

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE ECONOMIA

**Impacto Localizado de Novas Estações Meteorológicas na
Produtividade Agrícola: Uma Abordagem de Tratamento
Deslocado e Emparelhamento Dinâmico**

DANIEL CAVALLI
DRE 117038439

PROFESSOR: Prof.º. Eduardo Pontual Ribeiro

OUTUBRO 2024
[ACESSO AO CÓDIGO](#)

Resumo/Objetivo

Este artigo investiga o impacto causal da instalação de novas estações meteorológicas sobre a produtividade agrícola em regiões rurais. Partindo da hipótese de que informações climáticas mais precisas e localizadas melhoram a tomada de decisão no campo, adotamos um modelo de Diferenças em Diferenças (DiD) com adoção escalonada (staggered adoption), conforme proposto por Callaway & Sant'Anna (2020), para identificar o efeito do tratamento ao longo de múltiplos períodos e grupos. A análise considera o período pré-tratamento para validar o pressuposto de tendências paralelas e o período pós-tratamento para mensurar os ganhos em produtividade. Os resultados mostram, inicialmente, que antes da instalação das estações não havia diferenças sistemáticas entre os grupos, reforçando a credibilidade da estratégia identificadora. Após a intervenção, detecta-se um efeito positivo e estatisticamente significativo, situando o incremento da produtividade entre 9,4% e 17,7%. Desse modo, o artigo contribui para a compreensão do papel das informações agrometeorológicas na agricultura, orientando políticas públicas e decisões de investimento em infraestrutura meteorológica voltadas ao aumento sustentável da produção agrícola.

1. Introdução

A produtividade agrícola é influenciada pelas condições meteorológicas, que afetam diretamente o desenvolvimento das culturas e a eficácia das práticas agrícolas. Estudos como o de Monteiro et al. (2009) apontam que a variabilidade da produção agrícola global é amplamente atribuída às flutuações climáticas durante o ciclo de cultivo, o que evidencia a importância de um sistema de monitoramento climático robusto. Nesse sentido, a agrometeorologia, ciência que estuda as interações entre fatores meteorológicos e agrícolas, tem ganhado relevância como uma ferramenta estratégica para promover práticas mais eficientes e sustentáveis no campo. Com o aumento da população mundial e a consequente pressão sobre os recursos naturais, a adoção de práticas agrícolas que maximizem a produtividade enquanto minimizam os impactos ambientais tornou-se essencial (Monteiro et al., 2009).

1.1 O papel da agrometeorologia na produtividade agrícola

A agrometeorologia desempenha um papel crucial na agricultura ao fornecer informações meteorológicas aplicadas diretamente às necessidades dos cultivos. Esse campo integra dados climáticos e meteorológicos com parâmetros específicos das culturas, permitindo a antecipação dos efeitos do clima sobre as práticas agrícolas e possibilitando decisões mais informadas e eficientes. Como apontado por Rijks e Baradas (2000), os Serviços Nacionais de Meteorologia contribuem significativamente para a economia agrícola ao divulgar essas informações e facilitar seu uso eficiente, ajudando a mitigar riscos e aumentar a produtividade.

Segundo Mavi e Tupper (2004), as informações agrometeorológicas são aplicáveis em três áreas principais: no planejamento agrícola, na tomada de decisões táticas e na resiliência dos sistemas agrícolas. No planejamento, esses dados ajudam na escolha das épocas e locais mais adequados para o cultivo, considerando o macroclima e as condições específicas de cada região. Essa etapa é essencial para ajustar as atividades agrícolas ao contexto climático regional, reduzindo desperdícios e promovendo o uso sustentável dos recursos.

No contexto tático, as informações meteorológicas auxiliam na determinação dos melhores momentos para práticas agrícolas como a irrigação, semeadura e colheita, o que contribui para uma execução mais precisa e eficiente das operações. Essas decisões são ainda

mais importantes em áreas de cultivo de sequeiro, onde a dependência da precipitação é alta, e os dados sobre previsão de chuva e evapotranspiração são fundamentais para otimizar o uso dos recursos hídricos (Pereira et al., 2002).

Com a instalação de novas estações meteorológicas em áreas rurais, as informações climáticas se tornam mais precisas e localizadas, permitindo que os agricultores ajustem suas práticas de acordo com as condições específicas de suas regiões (Weiss et al., 2000). Sistemas de Informação Agrometeorológica, como o AGRITEMPO da EMBRAPA e o SISDAGRO do INMET, utilizam dados dessas estações meteorológicas para fornecer previsões de safra e orientações sobre manejo de recursos hídricos, ajudando os agricultores a tomar decisões informadas sobre a época de plantio, irrigação e controle fitossanitário (Weiss et al., 2000). Essas informações são cruciais para o presente estudo, pois evidenciam que dados meteorológicos detalhados e acessíveis podem impactar diretamente a produtividade agrícola.

1.2 O papel da econometria

Apesar da importância dos dados climáticos para a agricultura, ainda são poucos os estudos que investigam empiricamente o impacto direto da instalação de novas estações meteorológicas na produtividade agrícola. A hipótese subjacente ao presente estudo é que as estações meteorológicas, ao prover informações climáticas localizadas, geram benefícios mais significativos para as regiões próximas, onde as previsões se tornam mais precisas. Para validar a hipótese, este trabalho adota uma abordagem econométrica de Diferenças em Diferenças (DiD) com tratamento escalonada, para identificar os efeitos locais da instalação de estações meteorológicas em municípios tratados, isto é, localizados próximos a essas novas instalações.

A decisão pelo método escalonado serve para conferir rigor ao estudo por permitir construir grupos que compartilham características semelhantes antes do tratamento, isolando o efeito específico das novas estações na produtividade agrícola, enquanto minimiza possíveis vieses de seleção para os novos locais de instalação.

Ao fornecer evidências empíricas sobre o papel das estações meteorológicas na agricultura, este estudo pretende contribuir para o desenvolvimento de políticas públicas que considerem a expansão da infraestrutura meteorológica como um fator estratégico para o crescimento sustentável e econômico do setor agrícola. A instalação de estações

meteorológicas adicionais em áreas rurais pode, potencialmente, fortalecer a resiliência agrícola frente às variabilidades climáticas e fomentar práticas de Agricultura de Precisão em regiões onde tais tecnologias ainda são subutilizadas.

2. Especificação do Modelo

Para este trabalho, utilizaremos como principal referência o artigo de Callaway & Sant'Anna (2020), que apresenta uma extensão do modelo de Diferenças em Diferenças (DiD) para cenários com múltiplos períodos e momentos distintos de adoção do tratamento.

2.1 Staggered Adoption

A abordagem de Diferenças em Diferenças (DiD) proposta por Callaway & Sant'Anna (2020) oferece um arcabouço flexível para analisar situações em que unidades recebem um tratamento ao longo de múltiplos períodos, e não todas no mesmo instante. Diferente do DiD comum, que se concentra tradicionalmente em apenas dois períodos (pré e pós-tratamento) e dois grupos (tratado e controle), a extensão proposta considera diversos grupos que podem iniciar o tratamento em momentos distintos, bem como múltiplos períodos de observação. Esse modelo torna-se particularmente relevante em estudos que avaliam o impacto de intervenções ou choques econômicos que ocorram de forma escalonada (staggered adoption), permitindo lidar melhor com a heterogeneidade entre grupos e com problemas de interpretação associados ao uso de modelos de efeitos fixos bidimensionais (Two-Way Fixed Effects, TWFE).

2.2 Introdução ao Modelo

No DiD clássico, assume-se um grupo tratado que recebe a intervenção em um momento específico e um grupo controle que nunca é tratado. Sob essa configuração, a diferença no tempo entre pré e pós-tratamento e a diferença entre grupos tratado e controle fornecem a estimativa do efeito causal. Entretanto, para o caso que estamos tratando nesse trabalho existem múltiplos períodos e vários grupos recebendo o tratamento em momentos distintos ao longo dos 22 anos do período de tratamento. A abordagem de DiD tradicional, nesse caso, pode gerar estimativas enviesadas devido à heterogeneidade do tratamento ao longo do tempo, resultando em interpretação ambígua.

O modelo de Callaway & Sant'Anna (2020) surge como uma forma de permitir que esses cenários de tratamento escalonado, talvez muito mais comuns no mundo real do que experimentos naturais, possam ser avaliados. Por permitir a identificação de efeitos médios do tratamento específicos para cada grupo e período, acomoda a heterogeneidade do momento de adoção e suas dinâmicas, além de fornecer uma interpretação mais clara dos parâmetros causais.

2.3 Fundamentos do modelo

O modelo proposto pode ser entendido em três etapas conceituais:

1. **Identificação de parâmetros causais desagregados:** Primeiro, são obtidas estimativas do efeito causal para cada combinação de grupo tratado e período após a adoção (denotados por $ATT(g, t)$), focando em captar o efeito específico para um determinado conjunto de unidades tratadas em um dado momento do tempo.
2. **Agregação desses parâmetros:** Em seguida, esses parâmetros individuais, definidos para grupos e períodos específicos, podem ser combinados para produzir medidas resumidas de efeitos, como efeitos médios globais, ao longo do tempo, por coorte de tratamento ou segundo o tempo decorrido desde a intervenção.
3. **Estimação e inferência:** Por fim, procedimentos estatísticos são empregados para estimar esses parâmetros, bem como inferir sobre sua significância estatística.

2.3.1 Group-Time Average Treatment Effects $ATT(g,t)$

O parâmetro fundamental dessa abordagem é o $ATT(g, t)$, que representa o Efeito Médio do Tratamento para o grupo g no período t . Ao contrário do DiD tradicional, onde há um único efeito estimado, aqui obtemos uma coleção de efeitos, cada um refletindo o impacto do tratamento em um grupo que começou a ser tratado em um determinado momento e está sendo avaliado em um período específico após o início do tratamento.

Com isso é possível capturar heterogeneidades relacionadas:

- Ao grupo (unidades diferentes podem ter características e contextos distintos);
- Ao momento de início do tratamento (tratamentos iniciados em diferentes épocas podem ter efeitos variados devido a condições econômicas, políticas ou sociais);
- Ao tempo decorrido desde o tratamento (efeitos imediatos versus efeitos de longo prazo podem diferir).

2.3.2 Identificação

O artigo de Callaway & Sant'anna (2020) trás uma série de “Assumptions”(Pressupostos daqui pra frente) para identificação dos parâmetros causais. Boa parte delas não difere muito dos pressupostos do DiD tradicional. Abaixo destaco algumas importantes mudanças:

1. **Tendências Paralelas Condicionais:** A ideia central do DiD é que, na ausência de tratamento, as unidades tratadas seguiriam a mesma tendência de evolução dos resultados das unidades não tratadas. Existem diferenças conceituais entre o DiD tradicional e o DiD Staggered:
 - **Pressuposto 4 - "never-treated"** : Aqui, o grupo de comparação é formado por unidades que nunca recebem tratamento ao longo de todo o período observado. Pressupõe-se que, condicionalmente a covariáveis observáveis, esses "never-treated" representam a contrafactual apropriada para o que teria acontecido com os grupos tratados caso não tivessem sido tratados.
 - **Pressuposto 5 - "not-yet-treated"**: Nesse caso, o grupo de controle para um determinado período e grupo tratado é formado por unidades que ainda não foram tratadas até aquele momento, mas que virão a ser tratadas no futuro. Essa abordagem aproveita a natureza escalonada do tratamento para criar um grupo de comparação internamente consistente.
2. **Pressuposto 3 - Antecipação Limitada do Tratamento:** Admite-se que as unidades não são afetadas pelo tratamento antes de sua efetiva implementação, ou que se conheçam efeitos de antecipação limitados e controláveis. Caso haja antecipação, o modelo permite incorporar essa informação, desde que os períodos de antecipação sejam conhecidos e adequadamente modelados.
3. **Sobreposição (Overlap):** É necessário que haja sobreposição entre as características das unidades tratadas e as unidades de controle, garantindo que as diferenças observadas possam ser atribuídas ao tratamento e não a dessemelhanças estruturais entre grupos.

2.3.3 Estimação

Para estimar o $ATT(g,t)$, são propostas três abordagens principais:

1. **Outcome Regression (OR):** Modela-se diretamente o resultado nos grupos de controle, condicionando a covariáveis pré-tratamento. O efeito é então obtido

comparando a predição contrafactual com o resultado efetivo observado nas unidades tratadas.

2. **Inverse Probability Weighting (IPW):** Aqui, pondera-se cada unidade pela probabilidade condicional de tratamento. Ao ajustar esses pesos, obtém-se um contrafactual equilibrado, simulando um cenário onde o tratamento foi aplicado aleatoriamente.
3. **Doubly Robust (DR):** Combina OR e IPW, resultando em um estimador robusto a erros de especificação. Mesmo se um dos modelos (outcome ou probabilidade) estiver incorretamente especificado, a consistência pode ser mantida. Na prática, essa abordagem é muitas vezes recomendada por oferecer maior segurança em cenários reais, onde a especificação perfeita do modelo é incerta.

2.3.4 Agregação de Efeitos

Depois de estimar uma coleção de $ATT(g, t)$, é possível agregá-los de diferentes maneiras, fornecendo visões mais resumidas e interpretáveis:

- **Por tempo de exposição ao tratamento (Event Study):** Consolida-se o $ATT(g, t)$ em função do tempo decorrido após o início do tratamento. Isso permite visualizar a dinâmica do efeito: se ele cresce, diminui ou se mantém estável ao longo dos períodos pós-tratamento.
- **Por grupo de tratamento:** Agrupar $ATT(g, t)$ por coortes de adoção do tratamento, possibilitando examinar heterogeneidades entre diferentes grupos que adotaram o tratamento em momentos distintos.
- **Por tempo calendário:** Examina-se o impacto agregado em determinados períodos, independentemente do tempo de exposição, auxiliando a entender efeitos conjunturais.
- **Como um efeito médio global (Overall Treatment Effect):** Por fim, é possível sintetizar todos os efeitos $ATT(g, t)$ em um único parâmetro médio, oferecendo uma visão geral do impacto da intervenção ao longo do tempo e grupos.

2.4 Conclusão

A abordagem apresentada por Callaway & Sant'Anna (2020) não se reduz a uma única equação final, pois seu objetivo é oferecer uma estrutura flexível para estimar efeitos causais médios (ATT) específicos para cada grupo e período, além de permitir a agregação desses

efeitos de diferentes maneiras. No entanto, ela pode ser representada pela seguinte equação genérica:

$$\text{EQ [1]: } \theta = \sum_{g \in G} \sum_{t=2}^T w(g, t) \cdot ATT(g, t)$$

onde:

- o $ATT(g, t)$ é o Efeito Médio do Tratamento para a coorte g no período t
- $w(g, t)$ são funções de ponderação escolhidas pelo pesquisador (seção 3.1.1 do artigo), conhecidas ou estimáveis a partir dos dados, que determinam a importância relativa de cada $ATT(g, t)$ na composição do efeito agregado θ

3 Resultados Empíricos

A abordagem utilizada segue o arcabouço desenvolvido por Callaway e Sant’Anna (2020, 2021), que se baseia na estimação de efeitos médios do tratamento específicos para cada coorte de adoção (grupo que começa a ser tratado em um determinado período) e cada momento no tempo. Com esse método, primeiro são estimados os Efeitos Médios do Tratamento por Grupo e Tempo $ATT(g, t)$. Em seguida, tais efeitos são agregados de modo a fornecer uma medida do impacto médio em função do tempo desde o início do tratamento, criando assim o Estudo de Evento.

A ideia central é definir o “tempo relativo” ao tratamento. Se g representa o ano em que um grupo começa a receber o tratamento, e t é um ano qualquer após a implementação, então $e = t - g$ indica quantos períodos se passaram desde que o tratamento foi adotado. Por exemplo, se $e = 0$ representa o ano exato do início do tratamento, então $e = 1$ significa um período após a implementação.

3.1 Equação de Event Study

O modelo proposto por Callaway & Sant’anna pode tomar diversas formas. Para esse trabalho, foi utilizado o método de Event Study por termos a possibilidade efeitos contínuos e crescentes do tratamento. Ao contrário da análise de Diferenças em Diferenças (DiD) tradicional, que geralmente assume um efeito constante do tratamento após sua implementação, o Estudo de Evento permite examinar *se e como* o efeito evolui ao longo do

tempo. Essa abordagem permite identificar, por exemplo, se o impacto surge imediatamente, cresce de forma cumulativa, se estabiliza ou até mesmo desaparece após determinado período.

Nesse caso, $w(g, t)$ da equação 1 passa a ser denotado como:

$$\text{EQ [2]: } 1 \{g + e \leq T\} P(g | G + e \leq T)$$

Com isso, a agregação dos efeitos para o Estudo de Evento pode ser expressa da seguinte forma:

$$\text{EQ [3]: } \theta_{es}(e) = \sum_{g \in G} 1 \{g + e \leq T\} P(g | G + e \leq T) \cdot ATT(g, g + e)$$

Onde:

- $\theta_{es}(e)$ é um efeito médio do tratamento e períodos após sua introdução, considerando todas as coortes que podem ser observadas por pelo menos e períodos após o início do tratamento.
- O uso de pesos $P(g | G + e \leq T)$, onde G é o momento do início do tratamento na unidade, garante que cada grupo contribua de maneira proporcional à sua presença e tamanho na amostra
- O indicador $1 \{g + e \leq T\}$ assegura que apenas sejam incluídos grupos que tenham períodos observáveis menor do que o período final da amostra de dados ($T = \text{último ano dos dados}$).

3.2 Metodologia e Contextualização

Para implementação empírica, utilizou-se o software R (R Core Team, 2024) e o pacote 'did' desenvolvido pelos próprios autores Callaway e Sant'Anna, que disponibiliza as ferramentas necessárias para a estimação do DiD com múltiplos períodos e grupos de tratamento, bem como a computação do estimador DR.

É importante ressaltar que essa abordagem requer condições específicas para identificar efeitos. Em particular, a existência de um grupo controle apropriado (unidades nunca tratadas ou não tratadas até aquele momento) e a presença de períodos pré-tratamento suficientes e comparáveis são fundamentais. Caso essas condições não sejam atendidas, pode haver

dificuldades na identificação dos efeitos, resultando em valores ausentes (NA) ou indefinidos (NaN) nos parâmetros estimados.

Esta situação é observada nos resultados, especialmente para o grupo tratado em 2008 e 2016, onde não foi possível identificar os efeitos devido a limitações nos dados ou violação das condições necessárias para identificação.

3.3 Resultados

Os resultados são apresentados tanto na forma de efeitos específicos por grupo e tempo $ATT(g, t)$ quanto em uma análise de event-study, que agrega os efeitos em relação ao tempo até/depois do tratamento. O efeito médio geral estimado, na agregação por grupos, foi de 0,1274 com erro padrão de 0,015 e estatisticamente significativo a 95% de confiança, indicando um impacto positivo médio do tratamento sobre a produtividade.

3.3.1 Análise dos Efeitos por Grupo-Tempo

A análise por grupos de tratamento revela heterogeneidade importante na identificação e magnitude dos efeitos. Para o grupo de 2008, onde o tratamento iniciou muito cedo, e para o grupo 2016, que possui poucos exemplares, não foi possível identificar efeitos, resultando em estimativas com valores NA ou NaN em todos ou na maioria dos períodos.

A partir da coorte de 2017, observam-se resultados mais robustos. Para este grupo, destacam-se efeitos positivos e estatisticamente significantes nos períodos pós-tratamento, variando de 0,1724 a 0,2931. Em particular, o efeito imediato no ano de tratamento (2017) foi o mais expressivo (0,2931), seguido por uma estabilização em torno de 0,19-0,20 nos anos subsequentes, todos estatisticamente significantes.

A coorte de 2018 apresenta um padrão mais heterogêneo, com efeitos significativos em períodos específicos do pré-tratamento (2003: 0,1160; 2009: 0,2016) e efeitos negativos em outros (2004: -0,2121; 2010: -0,2403; 2015: -0,0936). Esta variabilidade sugere possível presença de tendências específicas a este grupo que merecem investigação adicional para identificar a causa dessas diferenças antes do tratamento.

3.3.2 Análise de Event Study

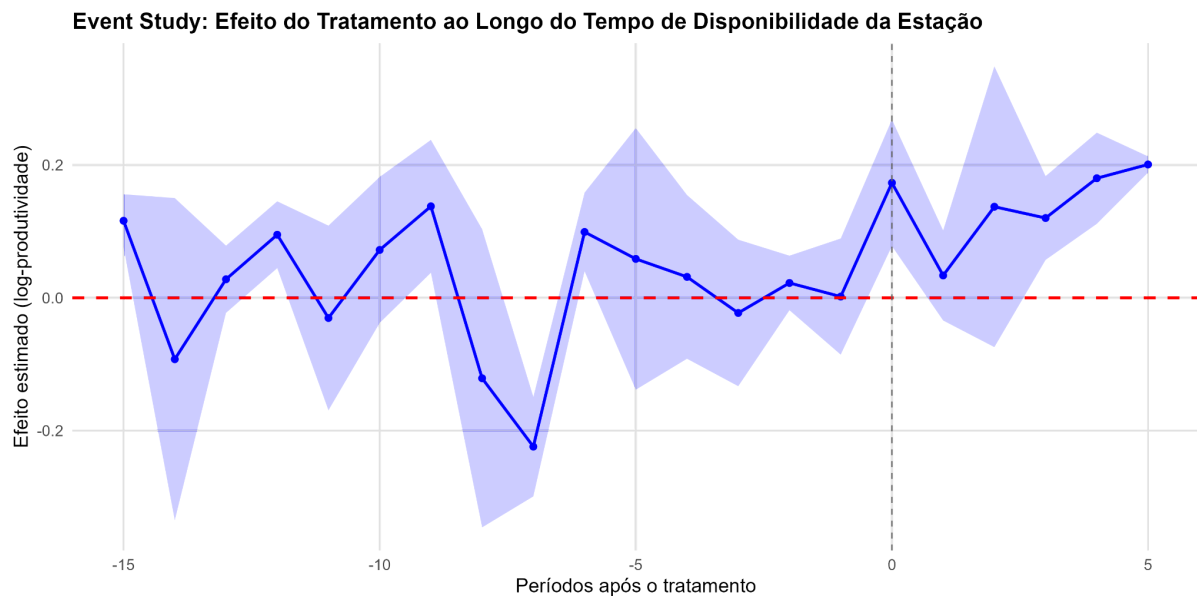


Figura 1: Efeitos do tratamento ao longo do tempo de disponibilidade da estação

A Figura 1 apresenta os resultados do event study, mostrando as estimativas pontuais (linha azul) e intervalos de confiança (área sombreada) para períodos antes e após o início do funcionamento da estação meteorológica automática. O tempo zero representa o momento de ativação da estação, com períodos negativos indicando anos anteriores à ativação e períodos positivos representando anos posteriores.

A análise detalhada dos resultados do event study revela padrões importantes em três momentos diferentes do tratamento:

1. Período Pré-Tratamento (-15 a -1):

- A média dos efeitos pré-tratamento é de 0.011, próxima a zero, com desvio padrão de 0.124, indicando substancial variabilidade
- Apenas dois períodos apresentam significância estatística robusta:
 - $e = -15$: (ATT: 0.116, IC: [0.026; 0.206])
 - $e = -7$: (ATT: -0.224, IC: [-0.394; -0.055])
- A ausência de um padrão sistemático nos efeitos pré-tratamento e sua média próxima a zero sugerem que, embora existam oscilações, não há evidência forte de violação da hipótese de tendências paralelas

2. Efeito Imediato e Transição (0 a 2):

- O período inicial mostra um efeito positivo de 0.1732, embora não estatisticamente significativo no intervalo de confiança simultâneo
- Os dois períodos seguintes apresentam variabilidade considerável:
 - $ATT\ e = 0$: 0.1732 (IC: [-0.0424; 0.3887])
 - $ATT\ e = 1$: 0.0337 (IC: [-0.1198; 0.1871])
 - $ATT\ e = 2$: 0.1372 (IC: [-0.3409; 0.6153])
- A amplitude dos intervalos de confiança neste período sugere maior incerteza durante a fase de adaptação ao tratamento, talvez pela adoção gradual da informação pelos produtores locais

3. Efeitos de Médio Prazo (3 a 5):

- A partir do terceiro ano de tratamento, os efeitos se estabilizam em valores positivos e crescentes:
 - $ATT\ e = 3$: 0.1202 (IC: [-0.0225; 0.2629])
 - $ATT\ e = 4$: 0.1801 (IC: [0.0248; 0.3354])
 - $ATT\ e = 5$: 0.2008 (IC: [0.1732; 0.2284])
- A redução na amplitude dos intervalos de confiança ao longo do tempo, sobretudo a partir do quinto ano, indica uma consolidação mais nítida do impacto do tratamento
- A tendência observada (Figura 1) reforça a interpretação de que o efeito positivo não apenas se sustenta, mas também se fortalece com o passar dos anos

O efeito médio agregado do tratamento é estimado em 0,1409 com erro padrão: 0,0604 e intervalo de confiança de 95% [0,0225; 0,2592], estatisticamente significativo. A comparação entre as médias dos períodos pré (0,011) e pós-tratamento (0,134) demonstra uma mudança relevante no nível da variável de resultado após a implementação da política.

3.3.3 Testes e conclusão

Para validar formalmente a hipótese de tendências paralelas e a estratégia de identificação e robustez do efeito do tratamento, foram realizados testes estatísticos formais sobre as médias dos efeitos médios do tratamento (ATT) nos períodos pré e pós-tratamento. O procedimento consistiu em:

1. **Teste t para efeito pré-tratamento:**

Antes da implementação da intervenção, calculou-se a média dos ATT ao longo do período pré-tratamento e seu desvio-padrão. Em seguida, aplicou-se um teste t bilateral para verificar se essa média difere estatisticamente de zero. Esse teste verifica a hipótese nula de que não há efeito (média igual a zero) contra a hipótese alternativa de que o efeito seja diferente de zero. Além disso, construiu-se um intervalo de confiança de 95% para essa média.

2. **Teste t para efeito pós-tratamento:**

O mesmo procedimento foi repetido para o período pós-tratamento. Ou seja, calculou-se a média dos ATT após a intervenção, aplicou-se um teste t bilateral para checar a significância estatística dessa média em relação a zero, e elaborou-se o intervalo de confiança de 95%.

Os resultados desses testes foram bastante esclarecedores. No período pré-tratamento, o *teste t* não detectou qualquer efeito estatisticamente significativo e o p-valor elevado (juntamente com o intervalo de confiança incluindo o zero) indicam que não havia diferença entre as tendências dos grupos antes do tratamento. Esse resultado **reforça a plausibilidade do pressuposto de tendências paralelas**. Já no período pós-tratamento, o teste t resultou em um p-valor extremamente baixo, permitindo rejeitar a hipótese nula de efeito zero com alto grau de confiança. O intervalo de confiança estritamente positivo, situando a média estimada entre aproximadamente 9,4% e 17,7%, sugere que a intervenção exerceu um impacto causal substancial sobre a produtividade.

A combinação de tendências paralelas validadas estatisticamente no período pré-tratamento com efeitos positivos e crescentes após a intervenção fornece evidência robusta para uma interpretação causal dos resultados. O resultado do modelo sugere que a instalação das estações meteorológicas automáticas foi bem-sucedida em seu objetivo de aumentar a produtividade agrícola, com impactos positivos e estatisticamente significantes

4 Comentários ao Professor

Os resultados do novo modelo foram muito diferentes do modelo na entrega anterior. Quatro coisas foram importantes para essa mudança:

1. Mudança de valores absolutos de produtividade para base log: As novas estações de 2016 em diante aumentaram a capilaridade da rede meteorológica brasileira e consequentemente microrregiões com baixa produtividade foram adicionadas. Isso significava que, para o período com a melhor qualidade dos dados, o modelo estava fazendo comparações fora de escala
2. Correções no dataset: Existia um erro no dataset original onde microrregiões que nunca tiveram estações estavam sendo excluídas. Na realidade existem diversos exemplos de microrregiões sem estações
3. Foco em um único tipo de plantação: Foquei apenas em cana de açúcar, um tipo de lavoura de sequeiro que é muito sensível a precipitação/umidade do solo, maximizando o impacto de uma boa previsão

A última e talvez mais importante foi o aumento da minha familiaridade com o pacote `did` do Callaway & Sant'anna. O pacote é razoavelmente complexo de usar, mal documentado e o artigo deles é bem complicado de entender também. Uma coisa que me ajudou muito foi que encontrei um [guia do próprio Callaway](#) explicando como o pacote funciona e como ele conversa com o artigo. Fiz algumas mudanças no estimador do modelo, trocando métodos e criando o event-study com as ferramentas do próprio pacote. Só isso já trouxe uma melhora dos resultados bem considerável, apesar de ainda estarem ruins antes das mudanças listadas acima.

Com isso o modelo parece ter performado bem melhor e realmente encontrei alguns resultados muito interessantes. Parece que existe sim uma correlação entre ter mais estações e aumento da produtividade mas tem alguns caveats nos dados que talvez voltem a jogar lama nessa ideia. Por exemplo, até ~2015 a decisão por implementar novas estações parece estar bem ligada com a produtividade local pois todos tratados durante essa época produzem muito mais que os não tratados. Após isso os dois grupos se aproximam mais.

Referências

- [1] MONTEIRO, José Eduardo B. A. (org.). *Agrometeorologia dos cultivos: o fator meteorológico na produção agrícola*. Brasília, DF: INMET, 2009. 530 p. il. ISBN 978-85-62817-00-7.
- [2] WEISS, A.; VAN CROWDER, L.; BERNARDI, M. *Communicating agrometeorological information to farming communities*. *Agricultural and Forest Meteorology*, v. 103, p. 185-196, 2000
- [3] MAVI, H. S.; TUPPER, G. J. *Agrometeorology – Principles and application of climate studies in agriculture*. New York: Food Products Press, 2004.
- [4] PEREIRA, A. R.; ANGELOCCI, L. R.; SENTELHAS, P. C. *Agrometeorologia – fundamentos e aplicações práticas*. Guaíba: Agropecuária, 2002.
- [5] CALLAWAY, Brantly; SANT'ANNA, Pedro H. C. Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, v. 225, n. 2, p. 200-230, 2021. DOI: <https://doi.org/10.1016/j.jeconom.2020.12.001>

ANEXO 1 - Resultados gerais do modelo

```
> summary(att_results)
```

Call:

```
att_gt(yname = "log_produtividade", tname = "ano", idname = "id_microrregiao",  
       gname = "gname", xformula = ~qtd_estacoes_ativas + precipitacao_por_area_plantada,  
       data = df, est_method = "dr")
```

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna. "Difference-in-Differences with Multiple Time Periods." Journal of Econometrics, Vol. 225, No. 2, pp. 200-230, 2021.

<<https://doi.org/10.1016/j.jeconom.2020.12.001>>, <<https://arxiv.org/abs/1803.09015>>

Group-Time Average Treatment Effects:

Group Time ATT(g,t) Std. Error [95% Simult. Conf. Band]

2008 2002	NA	NA	NA	NA
2008 2003	NaN	NaN	NaN	NaN
2008 2004	NaN	NaN	NaN	NaN
2008 2005	NaN	NaN	NaN	NaN
2008 2006	NaN	NaN	NaN	NaN
2008 2007	NaN	NaN	NaN	NaN
2008 2008	NaN	NaN	NaN	NaN
2008 2009	NaN	NaN	NaN	NaN
2008 2010	NaN	NaN	NaN	NaN
2008 2011	NaN	NaN	NaN	NaN
2008 2012	NaN	NaN	NaN	NaN
2008 2013	NaN	NaN	NaN	NaN
2008 2014	NaN	NaN	NaN	NaN
2008 2015	NaN	NaN	NaN	NaN
2008 2016	NaN	NaN	NaN	NaN
2008 2017	NaN	NaN	NaN	NaN
2008 2018	NaN	NaN	NaN	NaN
2008 2019	NaN	NaN	NaN	NaN
2008 2020	NaN	NaN	NaN	NaN

2008 2021	NaN	NaN	NaN	NaN
2008 2022	NaN	NaN	NaN	NaN
2016 2002	NA	NA	NA	NA
2016 2003	NaN	NaN	NaN	NaN
2016 2004	NaN	NaN	NaN	NaN
2016 2005	NaN	NaN	NaN	NaN
2016 2006	NaN	NaN	NaN	NaN
2016 2007	NaN	NaN	NaN	NaN
2016 2008	NaN	NaN	NaN	NaN
2016 2009	NaN	NaN	NaN	NaN
2016 2010	NaN	NaN	NaN	NaN
2016 2011	NaN	NaN	NaN	NaN
2016 2012	NaN	NaN	NaN	NaN
2016 2013	NaN	NaN	NaN	NaN
2016 2014	NaN	NaN	NaN	NaN
2016 2015	NaN	NaN	NaN	NaN
2016 2016	NaN	NaN	NaN	NaN
2016 2017	NaN	NaN	NaN	NaN
2016 2018	NaN	NaN	NaN	NaN
2016 2019	NaN	NaN	NaN	NaN
2016 2020	NaN	NaN	NaN	NaN
2016 2021	NaN	NaN	NaN	NaN
2016 2022	NaN	NaN	NaN	NaN
2017 2002	NA	NA	NA	NA
2017 2003	0.0271	0.0201	-0.0194	0.0737
2017 2004	0.0055	0.0296	-0.0631	0.0741
2017 2005	0.0707	0.0204	0.0235	0.1178 *
2017 2006	-0.1294	0.0374	-0.2160	-0.0427 *
2017 2007	0.0058	0.0328	-0.0701	0.0818
2017 2008	0.0740	0.0307	0.0029	0.1451 *
2017 2009	-0.0022	0.0286	-0.0685	0.0642
2017 2010	-0.2714	0.0210	-0.3202	-0.2227 *
2017 2011	0.0658	0.0514	-0.0533	0.1849
2017 2012	0.1778	0.0906	-0.0320	0.3877

2017 2013	0.0576	0.1013	-0.1771	0.2923
2017 2014	0.0480	0.0597	-0.0904	0.1864
2017 2015	NA	NA	NA	NA
2017 2016	NA	NA	NA	NA
2017 2017	0.2931	0.0107	0.2683	0.3180 *
2017 2018	0.1854	0.0018	0.1813	0.1894 *
2017 2019	0.2487	0.0081	0.2299	0.2675 *
2017 2020	0.1724	0.0066	0.1571	0.1878 *
2017 2021	0.1916	0.0017	0.1877	0.1955 *
2017 2022	0.2008	0.0097	0.1784	0.2233 *
2018 2002	NA	NA	NA	NA
2018 2003	0.1160	0.0203	0.0690	0.1631 *
2018 2004	-0.2121	0.0274	-0.2757	-0.1485 *
2018 2005	0.0503	0.0203	0.0033	0.0973 *
2018 2006	0.1193	0.0347	0.0389	0.1997 *
2018 2007	0.0686	0.0323	-0.0062	0.1434
2018 2008	0.1384	0.0402	0.0451	0.2316 *
2018 2009	0.2016	0.0280	0.1366	0.2665 *
2018 2010	-0.2403	0.0482	-0.3520	-0.1285 *
2018 2011	-0.1769	0.0540	-0.3021	-0.0517 *
2018 2012	0.1325	0.0504	0.0157	0.2493 *
2018 2013	-0.0606	0.0580	-0.1950	0.0737
2018 2014	0.0054	0.0507	-0.1121	0.1228
2018 2015	-0.0936	0.0146	-0.1274	-0.0598 *
2018 2016	0.0224	0.0214	-0.0273	0.0720
2018 2017	0.0018	0.0448	-0.1019	0.1056
2018 2018	0.0532	0.0262	-0.0076	0.1140
2018 2019	-0.1181	0.0291	-0.1855	-0.0506 *
2018 2020	0.0257	0.0410	-0.0694	0.1208
2018 2021	0.0679	0.0246	0.0110	0.1248 *
2018 2022	0.1686	0.0709	0.0043	0.3328 *

Signif. codes: '*' confidence band does not cover 0

Control Group: Never Treated, Anticipation Periods: 0

Estimation Method: Doubly Robust

ANEXO 2 - Resultados do modelo por estudo de eventos

```
> summary(agg_event) # Para o event study agregado
```

Call:

```
aggte(MP = att_results, type = "dynamic", na.rm = TRUE)
```

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna. "Difference-in-Differences with Multiple Time Periods." Journal of Econometrics, Vol. 225, No. 2, pp. 200-230, 2021.

<<https://doi.org/10.1016/j.jeconom.2020.12.001>>, <<https://arxiv.org/abs/1803.09015>>

Overall summary of ATT's based on event-study/dynamic aggregation:

ATT	Std. Error	[95% Conf. Int.]
0.1409	0.0604	0.0225 0.2592 *

Dynamic Effects:

Event time Estimate Std. Error [95% Simult. Conf. Band]

-15	0.1160	0.0204	0.0257	0.2064 *
-14	-0.0925	0.1239	-0.6412	0.4562
-13	0.0279	0.0259	-0.0867	0.1425
-12	0.0950	0.0257	-0.0189	0.2089
-11	-0.0304	0.0710	-0.3449	0.2841
-10	0.0721	0.0561	-0.1764	0.3206
-9	0.1378	0.0511	-0.0885	0.3641
-8	-0.1212	0.1146	-0.6289	0.3865
-7	-0.2242	0.0383	-0.3937	-0.0546 *
-6	0.0991	0.0303	-0.0349	0.2332
-5	0.0586	0.1005	-0.3868	0.5040

-4	0.0315	0.0629	-0.2473	0.3103
-3	-0.0228	0.0563	-0.2723	0.2267
-2	0.0224	0.0210	-0.0705	0.1153
-1	0.0018	0.0446	-0.1960	0.1996
0	0.1732	0.0487	-0.0424	0.3887
1	0.0337	0.0346	-0.1198	0.1871
2	0.1372	0.1079	-0.3409	0.6153
3	0.1202	0.0322	-0.0225	0.2629
4	0.1801	0.0351	0.0248	0.3354 *
5	0.2008	0.0062	0.1732	0.2284 *

Signif. codes: `*' confidence band does not cover 0

Control Group: Never Treated, Anticipation Periods: 0

Estimation Method: Doubly Robust

ANEXO 3 - Resultados dos testes estatísticos pré-tratamento

> # 3. Validando as Tendências Pré-Tratamento

> # Filtramos os períodos anteriores ao tratamento (time < group)

> att_pre <- att_table %>%

+ filter(time < group)

> # Calculamos a Média dos ATT pré-tratamento que devem estar, idealmente, pr\$

> pre_trend_mean <- mean(att_pre\$att, na.rm = TRUE)

> pre_trend_sd <- sd(att_pre\$att, na.rm = TRUE)

> cat("Média do ATT pré-tratamento:", pre_trend_mean,

+ "\nDesvio padrão do ATT pré-tratamento:", pre_trend_sd, "\n")

Média do ATT pré-tratamento: 0.007487044

Desvio padrão do ATT pré-tratamento: 0.124188

> # Número de observações pré-tratamento

> n <- nrow(att_pre)

> # Cálculo da estatística t

> t_stat <- pre_trend_mean / (pre_trend_sd / sqrt(n))

```

> # Cálculo do p-valor (teste bilateral)
> p_value <- 2 * pt(-abs(t_stat), df = n-1)
> # Para um intervalo de confiança de 95%
> error_margin <- qt(0.975, df = n-1) * (pre_trend_sd / sqrt(n))
> ci_lower <- pre_trend_mean - error_margin
> ci_upper <- pre_trend_mean + error_margin
> cat("Estatística t pré:", t_stat, "\nP-valor pré:", p_value,
+     "\nIntervalo de Confiança (95%) pré: [", ci_lower, ",", ci_upper, "]\n")
Estatística t pré: 0.4305424
P-valor pré: 0.6686514
Intervalo de Confiança (95%) pré: [ -0.02744139 , 0.04241547 ]

```

ANEXO 4 - Resultados dos testes estatísticos pós-tratamento

```

> # 4. Validando as Tendências Pós-Tratamento
> # Filtramos os períodos pós tratamento (time >= group)
> att_post <- att_table %>%
+   filter(time >= group)
> # Calculamos a Média e Desvio Padrão dos ATT pós-tratamento
> post_trend_mean <- mean(att_post$att, na.rm = TRUE)
> post_trend_sd <- sd(att_post$att, na.rm = TRUE)
> # Número de observações pós-tratamento
> n_post <- nrow(att_post)
> # Cálculo da estatística t
> t_stat_post <- post_trend_mean / (post_trend_sd / sqrt(n_post))
> # Cálculo do p-valor (teste bilateral)
> p_value_post <- 2 * pt(-abs(t_stat_post), df = n_post-1)
> # Para um intervalo de confiança de 95%
> error_margin_post <- qt(0.975, df = n_post-1) * (post_trend_sd / sqrt(n_post))
> ci_lower_post <- post_trend_mean - error_margin_post
> ci_upper_post <- post_trend_mean + error_margin_post
> cat("Estatística t pós:", t_stat_post, "\nP-valor pós:", p_value_post,
+     "\nIntervalo de Confiança (95%) pós: [", ci_lower_post, ",", ci_upper_post, "]\n")

```

Estatística t pós: 6.62824

P-valor pós: 1.783911e-07

Intervalo de Confiança (95%) pós: [0.09378836 , 0.1770068]

ANEXO 5 - Médias do efeito pós e pré tratamento

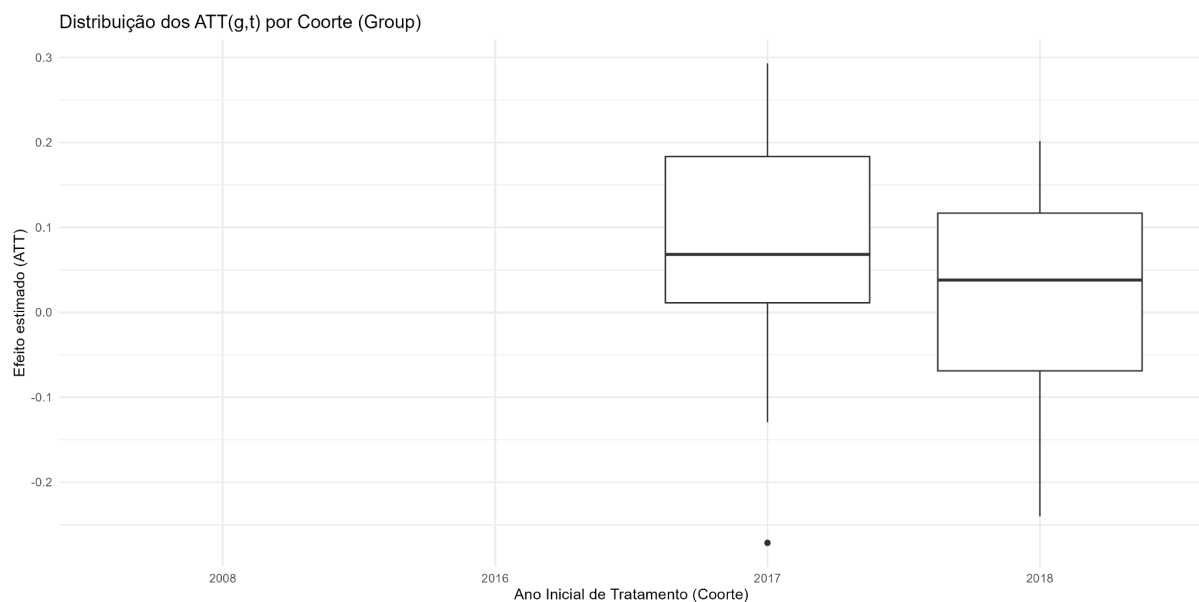
```
> # Verificamos se, em média, o efeito pós-tratamento é maior que o pré-tratam$
> post_event_att <- mean(event_df$att[event_df$event_time > 0], na.rm = TRUE)
> pre_event_att <- mean(event_df$att[event_df$event_time < 0], na.rm = TRUE)
> cat("Média do efeito pós-tratamento:", post_event_att,
+     "\nMédia do efeito pré-tratamento:", pre_event_att,
+     "\nSe a média pós > média pré e há evidência estatística (intervalos não$
+     "o tratamento parece ter impacto positivo ao longo do tempo.\n")
```

Média do efeito pós-tratamento: 0.1343858

Média do efeito pré-tratamento: 0.01141307

Se a média pós > média pré e há evidência estatística (intervalos não cruzam zero), o tratamento parece ter impacto positivo ao longo do tempo.

Gráfico de distribuição



ANEXO 6 - Código do modelo

A base de dados está disponível junto a esse script no meu perfil pessoal da plataforma Github, no repositório tcc-ie-ufrj-2024. Link:

<https://github.com/danielcavalli/tcc-ie-ufrj-2024>

```
library(did)
library(dplyr)
library(ggplot2)

# Leitura dos dados
df <- read.csv("data/csv/WightedMethod/sugar_cane_treated.csv")

# 1. Preparação dos dados -----

set.seed(42)

# Transformar a variável produtividade em log
df <- df %>%
  mutate(log_produtividade = log(produtividade))

# Ajustar a variável gname: se a unidade nunca for tratada, colocar 0
# ou NA.
# Aqui assumimos que todas as unidades são tratadas em algum momento.
df <- df %>%
  group_by(id_microrregiao) %>%
  mutate(gname = primeiro_ano_tratamento) %>%
  ungroup()

df <- df %>%
  rename(treated = tratado)

head(df)

# 2. Estimação ATT(g,t) com att_gt() -----

# São incluídas as covariáveis total_area_plantada e
# precipitacao_por_area_plantada
# Pelo método Doubly Robust (est_method = "dr")
att_results <- att_gt(
  yname = "log_produtividade",
  tname = "ano",
  idname = "id_microrregiao",
```



```

    gname = "gname",
    xformula = ~ qtd_estacoes_ativas + precipitacao_por_area_plantada,
    data = df,
    est_method = "dr"
)

# 3. Agregação dos Efeitos com aggte() -----

# (a) Efeito Médio Global (Overall Treatment Effect)
agg_overall <- aggte(att_results, type = "group", na.rm = TRUE)

# (b) Event Study (efeito por tempo de exposição)
# Definir type = "dynamic" cria uma análise de estudo de evento
agg_event <- aggte(att_results, type = "dynamic", na.rm = TRUE)

# 4. Visualização dos Resultados -----

# (a) Gráfico do Event Study
event_df <- data.frame(
  time_relative = agg_event$segt,
  att = agg_event$att.egt,
  se = agg_event$se.egt
)

ggplot(event_df, aes(x = time_relative, y = att)) +
  # Faixa de confiança (intervalo de confiança) usando geom_ribbon
  geom_ribbon(aes(ymin = att - 1.96*se, ymax = att + 1.96*se), fill =
"blue", alpha = 0.2) +

  # Linha do efeito estimado
  geom_line(color = "blue", size = 1) +

  # Pontos sobre a linha
  geom_point(color = "blue", size = 2) +

  # Linha horizontal em y=0 para referencia
  geom_hline(yintercept = 0, linetype = "dashed", color = "red", size =
1) +

  # Linha vertical em x=0 para marcar o início do tratamento
  geom_vline(xintercept = 0, linetype = "dashed", color = "gray40") +

  # Ajuste no tema

```

```

theme_minimal(base_size = 14) +
theme(
  panel.grid.minor = element_blank(),
  panel.grid.major.x = element_line(color = "gray90"),
  panel.grid.major.y = element_line(color = "gray90"),
  plot.title = element_text(face = "bold", size = 16)
) +

# Legenda
labs(
  title = "Event Study: Efeito do Tratamento ao Longo do Tempo de
Disponibilidade da Estação",
  x = "Períodos após o tratamento",
  y = "Efeito estimado (log-produtividade)"
)

ggsave(
  "event_study_plot.png",
  plot = last_plot(),
  width = 12,
  height = 6
)

#
=====
=====
# Interpretação dos resultados
#
=====
=====

# 1. Resumo dos ATT(g,t)
summary(att_results)
summary(agg_overall) # Para o efeito médio global
summary(agg_event)   # Para o event study agregado

# O objeto retornado por att_gt() possui componentes que armazenam os
ATT(g,t) individuais
# Para acessá-los individualmente, os coloquei na tabela abaixo:
att_table <- data.frame(
  group = att_results$group,
  time = att_results$t,
  att = att_results$att,

```

```

    se = att_results$se,
    n = att_results$n
  )

head(att_table)

# 2. Estatísticas Descritivas dos ATT
# Média, mediana, desvio padrão, mínimas e máximas dos efeitos
estimados
summary_att <- att_table %>%
  summarise(
    mean_att = mean(att, na.rm = TRUE),
    median_att = median(att, na.rm = TRUE),
    sd_att = sd(att, na.rm = TRUE),
    min_att = min(att, na.rm = TRUE),
    max_att = max(att, na.rm = TRUE)
  )

print(summary_att)

# 3. Validando as Tendências Pré-Tratamento
# Filtramos os períodos anteriores ao tratamento (time < group)
att_pre <- att_table %>%
  filter(time < group)

# Calculamos a Média dos ATT pré-tratamento que devem estar,
idealmente, próximo de zero # nolint
pre_trend_mean <- mean(att_pre$att, na.rm = TRUE)
pre_trend_sd <- sd(att_pre$att, na.rm = TRUE)

cat("Média do ATT pré-tratamento:", pre_trend_mean,
    "\nDesvio padrão do ATT pré-tratamento:", pre_trend_sd, "\n")

# Número de observações pré-tratamento
n <- nrow(att_pre)

# Cálculo da estatística t
t_stat <- pre_trend_mean / (pre_trend_sd / sqrt(n))

# Cálculo do p-valor (teste bilateral)
p_value <- 2 * pt(-abs(t_stat), df = n-1)

# Para um intervalo de confiança de 95%

```

```

error_margin <- qt(0.975, df = n-1) * (pre_trend_sd / sqrt(n))
ci_lower <- pre_trend_mean - error_margin
ci_upper <- pre_trend_mean + error_margin

cat("Estatística t pré:", t_stat, "\nP-valor pré:", p_value,
    "\nIntervalo de Confiança (95%) pré: [", ci_lower, ",", ci_upper,
    "]\n")

# 4. Validando as Tendências Pós-Tratamento
# Filtramos os períodos pós tratamento (time >= group)
att_post <- att_table %>%
  filter(time >= group)

# Calculamos a Média e Desvio Padrão dos ATT pós-tratamento
post_trend_mean <- mean(att_post$att, na.rm = TRUE)
post_trend_sd <- sd(att_post$att, na.rm = TRUE)

# Número de observações pós-tratamento
n_post <- nrow(att_post)

# Cálculo da estatística t
t_stat_post <- post_trend_mean / (post_trend_sd / sqrt(n_post))

# Cálculo do p-valor (teste bilateral)
p_value_post <- 2 * pt(-abs(t_stat_post), df = n_post-1)

# Para um intervalo de confiança de 95%
error_margin_post <- qt(0.975, df = n_post-1) * (post_trend_sd /
sqrt(n_post))
ci_lower_post <- post_trend_mean - error_margin_post
ci_upper_post <- post_trend_mean + error_margin_post

cat("Estatística t pós:", t_stat_post, "\nP-valor pós:", p_value_post,
    "\nIntervalo de Confiança (95%) pós: [", ci_lower_post, ",",
ci_upper_post, "]\n")

# 5. Interpretação do Efeito Médio Global
summary(agg_overall)

# O objeto agg_overall$att.agg fornece o efeito médio global,
# e agg_overall$se.agg o erro padrão.
# Podemos interpretar esse efeito da seguinte forma:

```

```

overall_effect <- agg_overall$att.agg
overall_se <- agg_overall$se.agg

cat("Overall Treatment Effect (ETE):", overall_effect,
    "\nSE:", overall_se,
    "\nInterpretação: Um valor positivo significa que, em média,\n",
    "a intervenção aumentou o log da produtividade. Por exemplo,",
    "um efeito de 0.1 implica um aumento aproximado de 10,5%\n",
    "na produtividade.\n")

# 6. Interpretação do Event Study (dinâmica do efeito)
summary(agg_event)

event_df <- data.frame(
  event_time = agg_event$egt,
  att = agg_event$att.egt,
  se = agg_event$se.egt
)

# Verificamos se, em média, o efeito pós-tratamento é maior que o
pré-tratamento
post_event_att <- mean(event_df$att[event_df$event_time > 0], na.rm =
TRUE)
pre_event_att <- mean(event_df$att[event_df$event_time < 0], na.rm =
TRUE)

cat("Média do efeito pós-tratamento:", post_event_att,
    "\nMédia do efeito pré-tratamento:", pre_event_att,
    "\nSe a média pós > média pré e há evidência estatística
(intervalos não cruzam zero),",
    "\no tratamento parece ter impacto positivo ao longo do tempo.\n")

ggplot(att_table, aes(x = factor(group), y = att)) +
  geom_boxplot() +
  theme_minimal() +
  labs(
    title = "Distribuição dos ATT(g,t) por Coorte (Group)",
    x = "Ano Inicial de Tratamento (Coorte)",
    y = "Efeito estimado (ATT)"
  )

ggsave(
  "distribution.png",
  plot = last_plot(),

```

```

width = 12,
height = 6
)

# ADICIONAL

#
=====
=====
# Algumas visualizações dos dados
#
=====
=====

library(dplyr)

desc_stats <- df %>%
  mutate(periodo = ifelse(pos_tratamento == 0, "Pre", "Pos"),
         grupo = ifelse(treated == 1, "Tratado", "Não Tratado")) %>%
  group_by(grupo, periodo) %>%
  summarise(
    mean_prod = mean(produtividade, na.rm = TRUE),
    median_prod = median(produtividade, na.rm = TRUE),
    mean_area = mean(total_area_plantada, na.rm = TRUE),
    median_area = median(total_area_plantada, na.rm = TRUE),
    mean_precip = mean(total_precipitacao, na.rm = TRUE),
    median_precip = median(total_precipitacao, na.rm = TRUE),
    n = n(),
    .groups = "drop"
  )

desc_stats

unidades_ano <- df %>%
  group_by(ano, treated) %>%
  summarise(num_unidades = n_distinct(id_microrregiao), .groups =
"drop")

print(n = 100, unidades_ano)

# Menor e maior ano na amostra
range_anos <- range(df$ano)

```

```

pre_trat <- df %>% filter(pos_tratamento == 0)
pos_trat <- df %>% filter(pos_tratamento == 1)

cat("Intervalo de anos:", range_anos, "\n")
cat("Anos pré-tratamento disponíveis:", length(unique(pre_trat$ano)),
"\n")
cat("Anos pós-tratamento disponíveis:", length(unique(pos_trat$ano)),
"\n")

summary_precip_area <- df %>%
  group_by(treated, pos_tratamento) %>%
  summarise(
    mean_ratio = mean(precipitacao_por_area_plantada, na.rm = TRUE),
    median_ratio = median(precipitacao_por_area_plantada, na.rm =
TRUE),
    sd_ratio = sd(precipitacao_por_area_plantada, na.rm = TRUE),
    .groups = "drop"
  )

summary_precip_area

estacoes_ano <- df %>%
  group_by(ano) %>%
  summarise(
    mean_estacoes = mean(qtd_estacoes_ativas, na.rm = TRUE),
    median_estacoes = median(qtd_estacoes_ativas, na.rm = TRUE),
    .groups = "drop"
  )

print(n = 100,estacoes_ano)

prod_ano <- df %>%
  group_by(ano, treated) %>%
  summarise(mean_prod = mean(log_produtividade, na.rm = TRUE), .groups
= "drop")

ggplot(prod_ano, aes(x = ano, y = mean_prod, color = factor(treated)))
+
  geom_line(size = 1) +
  theme_minimal() +
  labs(title = "Evolução da Produtividade ao Longo do Tempo",
x = "Ano",

```

```

    y = "Produtividade Média",
    color = "Tratado") +
  scale_color_discrete(labels = c("0" = "Não Tratado", "1" =
"Tratado"))

df_perodo_grupo <- df %>%
  mutate(perodo = ifelse(pos_tratamento == 0, "Pré", "Pós"),
         grupo = ifelse(treated == 1, "Tratado", "Não Tratado"),
         cat = paste(grupo, perodo))

ggplot(df_perodo_grupo, aes(x = cat, y = produtividade)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Distribuição da Produtividade por Grupo e Período",
       x = "Grupo e Período",
       y = "Produtividade")

ggplot(estacoes_ano, aes(x = ano, y = mean_estacoes)) +
  geom_line(size = 1, color = "blue") +
  theme_minimal() +
  labs(title = "Evolução do Número Médio de Estações Ativas ao Longo do
Tempo",
       x = "Ano",
       y = "Média de Estações Ativas")

```