

# Exercício 1

## Descrição

Considere para a análise os dados descritos no arquivos **ozone**, disponível no ‘R, em que as variáveis meteorológicas são observadas durante 330 dias na Bacia de Los Angeles, EUA, em 1976. As variáveis estão descritas na seguinte ordem:

- O3, máxima diária da concentração média de ozônio por hora em Upland, CA, medidas em partes por milhão (ppm)
- vh, altura da pressão 500 milibar medida na base da força aérea de Vandenberg
- wind, velocidade do vento em mph no aeroporto LAX
- humidity, umidade em porcentagem no LAX
- temp, temperatura na base aérea de Sandburg em graus Fahrenheit
- ibh, altura da inversão da temperatura base em pés
- dpg, gradiente da pressão a partir de Los Angeles para Daggert em mmHg
- ibt, inversão da temperatura base no LAX em graus Fahrenheit
- vis, visibilidade no LAX em milhas
- day, dia do ano

## Enunciado

Um dos objetivos do estudo com esse conjunto de dados é tentar explicar a concentração média de ozônio na atmosfera dadas as demais variáveis meteorológicas.

Como a resposta O3 é apresentada na forma de contagem (ppm), compare os ajustes de modelos com resposta Poisson e com resposta binomial negativa. Inicialmente, faça uma análise descritiva com os dados apresentando boxplots e os diagramas de dispersão de cada variável explicativa com a variável resposta. Em seguida, compare os ajustes de modelos log-lineares com resposta Poisson e binomial negativa. Para cada modelo faça uma seleção através do procedimento AIC. Compare os dois modelos segundo procedimentos de diagnóstico. Para o modelo selecionado interprete os resultados.

## Leitura dos dados

Os dados foram obtidos no R por meio dos comandos:

```
require(faraway)
data(ozone)
```

	O3	vh	wind	humidity	temp	ibh	dpg	ibt	vis	doy
	3	5710	4	28	40	2693	-25	87	250	33
	5	5700	3	37	45	590	-24	128	100	34
	5	5760	3	51	54	1450	25	139	60	35
	6	5720	4	69	35	1568	15	121	60	36
	4	5790	6	19	45	2631	-33	123	100	37

O3	vh	wind	humidity	temp	ibh	dpg	ibt	vis	doy
4	5790	3	25	55	554	-28	182	250	38

## Análise descritiva

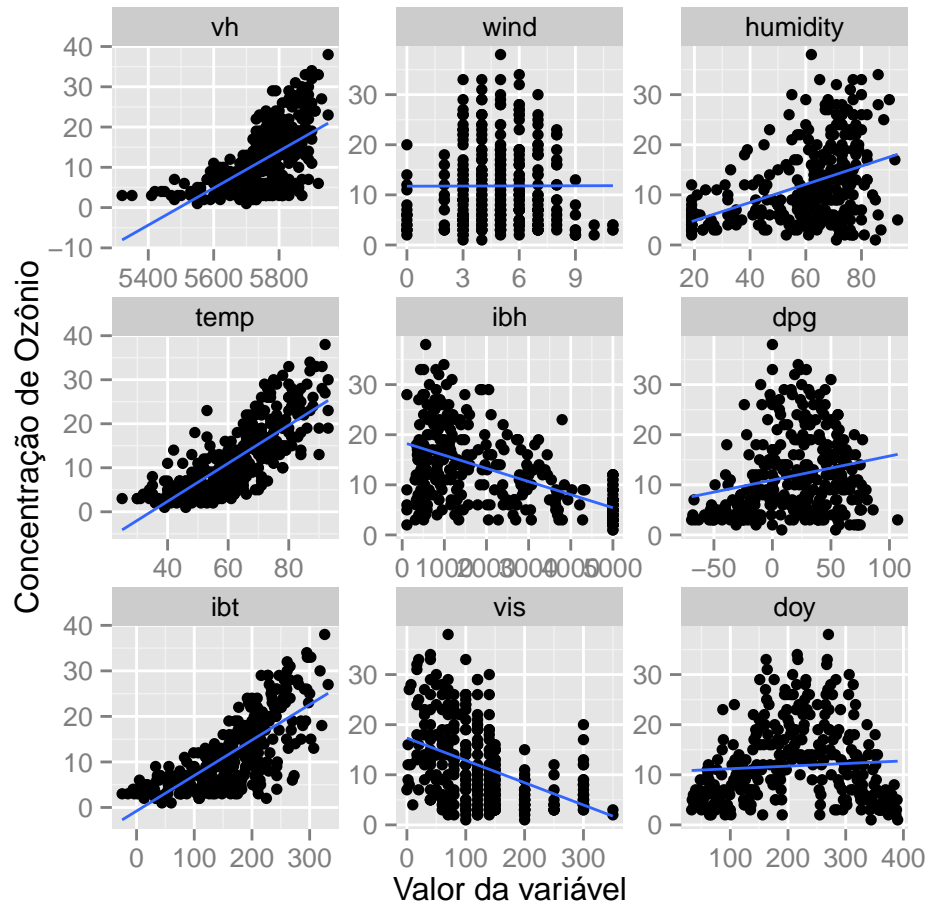


Figure 1: Gráfico de dispersão da variável resposta ‘O3’ (Concentração de Ozônio) por todas as variáveis explicativas.

Vemos na figura 1 que todas as variáveis apresentam relação linear com a variável resposta. As variáveis que apresentam a maior relação são a altura da pressão 500 milibar, a temperatura e a inversão da temperatura. Outras variáveis que também apresentaram relação com a concentração de ozônio foram a visibilidade e a altura da inversão da temperatura base.

## Modelo Poisson

Inicialmente vamos ajustar um modelo linear generalizado com distribuição de Poisson. Para selecionar as variáveis explicativas vamos usar o método AIC. Usando a função `step` do R, selecionaremos o modelo que tiver o menor AIC.

```

modelo.poisson.c <- glm(O3 ~ vh + wind + humidity + temp + ibh + dpq + ibt + vis + doy,
                        data = ozone, family = poisson)

modelo.poisson.r <- step(object = modelo.poisson.c, direction = "backward")

```

A tabela a seguir apresenta as estimativas dos parâmetros do modelo escolhido. O método AIC, depois de três passos, selecionou a velocidade do vento, a inversão da temperatura base, a umidade, a temperatura, a altura da inversão da temperatura, a visibilidade e o dia do ano como variáveis explicativas.

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.6305	0.1314	4.80	0.0000
wind	-0.0134	0.0094	-1.42	0.1566
humidity	0.0086	0.0012	7.36	0.0000
temp	0.0224	0.0029	7.80	0.0000
ibh	-0.0001	0.0000	-3.16	0.0016
ibt	0.0017	0.0007	2.39	0.0167
vis	-0.0007	0.0003	-2.54	0.0110
doy	-0.0009	0.0002	-4.38	0.0000

Apesar do parâmetro relacionado a variável ‘wind’ não ser significativo, continuaremos deixando-o no modelo, pois utilizamos o método AIC para selecionar as variáveis. A seguir fizemos alguns gráficos de diagnóstico a fim de verificar a qualidade do ajuste.

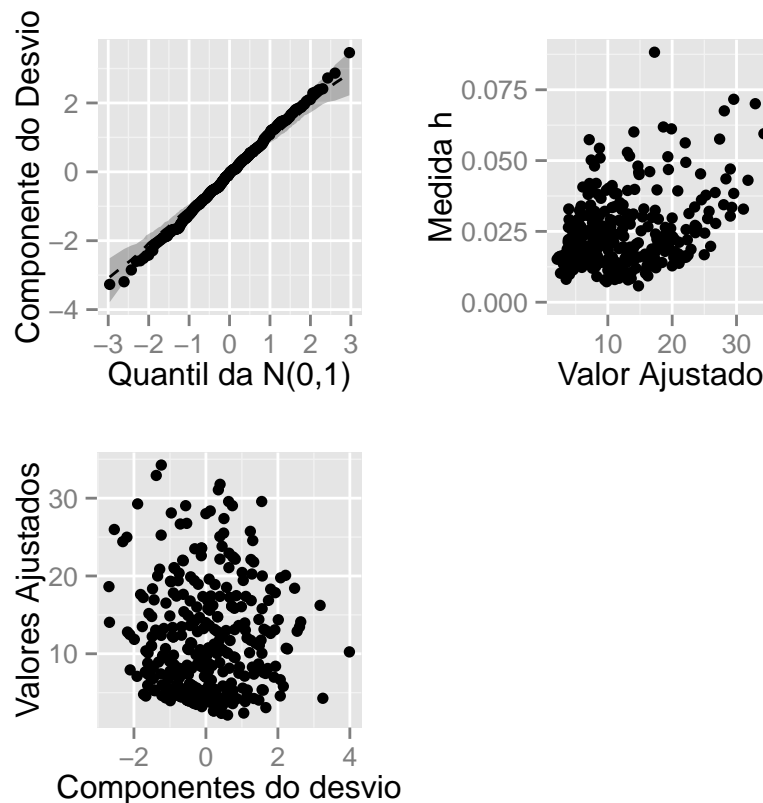


Figure 2: Gráficos de diagnóstico do modelo poisson reduzido.

Os gráficos de diagnóstico evidenciam um ajuste adequado do modelo Poisson. No entanto existe um leve desvio no gráfico quantil-quantil, alguns pontos estão fora da banda de confiança, o que não é recomendável. Por isso, vale a tentativa de ajuste de um outro modelo.

## Modelo Binomial-Negativo

O modelo binomial negativo foi ajustado no R usando a função `glm.nb` do pacote MASS. A seleção das variáveis explicativas também foi realizada utilizando o método AIC *backward*. Selecionamos aquele modelo com o menor AIC.

```
library(MASS)
modelo.binomialn.c <- glm.nb(O3 ~ vh + wind + humidity + temp + ibh + dpq + ibt + vis + doy,
                             data = ozone)
modelo.binomialn.r <- step(object = modelo.binomialn.c)
```

Pelo método stepAIC selecionamos as variáveis ibt, vis, ibh, doy, humidity e temp como variáveis explicativas do modelo. As estimativas dos parâmetros encontram-se na tabela a seguir.

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.6117	0.1497	4.09	0.0000
humidity	0.0079	0.0013	6.09	0.0000
temp	0.0225	0.0033	6.83	0.0000
ibh	-0.0001	0.0000	-3.02	0.0025
ibt	0.0017	0.0008	1.99	0.0463
vis	-0.0008	0.0003	-2.42	0.0154
doy	-0.0009	0.0002	-3.78	0.0002

Novamente, vamos fazer alguns gráficos de diagnóstico para verificar a adequabilidade do ajuste do modelo binomial negativo. Inicialmente fizemos o gráfico quantil-quantil, que verifica se a distribuição escolhida se ajusta aos dados.

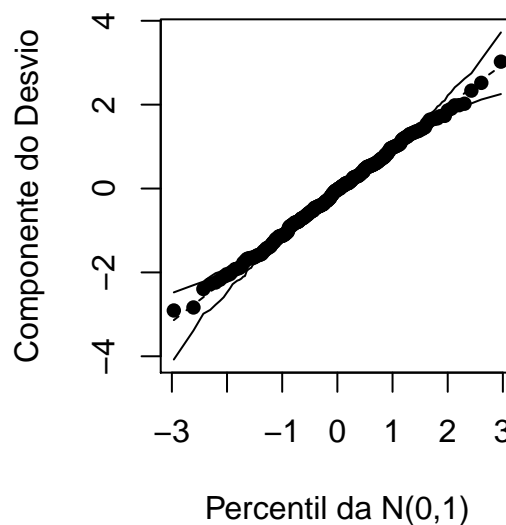


Figure 3: Gráfico Quantil-Quantil do modelo binomial negativo reduzido

Vemos no gráfico que todos os pontos estão dentro da banda de confiança, indicando que o ajuste está correto. Em seguida fizemos mais alguns gráficos, que ajudam a identificar pontos discrepantes, aberrantes ou influentes que podem causar problemas nas estimativas dos parâmetros do modelo.

Pelos gráficos de diagnóstico podemos concluir que o modelo binomial negativo mostrou-se adequado aos dados, inclusive seu ajuste foi melhor do que o ajuste do modelo Poisson. Não há pontos de alta influência, que poderiam causar problemas nas estimativas dos parâmetros. Portanto, escolhemos o modelo Binomial Negativo.

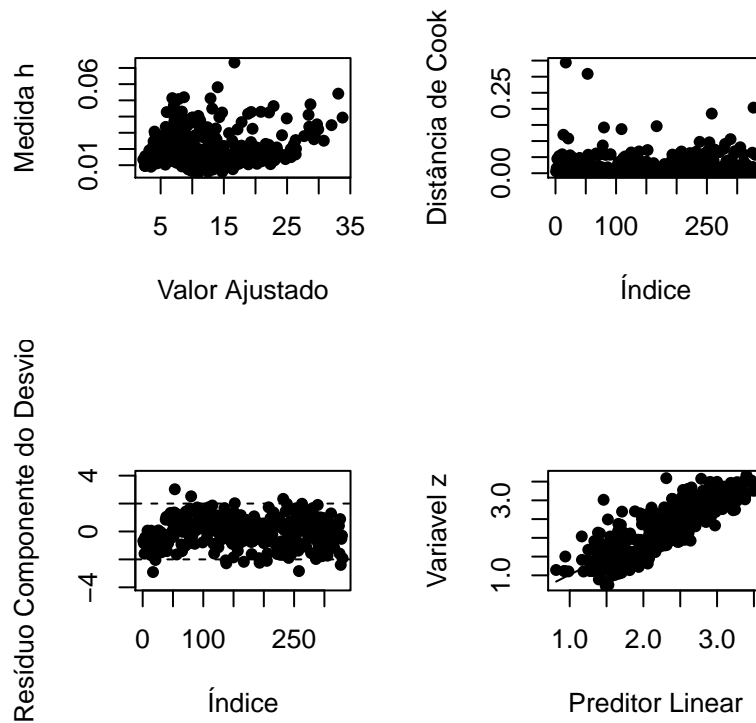


Figure 4: Gráficos de diagnóstico negativo binomial reduzido

## Interpretação dos parâmetros

As variáveis humidade, temperatura e inversão da temperatura base geram um aumento na concentração de gás O<sub>3</sub> na atmosfera, isto é, quanto maior a temperatura, humidade e inversão da temperatura, maior a concentração de O<sub>3</sub>. A variável temperatura é a que tem maior impacto neste aumento. Já as variáveis altura da inversão da temperatura, visibilidade e dia do ano tem um impacto negativo nesta concentração. Tanto uma como as outras conforme aumentam, fazem com que a concentração de O<sub>3</sub> diminua.