

Atividade Avaliativa 3 - Análise Extra

Alisson Rosa e Vítor Pereira

Sumário

1	Análise de Influência - DFBETAS	1
1.1	DFBETAS - Modelo Original	1
1.2	DFBETAS - Modelo Ajustado	2
2	Análise Descritiva	3
2.1	Correlação	4
3	Modelos de Regressão Linear	5
4	Conclusão	5

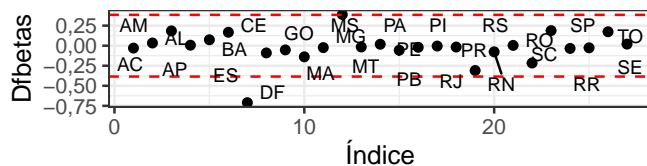
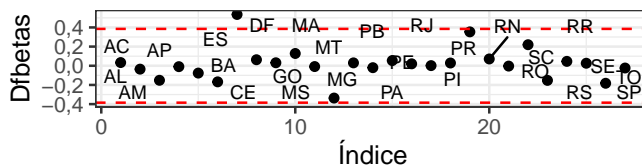
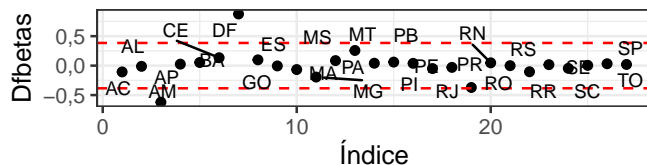
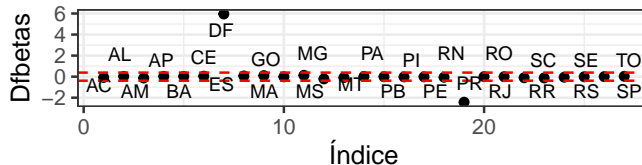
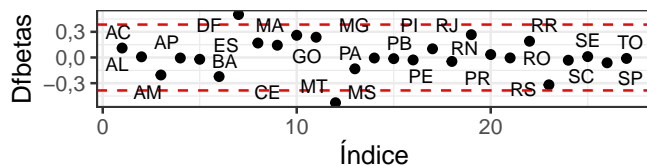
1 Análise de Influência - DFBETAS

No diagnóstico `dfbetas`, que informam o grau de influência que a observação i tem sobre o coeficiente \hat{x}_i , ou seja, sobre os parâmetros $\hat{\beta}_i$. Então é uma medida complementar ao `DFFITS`, que verifica a influência de i tem sobre o valor seu próprio valor ajustado \hat{y}_i .

1.1 DFBETAS - Modelo Original

Nesta seção teremos as análises dos DFBETAS para o modelo inicial com o DF.

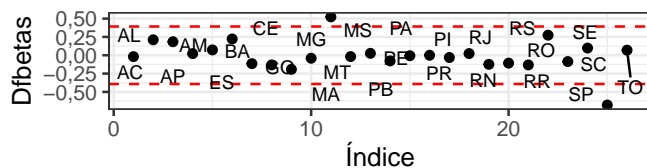
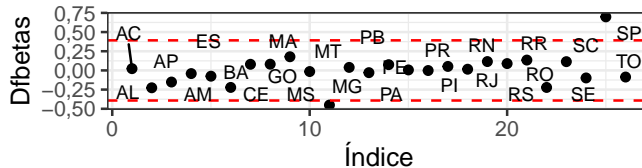
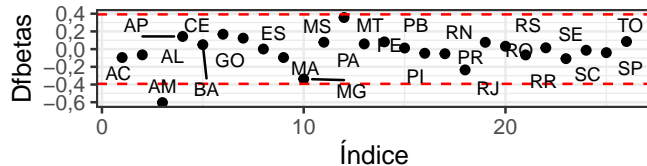
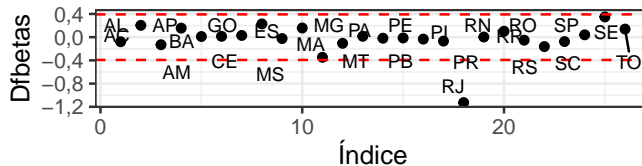
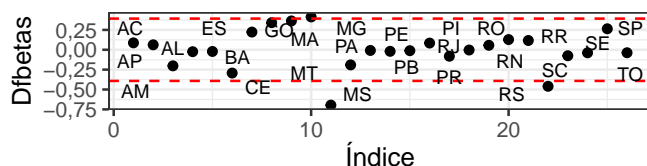
Nos gráficos abaixo podemos ver que o DF está fora de todos os limites do DFBETAS, principalmente no DFBETA do x4 em que realmente o DF achata muito o gráfico. Os estados de AM, RJ e MS são candidatos a ponto de influente nas covariáveis de Área, Densidade Demográfica e Taxa de Pobreza.

x1 Dfbeta 1 vs índice**x2 Dfbeta 2 vs índice****x3 Dfbeta 3 vs índice****x4 Dfbeta 4 vs índice****x5 Dfbeta 5 vs índice**

1.2 DFBETAS - Modelo Ajustado

Nesta seção teremos as análises dos DFBETAS para o modelo inicial sem o DF.

Nos gráficos abaixo podemos ver que não tem um estado que está fora de todos os limites do DFBETAS. MS está fora do DFBETA 1 e 5, SP está fora do DFBETA 1 e 2, o AM está fora do DFBETA 3 e RJ fora do DFBETA 4. Também podemos notar que nenhum dos possíveis pontos influentes achata o gráfico de uma forma extrema como acontecia com o DF no modelo original.

x1 Dfbeta 1 vs índice**x2 Dfbeta 2 vs índice****x3 Dfbeta 3 vs índice****x4 Dfbeta 4 vs índice****x5 Dfbeta 5 vs índice**

2 Análise Descritiva

Nesta seção veremos um breve resumo das variáveis de estudo, com medidas descritivas, medidas de dispersão e gráficos de dispersão do banco original com o DF e a comparação dessas medidas com o banco e o modelo ajustado

Começaremos por uma tabela resumo, com informações sobre as covariáveis, em comparação com a banco ajustado podemos notar uma diferença significativa matematicamente da média do PIB per capita 27172,676 no banco com DF e 25121,530, o desvio padrão no banco original é 46% maior, assim como podemos notar um aumento da média na Densidade Demográfica de 30% e 48% no desvio padrão, isso tudo com apenas uma observação a mais, podemos notar a influência do DF nas medidas descritivas.

Tabela 1: Resumo das variáveis:

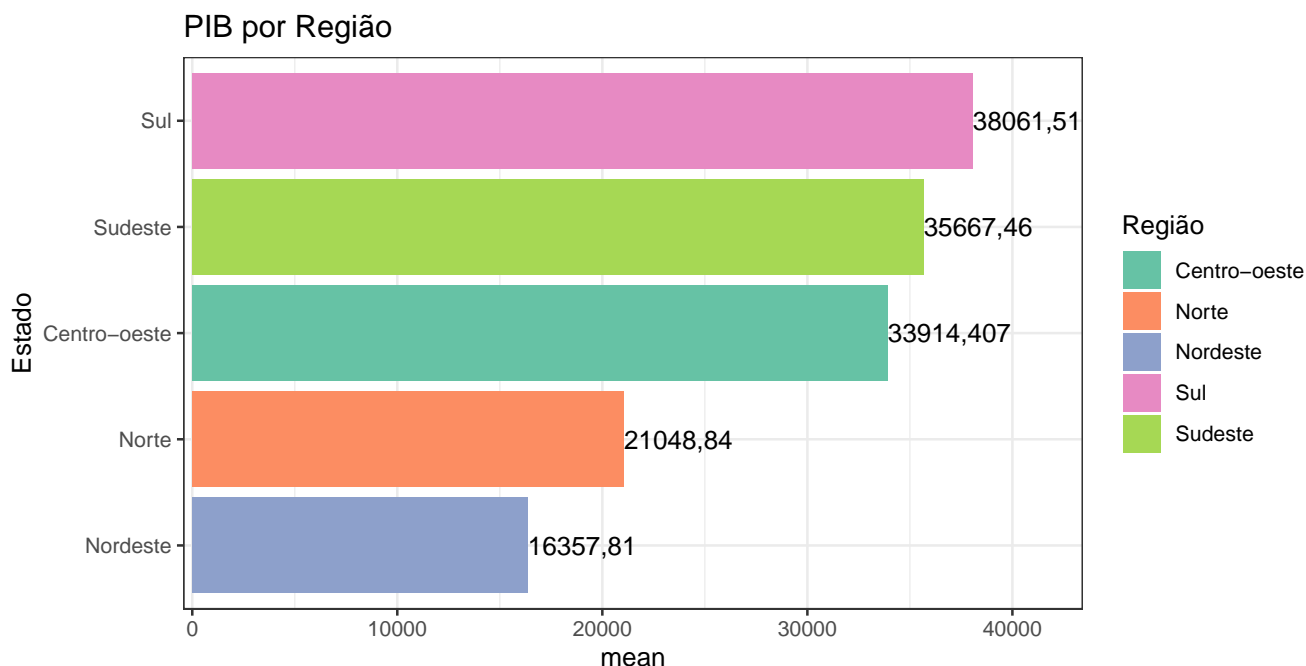
	n	Média	Desvio Padrão	Mediana	Minímo	Máximo
PIB	27	27172,676	14336,816	22936,280	12788,750	80502,470
IDHe	27	0,718	0,050	0,717	0,636	0,828
Área	27	315215,579	375101,871	224273,831	5760,783	1559168,117
Densidade Demográfica	27	75,917	120,552	36,090	2,660	523,410
Pobreza	27	0,116	0,075	0,128	0,017	0,263

A título de curiosidade vejamos o PIB por região:

Tabela 2: PIB per capita por Região

Região	Média PIB	Desvio padrão	Estados
Centro-oeste	45561	23649	4
Norte	21049	2643	7
Nordeste	16358	2070	9
Sul	38062	1328	3
Sudeste	35667	9566	4

Notamos que apenas o acréscimo do DF no banco de dados, faz com que a posição do Centro-Oeste de PIB per capita, saia do 3° lugar em média para o 1°, sendo a única região com maior de 40 mil. Assim como é a região disparadamente com maior desvio padrão, sendo 147% maior que o desvio padrão da região sudeste, a região com segundo maior desvio padrão.



2.1 Correlação

Nesta seção, será comparada a correlação dos banco de dados com o DF e sem o DF.

Temos que a matriz de correlação do banco original é:

Tabela 3: Correlação entre as variáveis

	PIB	IDHe	Área	Densidade Demográfica	Pobreza
PIB	1,000	0,773	-0,112	0,727	-0,759
IDHe	0,773	1,000	-0,003	0,405	-0,716
Área	-0,112	-0,003	1,000	-0,371	0,193
Densidade Demográfica	0,727	0,405	-0,371	1,000	-0,380
Pobreza	-0,759	-0,716	0,193	-0,380	1,000

Temos que a matriz de correlação do banco ajustado é:

Tabela 4: Correlação entre as variáveis

	PIB	IDHe	Área	Densidade Demográfica	Pobreza
PIB	1,000	0,822	0,016	0,391	-0,880
IDHe	0,822	1,000	0,058	0,237	-0,692
Área	0,016	0,058	1,000	-0,376	0,158
Densidade Demográfica	0,391	0,237	-0,376	1,000	-0,292
Pobreza	-0,880	-0,692	0,158	-0,292	1,000

Podemos notar que a correlação entre IDHe e PIB per capita cresce no modelo ajustado, inverte totalmente a relação linear entre a área e a variável desfecho, deixando de ser uma correlação positiva para uma correlação negativa. Como o DF tem a maior densidade demográfica e o maior PIB per capita, a correlação entre as duas variáveis aumenta quase 100%, assim também diminuindo a correlação da taxa de pobreza extrema e a variável resposta, isso demonstra novamente a influência da Distrito Federal nas nossas análises.

3 Modelos de Regressão Linear

Tem-se portanto como resumo do modelo final com o DF como observação é a seguinte tabela:

Tabela 5: Resumo do modelo original

Coeficientes	Estimativa	Erro Padrão	p-valor
(Intercept)	-29249,678	24898,553	0,253
IDHe	81293,014	32880,246	0,022
Área	0,006	0,003	0,075
‘Densidade Demográfica’	62,199	10,675	0,000
Pobreza	-73736,452	21224,661	0,002

Que informa que o intercepto não são significativos a 1%, tem-se um p-valor « 0.001 do teste F e R^2 dado por 0.876 que informa que aproximadamente 87.6% da variação do PIB per capita dos estados é explicada pelas covariáveis propostas.

E o modelo final sem o DF é dado pela tabela abaixo:

Tabela 6: Resumo do modelo final

Coeficientes	Estimativa	Erro Padrão	p-valor
(Intercept)	-18296,206	15527,248	0,252
IDHe	70391,625	20444,077	0,002
Área	0,004	0,002	0,046
‘Densidade Demográfica’	22,830	9,292	0,023
Pobreza	-80336,960	13190,734	0,000

Portanto podemos notar novamente a influência do DF, em que temos uma variação de 60% no intercepto, uma diminuição do beta 2 em 15%, uma pequena alteração no beta 3, uma diminuição de mais de 180% no beta 4 e um aumento da influência negativa de quase 10% no beta 5. Sendo assim o modelo ajustado com DF é consideravelmente diferente do modelo sem DF. Assim podemos notar que há um aumento inclusive na explicação geral das covariáveis no PIB per capita, pois o R^2 no modelo com DF é por 0.876, e no modelo final do projeto é 0.899, mesmo com uma observação a menos o modelo proposto pelo trabalho é superior em explicação da variável desfecho.

4 Conclusão

Podemos notar que o modelo proposto cumpre todos os testes de pressupostos, ao contrário do modelo original, tendo covariáveis que fazem sentido com realidade para explicar o PIB per capita, sendo esses de áreas realmente interessantes para o estudo desse componente social, como: População, Geografia, Condição de Vida e Educação.

Assim como percebemos a necessidade da remoção da observação do DF pelas análises de influência, análise descritiva e análise dos modelos ajustados, verificando uma clara interferência dessa observação no modelo final, que está bem justificada na parte principal do trabalho. Podemos confirmar que essa é uma análise de regressão acurada, pois possui covariáveis que explicam a variável dependente de forma significativa e correlata com os conhecimentos de outras áreas, com uma boa noção de causalidade.