

# Seleção de Modelos de Regressão - Parte 2

Fernando Bispo, Jeff Caponero

# Sumário

<b>Apresentação</b>	<b>2</b>
<b>Atividade 1</b>	<b>3</b>
Introdução . . . . .	3
Resultados . . . . .	3
Análise descritiva dos dados . . . . .	3

# Apresentação

O relatório desta semana está dividido em duas atividades. Na primeira foi analisado um banco de dados sobre uma indústria que realiza a oxidação de amônia, para o qual por meio de técnicas de regressão linear múltipla se elaborou um modelo para determinar a perda de amônia no processo. Na segunda atividade, se buscou determinar a quantidade de água perdida do solo, evaporação do solo, com base em um banco de dados sobre propriedades do solo e do ar associadas. Nesta segunda atividade também foram utilizadas técnicas de regressão linear múltipla.

# Atividade 1

## Introdução

Com base nos dados disponibilizados no *dataset* “stackloss” (do R base), que apresenta dados de 21 dias de operação de um indústria que realiza oxidação de amônia ( $NH_3$ ) em ácido nítrico ( $HNO_3$ ). O ácido nítrico produzido é absorvido na torre de absorção contracorrente. As informações disponíveis na base de dados referem-se a:

- **Air flow**: que representa a taxa de operação da indústria (corrente de ar refrigerado);
- **Water Temp**: é a temperatura de resfriamento da água que circula nos canos da torre de absorção;
- **Acid.Conc.**: é a concentração do ácido [em porcentagem, após tratamento]; e
- **stack.loss** (variável dependente) é o percentual (após tratamento) de amônia introduzida no processo industrial que escapa da absorção (representando uma medida(inversa) de eficiência total da indústria).

Com base nestes dados, objetiva-se:

1. Ajustar um modelo linear múltiplo completo para estes dados. Avaliando as estimativas dos parâmetros, os resíduos e a influência das observações no ajuste do modelo, incluindo leverage, distância de Cook, DFBETAs, DFFITs e COVRATIOs.
2. Avaliar a partir de regressão parcial e dos resíduos parciais as variáveis no modelo, bem como o pressuposto de normalidade do resíduos.

## Resultados

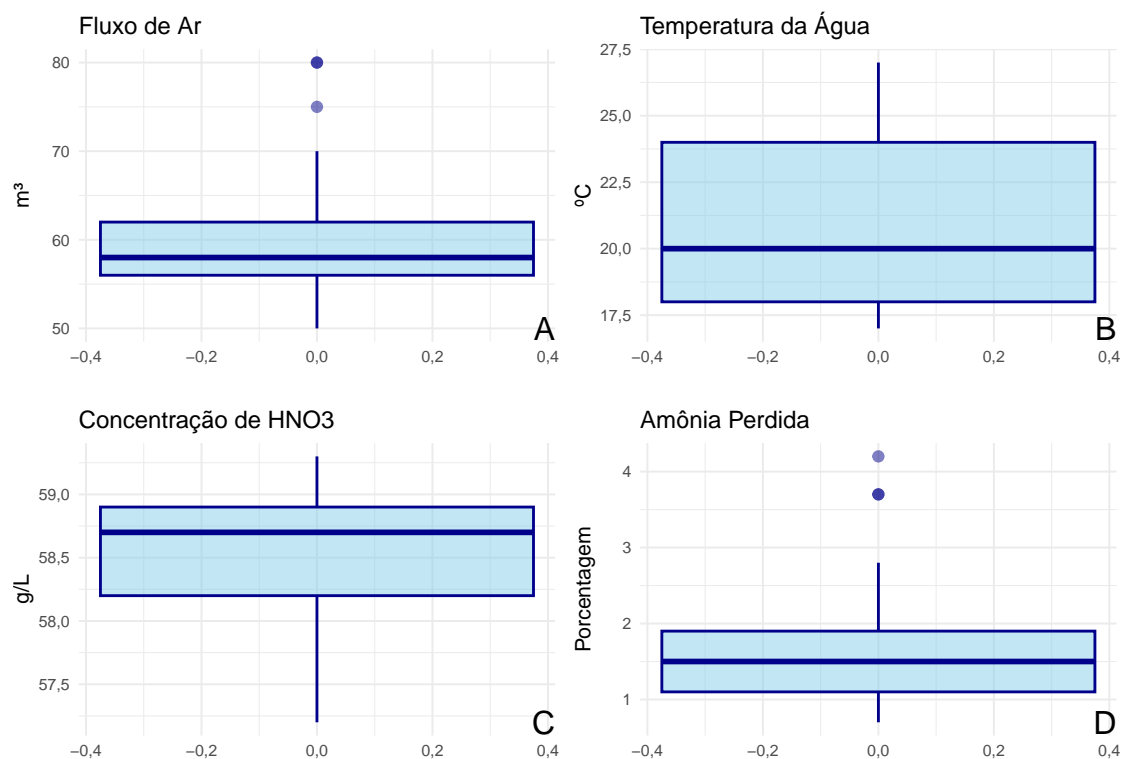
### Análise descritiva dos dados

É possível realizar uma descrição prévia dos dados por meio de medidas de resumo e de gráficos do tipo box-plot como vê-se a seguir:

Tabela 1: Medidas Resumo dos dados

	Mín	Q1	Med	Média	Q3	Máx	Desv.padrão	CV	Assimetria	Curtose
Amônia Perdida	0,7	1,1	1,5	1,75	1,9	4,2	1,02	0,58	1,16	0,13
Concentração de HNO <sub>3</sub>	57,2	58,2	58,7	58,63	58,9	59,3	0,54	0,01	-0,87	0,19
Fluxo de Ar	50,0	56,0	58,0	60,43	62,0	80,0	9,17	0,15	0,81	-0,26
Temperatura da Água	17,0	18,0	20,0	21,10	24,0	27,0	3,16	0,15	0,47	-1,23

Figura 1: BoxPlot das variáveis em análise.



Nota-se uma assimetria nos dados apresentados e algumas observações que podem ser descritas como *outliers*. Entretanto é possível propor um modelo de regressão como se segue.