

Aula 7 – Diagnóstico do Modelo e Análise de Resíduos (Parte 1)

Você já se perguntou se um mapa que você está usando realmente representa o terreno à sua frente? Ou se a previsão do tempo que você ouviu é confiável para planejar seu fim de semana? No mundo da modelagem estatística, especialmente com modelos de regressão, a sensação é muito parecida. Não basta apenas construir um modelo; precisamos ter certeza de que ele é um bom "mapa" da realidade e que suas "previsões" são dignas de confiança.

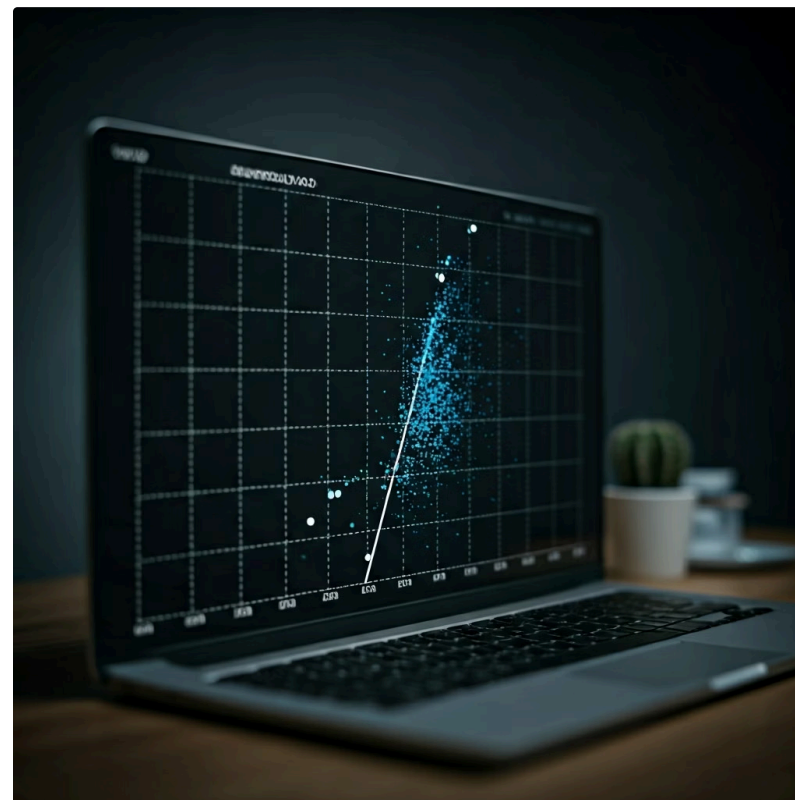
Nesta aula, embarcaremos em uma jornada crucial para qualquer cientista de dados ou estatístico: a arte e a ciência de diagnosticar modelos de regressão. Vamos desvendar os segredos por trás dos "resíduos", que são as pistas que o nosso modelo nos deixa sobre o que ele capturou e, mais importante, o que ele ainda não entendeu. É como ser um detetive que, após montar uma cena, verifica se todas as peças se encaixam perfeitamente e se não há pontas soltas.

Ao final deste encontro, você estará apto a compreender a importância fundamental da análise de resíduos, a verificar a suposição de normalidade dos erros utilizando ferramentas gráficas como histogramas e Q-Q plots, e a aplicar testes estatísticos como o de Shapiro-Wilk. Além disso, desenvolverá a habilidade de analisar graficamente os resíduos para identificar problemas de linearidade e começará a reconhecer a presença de outliers e pontos influentes, elementos que podem distorcer completamente as conclusões do seu modelo. Prepare-se para aprofundar sua compreensão e elevar a qualidade das suas análises.

O Que São Resíduos e Por Que Analisá-los?

Imagine que você está tentando prever o preço de um imóvel com base em seu tamanho. Você coleta dados de vários imóveis, ajusta um modelo de regressão e, para cada casa, o modelo te dá um preço estimado. No entanto, sabemos que o preço real de cada imóvel raramente será *exatamente* igual ao preço previsto pelo seu modelo. Essa diferença entre o valor observado (o preço real) e o valor previsto pelo modelo é o que chamamos de resíduo. Em essência, o resíduo é a "sobra", a parte da informação que o seu modelo não conseguiu explicar.

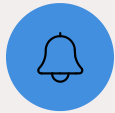
Pense nos resíduos como as "migalhas de pão" que o seu modelo deixa para trás. Se o seu modelo fosse um mapa perfeito, não haveria migalhas, ou elas seriam aleatoriamente espalhadas, sem formar nenhum padrão. Mas se as migalhas formam um caminho, ou se acumulam em certos lugares, isso é um sinal de que o mapa está incompleto ou distorcido em algumas áreas. Analisar essas migalhas é fundamental porque elas nos contam a história do que o modelo *não* capturou, revelando falhas nas suas suposições ou limitações na sua capacidade preditiva.



📌 **A análise de resíduos não é um mero detalhe técnico; é o coração da validação de um modelo de regressão.** Um modelo pode parecer ter um bom desempenho em métricas como o R^2 (coeficiente de determinação), mas se seus resíduos exibem padrões problemáticos, suas inferências e previsões podem ser completamente inválidas.

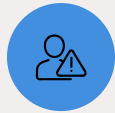
É como ter um carro que parece bonito por fora, mas que faz barulhos estranhos no motor: você pode até chegar ao seu destino, mas não com a segurança e a confiança que esperava. Ignorar os resíduos é como ignorar esses barulhos, colocando em risco a credibilidade de todo o seu trabalho.

Verificação da Suposição de Normalidade dos Erros



O que significa?

Os erros se distribuem simetricamente em torno de zero, com a maioria próximos a zero e menos erros extremos.



Por que importa?

Quando violada, a precisão das estimativas, p-valores e intervalos de confiança pode ser comprometida.



Como verificar?

Combinação de abordagens visuais (histogramas, Q-Q plots) e testes estatísticos formais.

Um dos pilares para a validade das inferências estatísticas em modelos de regressão linear (como testes de hipóteses para os coeficientes e construção de intervalos de confiança) é a suposição de que os erros do modelo seguem uma distribuição normal. Mas o que isso realmente significa na prática? Significa que, se pudéssemos observar os "erros verdadeiros" (que são inobserváveis, por isso usamos os resíduos como seus estimadores), eles se distribuiriam de forma simétrica em torno de zero, com a maioria dos erros próximos a zero e menos erros extremos.

Quando essa suposição é violada, as estimativas dos nossos parâmetros ainda podem ser válidas, mas a precisão dessas estimativas (e, conseqüentemente, os p-valores e intervalos de confiança) pode ser comprometida. É como tentar usar uma régua de borracha para medir algo: a medida pode até ser "próxima", mas você não terá certeza da sua exatidão. Por isso, verificar a normalidade dos resíduos é um passo crítico para garantir que as conclusões que tiramos do nosso modelo são estatisticamente robustas e confiáveis.

Existem diversas ferramentas para avaliar a normalidade dos resíduos, que vão desde abordagens visuais e intuitivas até testes estatísticos mais formais. A combinação dessas abordagens oferece uma visão completa e permite tomar decisões informadas sobre a adequação do modelo. Lembre-se, o objetivo não é que os resíduos sejam *perfeitamente* normais (o que é raro em dados reais), mas sim que não haja desvios *significativos* que possam invalidar as inferências.

Histograma dos Resíduos: Uma Primeira Olhada

A maneira mais direta e intuitiva de ter uma ideia da distribuição dos resíduos é através de um histograma. Depois de ajustar seu modelo de regressão, você calcula os resíduos para cada observação e, em seguida, plota a frequência com que esses resíduos caem em diferentes intervalos. O histograma nos oferece uma "fotografia" da forma da distribuição dos resíduos, permitindo-nos verificar visualmente se ela se assemelha a uma curva em forma de sino, característica da distribuição normal.



Procure por simetria

Os resíduos devem estar distribuídos simetricamente em torno de zero



Identifique o pico central

Deve haver uma concentração maior de resíduos próximos a zero



Observe as caudas

As frequências devem diminuir gradualmente nas extremidades

Ao observar o histograma, procuramos por simetria em torno de zero, um pico central e caudas que diminuem gradualmente. Se o histograma for assimétrico (enviesado para a direita ou para a esquerda), bimodal (com dois picos) ou tiver caudas muito pesadas (indicando muitos resíduos extremos), isso pode ser um sinal de que a suposição de normalidade está comprometida. É como olhar para a silhueta de uma pessoa: você pode rapidamente dizer se ela tem uma forma "típica" ou se há algo incomum.

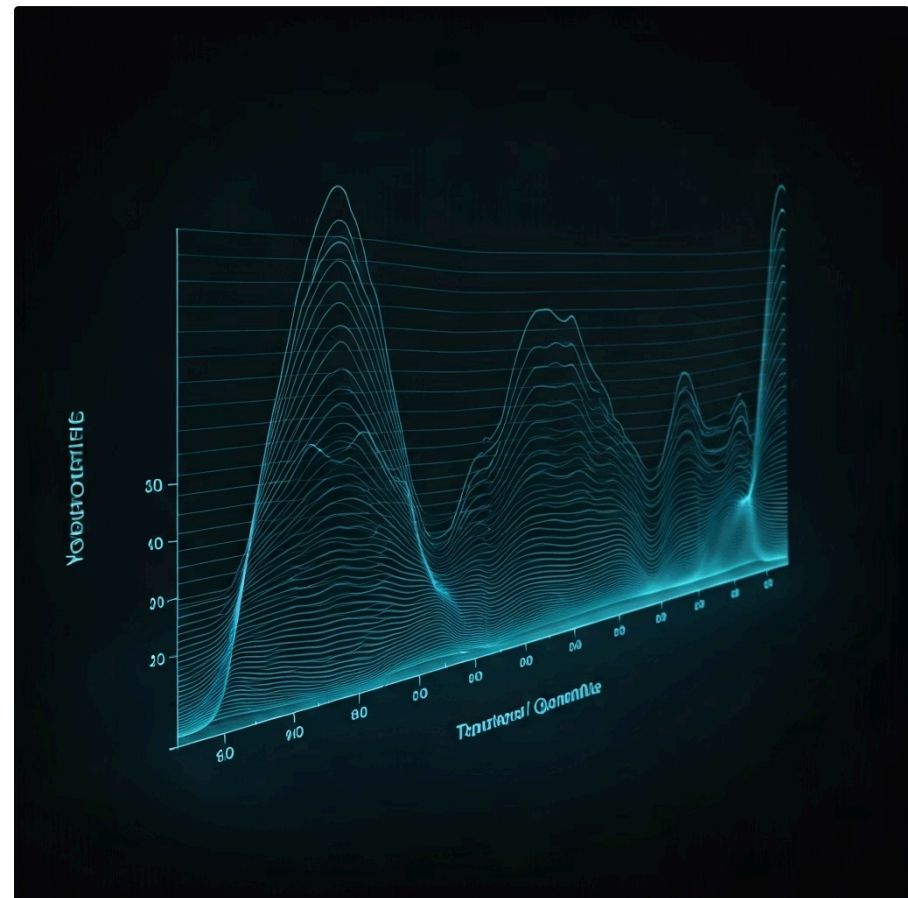
Por exemplo, se você está modelando o tempo de espera em uma fila e os resíduos do seu modelo de regressão mostram um histograma com uma cauda longa à direita, isso pode indicar que o modelo subestima frequentemente os tempos de espera mais longos. Essa visualização rápida é um excelente ponto de partida, mas é importante lembrar que a interpretação visual pode ser subjetiva e menos precisa para amostras pequenas.

Q-Q Plots: Comparando Quantis

A Ferramenta Mais Poderosa

Embora o histograma seja útil, o Q-Q plot (Quantile-Quantile plot) é uma ferramenta gráfica mais poderosa e precisa para avaliar a normalidade. Ele funciona comparando os quantis dos seus resíduos observados com os quantis teóricos de uma distribuição normal. Se os resíduos seguem uma distribuição normal, os pontos no Q-Q plot devem se alinhar aproximadamente ao longo de uma linha reta diagonal. Qualquer desvio significativo dessa linha reta sugere uma violação da normalidade.

Pense no Q-Q plot como um "teste de alinhamento". Se você tem uma fila de pessoas e quer saber se elas estão em ordem crescente de altura, você pode pedir para elas se alinharem em uma linha reta. Se a linha for perfeitamente reta e ascendente, elas estão em ordem. Da mesma forma, o Q-Q plot verifica se os resíduos estão "em ordem" de uma distribuição normal.



Curvatura nas extremidades superiores

Indica caudas mais pesadas do que o esperado para uma distribuição normal (mais resíduos extremos)

Formato em "S"

Sugere assimetria na distribuição dos resíduos

Pontos alinhados na diagonal

Indicam que a suposição de normalidade é razoável

Por exemplo, se você observa que os pontos no Q-Q plot se curvam para cima nas extremidades, isso indica que seus resíduos têm caudas mais pesadas do que uma distribuição normal (ou seja, há mais resíduos extremos do que o esperado). Essa informação é valiosa porque aponta para a natureza específica da não-normalidade, o que pode guiar a escolha de transformações nos dados ou a utilização de modelos mais robustos. A interpretação do Q-Q plot exige um pouco de prática, mas é uma habilidade essencial para qualquer analista de dados.

Teste de Shapiro-Wilk: A Prova Formal

Após as análises visuais, que são excelentes para identificar padrões e tendências, muitas vezes precisamos de uma "prova" mais formal. É aí que entram os testes de hipóteses para normalidade, como o Teste de Shapiro-Wilk. Este teste avalia a hipótese nula de que a amostra de resíduos foi extraída de uma população com distribuição normal. Em termos mais simples, ele nos ajuda a decidir se os desvios da normalidade que observamos nos gráficos são estatisticamente significativos ou apenas variações aleatórias.


$$\frac{f}{dx}$$

Cálculo da Estatística W

O teste calcula uma estatística W e um valor-p associado




Decisão Estatística

Se $p\text{-valor} < 0.05$: rejeitamos a normalidade. Se $p\text{-valor} \geq 0.05$: não rejeitamos



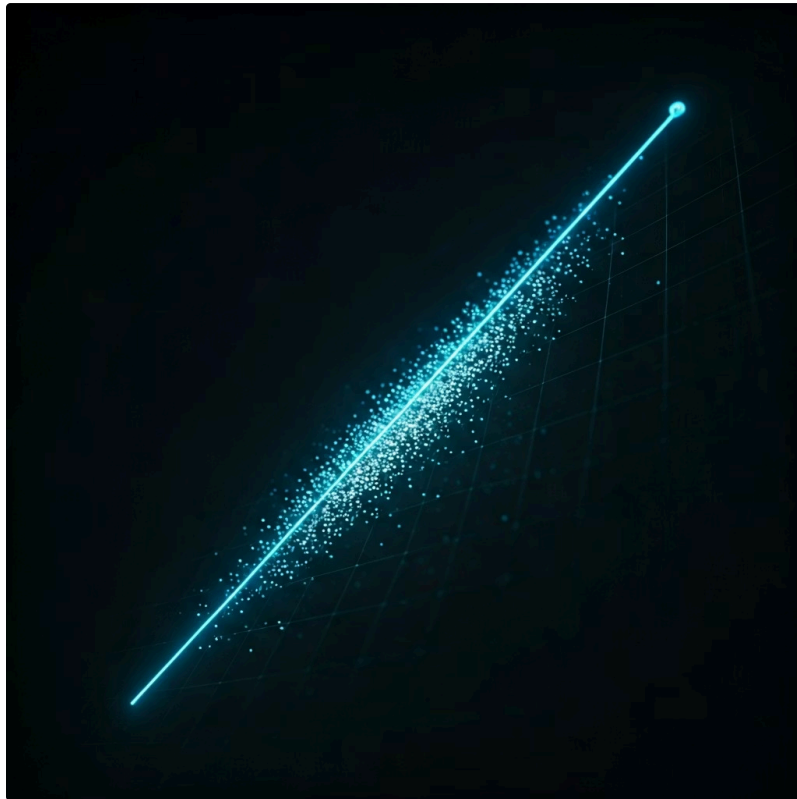
Interpretação Cuidadosa

Não rejeitar \neq confirmar normalidade, apenas indica evidências insuficientes

 **Atenção com amostras grandes:** O teste de Shapiro-Wilk pode ser muito sensível em amostras grandes e rejeitar a normalidade mesmo para pequenos desvios que não impactam significativamente as inferências. Por isso, a combinação com as análises gráficas é sempre a melhor abordagem.

É crucial entender que "não rejeitar a hipótese nula" não significa que os resíduos *são* normais, mas sim que não há evidências *suficientes* nos dados para afirmar o contrário. Pense nisso como um julgamento: se não há provas suficientes para condenar, o réu é considerado inocente, mas isso não significa que ele *não* cometeu o crime. Para amostras grandes, o teste de Shapiro-Wilk (e outros testes de normalidade) pode ser muito sensível e rejeitar a normalidade mesmo para pequenos desvios que não impactam significativamente as inferências. Por isso, a combinação com as análises gráficas é sempre a melhor abordagem.

Análise Gráfica de Resíduos para Verificar Linearidade



A Espinha Dorsal do Modelo

A suposição de linearidade é a espinha dorsal de um modelo de regressão linear: ela postula que a relação entre as variáveis preditoras e a variável resposta é, de fato, linear. Se essa suposição for violada, o modelo que ajustamos pode ser fundamentalmente inadequado para descrever os dados, levando a previsões imprecisas e inferências enganosas. É como tentar encaixar uma peça redonda em um buraco quadrado; não importa o quão forte você empurre, ela simplesmente não vai se ajustar corretamente.

Como podemos verificar essa suposição? Novamente, os resíduos são nossos melhores amigos. Se o modelo linear for apropriado, os resíduos devem ser distribuídos aleatoriamente em torno de zero, sem nenhum padrão discernível, quando plotados contra os valores preditos (ajustados) ou contra as próprias variáveis preditoras. Qualquer padrão sistemático nos resíduos é um sinal de alerta de que a linearidade pode não ser válida e que o modelo está perdendo alguma estrutura importante nos dados.

Sem padrões = Modelo adequado

Resíduos aleatoriamente espalhados indicam que o modelo linear é apropriado

Padrões sistemáticos = Alerta

Curvas ou formas específicas sugerem relações não lineares

Guia para melhorias

Os padrões sugerem transformações ou termos adicionais necessários

Essa análise gráfica é uma das mais importantes no diagnóstico de modelos, pois nos permite "ver" o que o modelo não está capturando. Ela vai além da simples verificação de suposições e nos ajuda a entender a natureza do relacionamento entre as variáveis, sugerindo possíveis melhorias no modelo, como a inclusão de termos não lineares ou a transformação de variáveis.

Resíduos vs. Valores Ajustados: O Mapa da Inadequação

Um dos gráficos mais informativos na análise de resíduos é o plot dos resíduos padronizados (ou brutos) contra os valores ajustados (ou preditos) do modelo. Neste gráfico, cada ponto representa uma observação, com seu resíduo no eixo vertical e o valor que o modelo previu para essa observação no eixo horizontal. Se o modelo de regressão linear for adequado, esperamos ver uma "nuvem" de pontos aleatoriamente espalhados em torno da linha horizontal zero, sem nenhum padrão óbvio.

1

Padrão em "U" ou "V"

Sugere que a relação entre as variáveis não é linear. O modelo está subestimando ou superestimando sistematicamente em diferentes faixas.

2

Formato de "cone" ou "funil"

Indica heteroscedasticidade - a dispersão dos resíduos aumenta ou diminui com os valores ajustados, violando a suposição de variância constante.

3

Dispersão aleatória

Pontos espalhados aleatoriamente em torno de zero indicam que o modelo está bem especificado e as suposições são atendidas.

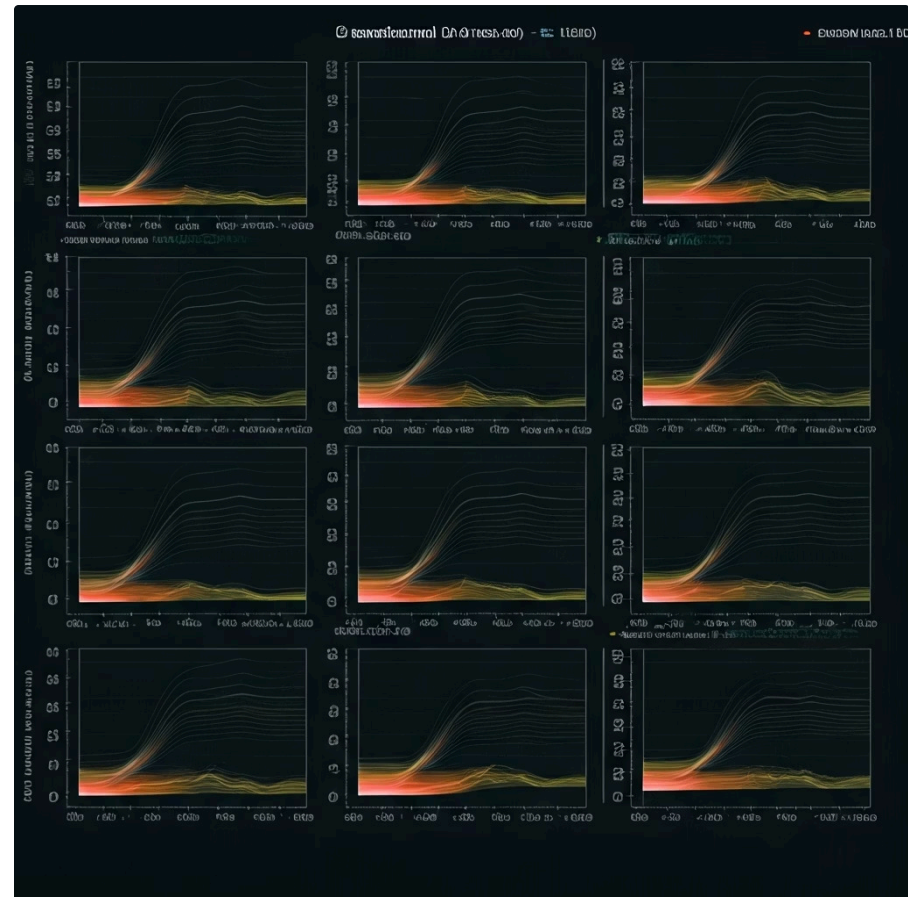
Imagine que você está jogando dardos em um alvo. Se seus dardos estão espalhados aleatoriamente ao redor do centro, você está acertando o alvo de forma consistente, mesmo que não seja perfeito. Mas se seus dardos formam um padrão específico (todos para cima e para a direita, ou em um círculo ao redor do centro), isso indica um problema sistemático na sua mira. Da mesma forma, padrões nos resíduos vs. valores ajustados revelam problemas sistemáticos no seu modelo.

Resíduos vs. Variáveis Predictoras: Olhando para Cada Peça

Análise Individual

Além de plotar os resíduos contra os valores ajustados, é igualmente importante plotá-los contra cada uma das variáveis predictoras individuais presentes no modelo. Essa análise nos permite investigar se a suposição de linearidade se mantém para cada preditor separadamente e se há alguma variável preditora que, apesar de estar no modelo, ainda apresenta uma relação não linear com os resíduos.

Se, ao plotar os resíduos contra uma variável preditora específica, você observar um padrão (como uma curva), isso pode indicar que a relação entre essa preditora e a variável resposta não é linear, e que talvez seja necessário incluir um termo quadrático ou logarítmico dessa variável no modelo. Ou, ainda, pode sugerir que uma variável importante foi omitida do modelo e que ela está correlacionada com a preditora em questão.



- ❏ **Exemplo prático:** Se você está modelando o desempenho de alunos e plota os resíduos contra a variável "horas de estudo", e vê um padrão curvilíneo, isso pode indicar que o efeito das horas de estudo não é puramente linear. Talvez, a partir de um certo ponto, mais horas de estudo não tragam o mesmo benefício incremental, ou até mesmo um benefício decrescente.

Essa análise detalhada, preditor por preditor, é como inspecionar cada componente de uma máquina complexa para garantir que todos estão funcionando como deveriam e que não há desgastes ou desajustes em partes específicas.

Identificação de Outliers e Pontos Influentes

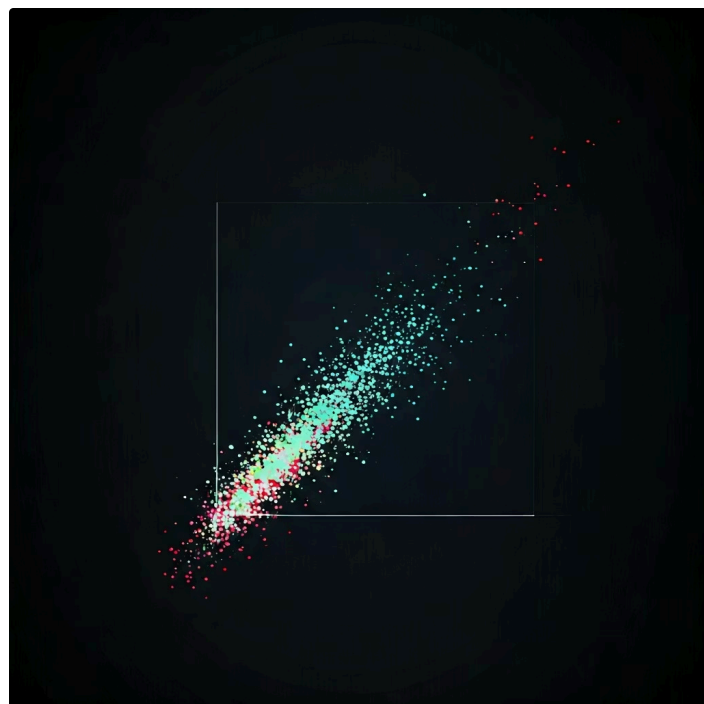
No universo dos dados, nem todas as observações são criadas iguais. Algumas se destacam, seja por serem incomuns em relação às demais ou por exercerem uma influência desproporcional sobre os resultados do nosso modelo. Essas observações especiais são conhecidas como outliers e pontos influentes, e identificá-las é crucial porque elas podem distorcer as estimativas dos parâmetros do modelo, inflar a variância e levar a conclusões errôneas. É como ter algumas peças de um quebra-cabeça que, embora pareçam pertencer a ele, na verdade são de outro jogo e, se forçadas, podem estragar a imagem final.



A presença de outliers e pontos influentes pode ser um sintoma de diversos problemas: erros de digitação ou medição, dados de uma população diferente da que estamos estudando, ou simplesmente eventos raros e legítimos. A decisão de como lidar com essas observações (seja corrigindo erros, removendo-as ou usando métodos mais robustos) deve ser tomada com cautela e baseada em um entendimento profundo do contexto dos dados. Ignorá-los pode significar construir um modelo que funciona bem para a maioria, mas falha dramaticamente para casos específicos, ou que é excessivamente sensível a poucas observações.

Nesta seção, vamos explorar como identificar esses "pontos de atenção" e entender a diferença entre um outlier (que tem um resíduo grande) e um ponto influente (que tem um impacto significativo na linha de regressão). Essa distinção é vital, pois nem todo outlier é influente, e nem todo ponto influente é um outlier no sentido tradicional.

Outliers: Os Pontos Fora da Curva



Definição e Identificação

Um **outlier** é uma observação que apresenta um resíduo muito grande em magnitude, ou seja, o valor observado está muito distante do valor previsto pelo modelo. Em termos simples, o modelo "errou feio" para essa observação. Pense em um outlier como um aluno que tirou uma nota muito acima ou muito abaixo da média da turma, destoando significativamente do desempenho geral. Embora um outlier possa ser um erro de entrada de dados, ele também pode representar um evento genuinamente raro ou uma observação de um subgrupo diferente.

Resíduos Padronizados

Resíduos brutos divididos pelo desvio padrão, tornando-os comparáveis

Resíduos Studentizados

Versão aprimorada que considera a variância individual de cada resíduo

Critério de Corte

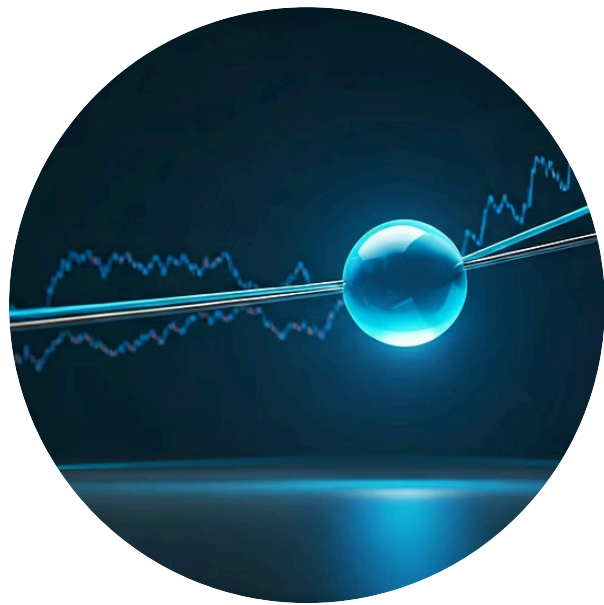
Valores que excedem ± 2 ou ± 3 desvios padrão são potenciais outliers

A identificação de outliers geralmente começa com a análise dos resíduos padronizados ou studentizados. Resíduos padronizados são os resíduos brutos divididos pelo seu desvio padrão, o que os torna comparáveis entre si. Resíduos studentizados são uma versão aprimorada, que leva em conta a variância de cada resíduo individualmente. Observações com resíduos padronizados ou studentizados que excedem um certo limite (por exemplo, ± 2 ou ± 3 desvios padrão) são consideradas potenciais outliers.

Importante: Um outlier não é necessariamente um "problema" a ser removido. Às vezes, ele pode conter informações valiosas que o modelo atual não consegue explicar. Por exemplo, se você está modelando o tempo de entrega de pacotes e encontra um outlier que representa um pacote entregue em tempo recorde, isso pode indicar uma otimização de rota ou um novo processo que vale a pena investigar, em vez de simplesmente descartar a observação. A decisão de como tratar um outlier deve ser sempre contextualizada e justificada.

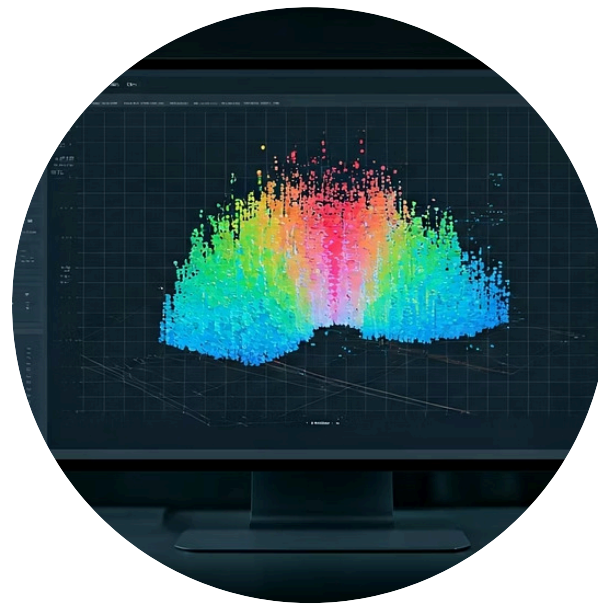
Pontos de Alavancagem (Leverage): A Força da Posição

Enquanto os outliers se destacam por terem resíduos grandes, os **pontos de alavancagem** (ou *leverage points*) são observações que possuem valores incomuns nas variáveis preditoras. Ou seja, eles estão "longe" do centro dos dados no espaço das preditoras. Imagine que você está tentando equilibrar uma gangorra: uma pessoa muito pesada sentada bem no meio não causa tanto desequilíbrio quanto uma pessoa de peso médio sentada na ponta. A posição na ponta é a "alavancagem".



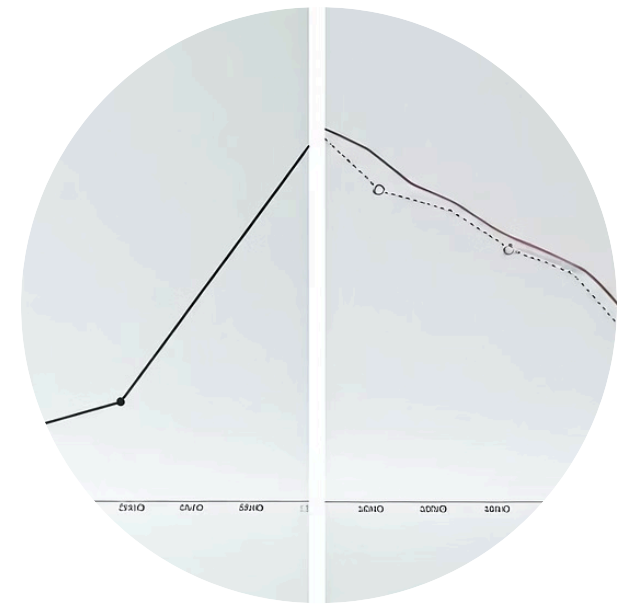
Posição Extrema

Valores incomuns nas variáveis preditoras, distantes do centro dos dados



Medida de Alavancagem

Calculada através da matriz de *hat values*, indica potencial de influência



Potencial de Influência

Pode puxar a linha de regressão mesmo sem resíduo grande

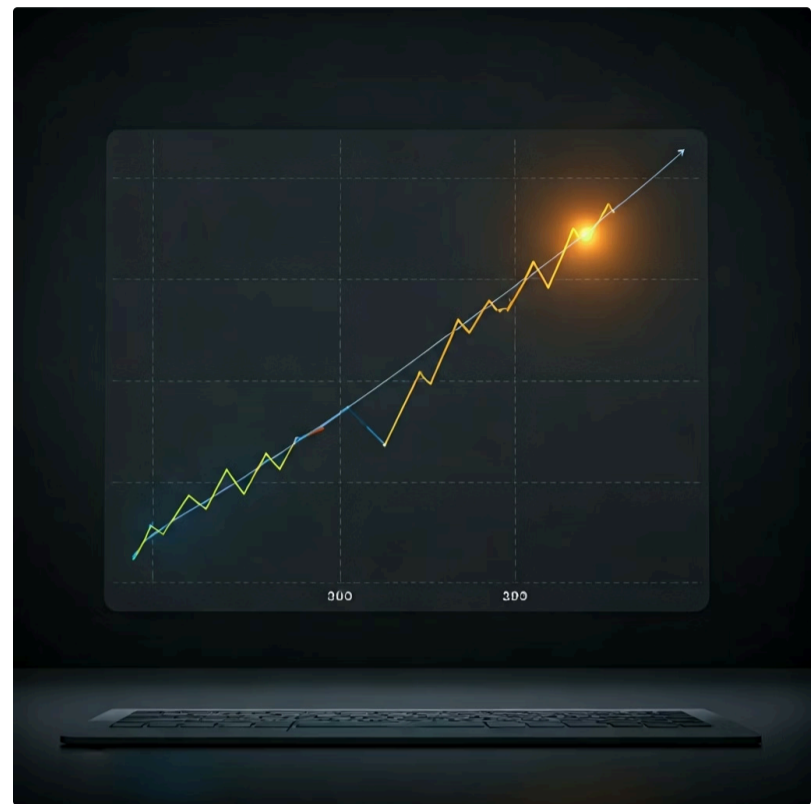
Um ponto de alavancagem tem o potencial de puxar a linha de regressão em sua direção, mesmo que seu resíduo não seja particularmente grande. Isso ocorre porque ele está em uma posição "extrema" no conjunto de dados das preditoras, dando-lhe uma influência desproporcional na determinação da inclinação da linha de regressão. A medida de alavancagem é calculada a partir da matriz de *hat values* (valores do chapéu), e valores de alavancagem altos indicam que a observação tem um grande potencial para influenciar o modelo.

Por exemplo, se você está modelando a relação entre a idade e a pressão arterial, e tem uma única observação de uma pessoa de 105 anos, enquanto a maioria dos seus dados está na faixa de 20 a 70 anos, essa observação de 105 anos terá alta alavancagem. Mesmo que a pressão arterial dessa pessoa não seja um outlier em si, sua idade extrema pode "puxar" a linha de regressão, alterando a estimativa do efeito da idade para todas as outras observações. Identificar pontos de alavancagem é crucial para entender a robustez do seu modelo.

Pontos Influentes: Outliers com Alavancagem

A Combinação Perigosa

A combinação de um resíduo grande (outlier) e uma alta alavancagem resulta em um **ponto influente**. Estes são os "verdadeiros pesos pesados" que podem alterar significativamente as estimativas dos coeficientes de regressão se forem removidos do conjunto de dados. Um ponto influente é uma observação que, se retirada, causaria uma mudança notável na linha de regressão ajustada. É como ter um jogador-chave em um time: sua presença ou ausência muda drasticamente o desempenho da equipe.



1

Distância de Cook

Métrica mais popular, quantifica impacto da remoção sobre todos os coeficientes

4/n

Regra de Bolso

Valores acima de 1 ou $4/n$ podem indicar pontos influentes

DFFITS

Outras Métricas

DFFITS e DFBETAS também ajudam a identificar influência

Para identificar pontos influentes, utilizamos medidas como a Distância de Cook, DFFITS e DFBETAS. A **Distância de Cook** é uma das métricas mais populares e quantifica o impacto de remover cada observação individualmente sobre as estimativas de todos os coeficientes do modelo. Um valor alto da Distância de Cook indica que a observação é influente. Regras de bolso sugerem que valores acima de 1, ou acima de $4/n$ (onde n é o número de observações), podem indicar pontos influentes.

- ❏ **Cuidado:** A identificação de pontos influentes não significa que você deve removê-los automaticamente. A remoção indiscriminada pode levar a um modelo que não representa a realidade completa. Em vez disso, a detecção de um ponto influente deve levar a uma investigação aprofundada: ele é um erro de dados? Ele representa uma condição única que não deveria ser generalizada? Ele indica que o modelo está mal especificado e precisa de mais variáveis ou de uma forma funcional diferente? A resposta a essas perguntas é fundamental para a construção de um modelo robusto e confiável.

Consolidação e Próximos Passos

Chegamos ao fim da primeira parte da nossa jornada pelo diagnóstico de modelos de regressão. Vimos que ir além da simples construção de um modelo e mergulhar na análise de seus resíduos é um passo indispensável para garantir a validade e a confiabilidade de nossas conclusões. Entendemos que os resíduos são as "impressões digitais" que o modelo deixa, revelando onde ele acertou e, mais importante, onde ele pode ter falhado em suas suposições. Exploramos ferramentas visuais como histogramas e Q-Q plots para verificar a normalidade dos erros, e o teste de Shapiro-Wilk para uma avaliação mais formal. Além disso, aprendemos a usar gráficos de resíduos contra valores ajustados e preditores para detectar problemas de linearidade e heteroscedasticidade, e começamos a desvendar a importância de identificar outliers, pontos de alavancagem e, principalmente, pontos influentes.

✓ **Análise de Resíduos**

Compreendemos a importância fundamental dos resíduos como diagnóstico do modelo

✓ **Verificação de Normalidade**

Dominamos ferramentas visuais e testes estatísticos para avaliar a distribuição dos erros

✓ **Linearidade**

Aprendemos a detectar problemas de linearidade através de gráficos de resíduos

✓ **Outliers e Influência**

Desenvolvemos habilidades para identificar observações problemáticas

Em prática: Ao construir seu próximo modelo de regressão, não se contente apenas com o R^2 ou os p-valores. Dedique tempo para analisar os resíduos. Use os histogramas e Q-Q plots para checar a normalidade, e os gráficos de resíduos vs. valores ajustados e preditores para verificar a linearidade. Esteja atento a padrões e pontos que se destacam, pois eles são as pistas para um modelo mais robusto e uma análise mais precisa.

📄 Na **Próxima Aula (Aula 8 – Diagnóstico do Modelo e Análise de Resíduos (Parte 2))**, aprofundaremos ainda mais neste tema. Abordaremos a detecção e tratamento de heteroscedasticidade, a multicolinearidade e suas consequências, e exploraremos estratégias para lidar com as violações das suposições, como transformações de variáveis e o uso de modelos de regressão robusta. Prepare-se para consolidar seu conhecimento e dominar as técnicas que farão de você um analista de dados mais completo e confiante.

Autoavaliação

Questão 1

Qual das seguintes ferramentas é mais eficaz para comparar os quantis dos resíduos observados com os quantis teóricos de uma distribuição normal?

- 1
- a) Histograma de Resíduos
 - b) Gráfico de Dispersão
 - c) Q-Q Plot
 - d) Box Plot

Questão 2

Um padrão em forma de "U" no gráfico de resíduos versus valores ajustados geralmente indica:

- 2
- a) Normalidade dos erros
 - b) Heteroscedasticidade
 - c) Problema de linearidade
 - d) Presença de outliers

Questão 3

Qual é a principal diferença entre um outlier e um ponto de alavancagem?

- 3
- a) Outliers têm resíduos pequenos, pontos de alavancagem têm resíduos grandes.
 - b) Outliers são incomuns na variável resposta, pontos de alavancagem são incomuns nas variáveis preditoras.
 - c) Outliers sempre influenciam o modelo, pontos de alavancagem nunca.
 - d) Não há diferença, são termos sinônimos.

Questão 4

O Teste de Shapiro-Wilk resulta em um valor-p de 0.01. Considerando um nível de significância de 0.05, qual a conclusão correta?

- 4
- a) Os resíduos seguem uma distribuição normal.
 - b) Não há evidências para rejeitar a normalidade dos resíduos.
 - c) Há evidências para rejeitar a normalidade dos resíduos.
 - d) O teste é inconclusivo.

Gabarito

1. c) Q-Q Plot
2. c) Problema de linearidade
3. b) Outliers são incomuns na variável resposta, pontos de alavancagem são incomuns nas variáveis preditoras.
4. c) Há evidências para rejeitar a normalidade dos resíduos.

Questão Discursiva

Explique a importância de analisar os resíduos de um modelo de regressão, detalhando como a detecção de padrões nos gráficos de resíduos pode guiar melhorias no modelo.

Recursos Adicionais



Livro "Análise de Regressão Linear" (Montgomery, Peck & Vining)

Para aprofundamento teórico e exemplos práticos sobre diagnóstico de modelos e análise de resíduos.



Documentação da biblioteca statsmodels ou scikit-learn (Python)

Para explorar a implementação prática das técnicas de diagnóstico apresentadas nesta aula.



Artigos recentes sobre "Robust Regression" (2023-2025)

Para entender abordagens mais modernas para lidar com violações de suposições em modelos de regressão.

- ❑ **NOTA IMPORTANTE:** As informações regulatórias/legais/técnicas desta aula estão atualizadas até 2025. Consulte sempre fontes oficiais para verificar alterações.

