

Intervalo de Confiança para a Estimativa do Produto Interno Bruto (PIB) do Estado de Alagoas

Superintendência de Informações e Cenários

Robson José Alves Brandão

Gerência de Indicadores e Cenários:

Juliana Carla da Silva Santos

Teresa Marcia da Rocha Lima Emery

Márcia Núbia Barbosa Lopes

Roberson Leite Silva Junior

Clerisvaldo Holanda dos Santos Junior

Introdução

A Secretaria de Estado do Planejamento, Gestão e Patrimônio (SEPLAG) tem como objetivo principal fornecer serviços de qualidade à sociedade, atendendo aos requisitos legais e às expectativas das partes interessadas. Uma das suas responsabilidades é produzir, gerir e disponibilizar dados e informações socioeconômicas e geoespaciais sobre o estado de Alagoas, incluindo o cálculo do Produto Interno Bruto (PIB) estadual, municipal e estimativa.

Nesse contexto, a SEPLAG confeccionou esta Nota Técnica cujo conteúdo trata da importância de estabelecer um intervalo de confiança para os resultados da estimativa do PIB de Alagoas. O intervalo de confiança é uma medida estatística essencial para obter estimativas precisas e confiáveis. Ele oferece uma faixa de valores na qual é provável que o verdadeiro parâmetro de interesse se encontre, com base em uma amostra observada.

O intervalo de confiança é uma ferramenta que, ao fornecer uma faixa de valores em torno de uma estimativa pontual, permite medir a incerteza associada a essa estimativa. Quanto mais amplo for o intervalo de confiança, maior será a incerteza em relação à localização do parâmetro. Por outro lado, um intervalo de confiança estreito indica uma estimativa mais precisa.

Outra vantagem do intervalo de confiança é a sua capacidade de controlar o nível de confiança desejado, que representa a probabilidade de que o intervalo

de confiança contenha o verdadeiro parâmetro. Normalmente, é utilizado um nível de confiança de 95%, o que significa que em 95% das amostras obtidas, o intervalo de confiança irá conter o parâmetro verdadeiro. No entanto, é possível ajustar o nível de confiança de acordo com as necessidades e a tolerância a erros do estudo.

Em resumo, o intervalo de confiança desempenha um papel crucial na inferência estatística. Ele fornece uma medida de incerteza, permitindo que pesquisadores e tomadores de decisão avaliem a precisão das estimativas e interpretem os resultados de forma mais robusta. Por meio do intervalo de confiança, é possível obter estimativas confiáveis e tomar decisões informadas com base em dados amostrais.

Objetivo

Determinar um intervalo de confiança que seja possível estimar o resultado do cálculo do PIB para o estado de Alagoas a partir do histórico dos resultados estimados anteriormente.

Metodologia

1. Levantamento de dados

Foram levantados os dados calculados pela SEPLAG e o resultado divulgado pelo IBGE, sendo possível ver na tabela a seguir.

Tabela 1 – Comparativo da estimativa do PIB e o resultado das Contas Regionais (CR)

ANO	SEPLAG	IBGE
2003	-1,29%	-1,15%
2004	6,04%	6,03%
2005	2,97%	3,56%
2006	2,39%	2,84%

ANO	SEPLAG	IBGE
2007	4,94%	5,24%
2008	6,84%	6,81%
2009	0,87%	0,96%
2010	4,87%	5,34%
2011	4,33%	4,65%
2012	1,60%	2,05%
2013	0,49%	0,38%
2014	4,65%	4,77%
2015	-2,59%	-2,88%
2016	-1,15%	-1,35%
2017	3,56%	3,33%
2018	0,99%	1,11%
2019	1,90%	1,95%
2020	-4,32%	-4,23%

Fonte: IBGE/CONAC – SEPLAG/SINC

2. Tratamentos de dados

O principal elemento para o cálculo do intervalo de confiança é a diferença absoluta entre o valor calculado pela SEPLAG e o valor divulgado pelo IBGE como amostra. Dessa forma, o importante não é se o erro está acima ou abaixo do resultado final, mas sim a magnitude do erro, conforme demonstrado na tabela 2 a seguir.

Tabela 2 – Diferença em pontos percentuais entre os resultados da estimativa SEPLAG e a CR

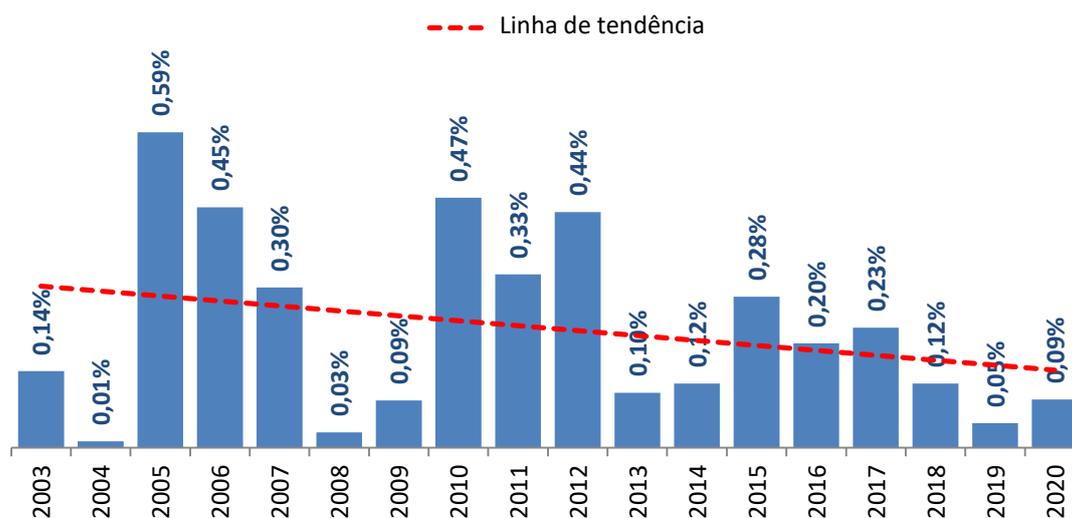
ANO	DIFERENÇA
2003	0,1440%
2004	0,0123%
2005	0,5919%
2006	0,4513%
2007	0,3011%
2008	0,0289%
2009	0,0891%

ANO	DIFERENÇA
2010	0,4690%
2011	0,3257%
2012	0,4423%
2013	0,1035%
2014	0,1204%
2015	0,2834%
2016	0,1964%
2017	0,2254%
2018	0,1206%
2019	0,0465%
2020	0,0909%

Fonte: IBGE/CONAC – SEPLAG/SINC

Com os dados acima, é possível observar que ao longo da série, a diferença entre o valor calculado pela SEPLAG e o calculado pelo IBGE tende a diminuir, visto que há pesquisas internas e testes estatísticos para que se aprimore o modelo e consiga convergir para o valor divulgado pelo IBGE, conforme observado na Figura 1 abaixo.

Figura 1 – Diferença em pontos percentuais entre os resultados da estimativa SEPLAG e a CR, considerando o valor absoluto do erro.



Fonte: IBGE/CONAC – SEPLAG/SINC

3. Validação temporal

Com a existência de tais condições, há a necessidade de usar a melhor forma de validação temporal para determinar o intervalo de confiança, sendo mais bem explicado nos tópicos a seguir.

Legenda de cores (referentes às figuras 2 a 4)

Cor	Descrição
	Ano estimado
	Dados utilizados
	Dados não utilizados

3.1 Validação temporal com janela expansiva

Para esta situação, à medida que se passa o tempo, o período anterior entra no conjunto de dados da amostra para que seja realizada uma estimativa posterior. Esta forma de validação é a mais utilizada. Conforme esquema a seguir:

Figura 2 – Esquema da validação temporal com janela expansiva.

	ANOS																	
Estimativa 1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	2
	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0
Estimativa 2	ANOS																	
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	2
Estimativa 3	ANOS																	
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	2	2	2
	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	1

Fonte: SEPLAG/SINC

3.2 Validação temporal com janela deslizante

Para esta situação, à medida que se passa o tempo, o período anterior entra no conjunto de dados da amostra que tem uma “memória” de períodos predeterminada para que seja realizada uma estimativa posterior. Conforme esquema a seguir.

Figura 3 – Esquema da validação temporal com janela deslizante.

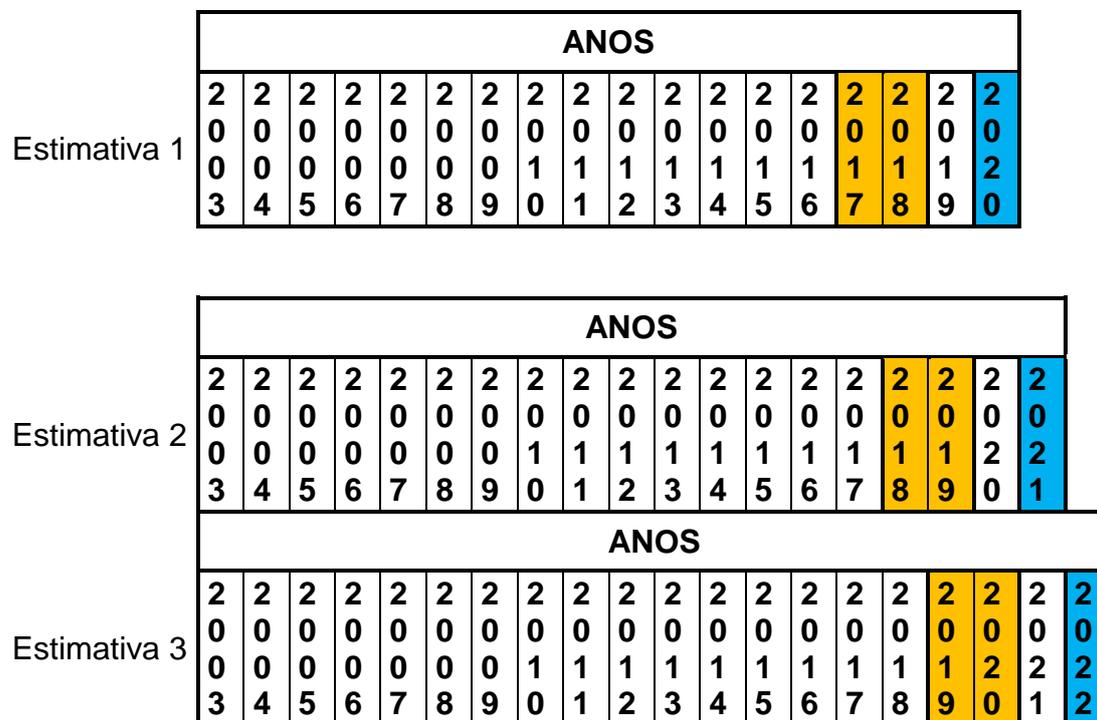
		ANOS																	
Estimativa 1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	2
	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	
		ANOS																	
Estimativa 2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	2
	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
		ANOS																	
Estimativa 3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	2	2
	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1

Fonte: SEPLAG/SINC

3.3 Validação temporal com janela deslizante e censura

Semelhante à validação temporal informada anteriormente, nesta há uma janela de “censura”. Esse tipo de validação é utilizado em casos em que existe a demora da liberação de dados para realizar tal validação, que é similar ao que acontece com o valor do PIB calculado pelo IBGE, o qual conta com uma defasagem de dois anos para sua liberação. Conforme esquema a seguir.

Figura 4 – Esquema da validação temporal com janela deslizante e censura.



Fonte: SEPLAG/SINC

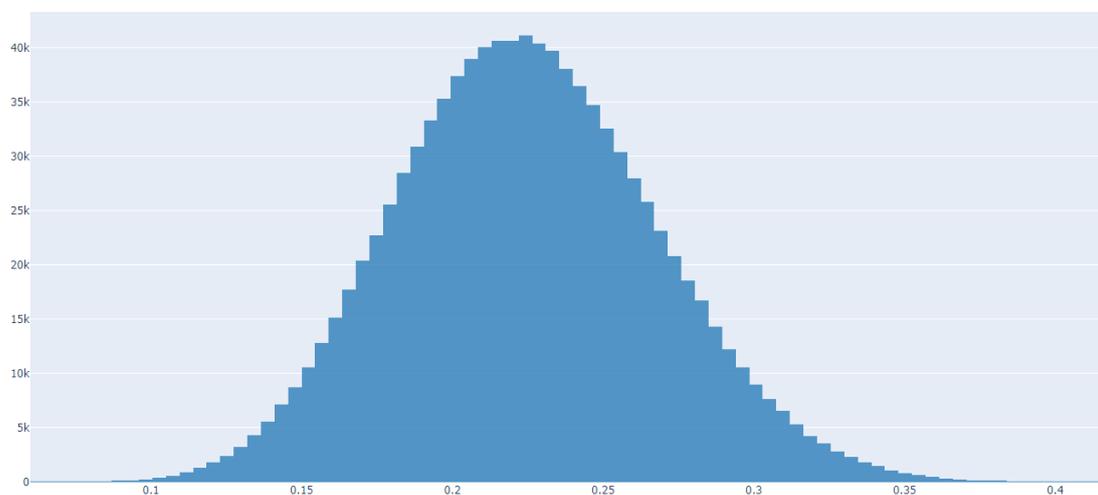
O método de censura pode ser realizado também com a validação temporal com janela expansiva pelo fato de que a censura é mais uma condição de disponibilização dos dados do que um método específico.

4. Definição do número de períodos da amostra

Para definir o número de períodos da amostra populacional, fez-se um teste numérico para ver qual a melhor janela de períodos, denominado *bootstrap*. Esta é uma técnica de reamostragem que assume que a distribuição de dados disponíveis da amostra é verdadeira, e a partir disso há uma seleção de dados aleatória do conjunto e a reposição da amostra com este dado.

Para este caso, foram feitos testes entre janelas com 2 períodos e 15 períodos, e estimado uma repetição de dados para os próximos um milhão de anos. A distribuição de dados pode ser visto na Figura 5 a seguir.

Figura 5 – Distribuição de dados utilizando *bootstrap* para uma janela com seis períodos.



Fonte: SEPLAG/SINC

É possível ver que a distribuição do erro é uma distribuição uniforme, com isso a probabilidade da diferença entre o valor calculado pela SEPLAG e o valor calculado pelo IBGE tende a ser entre 0,21% e 0,22% no infinito. E por ser também uma distribuição normal, à medida que aumenta o número de períodos da amostra há uma aproximação melhor da média, diminuindo o seu desvio padrão.

Para determinação do maior erro possível, usa-se a seguinte equação:

$$E = \bar{x} + \sigma \times T$$

Onde o E é o maior erro possível de acordo com o nível de confiança, o \bar{x} é a média da amostra de cada janela, o σ é o desvio padrão da amostra de cada janela e o T é o valor *t de student* de cada nível de confiança para cada tamanho de janela. Por exemplo, caso sejam utilizados os parâmetros de grau de confiança igual a 95%, há uma possibilidade de 95% de o erro estar entre 0 e E , sendo assim um teste unilateral.

Sabendo disso, foram feitos testes para observar como o desvio padrão se comportava com as mesmas janelas de período testadas anteriormente. Podendo ser visto o resultado na Tabela 3 a seguir.

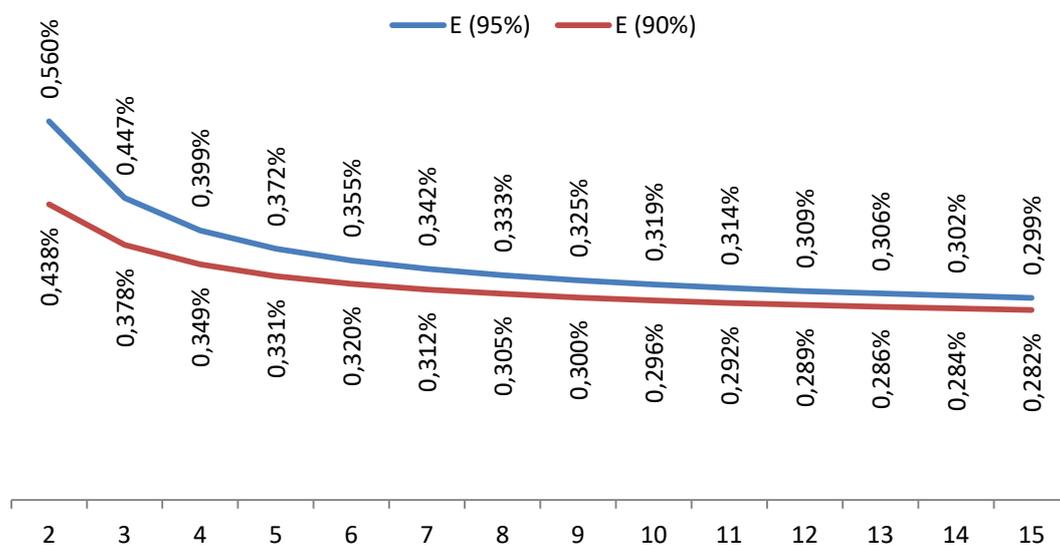
Tabela 3 – Teste para os valores de maior erro possível (*E*) por grau de confiança

N	MÉDIA	T (95%)	T (90%)	DESVIO PADRÃO	E (95%)	E (90%)
2	0,213700	2,920	1,886	0,118747	0,560%	0,438%
3	0,218533	2,353	1,638	0,097058	0,447%	0,378%
4	0,219850	2,132	1,533	0,084059	0,399%	0,349%
5	0,220560	2,015	1,476	0,075076	0,372%	0,331%
6	0,221467	1,943	1,440	0,068528	0,355%	0,320%
7	0,221929	1,895	1,415	0,063487	0,342%	0,312%
8	0,222312	1,860	1,397	0,059386	0,333%	0,305%
9	0,222422	1,833	1,383	0,056050	0,325%	0,300%
10	0,222650	1,812	1,372	0,053151	0,319%	0,296%
11	0,222891	1,796	1,363	0,050653	0,314%	0,292%
12	0,223008	1,782	1,356	0,048482	0,309%	0,289%
13	0,223162	1,771	1,350	0,046587	0,306%	0,286%
14	0,223279	1,761	1,345	0,044891	0,302%	0,284%
15	0,223353	1,753	1,341	0,043373	0,299%	0,282%

Fonte: SEPLAG/SINC

E os resultados calculados acima podem ser vistos melhor na figura abaixo.

Figura 6 – Comparação entre os maiores erros possível segundo grau de confiança.



Fonte: SEPLAG/SINC

A partir do gráfico acima, procurou-se o ponto de inflexão das curvas, sendo este o momento em que o aumento de períodos da janela começa a não diminuir tanto o erro máximo quanto com janelas menores. Sendo assim, foi encontrado um valor entre 4 e 6 períodos para o tamanho da janela deslizante para ambas as curvas.

5. Teste de hipóteses

O teste de hipótese foi utilizado para descobrir qual das opções teria o melhor desempenho e assim ser empregada para o cálculo do intervalo de confiança.

A seguir foram realizados testes do uso de janela expansiva (a mais utilizada), janela deslizante de 4 períodos, janela deslizante de 5 períodos e a janela deslizante de 6 períodos, como pode ser visto nos tópicos seguintes.

Para todos estes, foram utilizados os graus de confiança de 90% e 95%. Para isso, foi preciso utilizar a tabela da distribuição *t de student* para cada uma, lembrando que o teste é unilateral. Conforme tabela a seguir.

Tabela 4 – *T de student* (T) para cada tamanho de amostra (N) por grau de confiança

N	T (90%)	T (95%)
2	1,886	2,920
3	1,638	2,353
4	1,533	2,132
5	1,476	2,015
6	1,440	1,943
7	1,415	1,895
8	1,397	1,860
9	1,383	1,833
10	1,372	1,812
11	1,363	1,796
12	1,356	1,782
13	1,350	1,771
14	1,345	1,761

N	T (90%)	T (95%)
15	1,341	1,753
16	1,337	1,746
17	1,333	1,740
18	1,330	1,734

Fonte: MORETTIN, P. A.; BUSSAB, W. O - SEPLAG/SINC

5.1 Teste com janela expansiva com censura

Fazendo uso da validação temporal por janela expansiva com censura, devido ao fato da defasagem de dois anos pela disponibilização de dados pelo IBGE, foi necessário o cálculo da média e desvio padrão para a amostra de conjunto de anos anteriores ao ano estimado e multiplicado pelo valor de T da tabela anterior para 90% de grau de confiança. Com o resultado podendo ser visto na tabela a seguir.

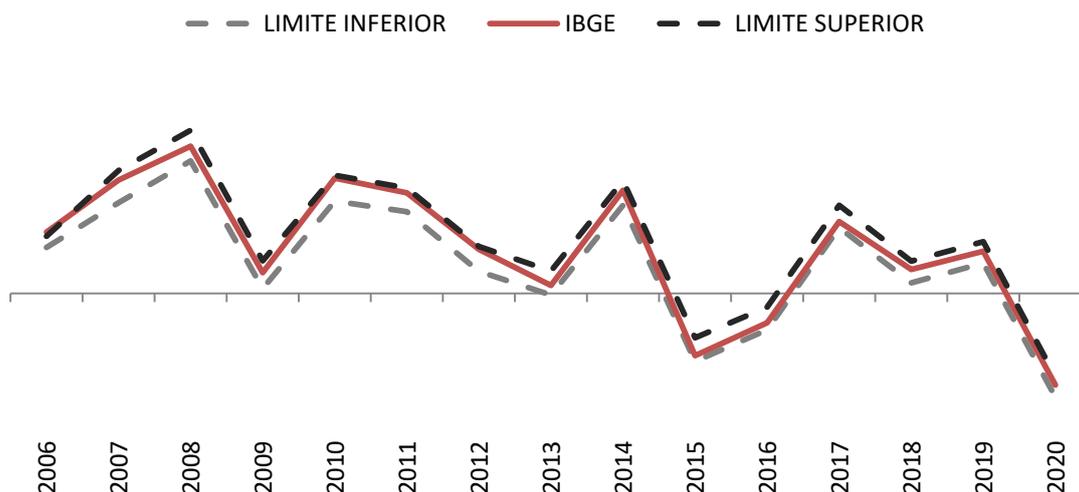
Tabela 5 – Cálculo do limite inferior e superior com grau de confiança de 90% para o teste com janela expansiva com censura

ANO	T (90%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2006	1,886	0,08	0,09	0,254	2,13	2,84	2,64
2007	1,638	0,25	0,30	0,747	4,19	5,24	5,69
2008	1,533	0,30	0,27	0,710	6,13	6,81	7,55
2009	1,476	0,30	0,23	0,643	0,23	0,96	1,52
2010	1,440	0,25	0,24	0,594	4,27	5,34	5,46
2011	1,415	0,23	0,22	0,548	3,78	4,65	4,87
2012	1,397	0,26	0,22	0,573	1,03	2,05	2,18
2013	1,383	0,27	0,21	0,559	-0,07	0,38	1,04
2014	1,372	0,29	0,21	0,568	4,08	4,77	5,21
2015	1,363	0,27	0,20	0,545	-3,14	-2,88	-2,05
2016	1,356	0,26	0,20	0,525	-1,68	-1,35	-0,63
2017	1,350	0,26	0,19	0,515	3,04	3,33	4,07
2018	1,345	0,25	0,18	0,500	0,49	1,11	1,49
2019	1,341	0,25	0,18	0,489	1,41	1,95	2,39
2020	1,337	0,24	0,17	0,476	-4,79	-4,23	-3,84

Fonte: SEPLAG/SINC

Os dados de limites inferior e superior podem ser vistos melhor na figura abaixo.

Figura 7 – Limite inferior e superior com grau de confiança de 90% para o teste com janela expansiva com censura



Fonte: SEPLAG/SINC

Com isso, todos os valores ficaram dentro do intervalo de confiança estimado com exceção de 2006 por ter uma amostra de apenas dois anos. Além disso, é percebida uma diminuição gradual do intervalo de confiança ao longo do tempo.

Analogamente, foi realizado também para um grau de confiança de 95%, podendo ter o resultado visto na tabela abaixo.

Tabela 6 – Cálculo do limite inferior e superior com grau de confiança de 95% para o teste com janela expansiva com censura

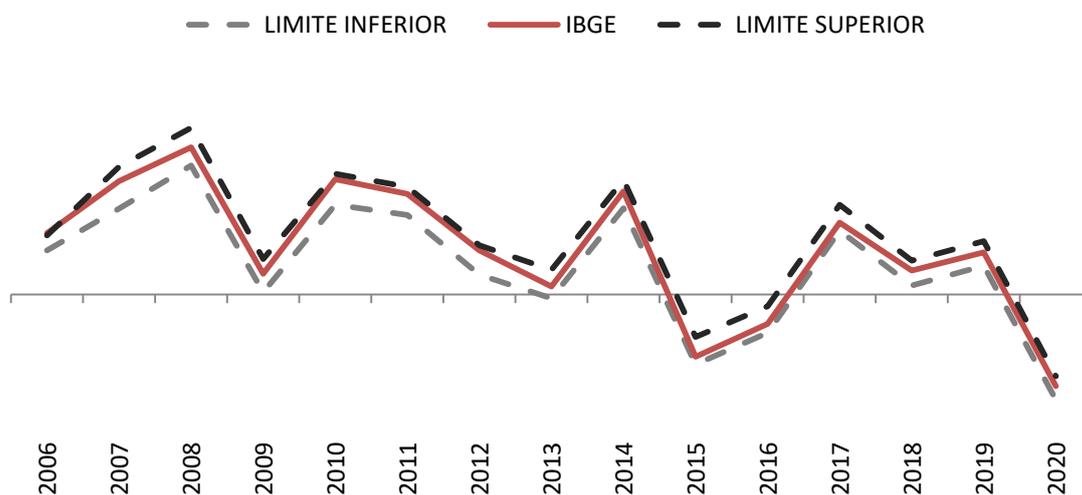
ANO	T (95%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2006	2,920	0,08	0,09	0,35	2,04	2,84	2,74
2007	2,353	0,25	0,30	0,96	3,97	5,24	5,90
2008	2,132	0,30	0,27	0,87	5,97	6,81	7,71
2009	2,015	0,30	0,23	0,77	0,11	0,96	1,64
2010	1,943	0,25	0,24	0,71	4,15	5,34	5,58

ANO	T (95%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2011	1,895	0,23	0,22	0,66	3,67	4,65	4,98
2012	1,860	0,26	0,22	0,68	0,93	2,05	2,28
2013	1,833	0,27	0,21	0,65	-0,17	0,38	1,14
2014	1,812	0,29	0,21	0,66	3,99	4,77	5,30
2015	1,796	0,27	0,20	0,63	-3,22	-2,88	-1,96
2016	1,782	0,26	0,20	0,61	-1,76	-1,35	-0,54
2017	1,771	0,26	0,19	0,59	2,96	3,33	4,15
2018	1,761	0,25	0,18	0,58	0,42	1,11	1,57
2019	1,753	0,25	0,18	0,56	1,34	1,95	2,46
2020	1,746	0,24	0,17	0,55	-4,86	-4,23	-3,77

Fonte: SEPLAG/SINC

Os dados de limites inferior e superior podem ser vistos melhor na Figura 8 a seguir.

Figura 8 – Limite inferior e superior com grau de confiança de 95% para o teste com janela expansiva com censura



Fonte: SEPLAG/SINC

É percebido o mesmo comportamento da figura anterior, porém com um intervalo maior devido ao grau de confiança ser maior.

5.2 Teste com janela deslizante de 4 períodos e censura

Fazendo uso da validação temporal por janela deslizante de 4 períodos e censura, foi necessário o cálculo do desvio padrão para a amostra de conjunto de anos anteriores ao ano estimado com deslocamento de um ano e multiplicado pelo valor de T da tabela somente para o valor de uma amostra de 4 períodos com um grau de confiança de 90%, pois para este caso o tamanho é único e igual a 1,533. Com o resultado podendo ser visto na tabela a seguir.

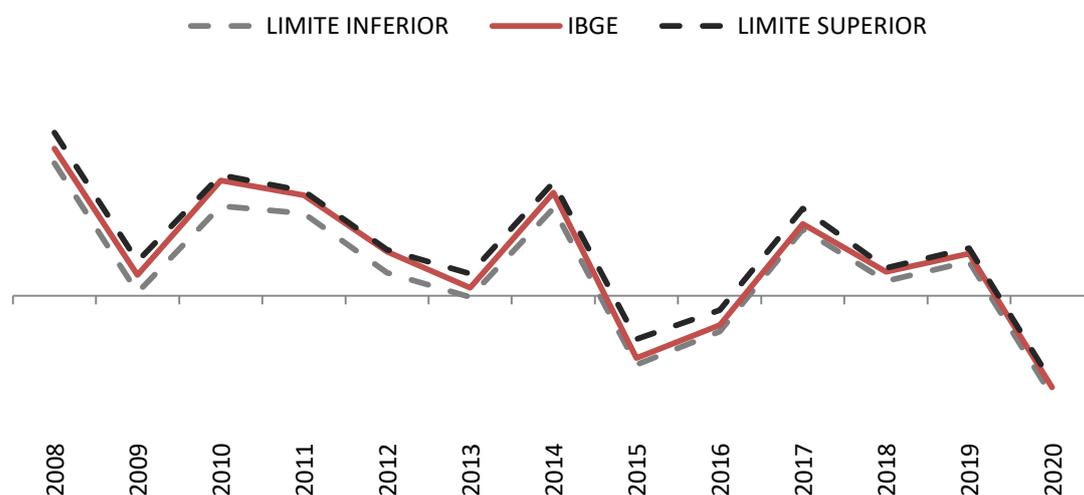
Tabela 7 – Cálculo do limite inferior e superior com grau de confiança de 90% para o teste com janela deslizante de 4 períodos e censura

ANO	T (90%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2008	1,533	0,30	0,27	0,710	6,13	6,81	7,55
2009	1,533	0,34	0,25	0,720	0,15	0,96	1,59
2010	1,533	0,34	0,24	0,713	4,15	5,34	5,58
2011	1,533	0,22	0,19	0,516	3,81	4,65	4,84
2012	1,533	0,22	0,20	0,531	1,07	2,05	2,14
2013	1,533	0,23	0,21	0,543	-0,06	0,38	1,03
2014	1,533	0,33	0,17	0,597	4,05	4,77	5,24
2015	1,533	0,34	0,17	0,590	-3,18	-2,88	-2,00
2016	1,533	0,25	0,16	0,500	-1,65	-1,35	-0,65
2017	1,533	0,24	0,16	0,481	3,07	3,33	4,04
2018	1,533	0,18	0,08	0,302	0,69	1,11	1,29
2019	1,533	0,21	0,07	0,310	1,59	1,95	2,21
2020	1,533	0,21	0,07	0,310	-4,63	-4,23	-4,01

Fonte: SEPLAG/SINC

Os dados de limites inferior e superior podem ser vistos melhor na Figura 9 a seguir.

Figura 9 – Limite inferior e superior com grau de confiança de 90% para o teste com janela deslizante de 4 períodos e censura



Fonte: SEPLAG/SINC

Observa-se que todos os valores calculados pelo IBGE ficaram dentro do intervalo definido pelo método e vai diminuindo o intervalo em momentos pontuais como 2012 e a partir de 2016.

Analogamente, foi calculado para um grau de confiança de 95%, sendo que o valor de T para uma janela de 4 períodos é igual a 2,132. Os valores calculados podem ser vistos na tabela a seguir.

Tabela 8 – Cálculo do limite inferior e superior com grau de confiança de 95% para o teste com janela deslizante de 4 períodos e censura

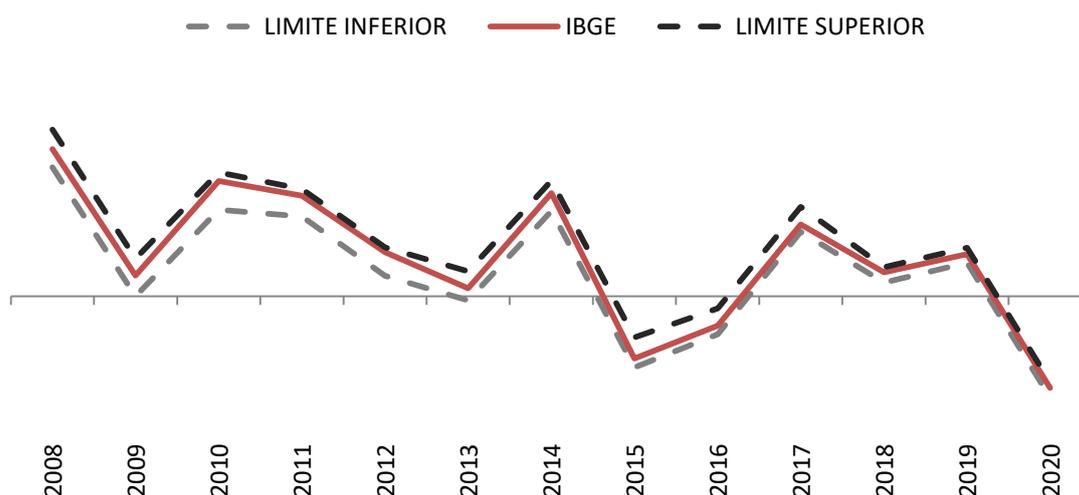
ANO	T (95%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2008	2,132	0,30	0,27	0,871	5,97	6,81	7,71
2009	2,132	0,34	0,25	0,868	0,00	0,96	1,74
2010	2,132	0,34	0,24	0,857	4,01	5,34	5,72
2011	2,132	0,22	0,19	0,633	3,69	4,65	4,96
2012	2,132	0,22	0,20	0,652	0,95	2,05	2,26
2013	2,132	0,23	0,21	0,666	-0,18	0,38	1,15
2014	2,132	0,33	0,17	0,701	3,94	4,77	5,35
2015	2,132	0,34	0,17	0,690	-3,28	-2,88	-1,90

ANO	T (95%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2016	2,132	0,25	0,16	0,598	-1,75	-1,35	-0,55
2017	2,132	0,24	0,16	0,576	2,98	3,33	4,13
2018	2,132	0,18	0,08	0,351	0,64	1,11	1,34
2019	2,132	0,21	0,07	0,351	1,55	1,95	2,25
2020	2,132	0,21	0,07	0,351	-4,67	-4,23	-3,97

Fonte: SEPLAG/SINC

Os dados de limites inferior e superior podem ser vistos melhor na figura abaixo.

Figura 10 – Limite inferior e superior com grau de confiança de 95% para o teste com janela deslizante de 4 períodos e censura



Fonte: SEPLAG/SINC

É percebido o mesmo comportamento da figura anterior, porém com um intervalo maior devido ao grau de confiança ser maior.

5.3 Teste com janela deslizante de 5 períodos e censura

Fazendo uso da validação temporal por janela deslizante de 5 períodos e censura, foi necessário o cálculo do desvio padrão para a amostra de conjunto de anos anteriores ao ano estimado com deslocamento de um ano e multiplicado pelo valor de T da tabela *t de student* somente para o valor de uma amostra de 5

períodos com um grau de confiança de 90%, pois para este caso o tamanho é único e igual a 1,476. Com o resultado podendo ser visto na tabela a seguir.

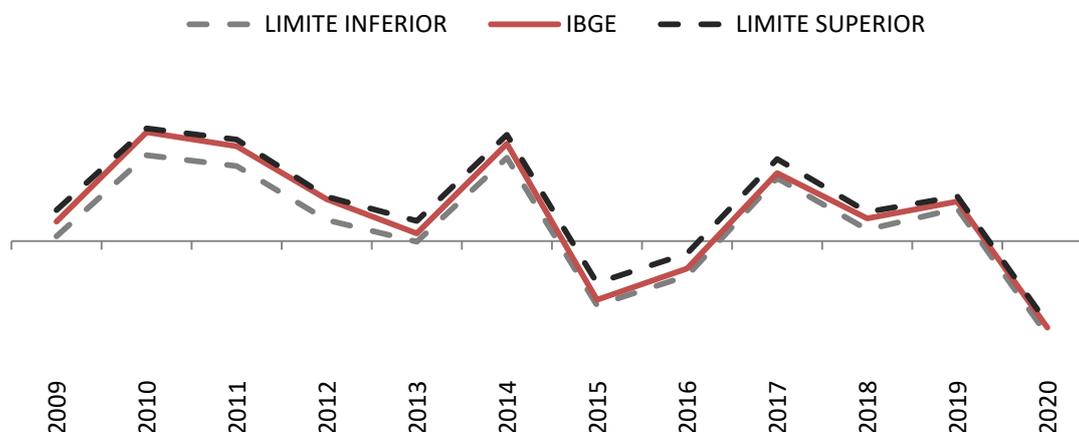
Tabela 9 – Cálculo do limite inferior e superior com grau de confiança de 90% para o teste com janela deslizante de 5 períodos e censura

ANO	T (90%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2009	1,476	0,30	0,23	0,643	0,23	0,96	1,52
2010	1,476	0,28	0,26	0,655	4,21	5,34	5,52
2011	1,476	0,29	0,24	0,643	3,68	4,65	4,97
2012	1,476	0,27	0,20	0,567	1,04	2,05	2,17
2013	1,476	0,24	0,18	0,510	-0,02	0,38	0,99
2014	1,476	0,27	0,20	0,569	4,08	4,77	5,21
2015	1,476	0,29	0,18	0,554	-3,15	-2,88	-2,04
2016	1,476	0,29	0,17	0,548	-1,70	-1,35	-0,60
2017	1,476	0,26	0,14	0,466	3,09	3,33	4,02
2018	1,476	0,23	0,14	0,434	0,56	1,11	1,43
2019	1,476	0,19	0,07	0,296	1,60	1,95	2,20
2020	1,476	0,19	0,07	0,293	-4,61	-4,23	-4,02

Fonte: SEPLAG/SINC

Os dados de limites inferior e superior podem ser vistos melhor na figura abaixo.

Figura 11 – Limite inferior e superior com grau de confiança de 90% para o teste com janela deslizante de 5 períodos e censura



Fonte: SEPLAG/SINC

É observado que todos os valores calculados pelo IBGE para o PIB ficaram dentro dos intervalos de confiança calculado, e que este intervalo veio diminuindo em momentos pontuais como 2013 e também a partir de 2017.

Analogamente, foi calculado para um grau de confiança de 95%, sendo que o valor de T para uma janela de 5 períodos é igual a 2,015. Os valores calculados podem ser vistos na tabela a seguir.

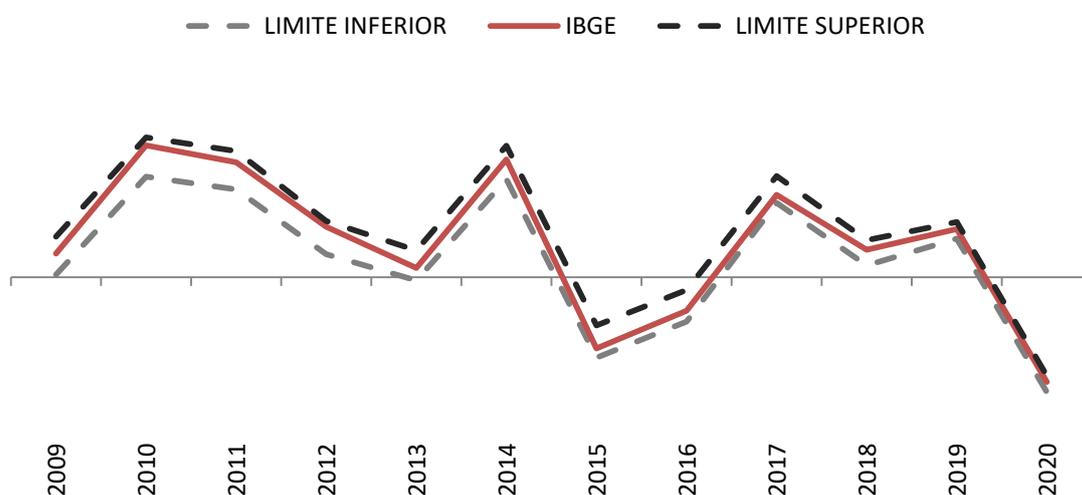
Tabela 10 – Cálculo do limite inferior e superior com grau de confiança de 95% para o teste com janela deslizante de 5 períodos e censura

ANO	T (95%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2009	2,015	0,30	0,23	0,768	0,11	0,96	1,64
2010	2,015	0,28	0,26	0,793	4,07	5,34	5,66
2011	2,015	0,29	0,24	0,771	3,56	4,65	5,10
2012	2,015	0,27	0,20	0,676	0,93	2,05	2,28
2013	2,015	0,24	0,18	0,607	-0,12	0,38	1,09
2014	2,015	0,27	0,20	0,678	3,97	4,77	5,32
2015	2,015	0,29	0,18	0,651	-3,24	-2,88	-1,94
2016	2,015	0,29	0,17	0,641	-1,79	-1,35	-0,51
2017	2,015	0,26	0,14	0,543	3,01	3,33	4,10
2018	2,015	0,23	0,14	0,509	0,48	1,11	1,50
2019	2,015	0,19	0,07	0,336	1,56	1,95	2,24
2020	2,015	0,19	0,07	0,331	-4,65	-4,23	-3,99

Fonte: SEPLAG/SINC

Os dados de limites inferior e superior podem ser vistos melhor na Figura 12 a seguir.

Figura 12 – Limite inferior e superior com grau de confiança de 95% para o teste com janela deslizante de 5 períodos e censura



Fonte: SEPLAG/SINC

É percebido o mesmo comportamento da figura anterior, porém com um intervalo maior devido ao grau de confiança ser maior.

5.4 Teste com janela deslizante de 6 períodos e censura

Fazendo uso da validação temporal por janela deslizante de 6 períodos e censura, foi necessário o cálculo do desvio padrão para a amostra de conjunto de anos anteriores ao ano estimado com deslocamento de um ano e multiplicado pelo valor de T da tabela t de student somente para o valor de uma amostra de 6 períodos com um grau de confiança de 90%, pois para este caso o tamanho é único e igual a 1,440. Com o resultado podendo ser visto na tabela a seguir.

Tabela 11 – Cálculo do limite inferior e superior com grau de confiança de 90% para o teste com janela deslizante de 6 períodos e censura

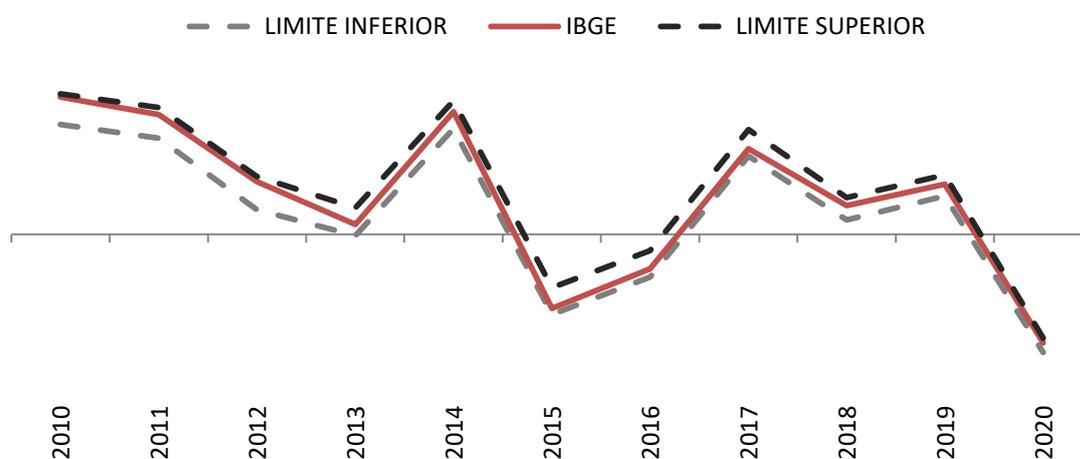
ANO	T (90%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2010	1,440	0,25	0,24	0,594	4,27	5,34	5,46
2011	1,440	0,25	0,24	0,593	3,73	4,65	4,92
2012	1,440	0,32	0,22	0,645	0,96	2,05	2,25
2013	1,440	0,28	0,18	0,541	-0,06	0,38	1,03

ANO	T (90%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2014	1,440	0,28	0,18	0,537	4,11	4,77	5,18
2015	1,440	0,24	0,19	0,521	-3,11	-2,88	-2,07
2016	1,440	0,26	0,18	0,511	-1,66	-1,35	-0,64
2017	1,440	0,29	0,15	0,514	3,04	3,33	4,07
2018	1,440	0,25	0,13	0,433	0,56	1,11	1,43
2019	1,440	0,23	0,12	0,407	1,49	1,95	2,31
2020	1,440	0,17	0,07	0,278	-4,60	-4,23	-4,04

Fonte: SEPLAG/SINC

Os dados de limites inferior e superior podem ser vistos melhor na figura abaixo.

Figura 13 – Limite inferior e superior com grau de confiança de 90% para o teste com janela deslizante de 6 períodos e censura



Fonte: SEPLAG/SINC

É observado que todos os valores calculados pelo IBGE para o PIB ficaram dentro dos intervalos de confiança calculado, e que este intervalo veio diminuindo em momentos pontuais como 2014 e também a partir de 2016.

Analogamente, foi calculado para um grau de confiança de 95%, sendo que o valor de T para uma janela de 6 períodos é igual a 1,943. Os valores calculados podem ser vistos na tabela a seguir.

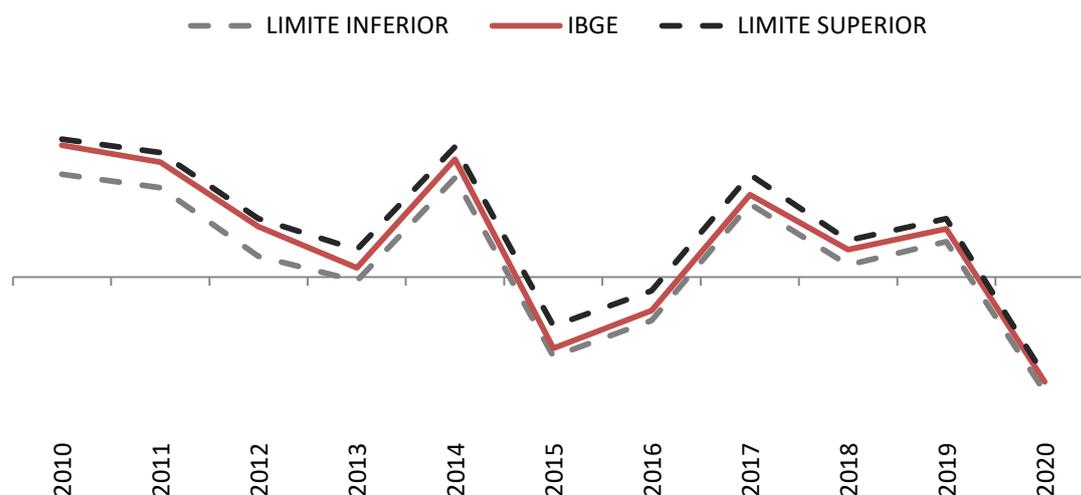
Tabela 12 – Cálculo do limite inferior e superior com grau de confiança de 95% para o teste com janela deslizante de 6 períodos e censura

ANO	T (95%)	MÉDIA	DESVIO PADRÃO	E	LIMITE INFERIOR	IBGE	LIMITE SUPERIOR
2010	1,943	0,25	0,24	0,712	4,15	5,34	5,58
2011	1,943	0,25	0,24	0,715	3,61	4,65	5,04
2012	1,943	0,32	0,22	0,758	0,85	2,05	2,36
2013	1,943	0,28	0,18	0,633	-0,15	0,38	1,12
2014	1,943	0,28	0,18	0,628	4,02	4,77	5,27
2015	1,943	0,24	0,19	0,618	-3,21	-2,88	-1,97
2016	1,943	0,26	0,18	0,600	-1,75	-1,35	-0,55
2017	1,943	0,29	0,15	0,592	2,96	3,33	4,15
2018	1,943	0,25	0,13	0,498	0,49	1,11	1,49
2019	1,943	0,23	0,12	0,470	1,43	1,95	2,37
2020	1,943	0,17	0,07	0,314	-4,63	-4,23	-4,00

Fonte: SEPLAG/SINC

Os dados de limites inferior e superior podem ser vistos melhor na Figura 14 a seguir.

Figura 14 – Limite inferior e superior com grau de confiança de 95% para o teste com janela deslizante de 6 períodos e censura



Fonte: SEPLAG/SINC

É percebido o mesmo comportamento da figura anterior, porém com um intervalo maior devido ao grau de confiança ser maior.

6. Comparação de intervalos de confiança

Para cálculo do tamanho do intervalo de confiança para um grau de confiança de 90%, foi subtraído o limite inferior do limite superior definido no capítulo anterior. Conforme tabela a seguir.

Tabela 13 – Comparação dos intervalos com grau de confiança de 90%

ANO	VALIDAÇÃO EXPANSIVA	VALIDAÇÃO JANELA 4 PERÍODOS	VALIDAÇÃO JANELA 5 PERÍODOS	VALIDAÇÃO JANELA 6 PERÍODOS
2006	0,508			
2007	1,494			
2008	1,421	1,421		
2009	1,285	1,439	1,285	
2010	1,187	1,425	1,309	1,187
2011	1,095	1,032	1,286	1,187
2012	1,146	1,063	1,134	1,290
2013	1,118	1,086	1,019	1,082
2014	1,135	1,194	1,138	1,074
2015	1,090	1,181	1,107	1,043
2016	1,050	1,000	1,096	1,023
2017	1,029	0,962	0,933	1,028
2018	1,001	0,604	0,868	0,866
2019	0,978	0,621	0,592	0,815
2020	0,952	0,620	0,585	0,557

Fonte: SEPLAG/SINC

Analisando a tabela anterior, percebe-se que a validação expansiva é melhor de 2006 até o ano de 2010, sendo o momento em que ela se mantém como o modelo de validação que tem o menor intervalo de confiança estimado. A partir de 2011, seu uso não é mais interessante.

Tendo em vista uma mudança de comportamento no erro entre os cálculos do IBGE e SEPLAG, o qual começou a ficar bem menor em comparação aos

primeiros anos desta série histórica. Sendo assim, o modelo de validação temporal com janela deslizante torna-se mais eficaz a partir deste ano, pelo fato de que ele usa somente dados mais recentes para estimar o intervalo de confiança, “esquecendo” os momentos em que os erros eram maiores.

Comparando as janelas deslizantes, observa-se que a de 4 períodos teve uma antecipação mais rápida e um número maior de menores intervalos de confiança do que as de 5 e 6 períodos. O tempo de resposta que esta janela tem é dado pelo alto peso atribuído à mudança pela quantidade reduzida da amostra, que para a primeira é de 25,0%, enquanto para as demais são de 20,0% e 16,7%, respectivamente.

Analogamente, foram realizados os cálculos para definição do tamanho do intervalo de confiança com um grau de confiança de 95%. Podendo ser visto na tabela a seguir.

Tabela 14 – Comparação dos intervalos com grau de confiança de 95%

ANO	VALIDAÇÃO EXPANSIVA	VALIDAÇÃO JANELA 4 PERÍODOS	VALIDAÇÃO JANELA 5 PERÍODOS	VALIDAÇÃO JANELA 6 PERÍODOS
2006	0,700			
2007	1,929			
2008	1,742	1,742		
2009	1,535	1,736	1,535	
2010	1,424	1,714	1,585	1,424
2011	1,310	1,265	1,542	1,429
2012	1,353	1,305	1,352	1,516
2013	1,307	1,332	1,214	1,265
2014	1,316	1,401	1,356	1,256
2015	1,266	1,380	1,303	1,237
2016	1,219	1,196	1,282	1,200
2017	1,189	1,152	1,087	1,184
2018	1,153	0,702	1,018	0,997
2019	1,123	0,702	0,672	0,939
2020	1,095	0,702	0,661	0,629

Fonte: SEPLAG/SINC

É percebido que o comportamento é parecido com o apresentado quando utilizado um grau de confiança de 90%, porém com um intervalo um pouco maior devido ao aumento do nível de confiança, pois é esperado que com 100 estimativas, tenha-se errado em 5 estimativas, ficando fora do intervalo estipulado, enquanto para o grau de 90% é esperado um erro em 10 estimativas.

7. Conclusões

Tendo em vista a sensibilidade para uma mudança de comportamento no erro, sendo positivo ou negativo, além da antecipação de tamanhos de intervalos de confiança menores que 1,0% mais rápida que as outras formas de validação temporal, e também pela maior quantidade de menores intervalos de confiança entre todas as formas de validação testadas, a forma de validação a ser escolhida será a de janela deslizante de 4 períodos com censura de 1 período.

Além disso, apesar de que a utilização de determinado grau de confiança interferir somente no momento em que o desvio padrão da amostra for alto, será utilizado o grau de confiança de 95%. A partir do momento em que o desvio padrão da amostra varia pouco, tendendo a zero, a definição do nível de confiança não se torna importante para definição deste erro, pois este aproximará da média amostral.

Sendo assim, a utilização desse conjunto de parâmetros para determinar o intervalo de confiança da estimativa do PIB estadual de Alagoas mostra-se eficaz por manter um intervalo relativamente curto, com uma alta confiança deste resultado, além de que se houver uma mudança de comportamento no erro, este intervalo se adaptará mais rápido do que outros métodos testados devido à alta sensibilidade da janela utilizada.

Secretaria de Estado do Planejamento, Gestão e Patrimônio

Secretário – Gabriel Albino Ponciano Nepomuceno

Secretária Especial de Planejamento, Orçamento e Governo Digital –

Nathalia Lavínia Farias de Araújo

Superintendência de Informações e Cenários (SINC):

Superintendente – Robson José Alves Brandão

Revisores:

Márcio de Mendonça Melânia

Nathalia Lavínia Farias de Araújo

Simone Craveiro Barros Pessôa

Equipe de apoio:

Alesson Santana Ferro

Antônio Bernardo Batista de Carvalho Soccol

Araken Barbosa da Silva

Klebson da Silva Lionaldo dos Santos

Luciano Soares Silvestre

Marcus Vinicius Sarmento de Souza

Mateus Henrique Gomes da Silva Bispo

8. Referências

MORETTIN, P. A.; BUSSAB, W. O. **Estatística Básica**. 9. Ed. São Paulo, SP, 2017.

FILHO, Mario. Validação de Séries Temporais Em Machine Learning Usando Python. **Mario Filho | Machine Learning**, 2023. Disponível em: <https://mariofilho.com/validacao-de-series-temporais-em-machine-learning-usando-python/>. Acesso em: 22 mar 2023.

MAYER, Fernando. Métodos de reamostragem: Bootstrap (não paramétrico). **CE089 - Estatística Computacional II**, 2022. Disponível em: http://cursos.leg.ufpr.br/ce089/10_bootstrap.html. Acesso em: 22 mar 2023.

CLAUDIO. A distribuição normal. EDTI, 2022. Disponível em: <https://www.escolaedti.com.br/a-distribuicao-normal>. Acesso em: 22 mar 2023.