

# Explorando a Relação entre Variáveis Ambientais e a Produção de Clorofila-a em Cianobactérias: Uma Abordagem com Modelos Lineares Generalizados

Thiago Tavares Lopes

15 novembro 2024

## Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Cianobactérias . . . . .	1
<b>2</b>	<b>Resultados</b>	<b>2</b>
2.1	Análise de Resíduos . . . . .	3

## 1 Introdução

Foi proposto um modelo linear generalizado para avaliar a produção de clorofila a em cianobactérias em diferentes condições climáticas. O *dataset* utilizado foi obtido do catálogo de dados público do governo dos Estados Unidos, disponível em *Data.Gov*. Esse *dataset* possui informações detalhadas sobre diferentes espécies de cianobactérias, sendo estas informações: Informações genéticas, condições climáticas do local de coleta das amostras e localização geográfica da coleta. Para a construção do modelo, foram consideradas exclusivamente as informações de clorofila a quantificada e condições ambientais (físicas e químicas) do local da coleta das amostras. O trabalho aqui desenvolvido foi fundamentado no artigo disponível no seguinte link.

### 1.1 Cianobactérias

Devido a atividade fotossintetizante das cianobactérias estima-se que as primeiras tiveram origem entre 2,6 a 3,5 bilhões de anos atrás e são uma das principais responsáveis pela atmosfera oxigenada como conhecemos hoje, participando do **Grande Evento de Oxigenação**, os estromatólitos são uma evidência da atividade de microrganismos que ocorreu aproximadamente 3,700 milhões de anos atrás. As cianobactérias são classificadas como microrganismo procariontes autotróficos com sistemas adaptativos particulares, como a capacidade de fixar nitrogênio do ar atmosférico devido a presença da enzima nitrogenase localizada nos heterócitos . São capazes de realizar a fotossíntese na presença ou ausência de oxigênio. Existem espécies que se desenvolvem na ausência de luz ou em condições anaeróbicas utilizando sulfetos como doadores de elétrons para a fotossíntese, além disso são bactérias gram – negativas e dispõem da estrutura chamada bainha mucilaginosa e tricoma, podem ou não apresentar o acineto e o heterócitos que são estruturas especializadas na sobrevivência da espécie em ambientes não favoráveis

Devido à grande complexidade metabólica destes organismos, uma variedade de espécies de cianobactérias produzem composto bioativos e de interesse econômico a partir do seu metabolismo como as citotoxinas e aminoácidos do tipo micosporina.

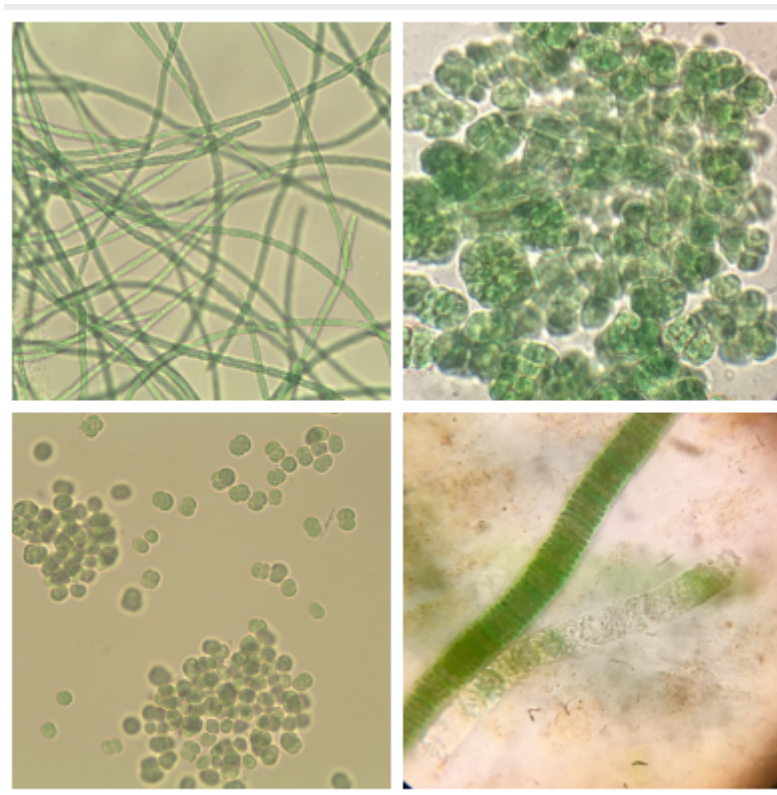


Figura 1: Cianobactérias

## 2 Resultados

Modelo 1

Tabela 1:

	Estimate	Std. Error	t value	Pr(> t )	Significance
(Intercept)	-7,17864	5,53834	-1,29617	0,20167	
total_nitrogen	0,00109	0,00054	2,02436	0,04903	*
total_phosphorus	0,00698	0,00441	1,58323	0,12053	
temp_water_celsius	-0,02700	0,05012	-0,53868	0,59282	
dissolved_oxygen	-0,25390	0,18544	-1,36920	0,17788	
pH_water	1,40239	0,61068	2,29644	0,02647	*
carbon_dioxide_water	0,24729	0,13331	1,85499	0,07031	.
total_nitrogen_water...8	0,28761	0,79233	0,36300	0,71834	
nitrite_water	-0,07889	5,14524	-0,01533	0,98784	
nitrate_water	0,18158	0,21271	0,85367	0,39791	
phosphorus_water	-1,14418	1,13485	-1,00823	0,31886	
sulfate_water	-0,00460	0,00567	-0,81218	0,42106	
total_nitrogen_water...13	-0,46093	0,84540	-0,54521	0,58836	
'ammonia(NH3 + NH4+)_water'	-1,31316	1,93832	-0,67747	0,50165	

<sup>a</sup> Nota:\*\*\* p<0.001; \*\* p<0.01; \* p<0.05; . p<0.1

Foi construído um novo modelo a partir do resultado da técnica de *stepwise*.

Nesta seção são apresentados os resultados da análise de resíduos para o modelo 2

## 2.1 Análise de Resíduos

```
# RESIDUAL ANALYSIS

# coeficiente de determinacao generalizado (Nagelkerke, 1991)

RsqGLM(fit2,plot=F)

## NOTE: Tjur R-squared applies only to binomial GLMs
## $CoxSnell
## [1] 0,41826
##
## $Nagelkerke
## [1] 0,41957
##
## $McFadden
## [1] 0,093991
##
## $Tjur
## [1] NA
##
## $sqPearson
## [1] 0,38694

# DIAGNOSTIC ANALYSIS

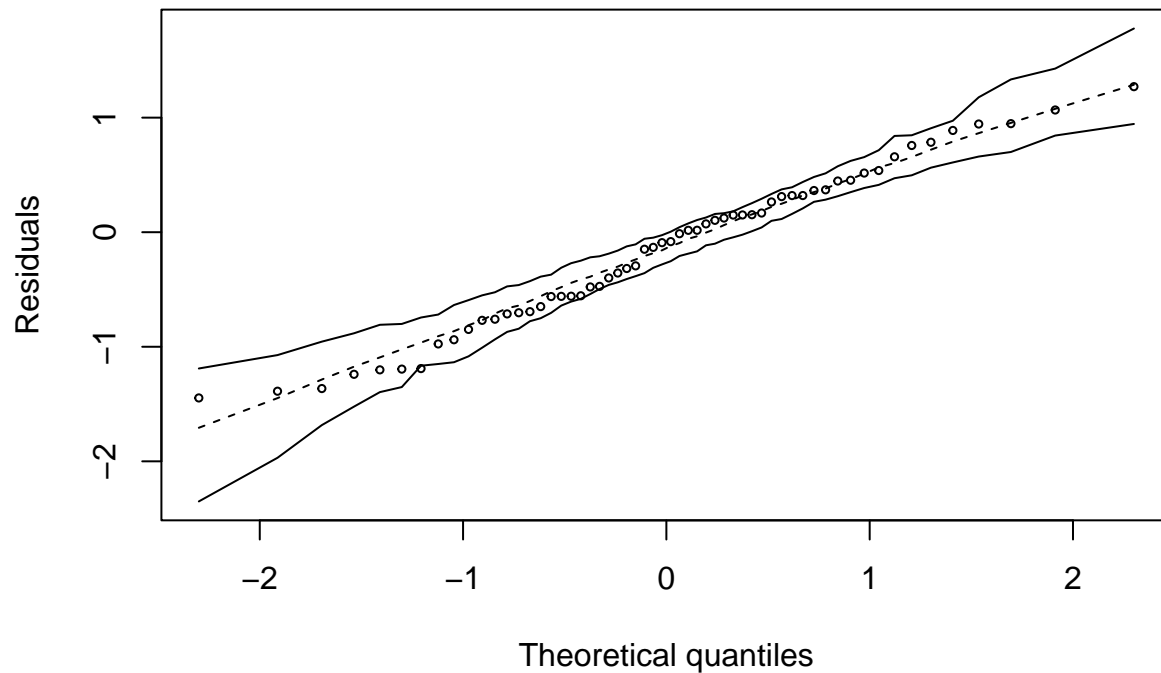
# td = resid(fit2,type="deviance")
#
# # RESIDUOS X INDICES
#
# abline(plot(td), lty=c(2,2,3), h=c(-2,2,0))
#
#
# abline(plot(fitted(fit2),td) , lty=c(2,2,3), h=c(-2,2,0))

#fitted(fit2) - aqui os valores já estão na escala original da variável resposta
#valores estimados para o preditor linear
#fit2$linear.predictors

#n1=2

#envelope simulado
#set.seed(2)
hnp(fit2, halfnormal = F, how.many.out = T, col.out = "red", main = "")

## Gamma model
```



```
## Total points: 58
```

```
## Points out of envelope: 1 ( 1,72 %)
```

```
## Alavancagem
## alavancagem vs valores ajustados
plot(fitted(fit2),hatvalues(fit2),ylab="Alavancagem")
##identify(fitted(fit),hatvalues(fit), n=n1) # identifica n1 valores no grafico
##which(hatvalues(fit2)>0.3)
#
## DFFIT
plot(dfrits(fit2),ylab="DFFITS")
##identify(dfrits(fit1), n=n1) # identifica n1 valores no grafico
#
## distancia de Cook
## cook vs valores ajustados
plot(fitted(fit2),cooks.distance(fit2),ylab="Distancia de Cook")
##identify(fitted(fit1),cooks.distance(fit1), n=n1)
##which(cooks.distance(fit2)>0.2)
```

```
#Grafico distancia de cook
```

```
n<-length(ciano)
```

```
k<-length(coef(fit2))
```

```
cooks_threshold<-4/(n-k)
```

```
fitted_values<-fitted(fit2)
```

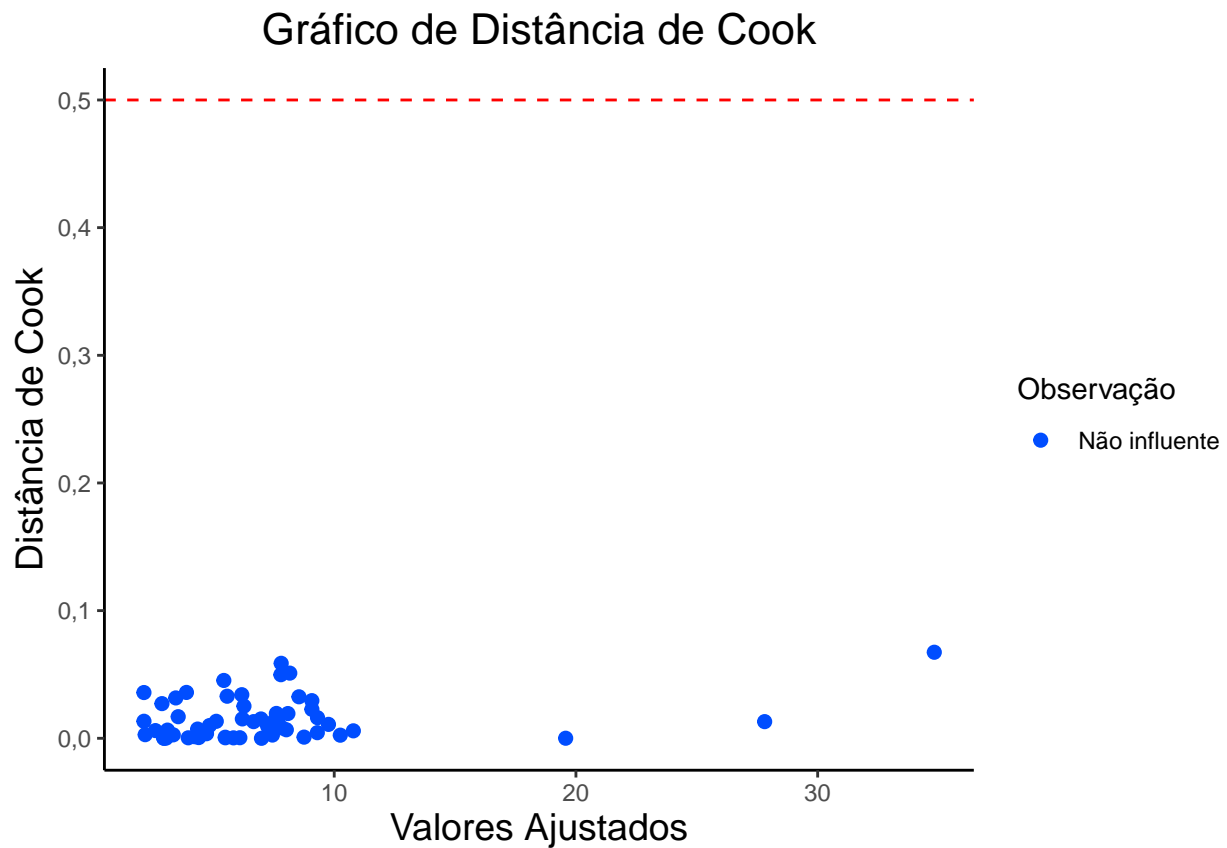
```
coocks_values<-cooks.distance(fit2)
```

```

cook_data<-data.frame(Fitted = fitted_values, CookDistance = cooks_values)

ggplot(cook_data, aes(x = Fitted, y = CookDistance))+
  geom_point(aes(color = CookDistance > cooks_threshold), size = 2)+
  geom_hline(yintercept = cooks_threshold, linetype = "dashed", color = "red")+
  labs(
    title = "Gráfico de Distância de Cook",
    x = "Valores Ajustados",
    y = "Distância de Cook"
  ) +
  theme_classic()+
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    plot.subtitle = element_text(hjust = 0.5, size = 12),
    axis.title = element_text(size = 14)
  ) +
  scale_color_manual(
    values = c("#004dff", "red"),
    labels = c("Não influente", "Influente"),
    name = "Observação"
  )
)

```



```

td <- resid(fit2, type = "deviance")
fitted_values <- fitted(fit2)

# Criando um data frame para o ggplot

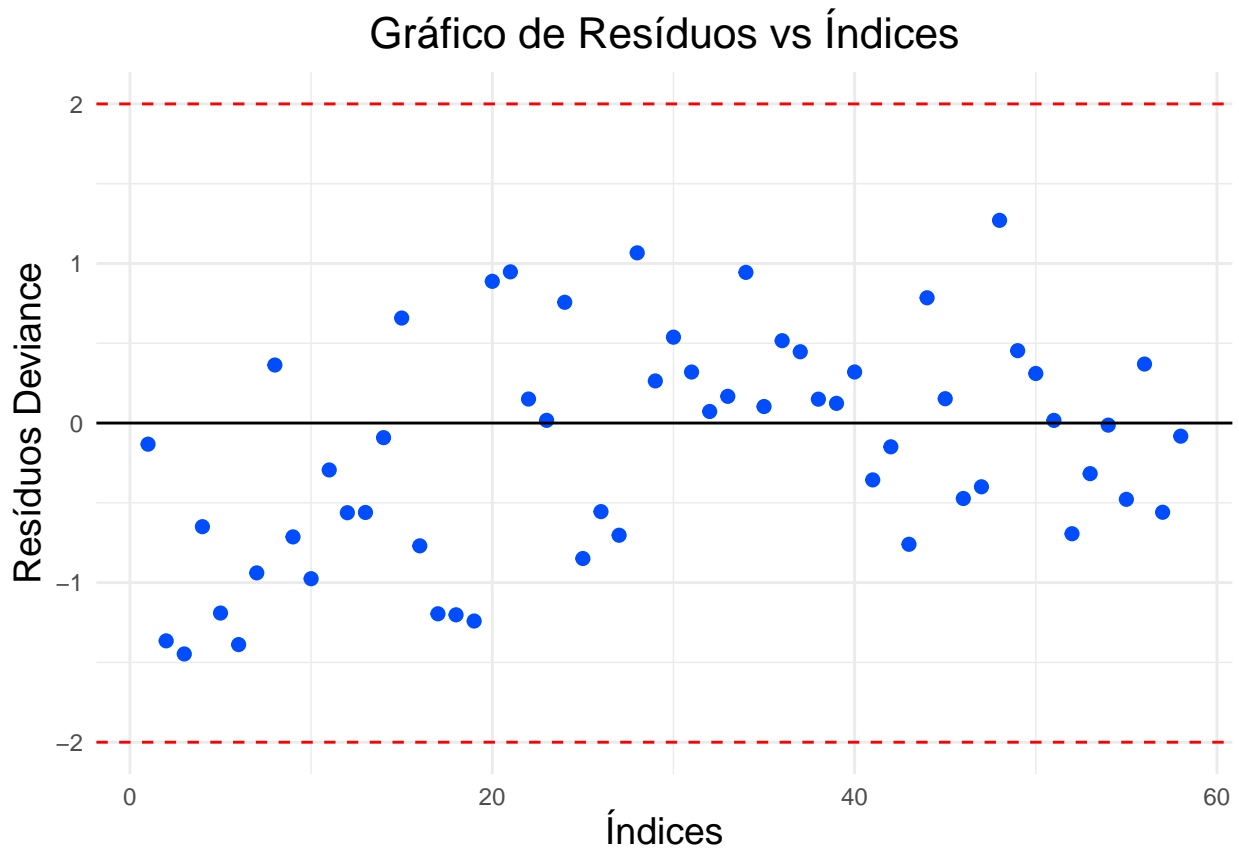
```

```

residuals_data <- data.frame(
  Index = seq_along(td),
  Residuals = td,
  Fitted = fitted_values
)

# Gráfico 1: Resíduos vs Índices
ggplot(residuals_data, aes(x = Index, y = Residuals)) +
  geom_point(color = "#004dff", size = 2) +
  geom_hline(yintercept = c(-2, 0, 2), linetype = c("dashed", "solid", "dashed"), color = c("red", "black", "red")) +
  labs(
    title = "Gráfico de Resíduos vs Índices",
    x = "Índices",
    y = "Resíduos Deviance"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    axis.title = element_text(size = 14)
  )

```



```

# Gráfico 2: Resíduos vs Valores Ajustados
ggplot(residuals_data, aes(x = Fitted, y = Residuals)) +
  geom_point(color = "#004dff", size = 2) +
  geom_hline(yintercept = c(-2, 0, 2), linetype = c("dashed", "solid", "dashed"), color = c("red", "black", "red")) +
  labs(
    title = "Gráfico de Resíduos vs Valores Ajustados",
    x = "Valores Ajustados",
    y = "Resíduos Deviance"
  )

```

```

) +
theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5, size = 16),
  axis.title = element_text(size = 14)
)

```

