um gráfico de dispersão suavizador pode ser usado para melhorar a aparência visual de um gráfico de dispersão de Y vs X, para nos ajudar a encontrar uma tendência no gráfico. O segundo uso é de estimar a dependência da esperança de Y com o seus preditores e nos servem como blocos de construção para os modelos aditivos.

O suavizador mais simples é o caso dos preditores categóricos, como sexo (masculino, feminino). Para suavizar Y podemos simplesmente realizar a médias dos valores de Y para cada categoria. Este processo captura a tendência de Y em X. Pode não parecer que simplesmente realizar as médias seja um processo de suavização, mas este conceito é a base para a configuração mais geral, já que a maioria dos suavizadores tenta imitar a média da categoria através da média local, ou seja, realizar a média dos valores de Y tendo os valores preditores próximos dos valores alvo. Esta média é feita nas vizinhanças em torno do valor alvo.

Nesse caso, tem-se duas decisões a serem tomadas:

- Como realizar a média dos valores da resposta em cada vizinhança;
- O quão grande esta vizinhança deve ser.

A questão de como realizar a média em uma vizinhança é a questão de qual tipo de suavizador utilizar, pois os suavizadores diferem principalmente pelo jeito de realizar as médias. O tamanho da vizinhança a ser tomada é normalmente expressa em forma de um parâmetro. Intuitivamente grandes vizinhanças irão produzir estimativas com variância pequena mas potencialmente com um grande viés e inversamente quando adotado vizinhanças pequenas. Portanto temos uma troca fundamental entre variância e viés estipulada pelo parâmetro suavizador. Este problema é análogo à questão de quantas variáveis preditoras colocar em uma equação de regressão.

3.3.1 Técnicas de suavização

Entre as principais técnicas de suavização estão a regressão paramétrica, vista anteriormente e que consiste em uma linha de regressão estimada por mínimos quadrados.

Essa abordagem pode ou não ser apropriada para dado conjunto de dados.

O suavizador bin, também conhecido como regressograma, imita um suavizador categórico, particionando os valores preditores em regiões disjuntas e então realizando a média da resposta em cada região. A estimativa final não tem uma forma bem suavizada, pois é possível ver um salto em cada ponto de corte.

A média móvel (*running mean*) é outra técnica que leva em conta o cálculo da média. É muito comum utilizar uma vizinhança/região de $(2k+1)$ observações, k para a esquerda e k para a direita de cada observação, onde o valor de k tem um comportamento de troca entre suavidade e qualidade do ajuste.

Um problema comum encontrado na média móvel é o viés. Uma saída é usar pesos para dar mais importância às vizinhanças mais próximas. Uma solução ainda melhor é utilizar a técnica de linha móvel (*running line*), na qual novamente são definidas as vizinhanças para cada ponto, tipicamente os k pontos mais próximos de cada lado. Nesse caso é mais interessante considerar a proporção de pontos em cada vizinhança, ou seja, $w = \frac{(2k+1)}{n}$, denominado *span*. Então ajusta-se uma linha de regressão aos pontos de cada região, que é usada para encontrar o valor predito suavizado para o ponto de interesse.

3.3.2 Loess

Também chamado de *Lowess*, essa técnica pode ser vista como uma linha móvel com pesos locais (*locally weighted running line*). Um suavizador desse tipo, seja denominado $s(x_0)$, usando k vizinhos mais próximos pode ser computada por meio dos seguintes passos:

- Os k vizinhos próximos de x_0 são identificados e denotados por $N(X_0)$;
- É computada a distância do vizinho-próximo mais distante de x_0:

$$\Delta(x_0) = max_{N_{x_0}}|X_0 - x_i|$$

- Pesos w_i são designados para cada ponto (N_{x_0}, usando a função de peso tri-cúbica:

$$W\left(\frac{|x_0 - x_i|}{\Delta(x_0)}\right)$$

onde

$$W(u) = \begin{cases} (1-u^3)^3, & 0 \leq u \leq 1 \\ 0, & \text{caso contrário} \end{cases}$$

- $s(x_0)$ é o valor ajustado no ponto x_0 do ajuste de mínimos quadrados ponderados de y para x contidos em $N(X_0)$ usando os pesos computados anteriormente.

As hipóteses em relação ao modelo *Loess* são menos restritivas se comparadas às do modelo de regressão linear, já que assume-se que ao redor de cada ponto x_0 o modelo deve ser aproximadamente **uma função local?**.

Destaca-se que nessa técnica deve-se ter atenção à escolha do valor do *span*. Um valor muito pequeno faz com que a curva seja muito irregular e tenha variância alta. Por outro lado, um valor muito grande fará com que a curva seja sobre-suavizada, podendo não se ajustar bem aos dados e resultando em perda de informações e viés alto. Nos passos mostrados anteriormente o valor do *span* foi escolhido através do método de vizinhos mais próximos.

3.3.3 Kernels

Um suavizador kernel usa pesos que decrescem suavemente enquanto se distancia do ponto de interesse x_0. Vários métodos podem ser chamados de suavizadores kernel através dessa definição. Porém na prática, o suavizador kernel representa a sequência de pesos descrevendo a forma da função peso através de uma função densidade com um parâmetro de escala que ajusta o tamanho e a forma dos pesos perto de x_0. Um suavizador Kernel pode ser definido da forma

$$\hat{y}_i = \frac{\sum_{j=1}^{n} y_i K\left(\frac{x_i - x_j}{b}\right)}{\sum_{j=1}^{n} K\left(\frac{x_i - x_j}{b}\right)}$$

onde b é o tamanho da vizinhança (parâmetro de escala), e K uma função kernel, ou seja, uma função densidade. Existem diferentes escolhas para K, geralmente usa-se a densidade de uma Normal, tendo-se assim um kernel Gaussiano.

3.3.4 Splines

REVER COMO FUNCIONA A ESTIMAÇÂO DOS SPLINES

Um *Spline* pode ser visto como uma função definida por um polinômio por partes. Pontos distintos são escolhidos no intervalo das observações (nós) e um polinômio é definido para cada intervalo, dessa forma é possível modelar com polinômios mais simples as curvas mais complexas. Os *splines* dependem principalmente do grau do polinômio e do número e localização dos nós.

Essa técnica é interessante pois tem uma maior flexibilidade para o ajuste dos modelos em comparação com o modelo de regressão polinomial ou linear e, após a de-

terminação da localização e quantidade de nós, o modelo é de fácil ajuste. Além disso, o *spline* permite modelar um comportamento atípico dos dados, o que não seria possível com apenas uma função.

3.3.5 Splines de regressão

Existem várias diferentes configurações para um *spline*, mas uma escolha popular é o *spline* cúbico, contínuo e contendo primeira e segunda derivadas contínuas nos nós. As *splines* cúbicas são as de menor ordem nas quais a descontinuidade nos nós são suficientemente suaves para não serem vistas a olho nu, então a não ser que seja necessário mais derivadas suavizadas, existe pouca justificativa para utilizar *splines* de maior ordem.

Para qualquer grupo de nós, o *spline* de regressão é ajustado a partir de mínimos quadrados em um grupo apropriado de vetores base. Esses vetores são as funções base representando a família do pedaço do polinômio cúbico, com valor dado a partir dos valores observados de X.

Uma variação do *spline* cúbico é o *spline* cúbico natural, que contêm a restrição adicional de que a função é linear além dos nós dos limites. Para impor essa condição, é necessário que, nas regiões dos limites: $f''' = f'' = 0$, o que reduz a dimensão do espaço de $K+4$ para K, se há K nós. Então com K nós no interior e dois nos limites, a dimensão do espaço do ajuste é de $K + 2$.

Quando trabalha-se com *splines*, existe uma dificuldade em escolher a localização e quantidade ideal dos nós, sendo mais importante o número de nós do que sua localização. Salienta-se que incluir mais nós que o necessário pode resultar em uma piora do ajuste do modelo. Existem algumas maneiras para fazer essas escolhas, como por exemplo colocar os nós nos quantis das variável preditora (três nós interiores nos três quartis).

Outro problema é a escolha de funções base para representar o *spline* para dados nós. Suponha que os nós interiores são denotados por $\xi_1 < ... < \xi_k$ e os nós dos limites são ξ_0 e ξ_{k+1}. Uma escolha simples de funções base para um *spline* cúbico é conhecida como base de séries de potência truncada, que deriva de:

$$s(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{j=1}^{K} \theta_j (x - \xi_j)_+^3$$

Onde s tem as propriedades necessárias: é um polinômio cúbico em qualquer subintervalo $[\xi_j, \xi_{j+1})$, possui duas derivadas contínuas e possui uma terceira derivada.

A função s pode ser escrita como uma combinação linear de $K + 4$ funções base $P_j(x) : P_1(x) = 1, P_2(x) = x$ e assim por diante. Cada função deve satisfazer as três condições e ser linearmente independente para ser considerada uma base. Então fica claro que são necessários $(K + 4)$ parâmetros para representar um *spline* cúbico.

As funções base B-*spline* fornecem uma alternativa numericamente superior para

a base de séries de potência truncada. A ideia principal é de que qualquer função base $B_j(x)$ é diferente de zero em um intervalo de no máximo cinco diferentes nós. Fica claro que as funções B_j são *splines* cúbicas, e são necessárias $K + 4$ delas para abranger o espaço.

A partir disso, observa-se que *splines* de regressão podem ser atrativos devido à sua facilidade computacional, quando os nós são dados. Porém a dificuldade em escolher o número e localização dos nós pode ser uma grande desvantagem da técnica.

3.3.6 Backfitting

REVER CASO FOR DEIXAR NO TRABALHO OU RETIRAR

No contexto dos modelos aditivos, quando mais de uma covariável está disponível para predizer a resposta, frequentemente utiliza-se o algoritmo retroajuste (HASTIE & TIBSHIRANI, 1987; BUJA et. al., 1989; HASTIE & TIBSHIRANI, 1990), para estimar cada função suave em um cenário não paramétrico, além do intercepto. A ideia do geral do algoritmo pode ser dado pelos seguintes passos:

1. Adotam-se os valores iniciais $\beta_0^{(0)} = \sum_{i=1}^{n}$ e $s_1^0(.) = s_2^0(.) = ... = s_p^0(.) = 0$;
2. Aplica-se um ciclo retroajuste, ou seja, para cada $j_Y = 1, 2, ..., p$, as funções $s_{jY}(.)$ são atualizadas, suavizando $y - \beta_0 - \sum_{jj \neq jY} s_{jj}^{(0)} | (z_{jj})$ por meio de algum suavizador do gráfico de dispersão, o que resulta em novas funções suaves $s_1^1(.), s_2^1(.), ..., s_p^1(.)$. Para acelerar a convergência, as funções suaves atualizadas podem ser utilizadas, por exemplo, $s_1^1(.)$ ao invés de $s_1^0(.)$ no cálculo de $s_2^1(.)$;
3. Repetem-se os passos 1 e 2 até que se obtenha a convergência.

Essa ideia é genérica, já que os detalhes diferem dependendo da técnica de suavização usada e do contexto no qual o algoritmo será utilizado.

3.4 Seleção de Modelos - Enfoque de Predição

Inserir aqui a parte do livro do Rafael sobre seleção de modelos, Data splitting, Validação cruzada.. a partir da pagina 12 do livro do Rafael - Resumir o conteúdo adequando a notação que você usou acima

4 Resultados e Discussão

4.1 Estudo de simulação

Neste seção será utilizada as simulações de dados, para gerar situações nas quais possam ser aplicadas as técnicas estudadas, analisando assim suas respectivas performances. A partir disso, iremos avaliar sob estes dados, aspectos visuais dos ajustes de todas as técnicas vistas até o momento. Realizaremos comparações , adotando diferentes tamanhos de janelas (*span*) para cada técnica e então para qual valor de *span*, temos o melhor ajuste. Em um segundo momento, classificaremos para qual técnica obtemos o melhor ajuste. Para esta etapa iremos utilizar as técnicas: Suavizadores com Kernel, *LOWESS* (Suavizador em diagrama de dispersão com pesos locais) e *Splines* de Regressão. Por fim, compararemos as performances dos ajustes do modelo linear e modelo aditivo, em relação a qualidade da predição. Para esta finalidade, utilizaremos as métricas apresentadas na Seção 3.4.

4.1.1 Cenário 1

Foram geradas 200 observações, sendo X uma sequência de 0 a 50 e Y definido pela função

$$y = 10 + 5sen\pi\frac{x}{24} + \epsilon$$

onde ϵ é um termo aleatório. Para ter uma ideia da variabilidade dos dados, foi calculado o coeficiente de variação, obtendo-se que $CV_x = 0.5817$ e $CV_y = 0.3645$. Na Figura 1 temos o comportamento dos dados juntamente com a curva:$10 + 5sen\pi\frac{x}{24}$.

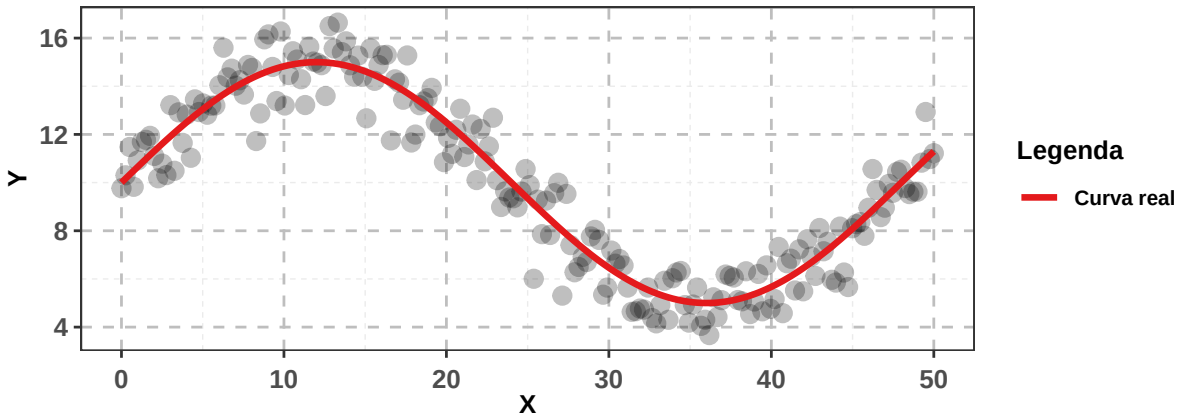

Figura 1: Gráfico de dispersão dos dados simulados e curva real

Considerando as técnicas de suavização, dependendo da escolhas dos parâmetros de suavização teremos uma curva ajustda mais irregular, ou seja, uma curva que ira interpolar uma quantidade grande de pontos, ou ainda, uma curva extremamente suavizada, que

tendera a ser uma reta. Na Figura 8, observa-se alguns ajustes realizados considerando a técnica *Bin Smoother*.

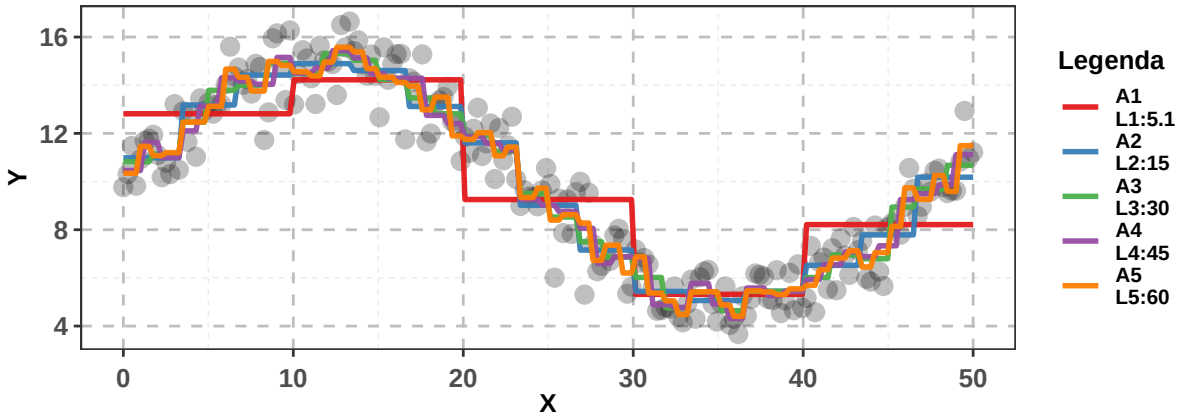

Figura 2: Alguns ajustes utilizando a técnica Bin Smoother

, tem-se o tamanho da vizinhança, ou seja, quantos pontos são considerados para fazer o ajuste em cada região, sendo indicado no gráfico por *tam*. Na Figura 2, observa-se que com o tamanho da vizinhança menor (0.5) a técnica interpola os dados, já com o valor maior nota-se que o suavizador tende a uma reta. Com o valor 11 tem-se um equilíbrio da curva, que capta a tendência dos dados. Logo, dentre os valores observados, pode-se concluir que nesse cenário os valores próximos de 11 tendem a apresentar um resultado melhor.

Nesse caso tem-se o valor do *span*, ou seja, a proporção de observações em cada vizinhança. Na Figura 3, com $span = 0.02$ é possível observar que a técnica fica bastante irregular, enquanto que com $span = 0.2$ a curva já fica mais suave e reflete bem o comportamento das observações. Nesse cenário, a técnica fica suave demais (tendendo a uma reta) quando o valor do *span* é igual a 1. Dessa forma, nota-se que o resultado é mais satisfatório para valores de *span* próximos de 0.2.

Nesse método tem-se a largura da região (*bandwidth*), indicado no gráfico por *tam*. Observando a Figura 4, como nas demais comparações até o momento, a técnica Kernel se comporta de maneira semelhante em relação à dimensão da largura nesse cenário: o menor valor resulta em uma curva extremamente irregular, enquanto o maior valor faz a curva ficar muito suave, tendendo a uma reta. Então dentro dos valores observados, percebe-se que valores próximos de 6 podem ser uma escolha interessante, resultando em uma curva suave que acompanha os dados.

Através da Figura 5, nota-se que o *spline* de regressão de grau um é extremamente desapropriado, pois a descontinuidade é muito grande. Já o *spline* cúbico segue a tendência dos pontos de uma forma suave, não sendo possível notar qualquer descontinuidade da curva nos nós. Logo, nesse cenário, o *spline* cúbico teve um resultado visualmente melhor.

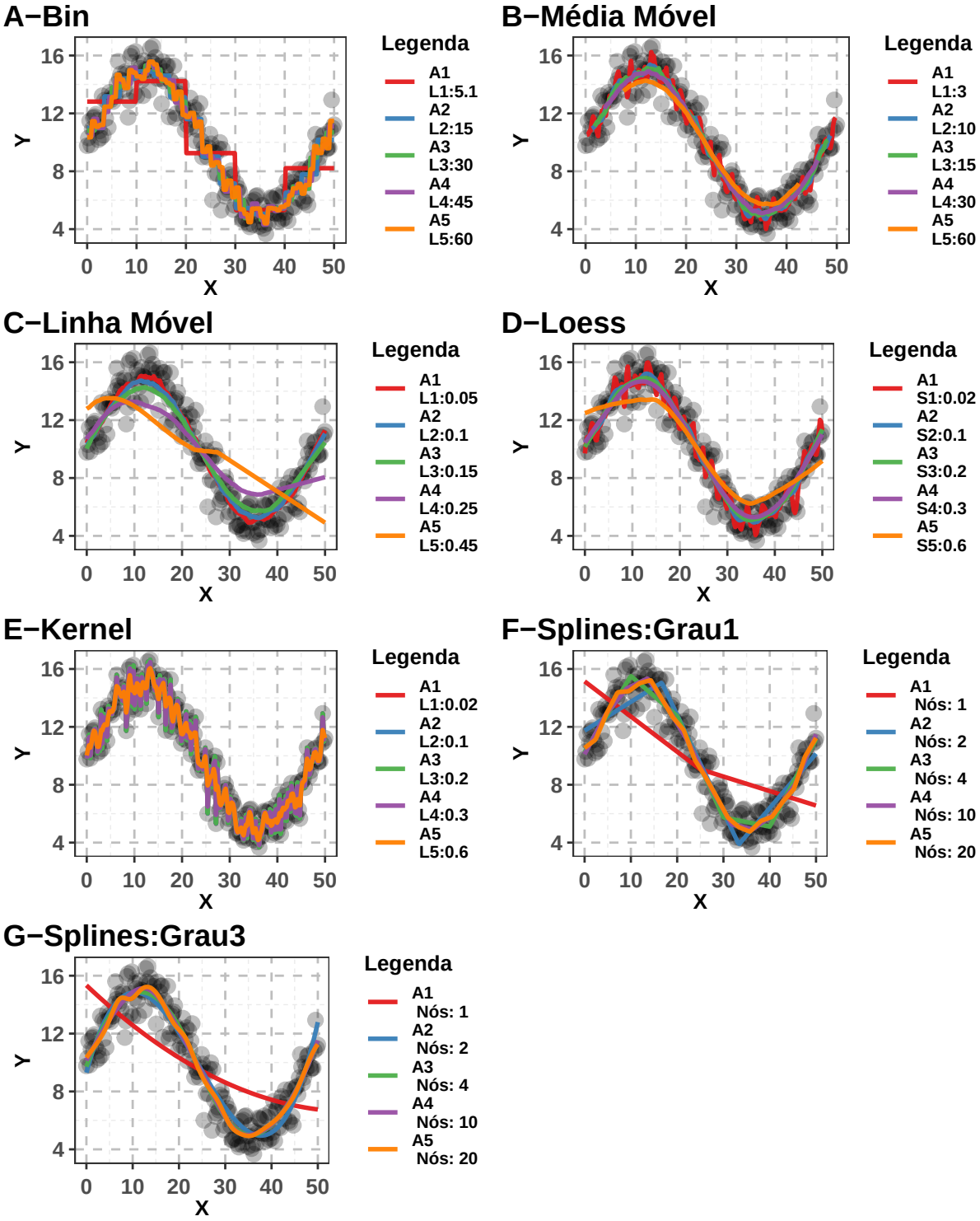

Figura 3: Comparação entre diferentes ajustes para valores de parâmetros distintos.

Vale ressaltar que em todos os métodos vistos é possível notar a relação de troca entre a variância e o viés. Para valores menores de *span* e largura/tamanho da vizinhança os pontos são interpolados (ou praticamente interpolados), o que resulta em uma curva com grande variância, porém viés pequeno, já que a curva está muito próxima de todas as observações. Por outro lado, quando esses valores aumentam, tem-se a situação inversa: a variância em relação à curva diminui, mas o viés aumenta.

Comparação entre os métodos

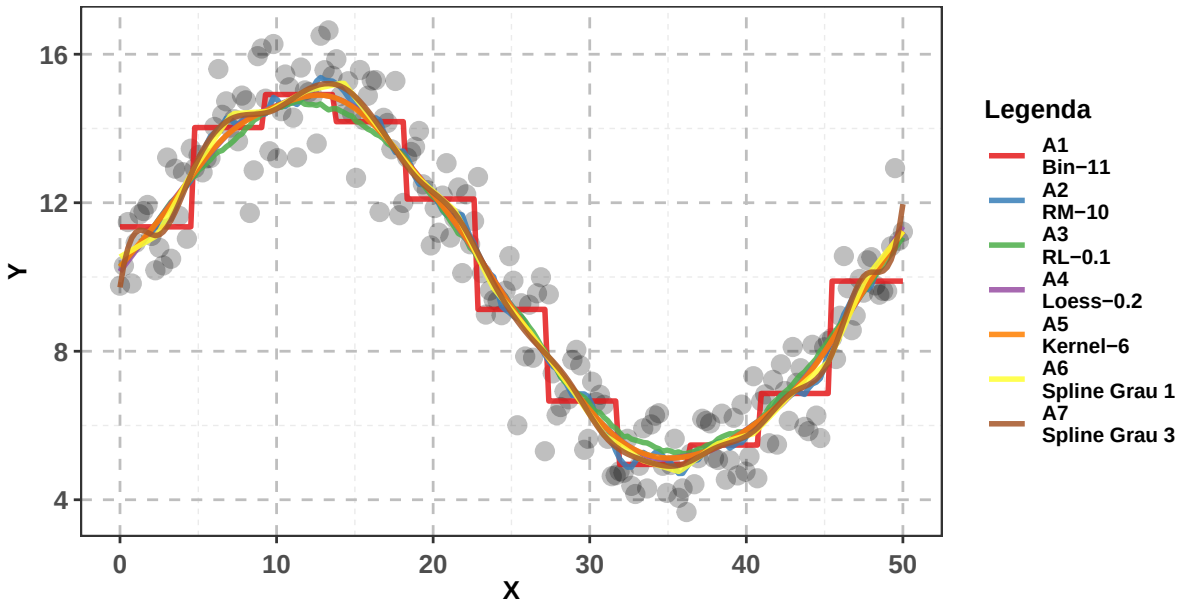

Figura 4: Comparação dos ajustes entre os métodos

Pela Figura 4, percebemos que o ajuste bin é extremamente irregular, sendo possível ver a descontinuidade da curva. A técnica kernel não tem um comportamento muito bom nas extremidades. É possivel notar que os ajustes das técnicas *loess* e *spline* cúbico ficaram muito próximas, se adequando muito bem aos dados, mas o *spline* ficou melhor, pois está captando melhor o comportamento dos dados. Portanto, para os dados gerados e dentre os métodos vistos, o *spline* cúbico mostrou um resultado mais satisfatório.

É importante notar que, para a escolha dos valores de *span* ou tamanho/largura da vizinhança e do método mais adequado para esse cenário, a análise feita foi estritamente gráfica, ou seja, visual. Existem medidas numéricas adequadas para fazer essa análise.

Tabela 2

4.1.2 Avaliando melhores parâmetros para os suavizadores

4.1.3 Leave One Out Cross Validation

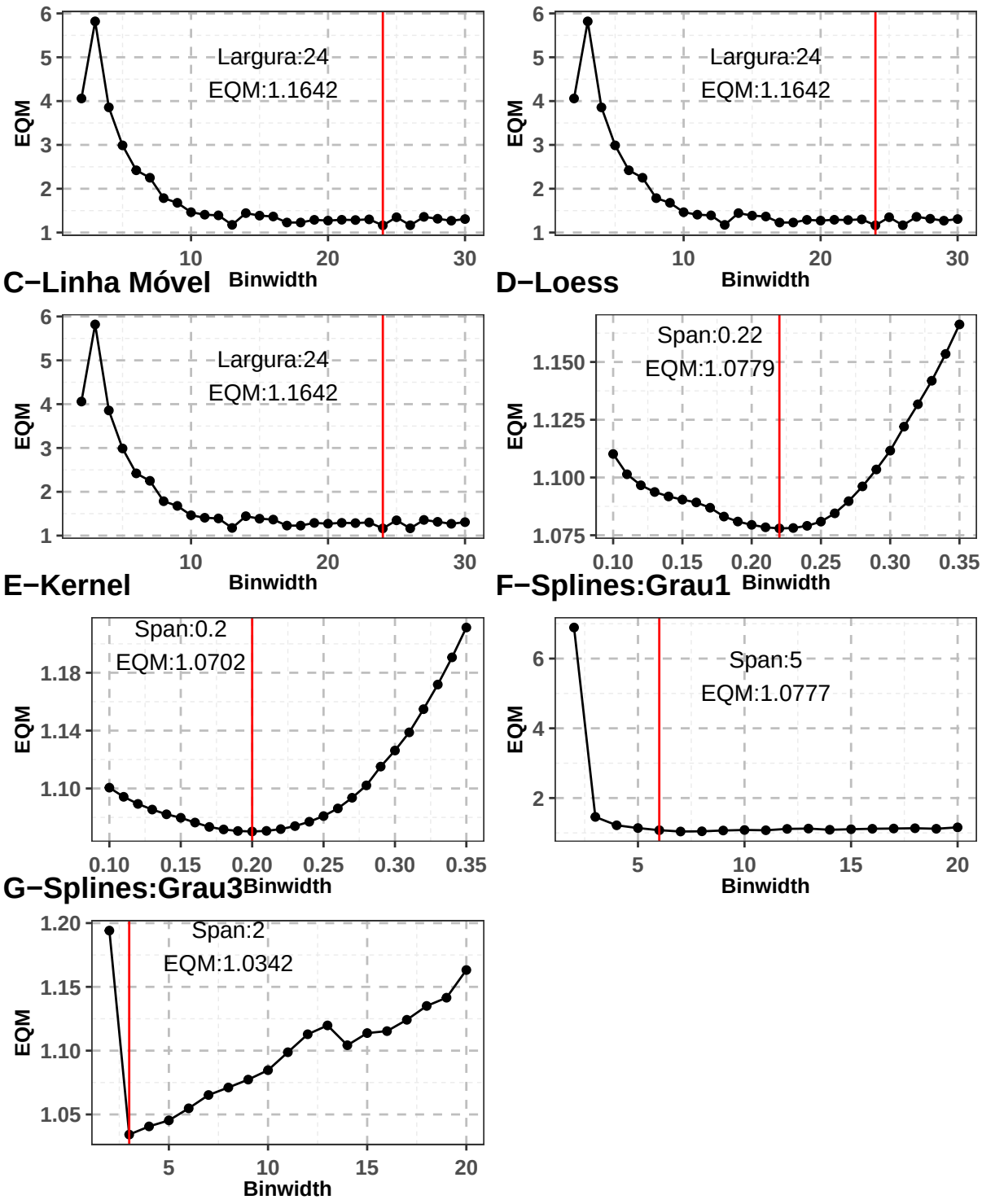

Figura 5: Comparação entre diferentes ajustes para valores de parâmetros distintos.

Tabela 1: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Smoother	EQM
Bin	1.1212351
RM	0.9594842
RL	1.0178423
Loess	0.9852521
Kernel	0.9813500
Splines Grau 1	0.9676263
Splines Grau 3	0.9564738

Tabela 2: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Smoother	EQM
Bin	1.164226
RM	1.000000
RL	1.000000
Loess	1.077945
Kernel	1.070248
Splines Grau 1	1.077697
Splines Grau 3	1.034217

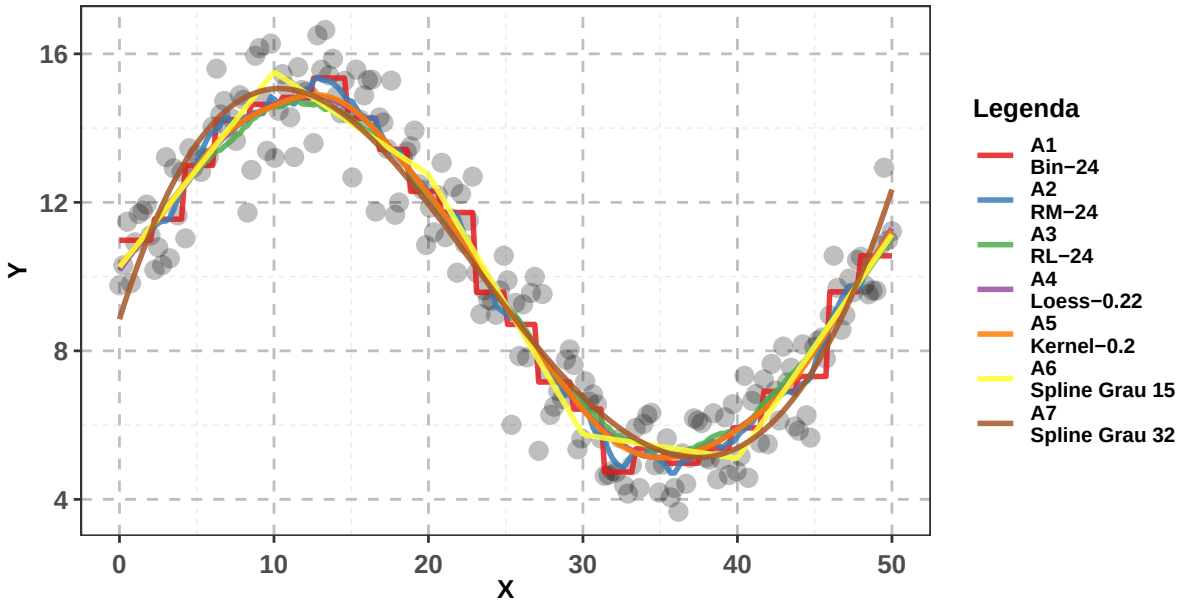

Figura 6: Comparação dos ajustes entre os métodos

4.1.4 Cenário 2

Para este cenário, foram geradas 201 observações, sendo x uma sequência de 0 à 2 com intervalos de 0.01. Ainda, temos que $y = f(x) + e$, com $f(x) \sim Gamma(6, 10)$ e e $N(0, 0.25)$. O gráfico de dispersão para estes dados pode ser verificado na Figura 7, onde podemos observar o seu comportamento.

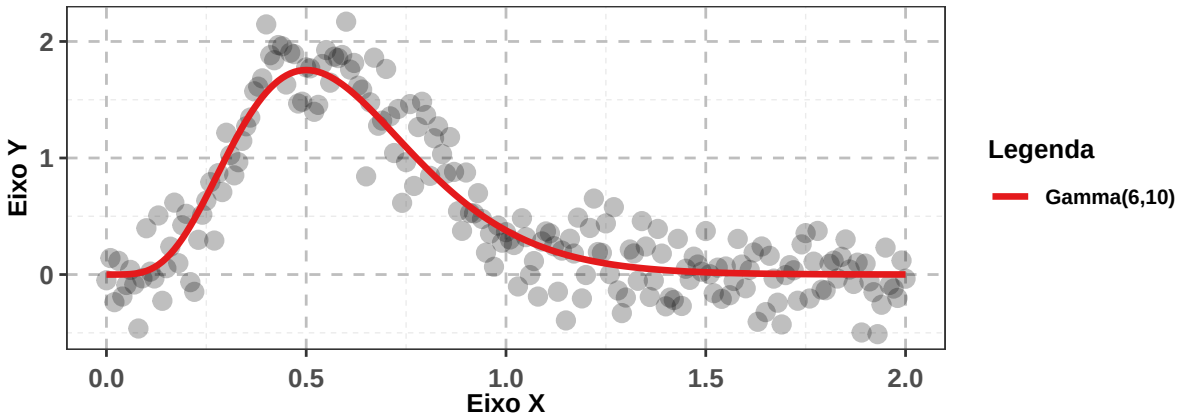

Figura 7: Gráfico de dispersão dos dados gerados para o estudo de simulação

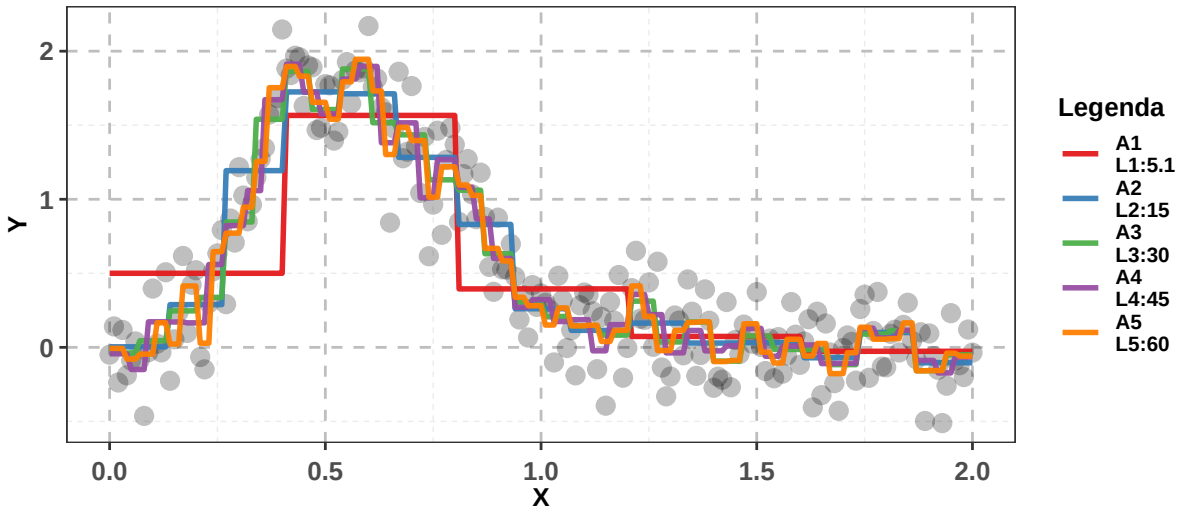

Figura 8: Alguns ajustes utilizando a técnica Bin Smoother

, tem-se o tamanho da vizinhança, ou seja, quantos pontos são considerados para fazer o ajuste em cada região, sendo indicado no gráfico por *tam*. Na Figura 2, observa-se que com o tamanho da vizinhança menor (0.5) a técnica interpola os dados, já com o valor maior nota-se que o suavizador tende a uma reta. Com o valor 11 tem-se um equilíbrio da curva, que capta a tendência dos dados. Logo, dentre os valores observados, pode-se

concluir que nesse cenário os valores próximos de 11 tendem a apresentar um resultado melhor.

Nesse caso tem-se o valor do *span*, ou seja, a proporção de observações em cada vizinhança. Na Figura 3, com $span = 0.02$ é possível observar que a técnica fica bastante irregular, enquanto que com $span = 0.2$ a curva já fica mais suave e reflete bem o comportamento das observações. Nesse cenário, a técnica fica suave demais (tendendo a uma reta) quando o valor do *span* é igual a 1. Dessa forma, nota-se que o resultado é mais satisfatório para valores de *span* próximos de 0.2.

Nesse método tem-se a largura da região (*bandwidth*), indicado no gráfico por *tam*. Observando a Figura 4, como nas demais comparações até o momento, a técnica Kernel se comporta de maneira semelhante em relação à dimensão da largura nesse cenário: o menor valor resulta em uma curva extremamente irregular, enquanto o maior valor faz a curva ficar muito suave, tendendo a uma reta. Então dentro dos valores observados, percebe-se que valores próximos de 6 podem ser uma escolha interessante, resultando em uma curva suave que acompanha os dados.

Através da Figura 5, nota-se que o *spline* de regressão de grau um é extremamente desapropriado, pois a descontinuidade é muito grande. Já o *spline* cúbico segue a tendência dos pontos de uma forma suave, não sendo possível notar qualquer descontinuidade da curva nos nós. Logo, nesse cenário, o *spline* cúbico teve um resultado visualmente melhor.

Vale ressaltar que em todos os métodos vistos é possível notar a relação de troca entre a variância e o viés. Para valores menores de *span* e largura/tamanho da vizinhança os pontos são interpolados (ou praticamente interpolados), o que resulta em uma curva com grande variância, porém viés pequeno, já que a curva está muito próxima de todas as observações. Por outro lado, quando esses valores aumentam, tem-se a situação inversa: a variância em relação à curva diminui, mas o viés aumenta.

Comparação entre os métodos

Pela Figura 10

Tabela 4

Tabela 3: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Smoother	EQM
Bin	0.2704799
RM	0.2269506
RL	0.2692138
Loess	0.2444978
Kernel	0.2435683
Splines Grau 1	0.2292791
Splines Grau 3	0.2269861

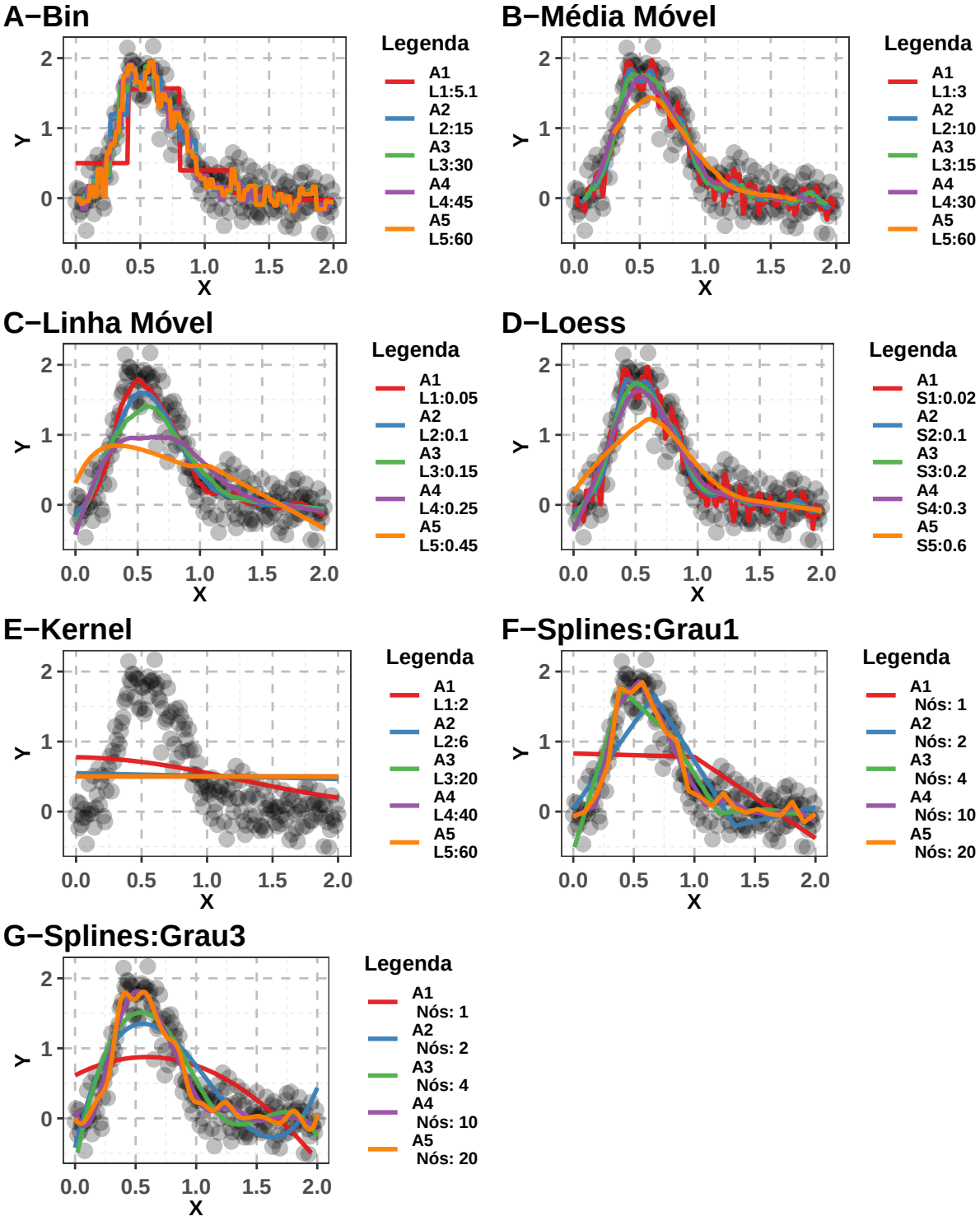

Figura 9: Comparação entre diferentes ajustes para valores de parâmetros distintos.

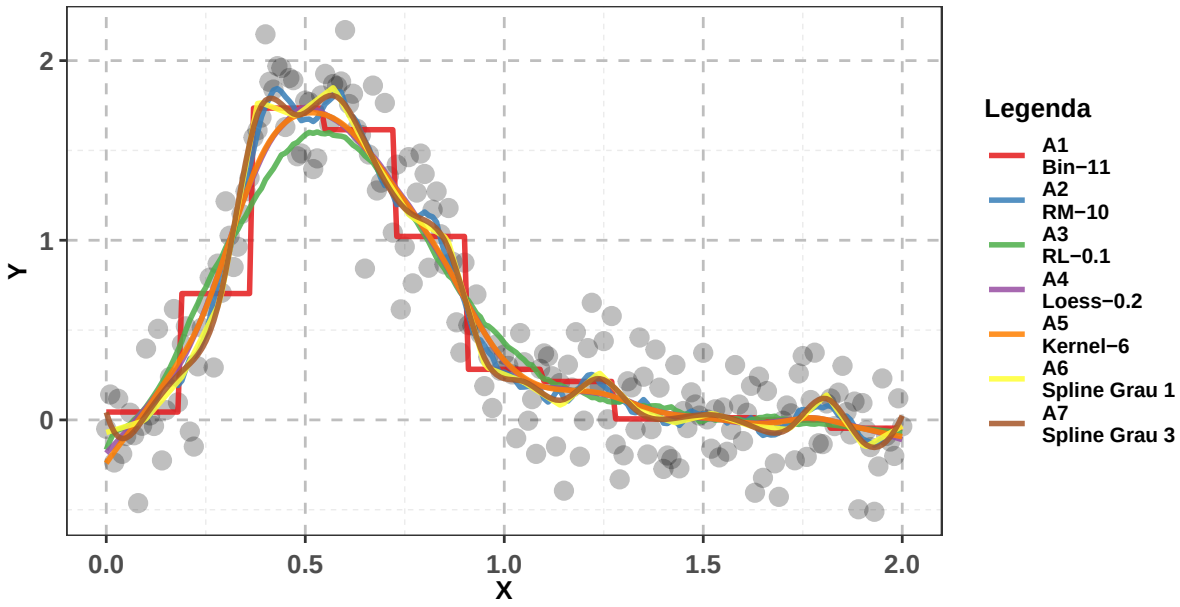

Figura 10: Comparação dos ajustes entre os métodos

4.1.5 Avaliando melhores parâmetros para os suavizadores

4.1.6 Leave One Out Cross Validation

Tabela 4: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Smoother	EQM
Bin	0.0667895
RM	1.0000000
RL	1.0000000
Loess	0.0630969
Kernel	0.0624000
Splines Grau 1	0.0639949
Splines Grau 3	0.0637379

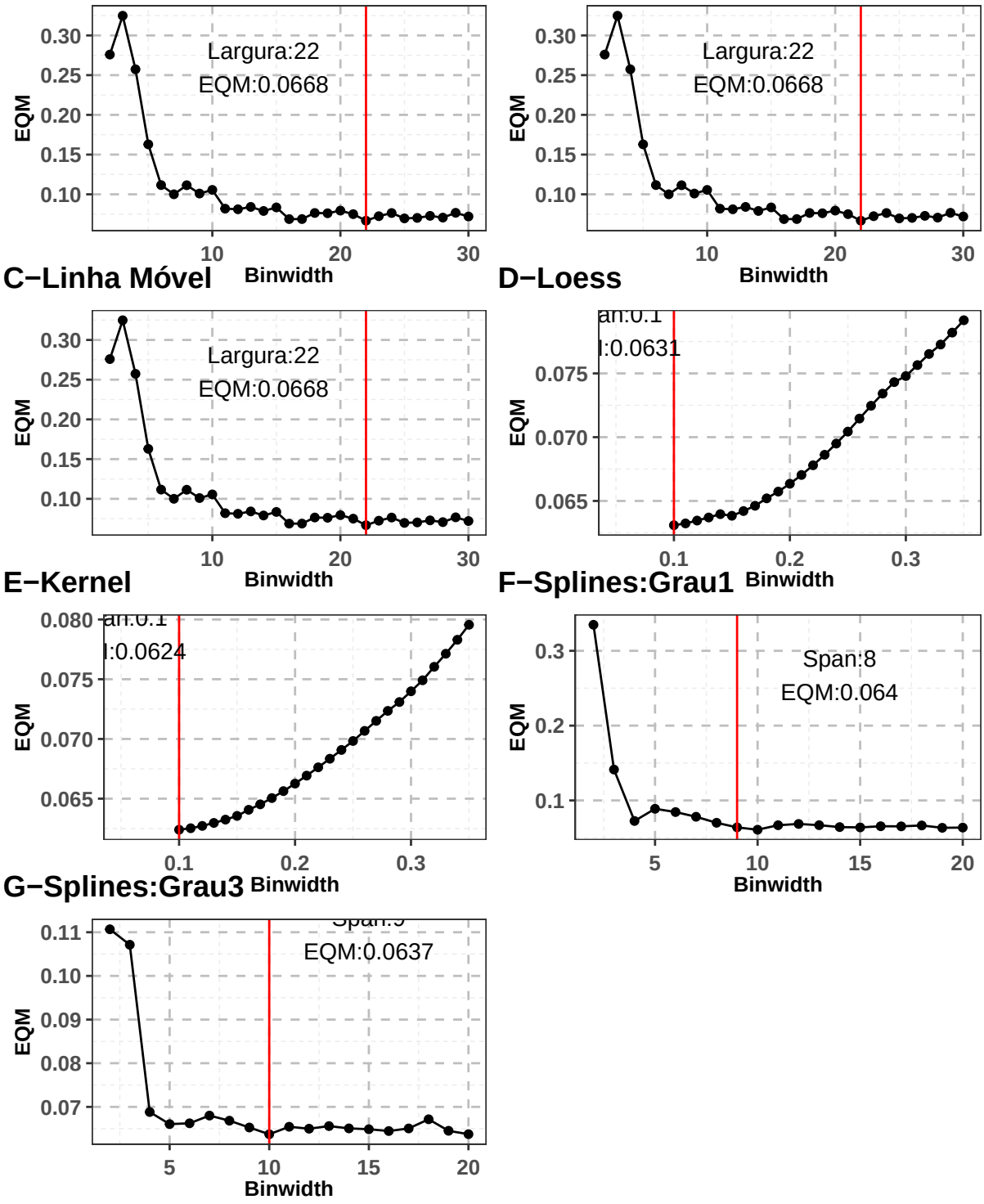

Figura 11: Comparação entre diferentes ajustes para valores de parâmetros distintos.

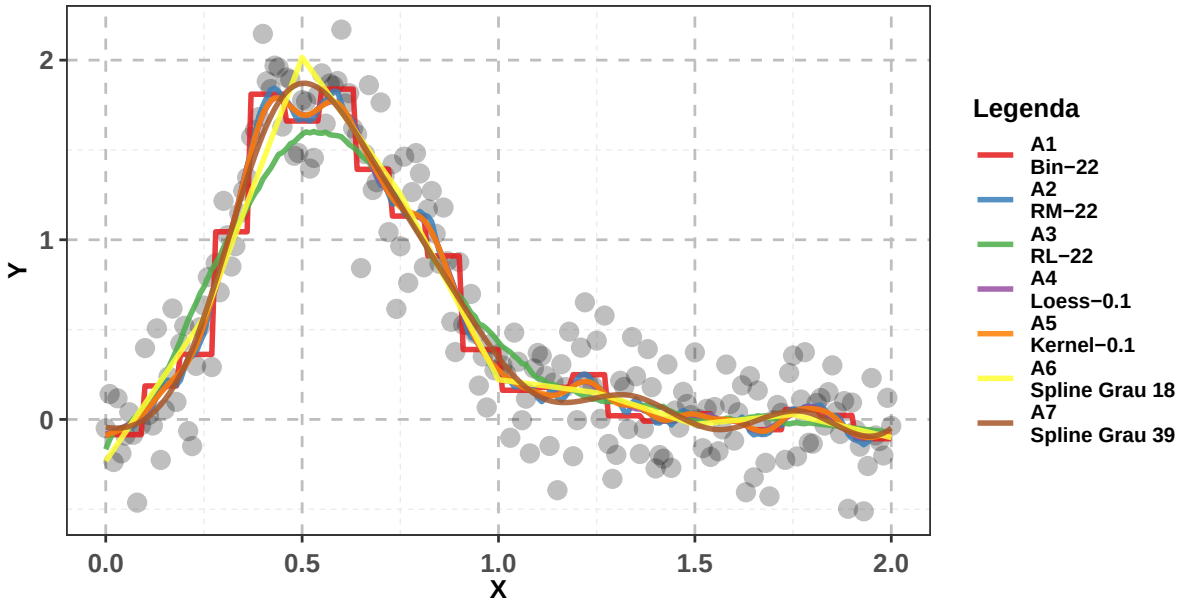

Figura 12: Comparação dos ajustes entre os métodos

5 Referências

BUJA, A., HASTIE, T. & TIBSHIRANI, R. (1989). **Linear smoothers and additive models**. The Annals of Statistics, 17, 453-510.

CLEVELAND, W. S. (1979). **Robust locally weighted regression and smoothing scatterplots**. Journal of the American Statistical Association, 74, 829-836.

HASTIE, T. J. & TIBSHIRANI, R. J. (1990). **Generalized additive models**, volume 43. Chapman and Hall, Ltd., London. ISBN 0-412-34390-8.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 5th Edition. John Wiley & Sons, 2012.

Izbicki, Rafael; Santos, Tiago Mendonça. **Aprendizado de máquina: uma abordagem estatística**. ISBN 978-65-00-02410-4.

TEAM, R. CORE. R: **A language and environment for statistical computing**. (2013).