NOTA TÉCNICA

1. AVALIAÇÃO DE IMPACTO

A avaliação de impacto tem como objetivo mensurar a influência do atendimento do SEBRAE na sobrevivência das empresas, eliminando efeitos externos que poderiam distorcer os resultados. Para isso, utilizamos um método quase-experimental, o *Propensity Score Matching (PSM),* uma técnica estatística amplamente utilizada para simular um experimento aleatório. Esse método permite comparar empresas atendidas e não atendidas de forma equivalente, garantindo que as diferenças observadas na sobrevivência sejam atribuídas ao atendimento do SEBRAE.

O PSM estima escores de propensão por meio de uma regressão logística multivariada, utilizando variáveis que influenciam tanto a decisão de buscar atendimento no SEBRAE quanto os resultados de interesse. As variáveis utilizadas no pareamento foram:

- Porte da empresa
- Opção pelo MEI
- Opção pelo Simples
- Setor de atividade
- Região
- Grupo de natureza jurídica
- Ano de abertura

Utilizamos o método de vizinho mais próximo (*nearest neighbor matching*) com *caliper* de 0,5. A qualidade do pareamento foi verificada por meio de gráficos de balanceamento, demonstrando que as distribuições das covariáveis ficaram bem ajustadas entre os grupos tratados e de controle. Os gráficos de validação do PSM são apresentados a seguir:

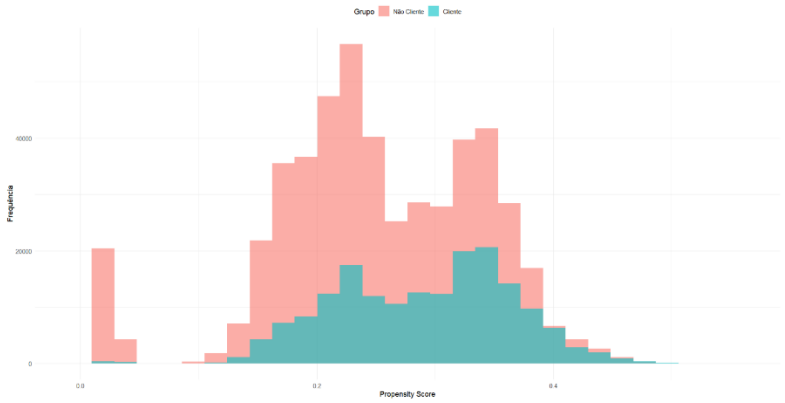

Figura 1 – Distribuição do Propensity Score ANTES do Macthing
Fonte: Resultados originais da pesquisa.

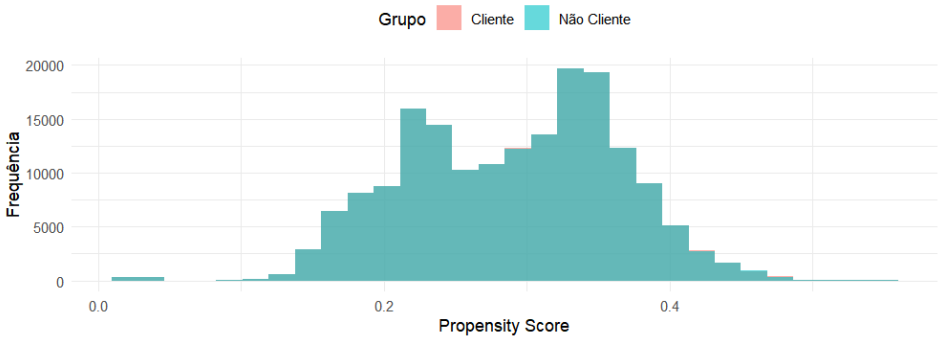

Figura 2 – Distribuição do Propensity Score DEPOIS do Macthing
Fonte: Resultados originais da pesquisa.

2. BASE DE DADOS

A base de dados utilizada neste estudo foi composta por informações do Cartão CPNJ da Receita Federal do Brasil (RFB) e dos registros de atendimento do SEBRAE Mato Grosso. O conjunto de dados inclui informações detalhadas sobre as empresas, como data de abertura, setor de atuação, regime tributário e situação cadastral ao longo do tempo. Os registros de atendimento incluem informações individuais de cada atendimento, como data do atendimento, instrumento, categoria do instrumento e tipo de atendimento.

O estudo inicialmente considerou 671.821 empresas da base da Receita Federal e 362.738 empresas atendidas pelo SEBRAE Mato Grosso entre 2021 e 2024. Para reduzir a variabilidade do perfil dos estabelecimentos, foram aplicados os seguintes critérios de filtragem:

- Apenas empresas nascidas a partir de 2014 foram consideradas.
- Foram categorizadas as empresas com situação cadastral ATIVA ou BAIXADA.
- Foram mantidas apenas empresas com natureza jurídica correspondente a MEI, ME, EPP e MGE. Empresas classificadas como "Demais" no porte declarado do CNPJ foram excluídas.

A Tabela 01 apresenta a distribuição das empresas selecionadas de acordo com o porte declarado no Cartão CNPJ da RFB.

Porte	Clientes		Não Clientes	
	N	%	N	%
	175.989	100%	175.989	100%
ME	81.703	46,43%	78.503	44,61%
MEI	81.195	46,14%	83.712	47,57%
EPP	12.459	7,08%	13.142	7,47%
MGE	632	0,36%	632	0,36%

Tabela 01 – Distribuição das empresas por Porte após o PSM

Fonte: Resultados originais da pesquisa.

A partir do universo inicial de 495.819 empresas, na base da RFB, nunca atendidas pelo SEBRAE, o modelo identificou 175.989 empresas semelhantes com base nas variáveis mencionadas. Outras 319.843 empresas não apresentaram pares comparáveis e, por isso, foram excluídas da análise. Dessa forma, a amostra final do estudo conta com um total de 351.978 empresas, sendo metade composta por empresas que foram atendidas pelo SEBRAE Mato Grosso entre 2021 e 2024, e a outra metade formada por empresas que nunca foram atendidas pelo SEBRAE.

A Tabela 02 exibe a distribuição das empresas por setor de atividade. Setores de Comércio e Serviços são os mais frequentes, representando em média 34% e 45%, da amostra, respectivamente.

Setor de Atividade	Clientes		Não Clientes	
	N	%	N	%
	175.989	100%	175.989	100%
Serviço	78.835	44,80%	81.876	46,52%
Comércio	63.024	35,81%	57.273	32,54%
Construção	16.663	9,47%	18.026	10,24%
Indústria	15.658	8,90%	16.097	9,15%
Agropecuária	1.809	1,03%	2.717	1,54%

Tabela 02 – Distribuição de empresas por Setor de Atividade após o PSM

Fonte: Resultados originais da pesquisa.

Além disso, o estudo abrange empresas de todas as regiões do Mato Grosso, permitindo uma análise detalhada das diferenças regionais na taxa de sobrevivência empresarial. A metodologia adotada no *Propensity Score Matching* possibilita a realização de análises comparativas mais robustas entre os negócios

atendidos e não atendidos pelo SEBRAE, fornecendo insights relevantes para o aprimoramento das políticas de apoio ao empreendedorismo local.

A Tabela 03 apresenta o número de empresas atendidas e não atendidas, organizadas por regionais do SEBRAE Mato Grosso. Essa distribuição reflete as características econômicas de cada região, considerando fatores como a concentração populacional, o dinamismo do mercado e a presença de setores estratégicos para o estado.

Região	Clientes		Não Clientes	
	N	%	N	%
	175.989	**100%**	**175.989**	**100%**
Metropolitana	66.073	37,54%	65.197	37,05%
Norte	44.712	25,41%	45.979	26,13%
Sudeste	25.143	14,29%	24.901	14,15%
Noroeste	16.466	9,36%	16.791	9,54%
Nordeste	15.028	8,54%	14.906	8,47%
Sudoeste	8.567	4,87%	8.215	4,67%

Tabela 03 – Distribuição de empresas por Regional após o PSM
Fonte: Resultados originais da pesquisa.

3. METODOLOGIA

Para falar sobre os modelos aplicados neste estudo é necessário deixar explicito que a resposta desse tipo de estudo é caracterizada pelas censuras, que são ausências de informações / desfecho. Para este caso, o tempo até a baixa do CNPJ nos registros da Receita Federal é considerado como uma censura.

A presença de censura ocorre quando o evento de interesse, baixa do CNPJ, ainda não foi observada na respectiva empresa durante o período de realização do estudo, causando observações incompletas. Esses dados censurados devem ser incluídos na análise pois eles fornecem informações sobre o tempo de vida das empresas. Sua omissão pode acarretar conclusões viciadas.

3.1 Estimador de *Kaplan-Meier*

O Estimador de *Kaplan-Meier* é um método amplamente utilizado para estimar a função de sobrevivência ao longo do tempo. Ele calcula a probabilidade de uma empresa continuar ativa após determinados períodos, permitindo uma comparação inicial entre os grupos analisados. Esse método é especialmente útil

para lidar com dados censurados, ou seja, casos em que não sabemos exatamente quando a empresa encerrará suas atividades.

3.2 Modelo de Regressão de Cox

No presente estudo, utilizou-se o Modelo de Regressão de Cox como abordagem principal para estimar a relação entre variáveis explicativas e o tempo de sobrevivência das empresas no estado de Mato Grosso. Este modelo é amplamente reconhecido por sua capacidade de analisar o tempo até a ocorrência de um evento — neste caso, o encerramento de atividades empresariais — sem exigir a especificação prévia da distribuição do tempo de falha.

Um dos principais parâmetros gerados pelo modelo é o *Hazard Ratio* (HR), que expressa a razão de risco de encerramento entre diferentes grupos. Valores de HR inferiores a 1 indicam menor risco de encerramento em relação ao grupo de referência, enquanto valores superiores a 1 sugerem maior vulnerabilidade à saída do mercado.

A modelagem dos dados foi conduzida de forma incremental, iniciando-se com uma estrutura sem interações, e, posteriormente, incluindo interações para capturar efeitos específicos de subgrupos. As especificações testadas foram:

- **Modelo sem interação, contendo como covariáveis:** situação de atendimento pelo Sebrae (cliente/não cliente); porte da empresa; região; setor de atividade;
- **Modelo com interações entre o atendimento e:** o porte da empresa; o setor de atividade econômica; a região de localização; a quantidade de atendimentos realizados; o tipo de atendimento (presencial ou online); e o instrumento de atendimento (ex: consultoria, palestra, cursos, etc.).

A inclusão dessas interações teve como objetivo avaliar se o impacto do atendimento varia de acordo com as características das empresas, o que confere maior granularidade e aplicabilidade aos resultados.

Cliente/Não Cliente	HR (Hazard Ratio)	Erro Padrão	p-valor
Não Cliente	Referência	Referência	Referência
Cliente	0.414	0.006	<0.001***

Tabela 4 - Modelo de Regressão de Cox – Atendimento do Sebrae (Cliente x Não Cliente)
Fonte: Resultados originais da pesquisa.

Variável	HR (Hazard Ratio)	Erro Padrão	p-valor
EPP	0.316	0.035	<0.001***
MGE	0.297	0.168	<0.001***
ME	0.420	0.008	<0.001***
MEI	0.376	0.012	<0.001***
AGROPECUÁRIA	0.744	0.066	<0.001***
INDÚSTRIA	0.349	0.022	<0.001***
CONSTRUÇÃO	0.501	0.021	<0.001***
COMÉRCIO	0.341	0.011	<0.001***
SERVIÇO	0.470	0.010	<0.001***
Sudeste	0.409	0.018	<0.001***
Noroeste	0.455	0.021	<0.001***
Nordeste	0.437	0.023	<0.001***
Norte	0.458	0.013	<0.001***
Sudoeste	0.447	0.029	<0.001***
Metropolitana	0.371	0.010	<0.001***

Tabela 5 - Modelo de Regressão de Cox – Comparativo Cliente x Não Cliente por Porte, Setor e Região
Fonte: Resultados originais da pesquisa.

Quantidade de Atendimento	HR (Hazard Ratio)	Erro Padrão	p-valor
2-5 Atendimentos	0,588	0,013	<0.001***
6-10 Atendimentos	0,435	0,018	<0.001***
11-20 Atendimentos	0,317	0,023	<0.001***
21+ Atendimentos	0,169	0,037	<0.001***

Tabela 6 - Modelo de Regressão de Cox – Comparativo por Quantidade de Atendimentos
Fonte: Resultados originais da pesquisa.

Tipo de Atendimento	HR (Hazard Ratio)	Erro Padrão	p-valor
Presencial	0,920	0,002	<0.001***
Online	0,981	0,002	<0.001***

Tabela 7 - Modelo de Regressão de Cox – Comparativo por Tipo de Atendimento
Fonte: Resultados originais da pesquisa.

Instrumento	HR (Hazard Ratio)	Erro Padrão	p-valor
Missão Caravana	0,700	0,103	<0.001***
Consultoria	0,804	0,003	<0.001***
Palestra	0,821	0,016	<0.001***
Feira	0,823	0,085	<0.05*
Seminário	0,884	0,013	<0.001***
Cursos	0,898	0,019	<0.001***
Informação	0,950	0,002	<0.001***
Orientação	0,971	0,002	<0.001***
Rodada de Negócios	0,996	0,1	0.96404
Oficina	1,010	0,011	0.34048
Clínica Tecnológica	1,035	0,167	0.83482

Tabela 8 - Modelo de Regressão de Cox – Comparativo por Instrumento de Atendimento
Fonte: Resultados originais da pesquisa.

De maneira geral, os modelos produziram estimativas estatisticamente significativas para a maioria das covariáveis testadas, demonstrando a relevância dos fatores considerados sobre o risco de encerramento empresarial. O *Hazard Ratio* permite interpretações intuitivas e comparações diretas entre grupos, o que torna este modelo especialmente útil em estudos aplicados como este.

Além disso, foram observados padrões consistentes entre os diferentes modelos, o que reforça a robustez dos efeitos identificados, mesmo com diferentes especificações.

Contudo, a verificação da suposição de proporcionalidade dos riscos, inerente ao modelo de Cox, indicou possíveis violações. Essa suposição pressupõe que a razão de riscos entre os grupos se mantém constante ao longo do tempo — uma condição que nem sempre é verificada em contextos reais de mercado. Diante dessa limitação, foi incorporado um modelo paramétrico complementar, o modelo de regressão de sobrevivência de *Weibull*, que permite lidar com variações no risco ao longo do tempo e oferece estimativas mais estáveis em cenários com mudanças de comportamento temporal.

A combinação entre os modelos de Cox (semiparamétrico) e Weibull (paramétrico) permite uma abordagem abrangente e robusta para a análise da sobrevivência empresarial, possibilitando a identificação de fatores críticos associados à trajetória dos pequenos negócios e o papel do Sebrae Mato Grosso nesse processo.

3.3 Modelo de *Weibull*

O Modelo de Regressão de *Weibull* é uma abordagem paramétrica que permite estimar diretamente a distribuição do tempo de sobrevivência das empresas. Esse modelo é particularmente útil quando há variação no risco de encerramento ao longo do tempo, o que foi identificado neste estudo por meio da violação da suposição de proporcionalidade dos riscos no modelo de Cox.

Ao contrário do Cox, que estima razões de risco entre grupos, o *Weibull* permite modelar diretamente o tempo até o encerramento das empresas. Ele também permite identificar como esse risco se comporta ao longo do tempo, por meio de um parâmetro chamado forma (ou shape).

A interpretação do shape é bastante intuitiva:

- Shape > 1 indica que o risco de fechamento aumenta com o tempo;
- Shape = 1 indica risco constante;
- Shape < 1 indica que o risco de encerramento diminui com o tempo

Os modelos de *Weibull* foram estimados em diferentes especificações, incluindo:

- Modelo com a variável Cliente x Não Cliente (Tabela 8);
- Modelos com interações com porte, setor e região (Tabela 9);
- Modelos por quantidade de atendimentos (Tabela 10);
- Por tipo de atendimento (presencial/online) (Tabela 11);
- E por instrumento de atendimento (Tabela 12).

Cliente/Não Cliente	Tempo Weibull	Log (Scale)	Shape	p-valor
Não Cliente	Referência	Referência	Referência	
Cliente	2,128	-0,15	1,161	<0.001***

Tabela 8 - Modelo de Weibull – Atendimento do Sebrae (Cliente x Não Cliente)
Fonte: Resultados originais da pesquisa.

Variável	Tempo_Weibull	Log (Scale)	Shape	p-valor
EPP	2.678	-0,162	1,176	<0,001***
MGE	2.752	-0,188	1,206	<0,001***
ME	2.507	0,071	0,931	<0,001***
MEI	1.729	-0,639	1,894	<0,001***
AGROPECUÁRIA	1.327	-0,044	1,045	<0,001***
INDÚSTRIA	2.317	-0,219	1,245	<0,001***
CONSTRUÇÃO	1.726	-0,235	1,264	<0,001***

COMÉRCIO	2.518	-0,145	1,156	<0,001***
SERVIÇO	1.935	-0,127	1,135	<0,001***
Sudeste	2.171	-0,142	1,153	<0,001***
Noroeste	2.019	-0,112	1,119	<0,001***
Nordeste	2.102	-0,103	1,108	<0,001***
Norte	2.012	-0,109	1,115	<0,001***
Sudoeste	2.075	-0,095	1,099	<0,001***
Metropolitana	2.222	-0,203	1,225	<0,001***

Tabela 9 - Modelo de Weibull – Comparativo Cliente x Não Cliente por Porte, Setor e Região
Fonte: Resultados originais da pesquisa.

Quantidade de Atendimentos	Tempo_Weibull	Log (Scale)	Shape	p-valor
2-5 Atendimentos	1,483	-0,306	1,358	<0,001***
6-10 Atendimentos	1,856	-0,306	1,358	<0,001***
11-20 Atendimentos	2,346	-0,306	1,358	<0,001***
21+ Atendimentos	3,74	-0,306	1,358	<0,001***

Tabela 10 - Modelo de Weibull – Comparativo por Quantidade de Atendimentos
Fonte: Resultados originais da pesquisa.

Tipo de Atendimento	Tempo_Weibull	Log (Scale)	Shape	p-valor
Presencial	1,064	-0,302	1,352	<0,001***
Online	1,015	-0,302	1,352	<0,001***

Tabela 11 - Modelo de Weibull – Comparativo por Tipo de Atendimento (Presencial x Online)
Fonte: Resultados originais da pesquisa.

Instrumento	Tempo_Weibull	Log (Scale)	Shape	p-valor
Missão Caravana	1,302	-0,314	1,368	<0,001***
Consultoria	1,174	-0,314	1,368	<0,001***
Palestra	1,156	-0,314	1,368	<0,001***
Feira	1,153	-0,314	1,368	0.02141
Seminário	1,094	-0,314	1,368	<0,001***
Cursos	1,084	-0,314	1,368	<0,001***
Informação	1,038	-0,314	1,368	<0,001***
Orientação	1,022	-0,314	1,368	<0,001***

Rodada de Negócios	1,005	-0,314	1,368	0.94244
Oficina	0,992	-0,314	1,368	0.33864
Clínica Tecnológica	0,972	-0,314	1,368	0.81649
Log(scale)	0,731	-0,314	1,368	<0,001***

Tabela 12 - Modelo de Weibull – Comparativo por Instrumento de Atendimento
Fonte: Resultados originais da pesquisa.

As estimativas demonstraram boa adequação estatística, com p-valores inferiores a 0,05 na maioria das variáveis. Como esperado, alguns subgrupos apresentaram maior variabilidade ou menor significância estatística, o que pode refletir amostras menores ou variabilidade interna não capturada pelos modelos, sendo recomendada cautela na leitura desses casos.

De modo geral, os modelos de *Weibull* confirmaram os achados do modelo de Cox, demonstrando:

- Consistência nos efeitos das covariáveis;
- Manutenção da significância estatística na maioria dos casos;
- E uma leitura complementar sobre o comportamento do risco ao longo do tempo.

Enquanto o modelo de Cox é útil para compreender o impacto relativo das variáveis sobre o risco de encerramento, o *Weibull* permite capturar a dinâmica temporal desse risco, identificando padrões como a vulnerabilidade inicial das empresas e sua posterior estabilização. Essa complementaridade fortalece a análise e confere maior confiabilidade aos resultados do estudo, especialmente no contexto da atuação do Sebrae Mato Grosso no apoio à sobrevivência dos pequenos negócios.