

Levantamento: Obras federais paralisadas de creches e escolas com maiores chances de serem concluídas

2021

Análise e redação: Jonas Coelho

Supervisão e edição: Manoel Galdino e Juliana Sakai

Direção executiva: Manoel Galdino

Fevereiro/2021



Introdução	3
Modelos preditivos	4
Tempo para conclusão da obra	5
Gastos	6
Resultados	7
O que influencia a chance de conclusão de uma obra?	8
Anexos	9
Anexo 1: Obras por probabilidade de término e gasto esperado	9
Anexo 2: Simulador de conclusão e investimento em obras	9
Anexo 3: Mapas da distribuição da probabilidade de conclusão de obras paralisadas	9
Anexo 4: Coeficientes das regressões	10
Anexo 5: O problema dos dados	11



Introdução

O objetivo deste relatório é apresentar quais as obras paralisadas de creches e escolas, dentre as financiadas pelo Fundo Nacional de Desenvolvimento da Educação (FNDE), têm maiores chances de serem concluídas, bem como estimar o gasto público necessário para a entrega dessas obras. Trata-se de uma atualização do relatório produzido pela Transparência Brasil ao final de 2019 a pedido da Fundação Maria Cecília Souto Vidigal. Nesta versão foram atualizados dados do andamento das obras para geração de nova versão do modelo preditivo.

A partir de dados disponibilizados pelo FNDE, foi gerado um modelo preditivo capaz de estimar a probabilidade de uma obra paralisada de creche ou escola ser concluída em um dado período, bem como o custo estimado para essa conclusão. O FNDE é uma autarquia federal ligada ao Ministério da Educação responsável pela execução financeira de programas como o Proinfância, que financia a construção de creches e escolas por todo o país por meio de convênios com entes subnacionais.

Estados e municípios financiados são obrigados a prestar contas regularmente para o Fundo, que divulga boa parte dessas informações em seu portal público [SIMEC Obras](#). Nele, é possível encontrar dados sobre o percentual de execução da obra, vigência do contrato, repasses já feitos, dados das últimas vistorias, entre outras. Com base nessas informações, o modelo preditivo foi construído.

Desde a criação do portal até dezembro de 2020, foram registradas 15.211 obras, todas financiadas, em todo ou em parte, com recursos do FNDE. De acordo com classificação oficial do FNDE, **metade (50%)** dessas obras já foi **concluída**, enquanto 16% estão em execução, 3% paralisadas, 10% inacabadas e 17% foram canceladas, além disso, há 4% de obras com outros status¹.

Como a classificação do FNDE não é a mais adequada para nossos propósitos, ampliamos o conceito de obra paralisada. Para a construção do modelo preditivo, consideramos tanto as paralisadas, quanto as inacabadas (paralisadas cujo convênio celebrado com o FNDE não está mais vigente) e ainda obras que paralisaram, tiveram seu processo administrativo retomado (voltaram à fase de contratação ou reformulação), mas estão fisicamente paradas.

Dentre as 3.343 obras paralisadas identificadas segundo nossa classificação, apenas 2.747 (82%) possuíam informações sobre valores pagos tanto em relação ao momento da paralisação quanto ao momento da conclusão (se entregue). Toda a análise de predição, do treino do modelo até a sua aplicação para fins preditivos, foi feita com base nessas 2.747 obras. Mais detalhes sobre os dados e metodologia para essa classificação podem ser encontrados no anexo 5 deste relatório.

¹ Todas as situações possíveis são: concluída, contratação, em reformulação, execução, inacabada, licitação, obra cancelada, paralisada ou planejamento pelo proponente



O modelo desenvolvido foi capaz de prever com 82% de precisão que de 2.747 obras paralisadas, apenas 68 possuem probabilidade acima de 60% de serem concluídas em até três anos e meio, o que equivale a 2,4% do total. O valor estimado para concluir essas obras é de R\$ 28,4 milhões.

Modelos preditivos

Assim como no relatório de 2019, foi adotado um modelo capaz de estimar a probabilidade de conclusão de uma obra paralisada. Para isso, foram analisadas informações de 15.211 obras municipais e estaduais financiadas pelo FNDE. Essas informações foram coletadas entre dezembro de 2016 e dezembro de 2020, a partir do portal SIMEC obras. Com isso, foi possível observar os padrões seguidos pelas obras em relação ao tempo tomado para sua conclusão.

Os dados fornecidos pelo FNDE não possuem histórico das obras, apenas o status atual delas. Entretanto, a partir do final de 2016 a Transparência Brasil coletou periodicamente as informações exibidas no portal, de forma que foi possível traçar um panorama melhor da evolução das obras ao longo dos últimos quatro anos. Esse histórico possibilita demarcar quais obras já concluídas foram paralisadas em algum momento.

A partir disso, usando as variáveis disponíveis acerca das obras (como custo da obra até o momento, percentual da execução, quantidade de paralisações etc.) foi possível estipular a correlação parcial dessas variáveis com a conclusão de uma obra paralisada em determinado período de tempo. Com essa correlação, passa a ser possível estipular a chance futura de obras atualmente paralisadas serem concluídas.

Para treinar e testar o modelo, as obras já concluídas que tiveram alguma paralisação no passado foram aleatoriamente divididas em dois grupos: treinamento (80%) e teste (20%). O grupo de treinamento serviu de base para estimar o impacto das variáveis. O grupo de teste, por sua vez, foi complementado também com obras que não haviam sido concluídas. Isso permitiu utilizar o modelo obtido a partir do treinamento e verificar o número de falsos positivos e falsos negativos de acordo com a taxa de acerto no grupo teste.

Dois algoritmos diferentes foram testados: um modelo de árvore de decisão e um modelo de regressão logística. Ambos foram testados estimulando a probabilidade de conclusão da obra em um ano e meio, dois anos e meio e três anos e meio. Entre as alternativas, o modelo que se provou mais preciso² foi o de regressão logística. Para os fins desse relatório, apresentamos aqui apenas o modelo preditivo para obras serem concluídas em 3,5 anos. O modelo para 1,5 anos e 2,5 anos pode ser visto no simulador (ver anexo 2).

² Precisão é um termo técnico definido como total de verdadeiros positivos dividido pela soma de verdadeiros positivos e falsos negativos. A intuição para a precisão pode ser pensada a partir do sistema de Justiça. Se o objetivo do Judiciário for condenar culpados, a precisão seria o total de condenados que são culpados (verdadeiro positivo), dividido pelo total de condenados culpados e condenados inocentes (falso positivo).

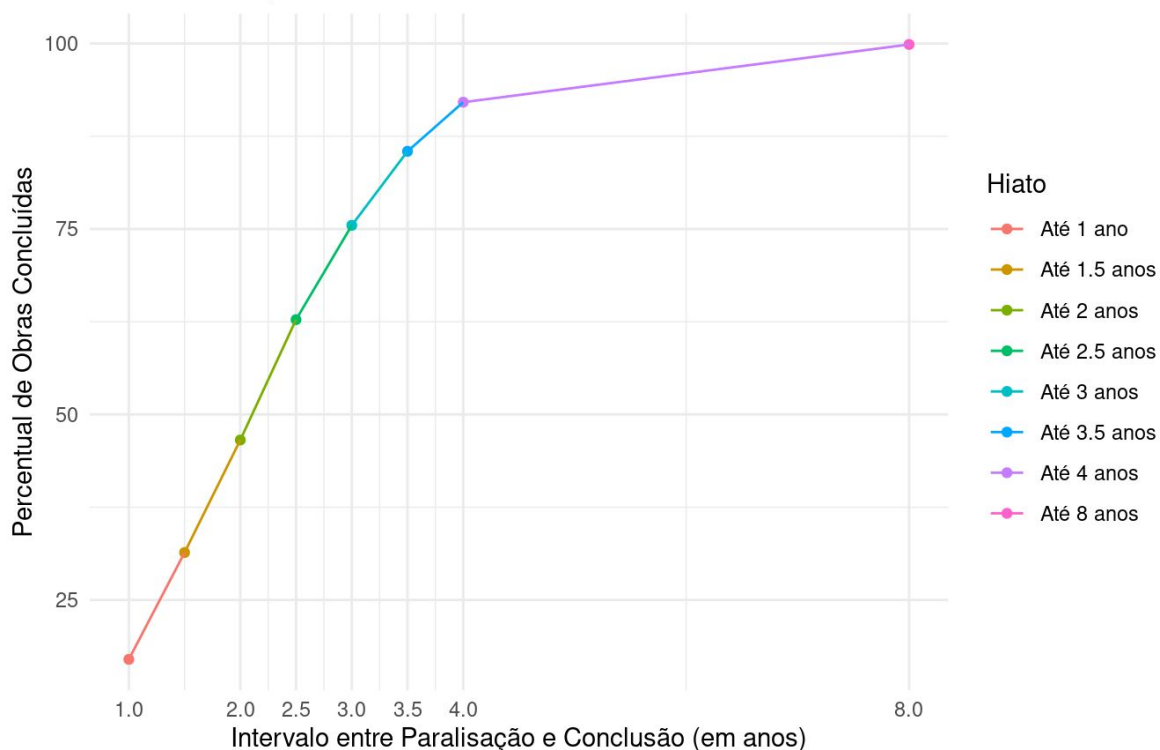


Tempo para conclusão da obra

Visando estipular o tempo para que uma obra paralisada seja concluída, foi observado o histórico das obras que foram efetivamente concluídas. Nos quatro anos de dados coletados (2016 a 2020), **771 obras foram paralisadas** em algum momento antes de sua conclusão. Esse valor representa um aumento significativo se comparado às 472 observadas até 2019.

Analisando as obras concluídas, é possível observar padrões do tempo máximo em que uma obra paralisada leva para se encerrar. Dentre as obras já concluídas, foram filtradas apenas aquelas que paralisaram em algum momento. Desse conjunto, calcula-se o tempo de conclusão em dias desde a primeira paralisação (algumas obras sofrem mais de uma paralisação e retomada). Para facilitar a compreensão dos dados, esses tempos de conclusão (hiato) foram separados em oito grupos, de acordo com a duração em anos. Como uma obra encerrada em até um ano também pode ser considerada encerrada em até 2 anos, os valores são cumulativos. A proporção de cada grupo entre as obras concluídas é ilustrada no gráfico 1 abaixo.

Gráfico 1 - Percentual de obras concluídas para cada intervalo de tempo



Elaborado por Transparência Brasil

Das 771 obras paralisadas que foram concluídas, 484 (62,8%) foram entregues em até dois anos e meio. Esse valor está consideravelmente abaixo do obtido até 2019, quando 78,6%



dentre as obras paralisadas concluídas foram entregues em até dois anos e meio. O principal motivo para essa queda consiste no aumento da proporção de obras paralisadas que levaram mais de quatro anos para serem entregues. Até 2019, menos de 2% das obras analisadas estavam nesse grupo. Até 2020, no entanto, 7,8% levaram mais de 4 anos para sua conclusão.

Esse comportamento é esperado conforme a base de dados é expandida cronologicamente, ou seja, mais obras paralisadas por 4 anos são obtidas a partir do momento em que se registra todas as obras. A tabela 1 compara a proporção entre os grupos analisados em 2019 e em 2020.

Tabela 1. Distribuição do tempo de conclusão das obras - comparação 2019 e 2020

Grupo	até dez/2019		até dez/2020	
	%	Total	%	Total
Até 1 ano	16.99	131	22.46	106
Até 1,5 ano	31.39	242	40.04	189
Até 2 anos	46.56	359	60.59	286
Até 2,5 anos	62.78	484	78.6	371
Até 3 anos	75.49	582	90.68	428
Até 3,5 anos	85.47	659	96.82	457
Até 4 anos	92.09	710	97.88	462
Até 8 anos	99.87	770	99.79	471

Elaborado por Transparência Brasil

É possível notar que se repete o mesmo padrão do ano anterior de redução na diferença da porcentagem cumulativa das obras concluídas conforme aumentamos o intervalo entre a paralisação e a conclusão: conforme os anos aumentam, uma quantidade cada vez menor de obras é concluída, com uma redução brusca a partir de quatro anos.

Com base nos padrões observados, manteve-se a escolha do modelo de regressão logística para três períodos como base para as previsões: conclusão em até 1,5 ano, 2,5 anos e 3,5 anos. A precisão do modelo cresce à medida em que o limite de tempo analisado também aumenta, o que significa que a chance de acerto da probabilidade de uma obra ser concluída em 3,5 anos é maior do que a chance de acerto da probabilidade de uma obra ser concluída em 1,5 ano. Os três cenários podem ser explorados no simulador (ver anexo 2).



Gastos

Nesta seção explicamos o cálculo estimado do gasto público necessário para concluir as obras. Essa informação é relevante por dois motivos. Primeiramente, esse valor é usado como variável do modelo, considerando que obras que já receberam mais repasses estão associadas a maiores chances de conclusão. Em segundo lugar, saber o montante necessário para concluir as obras atualmente paralisadas fornece uma importante identificação do estado atual dessas obras.

Como o repasse do governo federal está definido pelo convênio celebrado para a construção das obras, a estimativa do montante que falta ser investido é a diferença entre o montante total celebrado e o que já foi transferido. Ocorre que há uma ausência desses dados em diversas obras, e muitas vezes os valores apresentados indicam estar desatualizados, dificultando o controle social e complicando a estimativa.

Por esse motivo, não foi possível utilizar todas as 3.343 obras identificadas como paralisadas na análise, mas apenas as 2.747 (82%) que possuíam informações sobre valores pagos. Para treinar o modelo, foram consideradas apenas aquelas com valores disponíveis tanto em relação ao momento da paralisação quanto ao momento da conclusão (se entregue). Para a predição, como as obras ainda não estavam concluídas, foram consideradas todas com valores pagos apenas no momento da paralisação. Foram adotados os valores pagos segundo os dados do FNDE.

Para conseguir estimar os recursos faltantes mesmo para as obras que não foram analisadas pelo modelo, foi calculada, a partir das 2.747 obras, uma correlação entre o valor estimado e o valor realmente repassado. A correlação obteve a marca de 80%, indicando ser uma métrica adequada. Os dados apresentados a seguir tomam como base essa estimativa.

Cumpramos observar que esses gastos são estimativas do montante necessário para que as obras sejam concluídas a partir do que foi gasto com outras obras que sofreram paralisações. Esses montantes não refletem as obrigações de repasse do FNDE, tratam-se apenas de estimativas de investimento necessário para que uma obra interrompida seja entregue, com base nos dados disponíveis.

Resultados

Em dezembro de 2020, havia 3.343 obras paralisadas. O modelo preditivo para prever a possibilidade de conclusão das obras em três anos e meio foi aplicável a 2.747 obras (82%). Dessas, 68 (2,4%) possuem mais de 60% de chance de serem concluídas em três anos e meio, proporção pouco abaixo da obtida em 2019 (3%).

O custo estimado para conclusão dessas 60 obras é de aproximadamente R\$ 28,4 milhões, ou aproximadamente R\$ 418 mil por obra, em média. Esse valor é quase o dobro do valor médio estimado em 2019 para conclusão das obras em situação similar



(R\$ 223 mil por obra). Uma explicação possível para esse aumento é o estado de execução das obras paralisadas. Em 2019, a média de execução das obras paralisadas era 90%. Em 2020, esse valor caiu para 67%, e como consequência sobe o gasto médio estimado para entrega das obras.

A figura 1 abaixo apresenta um mapa com a distribuição dos municípios que possuem obras atualmente paralisadas com mais de 60% de chance de serem concluídas em até três anos e meio. Em azul, as obras com mais de 80% de chance de serem concluídas dentro desse limite. Em verde, obras com probabilidade entre 70% e 80% e em amarelo, obras com mais de 60% e menos de 70% de chance. O mapa abaixo também pode ser visualizado [de forma interativa](#). Similarmente, é possível também acessar uma [versão com todas as obras](#), incluindo aquelas com baixa probabilidade de conclusão.

Figura 1 - Obras com mais de 60% de chances de serem concluídas em até 3,5 anos



Elaborado por Transparência Brasil



O que influencia a chance de conclusão de uma obra?

Além de prever informações sobre obras e suas conclusões, o modelo também indica o quanto cada variável se correlaciona com a possibilidade de uma obra ser concluída. Obter dados sobre essa correlação podem auxiliar no estudo do que influencia uma maior ou menor chance de uma obra ser bem sucedida, bem como identificar fatores associados com obras que tomam mais tempo para serem concluídas.

Assim como no modelo de 2019, número de vistorias do fiscal do município influi negativamente na probabilidade de obras paralisadas serem concluídas. Obras com maior número de vistorias durante a execução também possuem tal correlação, ainda que em menor escala.

A inserção de novas obras com os dados de 2020 provocou alteração em outras variáveis quando comparado ao modelo anterior. Apesar de as obras urbanas ainda manterem correlação maior com a conclusão das obras se comparadas a obras rurais, a diferença entre ambas foi reduzida no modelo de previsão para conclusão em até três anos e meio. Obras em áreas quilombolas passaram a influenciar negativamente a probabilidade de conclusão da obra, ao contrário do visto no modelo de 2019. Obras em locais indígenas também tiveram seu efeito invertido conforme passaram a se correlacionar com maiores chances de conclusão, comparado ao ano passado, quando estavam relacionadas a menores chances.

Cumprir notar que tal comportamento não é inesperado. Em função do baixo número de obras com dados disponíveis para treinamento, especialmente as em áreas rurais, indígenas e quilombolas, o resultado leva a alterações consideráveis conforme novas obras são inseridas. Entretanto, com a expansão da base ao longo dos anos, os valores obtidos tendem a se estabilizar caso o comportamento se mantenha durante os anos. Apesar de correlacionados, cabe observar que a correlação indica características comuns nas obras atrasadas, sem apontar esses fatores como causa direta dos atrasos das obras.

Anexos

Anexo 1: [Obras por probabilidade de término e gasto esperado](#)

Este anexo contém a lista completa de obras atualmente paralisadas, com a probabilidade de conclusão em um ano e meio, dois anos e meio e três anos e meio para cada uma. Como há algumas obras cujo gasto na época da paralisação não foi informado, é impossível gerar a estimativa para essa obra. Por esse motivo, algumas linhas estão identificadas com “desconhecido” na coluna de probabilidade.



Anexo 2: [Simulador de conclusão e investimento em obras](#)

Anexo 3: Mapas da distribuição da probabilidade de conclusão de obras paralisadas

- [Mapa](#) apenas com as obras com mais de 60% de chances de serem concluídas;
- [Mapa](#) com todas as obras paralisadas e suas probabilidades.

Anexo 4: Coeficientes das regressões

Variável	3,5 anos	2,5 anos	1,5 ano
Intercept	0.91	2.38	1.62
Classificação da obra (base obra indígena)			
Obra urbana	-0.22	-0.21	-0.35
Obra rural	-0.23	-0.98	-1.20
Obra quilombo	-0.96	-1.94	0.53
Sem informação	-2.55	-5.63	-3.03
Percentual de execução 1a. paralisação	0.05	0.05	0.06
Diferença na previsão	0.00	0.00	0.00
Gasto para conclusão (em milhares de reais)	0.00	0.00	0.00
Valor do contrato (em milhões de reais)	0.00	0.00	0.00
Quantidade de mudanças no prazo de conclusão	0.02	0.01	0.02
Quantidade de vitorias em situação em planejamento pelo proponente	0.00	-0.01	0.04
Quantidade de vitorias em situação paralisada	-0.21	-0.25	-0.26
Quantidade de vitorias em situação execução	-0.14	-0.17	-0.17
Quantidade de vitorias em situação licitação	-0.08	-0.19	-0.12



Anexo 5: O problema dos dados

O SIMEC Obras fornece detalhes individuais sobre o estado de cada obra financiada pelo FNDE. Os dados podem ser acessados individualmente pelo portal ou baixados em um csv completo contendo todas as informações. Como os dados fornecem uma fotografia atual da situação da obra sem seu histórico, foi necessário adaptar de diversas formas os dados existentes para realizar a análise desse relatório.

Inicialmente, arquivos coletados ao longo dos quatro anos foram compilados de forma que cada obra pudesse ter um histórico de seu status desde sua licitação até conclusão. Com esses arquivos foi possível recuperar o histórico de obras concluídas que foram em algum momento paralisadas.

A Transparência Brasil também criou uma classificação própria de obras paralisadas, para obras que tiveram seus processos administrativos retomados após interrupção das obras. Nesses casos a obra estava, por exemplo, em processo de licitação, mas seu canteiro de obras continuava parado. A partir de dados de assinatura de contrato, passamos a identificar obras que estavam novamente em processo de contratação ou repactuação com a obra paralisada. Assim, toda obra que já tivesse contrato assinado, mas cujo status oficial não fosse “em execução”, “contratação” ou “cancelada”, foi classificada como paralisada. Essa metodologia foi primeiramente utilizada no [Relatório da Transparência Brasil de 2017](#) sobre obras paralisadas.

Por último, os dados fornecidos também não identificam a data de conclusão da obra. Visando contornar essa limitação, foi utilizada a data da vistoria que determinou que a obra estava concluída.